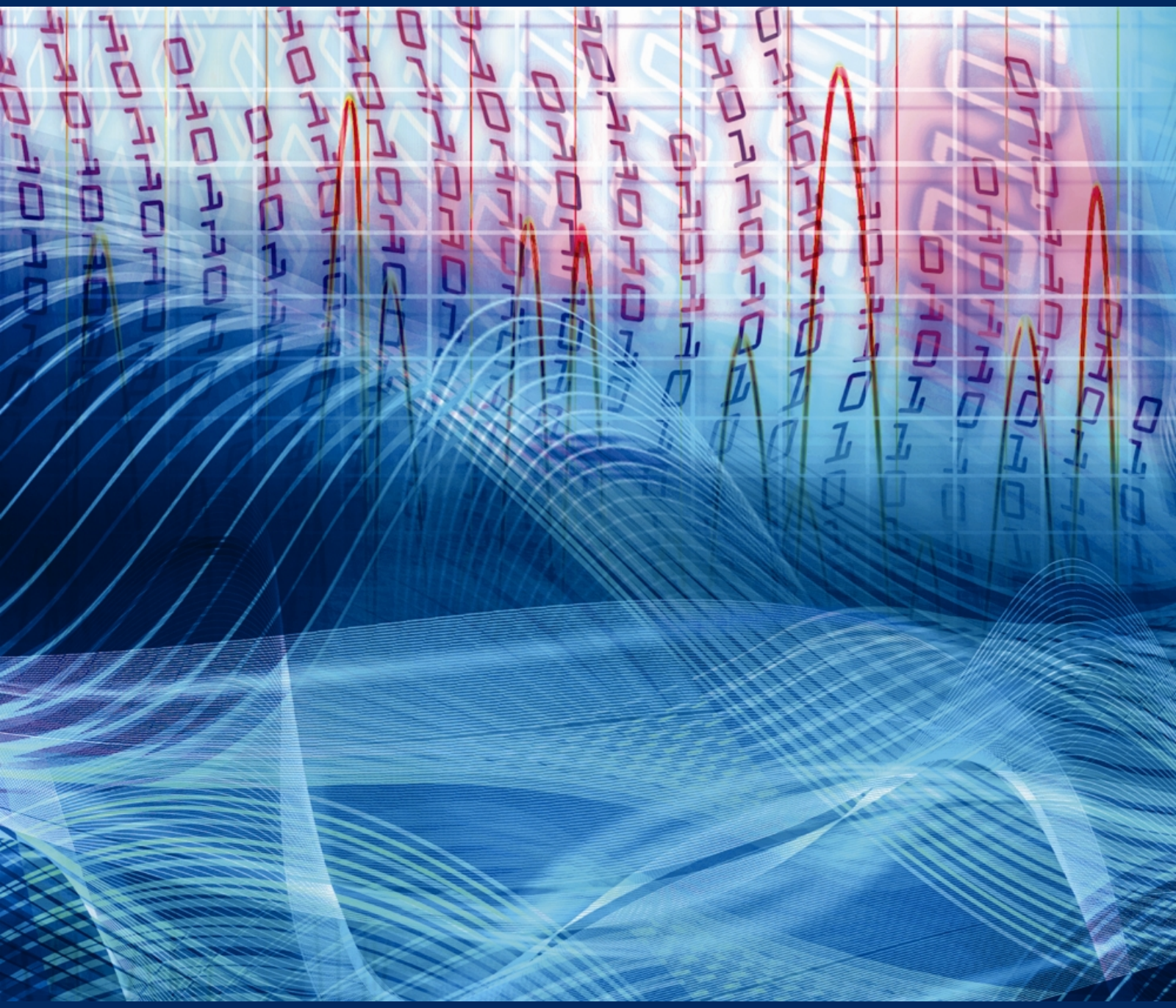
EURASIP Journal on Applied Signal Processing

Signal Analysis Tools for Optical Information Processing

Guest Editors: Christi K. Madsen, Daniela Dragoman, and José Azaña



EURASIP Journal on Applied Signal Processing

Signal Analysis Tools for Optical Information Processing

EURASIP Journal on Applied Signal Processing

Signal Analysis Tools for Optical Information Processing

Guest Editors: Christi K. Madsen, Daniela Dragoman, and José Azaña

Contents

Editorial, Christi K. Madsen, Daniela Dragoman, and José Azaña
Volume 2005 (2005), Issue 10, Pages 1449-1451

Active Optical Lattice Filters, L. Roberts Hunt, Vishnupriya Govindan, Issa Panahi, Jian Tong, Govind Kannan, Duncan L. MacFarlane, and Gary Evans
Volume 2005 (2005), Issue 10, Pages 1452-1461

Advanced Optical Processing of Microwave Signals, Beatriz Ortega, Daniel Pastor, José Mora, José Capmany, and Miguel V. Andrés
Volume 2005 (2005), Issue 10, Pages 1462-1484

DSP Approach to the Design of Nonlinear Optical Devices, Geeta Pasrija, Yan Chen, Behrouz Farhang-Boroujeny, and Steve Blair
Volume 2005 (2005), Issue 10, Pages 1485-1497

Fractional Transforms in Optical Information Processing, Tatiana Alieva, Martin J. Bastiaans, and Maria Luisa Calvo
Volume 2005 (2005), Issue 10, Pages 1498-1519

Applications of the Wigner Distribution Function in Signal Processing, Daniela Dragoman
Volume 2005 (2005), Issue 10, Pages 1520-1534

Wigner Distribution Moments Measured as Intensity Moments in Separable First-Order Optical Systems, Martin J. Bastiaans and Tatiana Alieva
Volume 2005 (2005), Issue 10, Pages 1535-1540

Concepts for the Temporal Characterization of Short Optical Pulses, Christophe Dorrer and Ian A. Walmsley
Volume 2005 (2005), Issue 10, Pages 1541-1553

Time-Frequency (Wigner) Analysis of Linear and Nonlinear Pulse Propagation in Optical Fibers, José Azaña
Volume 2005 (2005), Issue 10, Pages 1554-1565

A Novel Optical Vector Spectral Analysis Technique Employing a Limited-Bandwidth Detector, C. K. Madsen
Volume 2005 (2005), Issue 10, Pages 1566-1573

Optical Wavelet Signals Processing and Multiplexing, Gabriella Cincotti, Michela Svaluto Moreolo, and Alessandro Neri
Volume 2005 (2005), Issue 10, Pages 1574-1583

Adaptive Electronic Dispersion Compensator for Chromatic and Polarization-Mode Dispersions in Optical Communication Systems, Ut-Va Koc
Volume 2005 (2005), Issue 10, Pages 1584-1592

Linear and Nonlinear Crosstalk Evaluation in DWDM Networks Using Optical Fourier Transformers,
R. Llorente, R. Clavero, F. Ramos, and J. Marti
Volume 2005 (2005), Issue 10, Pages 1593-1602

Analysis of Optical CDMA Signal Transmission: Capacity Limits and Simulation Results,
Aminata A. Garba, Raymond M. H. Yim, Jan Bajcsy, and Lawrence R. Chen
Volume 2005 (2005), Issue 10, Pages 1603-1616

Design of Extended Depth-of-Focus Laser Beams Using Orthogonal Beam Expansions,
David P. Goren, Joseph Katz, and Leonard Bergstein
Volume 2005 (2005), Issue 10, Pages 1617-1623

EURASIP Journal on Applied Signal Processing 2005:10, 1449–1451

Editorial

Christi K. Madsen
Department of Electrical Engineering, Texas A&M University, 312 B Zachry Engineering Center, College Station, TX 77843, USA
Email: cmadsen@ee.tamu.edu

Daniela Dragoman
Faculty of Physics, University of Bucharest, P.O. Box MG-11, 077125 Bucharest, Romania
Email: daniela@solid.fizica.unibuc.ro

José Azaña
Institut National de la Recherche Scientifique (INRS) Énergie, Matériaux et Télécommunications, 800 de la Gauchetière Ouest, bureau 6900, Montréal, QC, Canada H5A 1K6
Email: azana@emt.inrs.ca

The application of advanced signal analysis tools (e.g., fractional Fourier transforms or joint time-frequency signal representations) to a wide variety of optics and photonics problems has led to a new and deeper understanding of several optical phenomena of fundamental and practical importance, including diffraction, holography, nonlinear optical processes, dispersion, and optical filtering. Signal analysis methods also constitute the basis of powerful techniques for the measurement and full characterization of ultrafast optical events or systems, which otherwise could not be characterized by conventional means. Moreover, novel analysis and synthesis methods for different photonics devices (e.g., fiber gratings, ring resonators, etc.) have been developed based on well-known continuous and discrete-time signal processing tools.

The use of different photonic technologies for processing spatial or temporal information in the optical domain is also a field of growing importance, with a strong potential for interesting applications in fields such diverse as optical telecommunications, ultrafast metrology, microwave engineering, image processing, and optical computing, to name only a few. Advantages of processing the information in the optical domain include the tremendous available bandwidth and the parallelism intrinsic to the optical approach, which translate into ultrahigh processing speeds, which otherwise are not possible.

The broad area of optical signal processing is becoming today one of the most active research areas in optics and photonics. Research in this area will have an important impact far beyond the conventional frontiers of photonic technologies. The present issue of EURASIP JASP is devoted to this increasingly important topic. Specifically, the aim of this special issue is to highlight innovative research in signal processing applied to optics and photonics problems, thus paving the way for future developments in the field. The present issue was thought of with the intention of providing an overview as complete as possible of the recent progress and current problematics in optical signal processing, while bringing the work in this area closer to the signal processing community. This was the philosophy behind the decision to prepare a special issue of the EURASIP JASP devoted to this area. In expressing this philosophy, we are very grateful to *Dr. Jacob Benesty*, who first suggested and encouraged us to proceed ahead with this special issue.

The special issue comprises both original research contributions and review papers by leaders in their respective arenas. This includes works ranging from applications of signal analysis tools to optical problems to the proposal and demonstration of innovative concepts, technologies, devices, and architectures for all-optical information processing. In particular, the current issue consists of fourteen contributions, namely, seven invited papers and seven regular contributions. The latter were selected by the Guest Editors following a suitable evaluation via a standard international peer-review process. As mentioned above, the intention was to cover most of the relevant topics in the area. Specifically, the *invited contributions* in this special issue are the following.

(1) "Active optical lattice filters" by L. R. Hunt et al.
(2) "Advanced optical processing of microwave signals" by B. Ortega et al.

(3) "Fractional transforms in optical information processing" by T. Alieva et al.
(4) "Applications of the Wigner distribution function in signal processing" by D. Dragoman.
(5) "Concepts for the temporal characterization of short optical pulses" by C. Dorrer and I. A. Walmsley.
(6) "Time-frequency (Wigner) analysis of linear and nonlinear pulse propagation in optical fibers" by J. Azaña.
(7) "A novel optical vector spectral analysis technique employing a limited-bandwidth detector" by C. K. Madsen.

Hunt et al. were invited to present and review their recent developments in active optical lattice filters. This work constitutes a relevant example of how well-known concepts of signal processing (i.e., adaptive lattice filtering) can be successfully applied in photonics. All-optical adaptive filtering devices are proposed and demonstrated. In their invited contribution, Ortega et al. give an extensive overview about their work on microwave signal processing based on photonics technologies. The authors review some recent, relevant approaches to implement high-performance transversal RF filters using optical devices such as fiber Bragg gratings, arrayed waveguide gratings, or interferometric structures. Experimental evidence of their proposals is also provided. Pasrija et al. propose the use of discrete-time signal processing tools for designing and synthesizing nonlinear optical devices. This proposal is based on the pioneer work by Madsen, where concepts of discrete-time signal analysis were applied for synthesizing linear allpass optical filters.

In their invited paper, Alieva et al. provide a comprehensive overview on the use of fractional linear integral transforms for different optical information processing applications, including phase retrieval, beam characterization, pattern recognition, adaptive filter design, encryption, watermarking, and motion detection. The contribution by Dragoman focuses on the application of phase-space representations, and in particular Wigner analysis, to a wide variety of signal processing problems with an emphasis on optical signals and systems. Her paper is a review of classical and relevant work on the use of advanced signal analysis tools in the context of optics and photonics. In their contributed paper, Bastiaans and Alieva elaborate further on the concept of Wigner distribution applied to optical systems.

In their invited contribution, Dorrer and Walmsley present an extensive review of signal analysis-based methods for the full (amplitude and phase) characterization of (ultra-) short optical pulses. It is discussed how an optical pulse can be analyzed and fully characterized through its representation in terms of correlation functions or time-frequency representations, and different methods to experimentally obtain these representations in the optical domain are discussed and demonstrated. In his work, Azaña makes use of joint time-frequency signal representations for investigating an optical problem of fundamental and practical significance, namely, the dynamics of picosecond pulse propagation through optical fibers in the linear and nonlinear regimes. A deeper insight into this problem is provided through this analysis.

The paper by Madsen introduces and analyzes a new and simple technique for characterizing both chromatic and polarization-mode dispersions in an optical channel. The technique is based on discrete-time signal analysis concepts and should prove to be very useful for applications in WDM optical communication systems. In their contribution, Cincotti et al. present a comprehensive overview of wavelet signal processing and multiplexing in the optical domain, using photonics integrated technologies. These developments are of interest for broadband multiple access networks. The work by Ut-Va Koc deals with improved adaptive equalization algorithms for the electronic compensation of chromatic and polarization-mode dispersions in fiber-optics communication links. In their paper, Llorente et al. propose and experimentally demonstrate an interesting application of the so-called real-time Fourier transformation technique, where the spectrum of an optical signal is mapped into the temporal domain via chromatic dispersion, for evaluating channel crosstalk in DWDM optical communication networks. The work by Garba et al. deals with the increasingly important topic of optical CDMA (OCDMA). In particular, different coding strategies for OCDMA are proposed and evaluated in terms of their capacity limits and noise performance for multiple-access networking.

Finally, the paper by Goren et al. introduces a novel signal analysis-based technique for synthesizing laser beams with extended depth of focus, of specific interest for scanning printed bar codes.

In the coming years, it is expected that the area of optical signal processing will become even more important from both fundamental and applied perspectives. We hope that this special issue will appeal to the signal processing community and will further stimulate work in this area. To finalize, we would like to thank all the people who have participated in the elaboration of this special issue, especially the authors of the published papers, the researchers who submitted their work for consideration, and last, but not least, the referees who helped in the revision and selection of the submitted works.

Christi K. Madsen
Daniela Dragoman
José Azaña

Christi K. Madsen received the B.S. degree from The University of Texas at Austin in 1986, the M.S. degree from Stanford University, Standford, Calif, in 1987, and the Ph.D. degree from Rutgers University, Piscataway, NJ, in 1996, all in electrical engineering. She joined AT&T Bell Laboratories in 1987 and worked for the submarine systems business unit. After completing her Ph.D., she transferred to the Integrated Photonics Research Department, Bell Laboratories. Her research has focused on the application of digital filter and signal processing techniques to optical filters for high-speed, high-capacity

optical communication systems. Madsen invented a class of tunable, multistage optical allpass filters that allow any phase response to be approximated and have application in chromatic and polarization-mode dispersion compensation. She has served on many conference committees and given short courses. In 2004, she was the General Chair for the Integrated Photonics Research (IPR) Conference. She was promoted to Distinguished Member of the Technical Staff at Bell Laboratories in 2002 and achieved Fellow ranking in the Optical Society of America in 2003. She holds 16 US patents and has given over 70 technical talks and papers. She is now a Professor at Texas A&M University, College Station, Tex.

Daniela Dragoman was born in 1965. She obtained the M.S. degree from the University of Bucharest, Romania, in 1989, and the Ph.D. degree from Limerick University, Ireland, in 1993. She was a Visiting Professor at the University Saint-Etienne, France, in 1997 and 2000 and was awarded the Alexander von Humboldt Fellowship in 1998, being a Visiting Professor at the University of Mannheim during 1998–1999 and 2001–2002. Presently, she is a Professor at the Physics Faculty, University of Bucharest, where she teaches integrated optoelectronic devices and the interaction of radiation with matter. She authored over 100 papers in the areas of phase space characterization of optical beams and systems, optical micromechanical devices, quantum physics, and mesoscopic devices. She is the coauthor of the books *Advanced Optoelecronic Devices*, Springer, 1999, *Optical Characterization of Solids*, Springer, 2002, and *Quantum-Classical Analogies*, Springer, 2004. She is a reviewer at several international journals of optics and photonics, and Editor for the Springer book series *The Frontiers Collection*. She was awarded in 1999 the Romanian Academy Prize "Gheorghe Cartianu."

José Azaña was born on December 8, 1972, in Toledo, Spain. He received the Ingeniero de Telecomunicación degree (a six-year engineering program) and the Ph.D. degree (in the areas of optical signal processing and fiber Bragg gratings) from the Universidad Politécnica de Madrid (UPM) in 1997 and 2001, respectively. He completed part of his Ph.D. work at the University of Toronto, Canada, and University of California, Davis, USA. From September 2001 to mid 2003, he worked as a Postdoctoral Research Associate in the Department of Electrical and Computer Engineering, McGill University, Montreal, Canada. Recently, he joined the Institut National de la Recherche Scientifique (INRS), Montreal, where he is an Assistant Research Professor in the Ultrafast Optical Processing Group. The research work of Dr. Azaña has resulted in more than 45 publications in top scientific and engineering journals and it has been recognized with several distinctions in Spain and Canada. His current research interests focus on fiber and integrated technologies for ultrafast optical signal processing and optical pulse shaping, for various applications, including optical telecommunications, ultrafast metrology, biomedical imaging, and microwave waveform manipulation.

EURASIP Journal on Applied Signal Processing 2005:10, 1452–1461

Active Optical Lattice Filters

L. Roberts Hunt,[1] Vishnupriya Govindan,[1] Issa Panahi,[1] Jian Tong,[1] Govind Kannan,[1] Duncan L. MacFarlane,[1] and Gary Evans[2]

[1] *Erik Jonsson School of Engineering and Computer Science, The University of Texas at Dallas, Richardson, TX 75083-0688, USA*
Emails: hunt@utdallas.edu, vishnupriyagovindan@hotmail.com, issa.panahi@utdallas.edu, jxt021000@utdallas.edu, govind@student.utdallas.edu, dlm@utdallas.edu

[2] *School of Engineering and Applied Science, Southern Methodist University, Dallas, TX 75275-0335, USA*
Email: gae@engr.smu.edu

Received 26 March 2004; Revised 20 October 2004

Optical lattice filter structures including gains are introduced and analyzed. The photonic realization of the active, adaptive lattice filter is described. The algorithms which map between gains space and filter coefficients space are presented and studied. The sensitivities of filter parameters with respect to gains are derived and calculated. An example which is relevant to adaptive signal processing is also provided.

Keywords and phrases: active, adaptive, filter, optical, lattice.

1. INTRODUCTION

Dowling and MacFarlane [1, 2, 3] viewed a lightwave lattice filter design using multistage etalons and resonators as a discrete-time linear system. As such they discussed the behavior of the system in terms of the transmission and reflection transfer functions. Given a desired transfer function denominator for an all-pole system in transmission they developed a layer-peeling algorithm to determine the appropriate reflection coefficients. The reflection transfer function has this same denominator polynomial. They proved that any denominator polynomial, and hence any autoregressive (AR) transfer function in transmission, can be generated by proper choice of the reflection coefficients. Transmission coefficients as well as reflection coefficients can be chosen, but these transmission coefficients do not affect the denominator polynomial. If an autoregressive moving average (ARMA) transfer function is desired in transmission, then it must first be approximated by an AR transfer function of higher order, and a higher-order lattice filter be designed.

The problem with the above optical lattice filter is that it is a passive device, and once it is built with a given set of reflection (also transmission) coefficients, its filtering characteristics cannot be changed. Hence this precludes programming in the field to change the filter as needed. Moreover, from the point of view of signal processing, the entire set of applications derived from adaptive signal processing is not accessible. Adaptive signal processing is used in communications, control, radar, sonar, seismology, biomedical, target tracking, and so forth [4]. Adaptation is needed if the statistics of a signal are unknown or if a system is time varying (nonstationary). Finite impulse response (FIR) filters have dominated the field, but infinite impulse response (IIR) filters can often provide a system of much lower order [5].

To add adjustment to these optical lattice filters we propose to add gains to each stage. These gains can be changed as needed and also serve to make our filters active instead of passive. Suppose we are given a target all-pole transfer function for transmission whose order is the same as the order (number of stages) of our optical filter. Once the reflection coefficients (by layer peeling) and transmission coefficients are determined and the device is built, then we tune the device by changing the gains. The question of interest is what all-pole transfer functions (or denominator polynomials) can we achieve by changing the gains? Certainly when the gains are all unity we have the denominator polynomial of the all-pole transfer function that we started. From the point of tunable signal processing it would be desirable to generate all "nearby" polynomials of the same degree.

It is interesting to note that after fixing these reflection and transmission coefficients, it is not possible to use the gains to generate an arbitrary all-pole transfer function in transmission. We present an example to illustrate this point. However, if we consider all gains as one when the reflection and transmission coefficients are fixed, then there is an open

neighborhood U of these ones in gains space and an open neighborhood V of the corresponding denominator polynomial coefficients in polynomial coefficients space, so that for each polynomial with coefficients in V there is a unique set of gains in U that produce that polynomial. All that needs to be checked is that the Jacobian matrix of the map from the gains (actually gains squared) to the polynomial coefficients is nonsingular when the gains are all one. If this matrix is nonsingular it is natural to solve for the gains as functions of the coefficients using the standard Newton-Raphson method [6, 7]. The equations expressing the relationship of the polynomial coefficients and the gains are easily generated and have a high degree of symmetry. Hence we have a procedure that will allow us to calculate all nearby (in terms of polynomial coefficients) denominator polynomials for our transfer functions. Thus we can use gains to adapt our filter. Of course we are only interested in stable filters when applications are considered.

For the map from the gains to the denominator polynomial coefficients we stress the case where the gains are all one. However, for a given denominator polynomial there may be other sets of gains that yield the same polynomial coefficients. If for such a point in gain space, the Jacobian matrix of the gains to polynomial coefficients map is nonsingular, then there is an open neighborhood U^* of the gain point and an open neighborhood V^* of the corresponding polynomial coefficient point in polynomial coefficient space on which the map is one-to-one and onto. Of course U and U^* cannot intersect, so we can adapt in the set U (or U^*) without disturbing the other set.

The main purpose of this paper is to provide design tools for active lattice filters that contain active gain elements. This discussion is particularly relevant to an optical architecture that is currently under early development. This photonic realization of the active lattice filter is described in Section 2 of this paper. In particular we advocate a semiconductor laser amplifier structure in which coupling between gain and delay stages is accomplished by surface gratings [8]. While there is a rich tradition of active electronic filters, optical filters with gain are not generally used. Gain elements allow filters with high-quality factors, and filters with programmable transfer functions. Further, gain elements are essential to providing an architecture that may be scaled to reasonable sophistication.

In Section 3 of this paper we discuss the Dowling-MacFarlane optical lattice filter and the layer-peeling procedure for computing the reflection coefficients. We indicate how gains are added to their structure. We then compute the transfer function denominator polynomials as functions of the gains assuming that all reflection and transmission coefficients have been fixed. In Section 4 we provide an example showing that an arbitrary transmission all-pole denominator cannot be generated using gains once the reflection and transmission coefficients have been chosen. Here we work with a two-stage filter and a second-order transfer function denominator polynomial. We generate the map from two-dimensional gains space to two-dimensional polynomial coefficients space. For given polynomial coefficients this map can be represented by two curves in gains space and there are three possible outcomes: transversal intersection where the Jacobian matrix is nonsingular, no intersection and no solution, and tangential intersection where the Jacobian matrix is singular. It is the first case that is of interest to us. It is important to understand that we start with a given polynomial and fix the reflection coefficients so that we have this case if the Jacobian matrix is nonsingular when all gains are unity. Then all "nearby polynomials" can be implemented by a proper choice of gains.

Section 5 contains a development of the map from gains (squared) space to denominator polynomial coefficient space in the general case. Then we state and prove our main result using the inverse function theorem [9]. The assumption on the nonsingularity of the Jacobian matrix is the exact condition needed to employ the Newton-Raphson algorithm to solve our equations for the gains. We return to the example in Section 4 for the nonsingular case, vary the coefficients of the polynomial, and show the computations of the gains. We also provide an example that is relevant to adaptive signal processing, assuming that a feedback loop may be implemented that is fast enough to respond to a changing environment. We start with an output response of an AR system. Next we identify the all-pole transfer function and determine the appropriate reflection coefficients (through layer peeling), which we fix in our lattice filter. Then we allow the time response to change, identify the corresponding denominator polynomials for the new transfer functions, and compute the gains to deliver those polynomials. Here the time response could be replaced by a frequency response or by autocorrelations. In Section 6 we derive and calculate the sensitivities of filter parameters with respect to gains. Section 7 contains our conclusions and a discussion of future research.

2. A REALIZATION ARCHITECTURE

The optical gains that are modeled in this paper may be realized in a number of ways. For example, it is possible to imagine using erbium-doped fibers or waveguides separated by fiber Bragg gratings or an equivalent coupling interface. The passive version of this fiber based-filter architecture is very nicely described in the classic paper by Moslehi et al. [10]. The active fiber filter with gain would have particular interest for fiber sensors and instrumentation [11]. In the interests of faster updating speeds and adaptive operation we are envisioning a semiconductor laser amplifier structure in which coupling between gain and delay stages is accomplished by surface gratings. This architecture also offers faster clock frequencies, higher efficiencies and integrated manufacture advantages.

In Figure 1 is shown a schematic of the active optical lattice filter that uses semiconductor laser amplifier stages to provide gain and delays. By way of example, the figure shows a two-stage active lattice filter with two gain/delay regions and three interstage couplers. The figure is drawn in side view and shows a substrate with an epitaxial-grown quantum well active region. On the surface of this substrate are two electrodes through which injection current may enter the

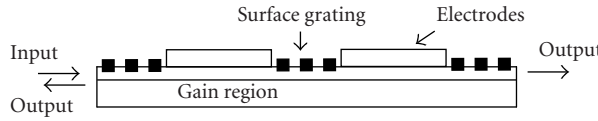


Figure 1: Block diagram of an active lattice filter based on surface grating interstage couplers and semiconductor optical amplifiers.

quantum well. The filter stage gain is controlled by the injection current to the individually addressable gains in the amplifying regions. These regions also provide the phase delays that lead to filter responses.

Also on the surface of the substrate are three grating couplers that provide transmission and reflection of optical signals between the two stages under the two electrodes. These gratings may be written by holographic lithography [8, 12] or by focused ion beam micromachining.

In operation, an optical signal is injected into the heterojunction of the structure where it undergoes gain and delay. At the grating couplers part of the signal is reflected and part is transmitted. Each of these components combines with signals from the adjacent stage, and this delay and mix action is the same as in coupled etalons, or in thin film filters. The additional presence of gain provides advantages in filter performance including a tunable response.

In practice, the semiconductor optical amplifier may suffer from two impairments: saturation and phase distortion. Since the theory presented herein applies to linear, time-invariant (LTI) systems, either impairment will render the results presented in this paper an approximation valid over a finite region of operation. Since the filter will be operated in a stable region, well below lasing threshold, the overall gains needed for useful filters will be small, and will vary over a limited range. Hence the linearity condition will often be justified. Similarly, it may also be argued that in this operating region, the injected current will not substantially impact the refractive index of the inversion region, particularly if the effective gain length is kept short and the epitaxy is properly designed. An additional possibility for adjusting and maintaining the correct phase is to add an additional phase control section to the filter as is done in many tunable lasers so as to allow an additional adjustment for phase control.

3. DOWLING-MACFARLANE FILTER

In 1944 MacFarlane and Dowling studied the analysis and design of purely passive coupled Fabry-Perot etalons and thin film filters using digital signal processing techniques [1, 2]. This approach allowed certain design optimizations, and brought an ease of use that proved helpful to electrical engineers working in the photonic telecommunications industry over the last decade. In Figure 2 is shown a z-transform-based block diagram for an exemplary three-stage photonic lattice filter. The interfaces are characterized by reflection and transmission coefficients, and the signal transit time between interfaces is described by a delay block. Considering the realization architecture discussed above, we follow the notation of [2]. For a lossless interface, the field reflection and transmission coefficients obey an energy conservation condition:

$$r_i^2 + t_i^2 = 1. \tag{1}$$

The transmission transfer function denominator polynomial for one stage is

$$1 + r_0 r_1 z^{-1}, \tag{2}$$

for two stages is

$$1 + (r_0 r_1 + r_1 r_2) z^{-1} + r_0 r_2 z^{-2}, \tag{3}$$

and for three stages is

$$1 + (r_0 r_1 + r_1 r_2 + r_2 r_3) z^{-1} + (r_0 r_2 + r_0 r_1 r_2 r_3 + r_1 r_3) z^{-2} + r_0 r_3 z^{-3}. \tag{4}$$

Simple induction can derive the polynomial for n stages.

In this paper we include gain as well as delay between the interfaces, and this block diagram for an active lattice filter is shown in Figure 3. Equations (2), (3), and (4) may be readily generalized to include this gain. For example, the transmission transfer function denominator polynomial for one stage with gains is

$$1 + r_0 r_1 G_1^2 z^{-1}, \tag{5}$$

for two stages with gains is

$$1 + (r_0 r_1 G_1^2 + r_1 r_2 G_2^2) z^{-1} + r_0 r_2 G_1^2 G_2^2 z^{-2}, \tag{6}$$

and for three stages with gains is

$$\begin{aligned} 1 &+ (r_0 r_1 G_1^2 + r_1 r_2 G_2^2 + r_2 r_3 G_3^2) z^{-1} \\ &+ (r_0 r_2 G_1^2 G_2^2 + r_0 r_1 r_2 r_3 G_1^2 G_3^2 + r_1 r_3 G_2^2 G_3^2) z^{-2} \\ &+ r_0 r_3 G_1^2 G_2^2 G_3^2 z^{-3}. \end{aligned} \tag{7}$$

Again, induction can be used to derive the polynomial for n stages. The gain for every stage appears to a power 2, regardless of the number of stages used.

Given a desired polynomial of degree n we can use the layer-peeling process [1, 2, 3] to compute $r_0, r_1, \ldots, r_n$ (actually r_0 can be chosen and the other reflection coefficients are computed). Fixing these reflection coefficients we can simply write the one-stage polynomial with gains as

$$1 + c_1 G_1^2 z^{-1}, \tag{8}$$

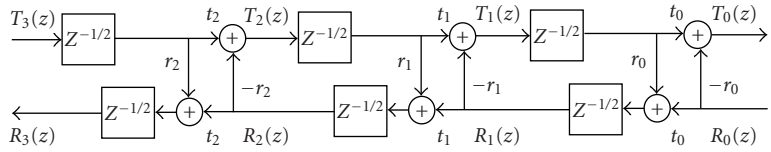


Figure 2: Signal flow diagram for three stages of an optical lattice filter.

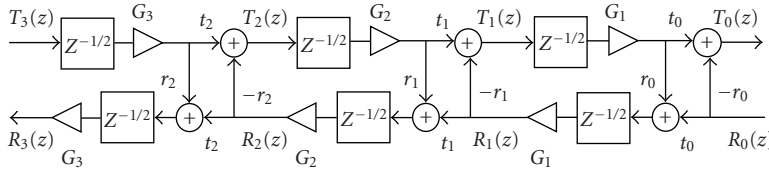


Figure 3: Signal flow diagram for three stages of an active optical lattice filter.

the two-stage polynomial with gains as

$$1 + (c_1 G_1^2 + c_2 G_2^2) z^{-1} + c_{12} G_1^2 G_2^2 z^{-2}, \tag{9}$$

and the three-stage polynomial with gains as

$$\begin{aligned} &1 + (c_1 G_1^2 + c_2 G_2^2 + c_3 G_3^2) z^{-1} \\ &\quad + (c_{12} G_1^2 G_2^2 + c_{13} G_1^2 G_3^2 + c_{23} G_2^2 G_3^2) z^{-2} + c_{123} G_1^2 G_2^2 G_3^2 z^{-3}. \end{aligned} \tag{10}$$

Here the c's are computed from the r's in the obvious way. The nth-degree polynomial with gains is

$$\begin{aligned} &1 + \left(\sum_{i=1}^{n} c_i G_i^2 \right) z^{-1} + \left(\sum_{i,j=1, i<j}^{n} c_{ij} G_i^2 G_j^2 \right) z^{-2} \\ &\quad + \cdots + c_{12\cdots n} G_1^2 G_2^2 \cdots G_n^2 z^{-n}. \end{aligned} \tag{11}$$

Again the c coefficients are known numbers computed from the r's, and the gains G_i are allowed to vary. Since we are interested in only nonnegative real gains G_i we set $x_i = G_i^2$ and replace (11) by

$$\begin{aligned} &1 + \left(\sum_{i=1}^{n} c_i x_i \right) z^{-1} + \left(\sum_{i,j=1, i<j}^{n} c_{ij} x_i x_j \right) z^{-2} \\ &\quad + \cdots + c_{12\cdots n} x_1 x_2 \cdots x_n z^{-n}. \end{aligned} \tag{12}$$

If we determine $x_1, x_2, \ldots, x_n$ then we know $G_1, G_2, \ldots, G_n$. We first consider a two-stage filter and return to the general case later.

4. TWO-STAGE EXAMPLE

In the case of a two-stage active lattice filter, (12) becomes

$$1 + (c_1 x_1 + c_2 x_1) z^{-1} + c_{12} x_1 x_2 z^{-2}, \tag{13}$$

where the c's are fixed and only the x_i can vary. If a desired denominator polynomial is

$$1 + a_1 z^{-1} + a_2 z^{-2}, \tag{14}$$

then the equations to determine the gains as function of the polynomial coefficients are

$$\begin{aligned} c_1 x_1 + c_2 x_1 &= a_1, \\ c_{12} x_1 x_2 &= a_2. \end{aligned} \tag{15}$$

Geometrically, these equations represent a straight line and a hyperbola in the closed first quadrant in (x_1, x_2) space. With c_1, c_2, c_{12} determined and a_1, a_2 given, a solution is a point (x_1, x_2) where the line and the hyperbola intersect. These equations can also be thought of as a map from the gains (actually gains squared) space (x_1, x_2) to the desired polynomial coefficient space(a_1, a_2).

For example, if we set $r_0 = 1$, $r_1 = 1/4$, and $r_2 = 1/4$, (3) for the transfer function denominator becomes

$$1 + \left(\frac{5}{16} \right) z^{-1} + \left(\frac{1}{4} \right) z^{-2}. \tag{16}$$

This discussion could have begun from the point of view of choosing coefficients $a_1 = 5/16$, $a_2 = 1/4$, and $r_0 = 1$, and using the Dowling-MacFarlane layer-peeling algorithm to find $r_1 = 1/4$ and $r_2 = 1/4$. We have stable systems in transmission and reflection since both poles are inside the unit circle. Equations (15) now becomes

$$\begin{aligned} \left(\frac{1}{4} \right) x_1 + \left(\frac{1}{16} \right) x_2 &= a_1, \\ \left(\frac{1}{4} \right) x_1 x_2 &= a_2. \end{aligned} \tag{17}$$

With $a_1 = 5/16$ and $a_2 = 1/4$ we have two solutions $(1, 1)$ and $(1/4, 4)$ for (x_1, x_2). The solution of interest to us

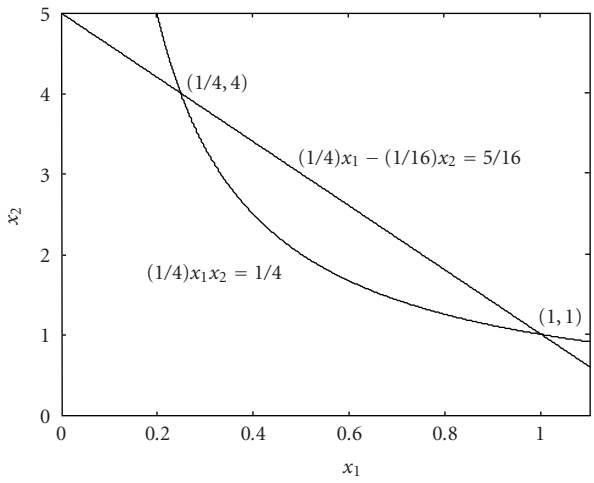


FIGURE 4: Intersection in (x_1, x_2) space for $a_1 = 5/16$ and $a_2 = 1/4$.

is $(1, 1)$. Considering (17) as a map from (x_1, x_2) space to (a_1, a_2) space, the Jacobian matrix is $\left[\begin{smallmatrix} 1/4 & 1/16 \\ (1/4)x_2 & (1/4)x_1 \end{smallmatrix}\right]$. We note that this matrix, evaluated at $(x_1, x_2) = (1, 1)$, is nonsingular. Thus by the inverse function theorem [9] there is an open set U containing $(1, 1)$ in (x_1, x_2) space and open set V containing $(5/16, 1/4)$ in (a_1, a_2) space. Hence for each coefficient pair (a_1, a_2) in V there is exactly one set of gains (squared) (x_1, x_2) in U that solves (17). Geometrically, for $a_1 = 5/16$ and $a_2 = 1/4$ the line and the hyperbola intersect transversally (nontangentially) at $(1, 1)$ in (x_1, x_2) space. If we change a_1 and a_2 slightly we still have a transverse intersection near $(1, 1)$. This intersection is shown graphically in Figure 4.

We now show that it is not possible to find the gains in (17) to do arbitrary pole placement for the same values of r_0, r_1, and r_2. If we let $a_1 = 5/16$ and $a_2 = 1$ in (17), then

$$\begin{aligned} \left(\frac{1}{4}\right)x_1 + \left(\frac{1}{16}\right)x_2 &= \frac{5}{16}, \\ \left(\frac{1}{4}\right)x_1x_2 &= 1 \end{aligned} \tag{18}$$

have only complex solutions for x_1 and x_2. Of course the straight line and hyperbola in (17) do not intersect in (x_1, x_2) space. Hence arbitrary denominator polynomials cannot be realized for the transmission and reflection transfer functions if the reflection coefficients are held fixed and the gains varied, for given order (number of lattice stages).

There is one more case of interest. If $a_1 = 1/2$ and $a_2 = 1/4$ in (17), then

$$\begin{aligned} \left(\frac{1}{4}\right)x_1 + \left(\frac{1}{16}\right)x_2 &= \frac{1}{2}, \\ \left(\frac{1}{4}\right)x_1x_2 &= 1 \end{aligned} \tag{19}$$

have only the solution $x_1 = 1$ and $x_2 = 4$. Hence the straight line and the hyperbola from (17) intersect at only one point, a point of tangency. Slight changes in a_1 from 1/2 can result in either no points of intersection or two points of intersection. It is very important to note that the Jacobian matrix $\left[\begin{smallmatrix} 1/4 & 1/16 \\ (1/4)x_2 & (1/4)x_1 \end{smallmatrix}\right]$ of (17) is singular at the point $(x_1, x_2) = (1, 4)$.

5. MAIN RESULTS

We return to the general case for n stages with gains in Section 3. Equation (12) of interest is

$$\begin{aligned} 1 + \left(\sum_{i=1}^{n} c_i x_i\right)z^{-1} + \left(\sum_{i,j=1, i<j}^{n} c_{ij}x_ix_j\right)z^{-2} \\ + \cdots + c_{12\ldots n}x_1x_2\cdots x_nz^{-n}. \end{aligned} \tag{20}$$

Here the c's (through the r's) have been chosen by using layer peeling to deliver a specified transmission transfer function denominator without using the gains (or equivalently, when all the $x_i = 1$). If another desired denominator polynomial to be achieved by computing only the gains is

$$1 + a_1z^{-1} + a_2z^{-2} + \cdots + a_nz^{-n}, \tag{21}$$

then the equations to determine the gains as functions of the polynomial coefficients are

$$\begin{aligned} \left(\sum_{i=1}^{n} c_i x_i\right) &= a_1, \\ \left(\sum_{i,j=1, i<j}^{n} c_{ij}x_ix_j\right) &= a_2, \\ &\vdots \\ c_{12\ldots n}x_1x_2\cdots x_n &= a_n. \end{aligned} \tag{22}$$

The n by n Jacobian matrix of the gains (or gains squared) to polynomial coefficient map is

$$\mathbf{J} = \begin{bmatrix} c_1 & c_2 & \cdots & c_n \\ \sum_{j=2}^{n} c_{1j}x_j & c_{12}x_1 + \sum_{j=3}^{n} c_{2j}x_j & \cdots & \sum_{i=1}^{n-1} c_{in}x_i \\ \vdots & \vdots & \ddots & \vdots \\ c_{12\ldots n}x_2x_3\cdots x_n & c_{12\ldots n}x_1x_3\cdots x_n & \cdots & c_{12\ldots n}x_1x_2\cdots x_{n-1} \end{bmatrix} \tag{23}$$

which is easily computed by taking partial derivatives in (22). When all gains (squared) x_i in (22) are unity, we denote the corresponding polynomial coefficients by $\hat{a}_1, \hat{a}_2, \ldots, \hat{a}_n$. By the inverse function theorem [9], these are the coefficients of the transfer function denominator polynomial for which the reflection coefficients are fixed.

If we move the a_i from the right-hand side to the left-hand side in (22), we have a system of equations of the form

$$\begin{aligned} f_1(x_1, x_2, \ldots, x_n) &= 0, \\ f_2(x_1, x_2, \ldots, x_n) &= 0, \\ &\vdots \\ f_n(x_1, x_2, \ldots, x_n) &= 0. \end{aligned} \tag{24}$$

Equations (24) may be solved by the Newton-Raphson algorithm to find the $(x_1, x_2, \ldots, x_n)$ as a function of the $(a_1, a_2, \ldots, a_n)$, so long as the Jacobian matrix of this system is nonsingular.

As an example, we now return to the two-stage example in Section 4 and apply the Newton-Raphson algorithm to compute the correct gains as we vary the desired polynomial coefficients. We use Matlab to implement Newton-Raphson [13]. Equations (17) of interest are

$$\begin{aligned} \left(\frac{1}{4}\right)x_1 + \left(\frac{1}{16}\right)x_2 &= a_1, \\ \left(\frac{1}{4}\right)x_1 x_2 &= a_2. \end{aligned} \tag{25}$$

The starting values are $a_1 = 5/16$ and $a_2 = 1/4$ for which the solutions are $x_1 = 1$ and $x_2 = 1$. Of course this implies that $G_1 = 1$ and $G_2 = 1$. To make the changes in a_1 and a_2 somewhat random we choose

$$\begin{aligned} a_1 &= \frac{5}{16} \pm 0.05 \times (\text{rand} - 0.5), \\ a_2 &= \frac{1}{4} \pm 0.05 \times (\text{rand} - 0.5), \end{aligned} \tag{26}$$

where rand is a Matlab command for a uniformly distributed random variable with values between 0 and 1. Then we solve for x_1, x_2, G_1, and G_2. Our computations yield the results in Table 1.

The Newton-Raphson algorithm actually provides two solutions (x_1, x_2) for each a_1, a_2, but we take only that solution which is on the same side of the straight line $x_2 = 4x_1$ as the point $(1, 1)$ (recall that we work in an open neighborhood of $(1, 1)$). The line $x_2 = 4x_1$ represents the set of points where the Jacobin matrix $\begin{bmatrix} 1/4 & 1/16 \\ (1/4)x_2 & (1/4)x_1 \end{bmatrix}$ is singular.

These results follow from the point of view that given the desired second-degree denominator polynomial, we program the gains to achieve that polynomial. We next merge our gain computation procedure with an adaptive process so that we start with data, identify the AR transfer function, and compute the appropriate gains.

Table 1: Filter coefficients and gains.

a_1	a_2	x_1	x_2	G_1	G_2
5/16	1/4	1	1	1	1
0.3350	0.2366	1.13	0.83	1.063	0.910
0.31780	0.2493	1.028	0.969	1.013	0.984
0.3321	0.2631	1.086	0.968	1.042	0.983
0.3103	0.2259	1.019	0.886	1.009	0.9412
0.3286	0.2472	1.08	0.909	1.039	0.953
0.3067	0.2646	0.947	1.116	0.973	1.056
0.2914	0.2619	0.861	1.215	0.9279	1.1022
0.3172	0.2338	1.045	0.894	1.022	0.945
0.2907	0.2708	0.840	1.28	0.916	1.131
0.3170	0.2697	0.997	1.081	0.998	1.039
0.3346	0.2574	1.10	0.931	1.048	0.964
0.2968	0.2245	0.872	1.25	0.933	1.118
0.3306	0.2649	1.076	0.984	1.037	0.991
0.3276	0.2448	1.084	0.902	1.041	0.949
0.3239	0.2651	1.04	1.018	1.019	1.008
0.3152	0.2623	0.998	1.051	0.998	1.025
0.2909	0.2483	0.882	1.125	0.939	1.06
0.3166	0.2673	0.998	1.07	0.998	1.034
0.3112	0.2351	1.01	0.928	1.004	0.963

We generate data according to the linear time-varying difference equation

$$\begin{aligned} y(n) + \left(\frac{13}{24}\right)y(n-1) + \left(\frac{5}{8}\right)y(n-2) \\ + \left[\frac{8}{27} + \frac{1}{27}\cos(n-3)\right]y(n-3) = u(n) \end{aligned} \tag{27}$$

with zero initial conditions $y(-1) = y(-2) = y(-3) = 0$ and input $u(n) = \delta(n)$, the discrete delta function. We assume that the data is associated with the time-invariant linear system

$$y(n) + a_1 y(n-1) + a_2 y(n-2) + a_3 y(n-3) = u(n) \tag{28}$$

and we identify a_1, a_2, a_3 from the data as n moves. These a_1, a_2, a_3 are the coefficients of the desired AR denominator polynomial $1 + a_1 z^{-1} + a_2 z^{-2} + a_3 z^{-3}$. Then we compute the x_1, x_2, x_3 (and hence the gains G_1, G_2, G_3) from the equations

$$\begin{aligned} c_1 x_1 + c_2 x_2 + c_3 x_3 &= a_1, \\ c_{12} x_1 x_2 + c_{13} x_1 x_3 + c_{23} x_2 x_3 &= a_2, \\ c_{123} x_1 x_2 x_3 &= a_3. \end{aligned} \tag{29}$$

For the starting point we take the difference equation

$$y(n) + \left(\frac{13}{24}\right)y(n-1) + \left(\frac{5}{8}\right)y(n-2) + \left(\frac{1}{3}\right)y(n-3) = u(n) \tag{30}$$

and thus the polynomial $1 + (13/24)z^{-1} + (5/8)z^{-2} + (1/3)z^{-3}$ results. This difference equation can be found by substituting $u(n) = \delta(n)$ into (27) with zero initial conditions, evaluating $y(0)$, $y(1)$, $y(2)$, $y(3)$, and computing a_1, a_2, a_3 from (28) using 3 linear equations (for $y(n) = y(1), y(2), y(3)$) in 3 unknowns. Choosing $r_0 = 1$ we compute the reflection coefficients $r_1 = 1/4$, $r_2 = 1/2$, $r_3 = 1/3$ by layer peeling. Thus (29) becomes (see (7) and (10))

$$\begin{aligned} \left(\frac{1}{4}\right)x_1 + \left(\frac{1}{8}\right)x_2 + \left(\frac{1}{6}\right)x_3 &= a_1, \\ \left(\frac{1}{2}\right)x_1x_2 + \left(\frac{1}{24}\right)x_1x_3 + \left(\frac{1}{12}\right)x_2x_3 &= a_2, \\ \left(\frac{1}{3}\right)x_1x_2x_3 &= a_3. \end{aligned} \tag{31}$$

The Jacobian matrix is

$$\begin{bmatrix} \left(\frac{1}{4}\right) & \left(\frac{1}{8}\right) & \left(\frac{1}{6}\right) \\ \left(\frac{1}{2}\right)x_2 + \left(\frac{1}{24}\right)x_3 & \left(\frac{1}{2}\right)x_1 + \left(\frac{1}{12}\right)x_3 & \left(\frac{1}{24}\right)x_1 + \left(\frac{1}{12}\right)x_2 \\ \left(\frac{1}{3}\right)x_2x_3 & \left(\frac{1}{3}\right)x_1x_3 & \left(\frac{1}{3}\right)x_1x_2 \end{bmatrix}. \tag{32}$$

As discussed above, this matrix is nonsingular when x_1, x_2, x_3 are all unity. We then substitute in the varying values for a_1, a_2, a_3 and compute the corresponding x_1, x_2, x_3 (from (31)) and G_1, G_2, G_3. The first computation is for $a_1 = 13/24$, $a_2 = 5/8$, and $a_3 = 1/3$, and the solution is $x_1 = 1$, $x_2 = 1$, and $x_3 = 1$.

The second computation involves a_1, a_2, a_3 calculated by substituting $u(n) = \delta(n)$ into (27) with zero initial conditions, evaluating $y(0)$, $y(1)$, $y(2)$, $y(3)$, $y(4)$, and computing a_1, a_2, a_3 from (28) using 3 linear equations (for $y(n) = y(2), y(3), y(4)$) in 3 unknowns. The third and higher computations proceed in the obvious way. We record our results in Table 2

For this example we could have driven the system by noise, computed the autocorrelations, arrived at the polynomial coefficients through standard algorithms, and then computed the gains using our technique. Moreover, we also could have estimated the polynomial coefficients using a recursive least squares method based on the Kalman filter. However, we decided to keep things simple and work straight with the data from the difference equation, using sets of 3 equations in 3 unknowns. With noisy data we also could have used more linear equations in the 3 unknowns and found pseudoinverse solutions.

6. PARAMETER SENSITIVITY ANALYSIS

The goal is to quantify the variation of denominator coefficients as the gains vary. Since the gains are physically realized their numerical value is prone to change. This might result in a significantly different transfer function. Hence the perturbation of the coefficients a_i's due to error in the realization of the gains G_j's is worth analyzing. We quantify the perturbation relations by a measurable parameter called sensitivity. For ease of notation we analyze the sensitivity of the coefficients a_i with respect to the gains squared x_j.

The traditional definition of sensitivity of a parameter a with respect to a parameter x is given by

$$S = \frac{da/a}{dx/x}. \tag{33}$$

This can be viewed as the relative change of a with respect to change in x. Low values of S means the coefficients are somewhat insensitive to gain changes. The next step is to see how S fits into the nonlinear gains squared-coefficients relations.

The general gain squared-coefficients relation according to (22) can be expressed in the form

$$\begin{aligned} f_1(x_1, x_2, \ldots, x_n) &= a_1, \\ f_2(x_1, x_2, \ldots, x_n) &= a_2, \\ &\vdots \\ f_n(x_1, x_2, \ldots, x_n) &= a_n. \end{aligned} \tag{34}$$

The above set of equations can be represented in a vector form as

$$f(\mathbf{x}) = \mathbf{a}, \tag{35}$$

where $\mathbf{x} = [x_1, \ldots, x_n]^T$ and $\mathbf{a} = [a_1, \ldots, a_n]^T$ and T is the matrix-transpose operator.

Table 2: Filter coefficients and gains.

a_1	a_2	a_3	x_1	x_2	x_3	G_1	G_2	G_3
13/24	5/8	1/3	1	1	1	1	1	1
13/24	5/8	0.3163	1.06948	0.95293	0.93108	1.0342	0.9762	0.9649
13/24	5/8	0.2809	1.19	0.883	0.8006	1.0909	0.9397	0.8948
13/24	5/8	0.2596	1.255	0.852	0.7275	1.1203	0.9230	0.8529
13/24	5/8	0.2721	1.2182	0.8702	0.7699	1.1037	0.9328	0.8774
13/24	5/8	0.3068	1.1045	0.93142	0.89466	1.0509	0.9651	0.9459
13/24	5/8	0.3319	1.00627	0.9954	0.9938	1.0031	0.9977	0.9969
13/24	5/8	0.3242	1.03851	0.9731	0.96238	1.0191	0.9865	0.981
13/24	5/8	0.2909	1.1589	0.90059	0.83612	1.0765	0.94899	0.91439
13/24	5/8	0.2626	1.24651	0.85683	0.73761	1.1130	0.9276	0.8639
13/24	5/8	0.2652	1.23887	0.86039	0.74641	1.0678	0.9544	0.9254
13/24	5/8	0.2965	1.0.1487	0.98940	0.98565	1.0074	0.9946	0.9928
13/24	5/8	0.3299	1.12353	0.92031	0.87447	1.0599	0.9607	0.93513
13/24	5/8	0.3014	1.22996	0.8646	0.75661	1.1090	0.9298	0.8698
13/24	5/8	0.2682	1.25177	0.85441	0.73155	1.1188	0.9243	0.8553
13/24	5/8	0.2608	1.17455	0.89230	0.81895	1.0838	0.9446	0.9049
13/24	5/8	0.2861	1.05207	0.96415	0.94878	1.0257	0.9819	0.9924
13/24	5/8	0.3329	1.08782	0.94150	0.91214	1.0429	0.9703	0.955
13/24	5/8	0.3114	1.20631	0.8761	0.78347	1.0983	0.9360	0.8851
13/24	5/8	0.2760	1.25612	0.85242	0.72651	1.1207	0.9233	0.8523

Writing (35) in differential form we get

$$\mathbf{J}\mathbf{d}_{\mathbf{x}} = \mathbf{d}_{\mathbf{a}}, \tag{36}$$

where $\mathbf{d}_{\mathbf{x}} = [dx_1, \ldots, dx_n]^T$, $\mathbf{d}_{\mathbf{a}} = [da_1, \ldots, da_n]^T$, and

$$\mathbf{J} = \begin{bmatrix} \dfrac{\partial f_1(\mathbf{x})}{\partial x_1} & \dfrac{\partial f_1(\mathbf{x})}{\partial x_2} & \cdots & \dfrac{\partial f_1(\mathbf{x})}{\partial x_n} \\ \dfrac{\partial f_2(\mathbf{x})}{\partial x_1} & \dfrac{\partial f_2(\mathbf{x})}{\partial x_2} & \cdots & \dfrac{\partial f_2(\mathbf{x})}{\partial x_n} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \dfrac{\partial f_n(\mathbf{x})}{\partial x_1} & \dfrac{\partial f_n(\mathbf{x})}{\partial x_2} & \cdots & \dfrac{\partial f_n(\mathbf{x})}{\partial x_n} \end{bmatrix}, \tag{37}$$

where $\mathbf{J}$ is the Jacobian matrix. The entries of the matrix $\mathbf{J}$ are $\mathbf{J}_{i,j}$. It is easy to see from (36) that if only x_j changes, then

$$\mathbf{J}_{i,j}\mathbf{d}_{\mathbf{x}_j} = \mathbf{d}_{\mathbf{a}_i}. \tag{38}$$

Plugging (38) into (33) and a little manipulation gives

$$\mathbf{S}_{i,j} = \frac{\mathbf{J}_{i,j}\mathbf{x}_j}{\mathbf{a}_i}, \tag{39}$$

where $\mathbf{S}_{i,j}$ relates the sensitivity of $\mathbf{a}_i$ to changes in $\mathbf{x}_j$. Thus we can find a sensitivity matrix $\mathbf{S}$ with $\mathbf{S}_{i,j}$ as its elements. $\mathbf{S}$ gives the one-to-one sensitivity relations between the square of the gains and the coefficients.

An upper bound on the sensitivity matrix, when one or more gains vary, is estimated using the 2-norm as follows. Equation (36) can be written as

$$\mathbf{J}\mathbf{d}_{\mathbf{x}} = \mathbf{d}_{\mathbf{a}}. \tag{40}$$

Applying norm to both sides of the equation we get

$$||\mathbf{J}\mathbf{d}_{\mathbf{x}}|| = ||\mathbf{d}_{\mathbf{a}}|| \Longrightarrow ||\mathbf{d}_{\mathbf{a}}|| \leq \|\mathbf{J}\|\|\mathbf{d}_{\mathbf{x}}||. \tag{41}$$

Here $\|\cdot\|$ denotes $\sqrt{\max(\text{eigenvalues of } \mathbf{J}^T\mathbf{J})}$ for matrices and 2-norm for vectors. Thus we get

$$\frac{||\mathbf{d}_{\mathbf{a}}||}{||\mathbf{d}_{\mathbf{x}}||} \leq \|\mathbf{J}\| \tag{42}$$

or equivalently,

$$\frac{||\mathbf{d}_{\mathbf{a}}||/\|\mathbf{a}\|}{||\mathbf{d}_{\mathbf{x}}||/\|\mathbf{x}\|} = s \leq \frac{\|\mathbf{J}\|\|\mathbf{x}\|}{\|\mathbf{a}\|}. \tag{43}$$

Here s is sensitivity with respect to gain squared. Thus (43) gives an upper bound on sensitivity. The sensitivity bound (s) is tabulated in Table 3 for different gains squared.

Table 3: Sensitivity of filter parameters respective to gains.

a_1	a_2	a_3	x_1	x_2	x_3	s
13/24	5/8	1/3	1	1	1	1.9702
13/24	5/8	0.3163	1.06948	0.95293	0.93108	1.9359
13/24	5/8	0.2809	1.19	0.883	0.8006	1.9109
13/24	5/8	0.2596	1.255	0.852	0.7275	1.9139
13/24	5/8	0.2721	1.2182	0.8702	0.7699	1.9109
13/24	5/8	0.3068	1.1045	0.93142	0.89466	1.9243
13/24	5/8	0.3319	1.00627	0.9954	0.9938	1.9664
13/24	5/8	0.3242	1.03851	0.9731	0.96238	1.9492
13/24	5/8	0.2909	1.1589	0.90059	0.83612	1.9135
13/24	5/8	0.2626	1.24651	0.85683	0.73761	1.9129
13/24	5/8	0.2652	1.23887	0.86039	0.74641	1.9121
13/24	5/8	0.2965	1.0.1487	0.9894	0.98565	1.9162
13/24	5/8	0.3299	1.12353	0.92031	0.87447	1.9615
13/24	5/8	0.3014	1.22996	0.8646	0.75661	1.9196
13/24	5/8	0.2682	1.25177	0.85441	0.73155	1.9115
13/24	5/8	0.2608	1.17455	0.8923	0.81895	1.9135
13/24	5/8	0.2861	1.05207	0.96415	0.94878	1.9119
13/24	5/8	0.3329	1.08782	0.9415	0.91214	1.969
13/24	5/8	0.3114	1.20631	0.8761	0.78347	1.9294
13/24	5/8	0.276	1.25612	0.85242	0.72651	1.9106

The results in Table 3 show that the active lattice is reasonably robust with respect to gain tolerances. That is to say that the filter response does not appreciably change with small injection current uncertainties. This advantageous result is somewhat expected given the traditional lattice filter's robustness with respect to round off error or the reflection coefficient variations [2, 4, 13].

7. CONCLUSIONS AND FUTURE RESEARCH

We have added gains to the optical lattice filters [1, 2, 3] in order to make the response functions programmable. To determine the denominator polynomials achievable by gains only we set up a map from the gains (squared) to the polynomial coefficients. Applying the inverse function theorem we provide a result giving sufficient conditions that all polynomials in an open set can be generated by appropriated choices of gains. This result naturally leads to a method for computing the gains using the Newton-Raphson algorithm. Two interesting examples are given, one of which stressed the adaptive signal processing point of view. We also performed a sensitivity analysis, measuring how the denominator coefficients vary as the gains vary.

We have shown that the presence of gains in the lattice filter can provide additional flexibility in the filter response. These gains may be implemented in a semiconductor laser amplifier, and hence these will be adjustable at GHz rates. In accompanying experimental work, we will be building and testing these devices, and exploring their range of operation both theoretically and empirically. This range will be bounded on one end by the noise figure and stability on the other. In addition, theoretical and experimental work is also underway in two-dimensional active lattice filters based on four-directional couplers [14].

REFERENCES

[1] D. L. MacFarlane and E. M. Dowling, "Z-domain techniques in the analysis of Fabry-Perot etalons and multilayer structures," *Journal of the Optical Society of America* {*A*}, vol. 11, no. 1, pp. 236–245, 1994.

[2] E. M. Dowling and D. L. MacFarlane, "Lightwave lattice filters for optically multiplexed communication systems," *J. Lightwave Technol.*, vol. 12, no. 3, pp. 471–486, 1994.

[3] D. L. MacFarlane, E. M. Dowling, and V. Narayan, "Ring resonators with $N \times M$ couplers," *Fiber and Integrated Optics*, vol. 14, no. 3, pp. 195–210, 1995.

[4] S. Haykin, *Adaptive Filter Theory*, Prentice Hall, Englewood Cliffs, NJ, USA, 4th edition, 2002.

[5] P. A. Regalia, *Adaptive IIR Filtering in Signal Processing and Control*, Marcel Dekker, New York, NY, USA, 1995.

[6] M. J. Maron and R. J. Lopez, *Numerical Analysis: A Practical Approach*, Wadsworth, Belmont, Calif, USA, 3rd edition, 1991.

[7] J. E. Dennis Jr. and R. B. Schnabel, *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, Prentice Hall, Englewood Cliffs, NJ, USA, 1983.

[8] G. Evans, N. W. Carlson, J. M. Hammer, and J. K. Butler, "Grating surface emitting lasers," in *Chapter 4 of Surface Emitting Semiconductor Lasers and Arrays*, G. A. Evans and J. M. Hammer, Eds., Academic Press, New York, NY, USA, 1993.

[9] W. Rudin, *Principles of Mathematical Analysis*, McGraw-Hill, New York, NY, USA, 2nd edition, 1964.

[10] B. Moslehi, J. W. Goodman, M. Tur, and H. J. Shaw, "Fiber-optic lattice signal processing," *Proc. IEEE*, vol. 72, no. 7, pp. 909–930, 1984.

[11] G. A. Ball, W. W. Morey, and P. K. Cheo, "Single- and multipoint fiber-laser sensors," *IEEE Photon. Technol. Lett.*, vol. 5, no. 2, pp. 267–270, 1993.

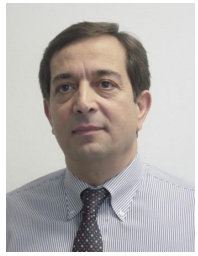
[12] H. A. Zarem, M. E. Hoenk, W. B. Bridges, K. Vahala, and A. Yariv, "Generation of 1180 Å period gratings with a Xe ion laser," *Electronics Letters*, vol. 24, no. 22, pp. 1366–1367, 1988.

[13] G. Recktenwald, *Numerical Methods with MATLAB*, Prentice Hall, Englewood Cliffs, NJ, USA, 2000.

[14] D. L. MacFarlane, J. Tong, C. Fafadia, V. Govindan, L. R. Hunt, and I. Panahi, "Extended lattice filters enabled by four-directional couplers," *Applied Optics*, vol. 43, no. 33, pp. 6124–6133, 2004.

L. Roberts Hunt received his B.S. degree from Baylor University in 1964 and his Ph.D. degree from Rice University in 1970, both in mathematics. He was a faculty member in the Department of Mathematics at Texas Tech University from 1969 to 1984, taking a two-year leave at the NASA Ames Research Center from 1980 to 1982. Since 1984 he has been a faculty member at the University of Texas at Dallas, where he is currently a Professor of electrical engineering. He is a Fellow of the IEEE and a licensed Professional Engineer in the State of Texas. His research interests are in nonlinear systems and control and signal processing. A previous application area was automatic flight control with the NASA Ames Research Center from 1980 to 2001. Present research projects are design of active optical lattice filters and design of power converters.

Vishnupriya Govindan received the B.E. degree in electronics and communication engineering from the University of Madras, India, in 2002 and the M.S.E.E. degree from the University of Texas at Dallas in 2004. She is currently working towards her Ph.D. degree from the University of Utah, Salt Lake City.



Issa Panahi received the Ph.D. degree in electrical engineering from the University of Colorado at Boulder. Dr. Panahi is currently an Assistant Professor of electrical engineering at the University of Texas at Dallas (UTD) and Director of the Statistical Signal Processing Research Laboratory. His main research interests and experience are in the areas of digital signal processing, adaptive filtering, system identification, detection and estimation, active noise cancellation, and embedded digital systems and processors. Dr. Panahi worked as a Research Scientist on the multidimensional and statistical signal processing at Bellaire Research Center, Shell Oil Company, Houston, for over three years before joining the DSP Division of Texas Instruments (TI) in Houston, Texas. Dr. Panahi worked at TI for over ten years as the DSP Chief Architect for TMS320C/F24xx and 28xx family of devices, Chief Technology Officer, Senior Member of the Technical Staff, and the Worldwide Applications Manager of the DSP-embedded control systems business unit. He was an Applications Manager with the Wireless OMAP Business Unit at TI before Joining UTD in 2001. Dr. Panahi holds patents and has authored/coauthored 3 DSP-based TI's books and over 35 technical papers and articles.

Jian Tong was born in China in 1976. He received the B.S. degree in electrical engineering from Shanghai Jiao Tong University, China, in 1998. From 1998 to 2001, he worked as a Technical Engineer at Lucent Technologies of Shanghai, China. Since 2001, he has been a graduate student in the University of Texas at Dallas (UTD) where he received the M.S.E.E. degree in 2003. He currently is a Ph.D. candidate and works as a Research Assistant at UTD.

Govind Kannan was born in Chennai, India. He received the B.E. degree in electronics and communication engineering from the University of Madras, India, in 2003.He is currently working towards his M.S. degree in electrical engineering at the University of Texas, Dallas, Tex. His primary research interests are in digital signal processing and filter design.

Duncan L. MacFarlane is a Full Professor of electrical engineering at the University of Texas at Dallas. Dr. MacFarlane has written more than 100 technical papers or patents in the general area of photonic systems and components. Specific research projects have included micro-optics, ultrafast lasers, photonic integrated filters, nonlinear optics, semiconductor lasers, ellipsometry, and advanced displays. Dr. MacFarlane received his B.S.E.E. and M.S.E.E. degrees from Brown University, his Ph.D. degree from Portland State University, and his MBA from SMU. He has worked at Schafer Associates, Texas Instruments, and at JDS Uniphase, and helped start Celion Networks, a telecommunication system start-up backed by Sequoia. He teaches graduate and undergraduate courses in electromagnetics, microwave engineering, optics, probability and statistics, and electronics, and is an Associate Dean for interdisciplinary programs. Dr. MacFarlane is a registered professional engineer in the State of Texas.

Gary Evans was born in Omak, Wash, and received the B.S.E.E. degree from the University of Washington, Seattle, in 1970 and the M.S.E.E. and Ph.D. degrees in electrical engineering and physics from the California Institute of Technology, Pasadena, in 1971 and 1975. After a postdoctoral year at Caltech, he worked for R&D Associates, Marina Del Rey, Calif, and was a Visiting Assistant Professor in the Electrical Engineering Department at the University of Washington (1977–1979). He worked at the Aerospace Corporation, El Segundo, Calif (1979–1981), TRW, Redondo Beach, Calif (1981–1984), and RCA Laboratories (now Sarnoff Corporation), Princeton, NJ (1984–1992). In 1992, he joined Southern Methodist University, Dallas, Tex, as a Professor in the Electrical Engineering Department. He is a founder and member of the board of directors of Photodigm. His work on the design, growth, and fabrication of semiconductor lasers and photonic devices is described in more than 240 publications, 20 patents, and the book *Surface Emitting Semiconductor Lasers* Academic Press, New York, 1993. He is a licensed professional engineer and Fellow of the IEEE.

EURASIP Journal on Applied Signal Processing 2005:10, 1462–1484

Advanced Optical Processing of Microwave Signals

Beatriz Ortega
Grupo de Comunicaciones Opticas, ITEAM, Universidad Politécnica de Valencia, Camino de Vera, s/n, 46022 Valencia, Spain
Email: bortega@dcom.upv.es

Daniel Pastor
Grupo de Comunicaciones Opticas, ITEAM, Universidad Politécnica de Valencia, Camino de Vera, s/n, 46022 Valencia, Spain
Email: dpastor@dcom.upv.es

José Mora
Grupo de Comunicaciones Opticas, ITEAM, Universidad Politécnica de Valencia, Camino de Vera, s/n, 46022 Valencia, Spain
Email: jmalmer@iteam.upv.es

José Capmany
Grupo de Comunicaciones Opticas, ITEAM, Universidad Politécnica de Valencia, Camino de Vera, s/n, 46022 Valencia, Spain
Email: jcapmany@dcom.upv.es

Miguel V. Andrés
Grupo de Semiconductores y Fibras Ópticas, ICMUV, Universidad de Valencia, Dr. Moliner 50, 46100 Burjassot (Valencia), Spain
Email: miguel.andres@uv.es

Received 16 April 2004; Revised 20 September 2004

The authors present a review on the recent approaches proposed to implement transversal RF filters. Different tunable transversal filters consisting of wavelength tunable optical taps and those employing the tunability of dispersive devices are presented showing their high-performance characteristics. A comprehensive review of the fundamentals and a discussion on the main limitation of these structures are also included.

Keywords and phrases: microwave photonics, transversal filters, optical fibers, delay lines.

1. INTRODUCTION

Over the last 25 years microwave photonics has been a discipline under constant active research because of the unique properties that photonic devices and systems bring to the generation, transport, processing, and detection of microwave and millimeter wave signals [1].

The possibility of using photonic devices to implement flexible filters for microwave and radiofrequency (RF) signals with larger bandwidth is one of the applications that first attracted the interest of the researchers [2, 3, 4, 5, 6] since traditional approaches for RF signal processing are fraught with the *electronic bottleneck* [2] and other sources of degradation as electromagnetic interference (EMI) and frequency dependent losses.

An interesting approach to overcome the above limitations involves the use of photonics technology and especially fiber and integrated photonic devices and circuits to perform the required signal processing tasks of RF signals conveyed by an optical carrier directly in the optical domain. The discrete-time optical processing of microwave signals (DOPMS) approach is shown in Figure 1. The RF-to-optical conversion is achieved by direct (or externally) modulating a laser. The RF signal is conveyed by an optical carrier and the composite signal is fed to a photonic circuit that samples the signal in the time domain, weights the samples, and combines them using optical delay lines and other photonic elements. At the output(s) the resulting signal(s) are optical-to-RF converted by means of an/various optical receiver(s).

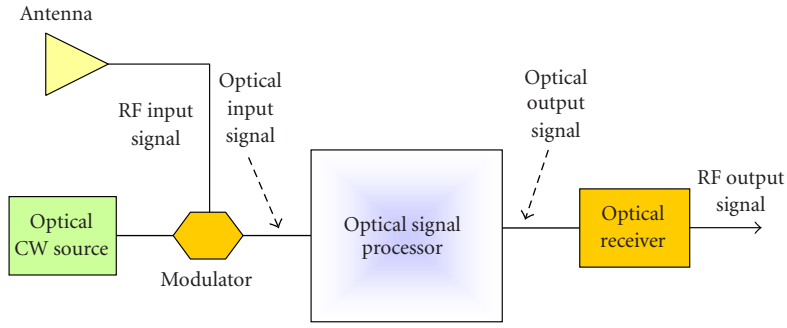


Figure 1: Discrete-time optical processing approach.

The DOPMS approach is of interest, for example, in radio-over-fiber systems, both for channel rejection and channel selection applications [7]. Another application example is for noise suppression and channel interference mitigation in the front-end stage after the receiving antenna of a UMTS base station prior to a highly selective SAW filter [8, 9]. Photonic filters for RF signals can also be of interest for applications where lightweight is a prime concern, for example, analog notch filters are also needed to achieve cochannel interference suppression in digital satellite communications systems [71]. Moreover, in moving target identification (MTI) radar systems [11] the filtering of clutter and noise (the unwanted signals) is performed using a digital notch filter placed after frequency downconversion to baseband and analog-to-digital (ADC) conversion.

Research contributions within this area extend over the last 25 years starting with the seminal paper of Wilner and Van den Heuvel [3] who noted that the low-loss and high-modulation bandwidth of optical fibers made then an ideal candidate as a broadband delay line. Several contributions during the 70s addressed experimental work on DOPMS using multimode fibers [4, 5]. An intensive theoretical and experimental research work on incoherent DOPMS using single-mode fiber delay lines was carried by researchers at the University of Stanford during the period between 1980 and 1990. Multiple configurations, applications, and potential limitations of these structures were considered and the main results can be found summarized in [2, 6]. However, the DOPMS demonstrated serious limitations arising from losses and lack of reconfiguration since the technology status regarding optical fiber and integrated components was at the time at its infancy. The advent of the optical amplifier at the end of the 80s and the development of optical components (variable couplers, modulators, electro-optic switches) and specific purpose instrumentation fuelled the activity towards more flexible structures employing these components [12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33]. Nevertheless, the availability of novel components, such as the fiber Bragg grating (FBG) and the arrayed waveguide grating (AWG) has opened a new perspective towards the implementation of fully reconfigurable and tunable DOPMS [34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62].

In this paper, we present the recent advances in photonic processing of radiofrequency signals, focusing on different alternatives for implementing transversal filters. The outline of this paper is as follows. In Section 2 we describe the fundamental concepts and limitations related to the photonic RF processors, paying special attention to incoherent signal processing techniques. Section 3 reviews the recent advances in the implementation of DOPMS from a structural point of view. We discuss different alternatives for two main consolidated technical approaches towards the practical implementation of transversal filters. The first one corresponds to tunable transversal filters consisting of wavelength tunable optical taps, whereas the second one employs the tunability of dispersive devices. In the former one, we include the efforts made so far to overcome the limitations in incoherent filters imposed by the positive nature of their coefficients. Finally, a summary and conclusions are presented in Section 4.

2. FUNDAMENTAL CONCEPTS AND LIMITATIONS

2.1. Fundamental concepts

Any filter implemented using DOPMS tries to provide a system function for the RF signal given by [63]

$$H(z^{-1}) = \frac{\sum_{r=0}^{N} a_r z^{-r}}{1 - \sum_{k=1}^{M} b_k z^{-k}}, \tag{1}$$

where z^{-1} represents the basic delay between samples and a_r, b_k the filter coefficients which are implemented by optical components. The numerator represents the finite impulse part (i.e., nonrecursive or FIR) of the system function, whereas the denominator accounts for the infinite impulse part (i.e., recursive or FIR) of the system function. N and M stand for the order of the FIR and IIR parts, respectively. If $b_k = 0$ for all k, filter is nonrecursive and is also known as transversal filter. Otherwise the filter is recursive and it is common to use the term recirculating delay line. Figure 2 shows how (1) is implemented for an N-tap transversal incoherent filter using a single optical source. The optical carrier is amplitude-modulated by an RF signal. Once the optical-modulated signal is tapped, each tap is differently delayed and weighted. After combining the samples, the receiver gets

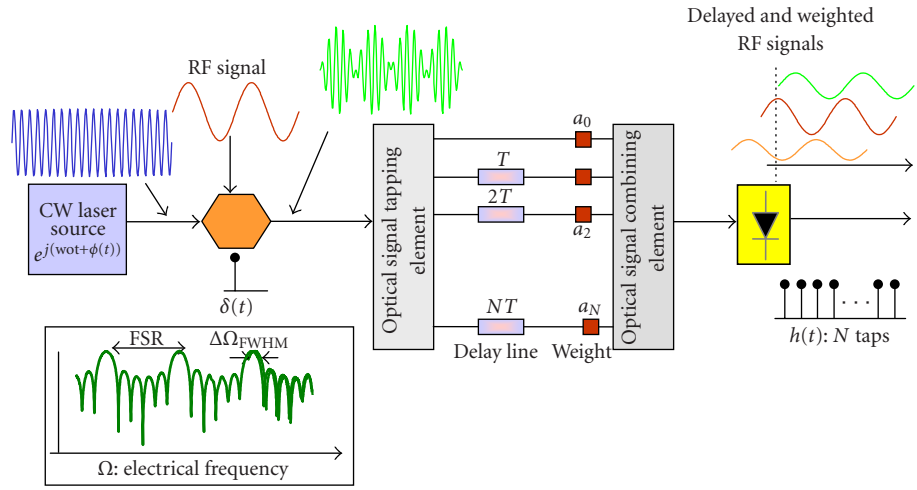


FIGURE 2: Implementation layout of an N-tap incoherent transversal filter using a single optical source.

the sum of differently delayed samples in order to give the system response shown in the inset of Figure 2. The end-to-end (electrical) impulse response corresponding to this situation can be directly derived from (1) yielding

$$h(t) = \sum_{r=0}^{N} a_r \delta(t - rT), \tag{2}$$

which convolved with the input RF signal $s_i(t)$ yielding the following output signal $s_o(t)$:

$$s_o(t) = \sum_{r=0}^{N} a_r s_i(t - rT). \tag{3}$$

Similar expressions to (2) and (3) apply for IIR or recirculating delay line filters with $N \to \infty$.

According to (2) and (3) the standard electrical/optical impulsive response of the optical processor is represented by an equally time-spaced (T) pulse train where pulses implement the filter taps. If all the samples have the same amplitude, the filter is called uniform; if the samples have different amplitudes, the filter is termed as apodized or windowed [63]. The electrical frequency response $H(\Omega)$ of such a structure can be obtained by Fourier transformation of the impulsive response (2):

$$H(\Omega) = \sum_{r=0}^{N} a_r e^{-jr\Omega T}. \tag{4}$$

The above expression identifies a transfer function with a periodic spectral characteristic (see inset of Figure 2). The frequency period is known as the filter *free spectral range* or FSR which is inversely proportional to the time spacing T between adjacent samples in the impulse response. The resonance full-width half-maximum is denoted as $\Delta\Omega_{FWHM}$. The filter selectivity is given by its *quality or* Q factor which is given by

$$Q = \frac{FSR}{\Delta\Omega_{FWHM}}. \tag{5}$$

The value of the Q factor is related to the number of samples (taps) used to implement it. If the number of taps is high (greater than 10), Q factor can be approximated for uniform filters by the number of taps $Q \cong N$. This relation can be slightly corrected ($Q < N$) for apodized filters [63].

As shown in Figure 2, the implementation of the DOPMS requires specific optical components to provide (a) signal tapping, (b) optical delay lines, (c) optical weights, and (d) optical signal combination, as detailed in Table 1.

Some further definitions are now introduced to complete our general description of these filters.

Coherent and incoherent operation regime

The multiple optical contributions carrying the RF signal are mixed at the detector end of the structure under two different regimes. The former regime is verified when the light arising from each tap of the filter has a deterministic optical phase relationship with the rest at the input of the photodetector. In this case, the optical power-to-electric current conversion operation at the photodetector generates an interference term. This situation can only take place when a single optical source is employed and its "coherence time" τ_c is much longer than the time delay T between adjacent samples or taps of the optical. Under coherent regime operation, the optical phase of the taps plays a predominant role in the overall time and frequency response of the processor, and filters with negative and complex coefficients can be implemented. On the other hand, since the filter operation relies on optical interferences, any slight change in the propagation characteristics of any part of the optical processor drastically affects the filter response and its properties.

Table 1: Components needed in a DOPMS.

Function	Components
Signal tapping	2×2 and $1 \times N$, $N \times 1$ star couplers
Optical signal combination	2×2 and $1 \times N$, $N \times 1$ star couplers
Optical weights	Variable 2×2 couplers, optical amplifiers (both EDFAs and SOAs), electro-optic and electroabsorption modulators
Optical delay lines	Standard, high-dispersion single-mode fiber coils and fiber Bragg gratings

This circumstance constitutes a very serious practical limitation for the implementation of these filters since a very stable platform and considerable electronic feedback loops must be provided for successful operation. The incoherent regime is due to completely random optical phase relationship between the filter taps (i.e., limited source coherence time, $\tau_c \ll T$). The interference is lost and the optical power at the photodetector input is the sum of the optical powers of the filter samples. In this case the filter structure is free from environmental effects and thus is very stable and its performance is quite repeatable. The main drawback of this approach is that filter coefficients can only be positive in principle and this leads to a serious limitation of the range of transfer functions that can be implemented. Fortunately, there are solutions available for the implementation of incoherent filters with negative coefficients and these will be presented in a following section.

Filter tunability

This property makes reference to the possibility to tune the RF bandpass position in a sufficiently fast way. Tunability can be achieved either in a step by step or in a continuous way, and is a key feature required for high-performance flexible filters. In order to tune the RF response of the filter, the FSR has to be modified and therefore also the basic time delay T between samples or taps. The techniques to produce a true time delay can be classified as follows.

(i) *Wavelength tuning of one or multiple sources combined with dispersive optical devices.* This technique takes advantage of currently available modern tunable sources. The dispersive devices can be standard fiber, high dispersive (dispersion compensating) fiber, and linearly chirped fiber Bragg gratings (LCFBG). It can provide continuous or step tunability at high speed, limited by the tuning speed of the sources (depending on the tunable source technology from 100 nanoseconds to greater than 100 milliseconds).

(ii) *Fixed wavelength multiple sources or sliced broadband sources combined with tunable dispersive devices.* This approach is based on the utilization of novel devices with tunable dispersion properties as special chirped FBGs with actuators to change their dispersion properties. It can provide continuous and step tunability, but in this case the time and accuracy to perform a dispersion change on the Fiber device is not so well controlled (100 milliseconds –1 second). Discrete tunability can be achieved by using switched delay lines since different paths providing different basic propagation delays can be chosen by means of an optical space switch. Only step by step tunability is allowed, and the tuning speed is limited by the switching time (1–10 milliseconds).

Filter reconfiguration

This property makes reference to the possibility of changing dynamically the amplitudes of the filter taps (a_k, b_r coefficients) to reshape its spectral response: the weighting or apodization of the amplitude of the filter taps is also a fundamental aspect to ensure enough rejection of the avoided bands. The uniform tap apodization (equal amplitude of the taps) provides a rejection ratio or main-to-secondary lobe ratio (MSLR) that increases linearly with the number of taps. This can be insufficient for certain applications. Different apodization functions have been demonstrated for MSLR improvement, either by adjusting the power of the optical sources or by controlling the attenuation/gain suffered by the taps when they travel through the optical processor.

2.2. Limitations

DOPMS must overcome a series of potential limitations prior to their practical realization as pointed out in previous papers. The main limitations arise from the following.

Source coherence

The source(s) spectral characteristics must be carefully chosen attending to the desired working regime. While coherent operation provides the possibility of implementing any kind of desired transfer function, these structures are very sensitive to environmental conditions [4]. Thus in the majority of cases, incoherent operation is employed since the filters are very compact and robust. Yet coherent effects can appear even under incoherent operation. These undesirable coherent effects may be overcome, for instance, by the use of birefringent fiber delay lines [64].

Polarization

Polarization effects are mainly important under coherent operation [4]. However, it has been outlined and experimentally demonstrated that even under incoherent operation, the filter can be sensitive to signal polarization [65]. The main cause for this apparent contradiction is that some signal samples experience exactly the same delay within the filter leading to coherent interference between them even if a broadband source is employed [65]. Also, when using laser sources

and external modulators, care must be taken to adjust the source polarization to that required by the modulator. The use of polarization preserving fiber pigtails at the modulator input helps to overcome this limitation.

Positive coefficients

Filters working under incoherent regime are linear in optical intensity, thus the coefficients of their impulse responses are always positive. This has two important implications as derived from the theory of positive systems [5]. The first one and more important is that the range of transfer functions that can be implemented is quite limited. The second one is that regardless of its spectral period, the transfer function always has a resonance place at baseband. This is not a serious limitation since a DC blocking filter can be inserted at the optical receiver output. Nevertheless, incoherent filters with negative coefficients can be implemented by means of differential detection [5, 21] and cross-gain modulation in an SOA [66] and other recently developed techniques which are further discussed in Section 4.

Limited spectral period or free spectral range

DOPMSs are periodic in spectrum since they sample the input signal at a time rate given by T. Thus the spectral period or FSR is given by $1/T$. If the DOPMS is fed by only one optical source, then the source coherence time (which is inversely related to the source linewidth) limits the maximum (minimum) value of the attainable FSR under incoherent (coherent) operation. To overcome this limitation, it has been proposed to feed the DOPMSs with source arrays [65].

Noise

As far as the optical source is concerned, passive DOPMSs behave as frequency discriminators and thus convert the optical source phase noise into intensity noise which materializes into RF baseband noise at the filter output [67, 68, 69, 70, 71, 72]. This conversion is dependent on the operation regime. For incoherent operation, the noise is periodic in spectrum showing notches at zero frequency and multiples of the filter FSR. Under active operation (i.e., when incorporating optical amplifiers) new RF noise sources appear as a direct consequence of the beating between the signal and the spontaneous emission [71, 72]. It has been proved however, that the converted phase noise is still the dominant noise source [72]. The use of source arrays to feed the DOPMS is an attractive solution to overcome noise limitations [58]. This is due to the fact that signals recombining at the photodetector at different wavelengths will generate the intensity noise centered at the frequency resulting from the beats of the optical carriers. Since these have very high values, they will be filtered out by the receiver.

Reconfigurability

As defined previously, this property refers to the possibility to dynamically change the values of a_r and b_k in (1). Passive structures are incapable of this feature. Several solutions have been proposed to overcome this limitation including the use of optical amplifiers [7, 8, 9], modulators [10, 34, 35], fiber gratings, and laser arrays [56, 57, 58, 59, 60].

Tunability

As defined previously, this property refers to the possibility to dynamically change the position of filter resonances or notches. To provide tunability, it is necessary to alter the value of the sampling period T. Solutions that include the use of switched fiber delay lines, fiber Bragg gratings, or other tunable tap schemes have been proposed, as will be described below.

3. HIGH-PERFORMANCE INCOHERENT FILTERS

3.1. Transversal filters based on the optical tap wavelength tunability

3.1.1. Positive coefficient transversal filter

High-performance and programmable RF transversal filters can be obtained employing both LCFBGs or fiber coils as dispersive media in combination with an array of optical sources [56, 57, 58, 59]. The layout of the filter for a specific case of a laser array of 5 elements is shown in Figure 3, although in general it is composed of N sources. The advantage of using a laser array to feed the delay line is twofold. On one hand, the wavelengths of the lasers can be independently adjusted. Thus spectrally equally spaced signals representing RF signal samples can be fed to the fiber grating suffering different delays, but keeping constant the incremental delay T between two adjacent wavelengths emitted by the array if the delay line is implemented by means of a linearly chirped fiber grating. This means, for instance, and referring to Figure 3b, that the delay between the signals at λ_1 and $\lambda_2, \lambda_3, \lambda_4, \lambda_5, \ldots, \lambda_N$ is, respectively, T, $2T$, $3T$, $4T$, and $(N-1)T$. Hence the configuration can act as a transversal filter, where the basic delay is given by T. Furthermore, T can be changed by proper tuning of the central wavelengths emitted by the laser array. Thus, this structure provides the potential for implementing tunable positive coefficient RF filters.

The second advantage stems from the fact that the output powers of the lasers can be adjusted independently at high speed. This means that the time impulse response of the filter can be apodized or in other words, temporal windowing can be easily implemented and therefore the filter transfer function can be reconfigured at high speed. We have experimentally succeeded in the demonstration of both tunability and reconfigurability. For instance, Figure 4 shows the results of the samples of the five-stage uniform filter where weights are given by a truncated Gaussian window. The upper trace in Figure 4a, shows the spectrum corresponding to the uniform filter (i.e., unapodized) where the normalized output powers from the lasers in the array are [1 1 1 1 1]. The intermediate trace corresponds to a five-stage Gaussian windowed filter where the normalized output powers from the lasers in the array are given by [0.46 0.81 1 0.81 0.46]. Figure 4b demonstrates the resonance tunability, increasing the resonance position from approximately 2 GHz up to 4 GHz (i.e.,

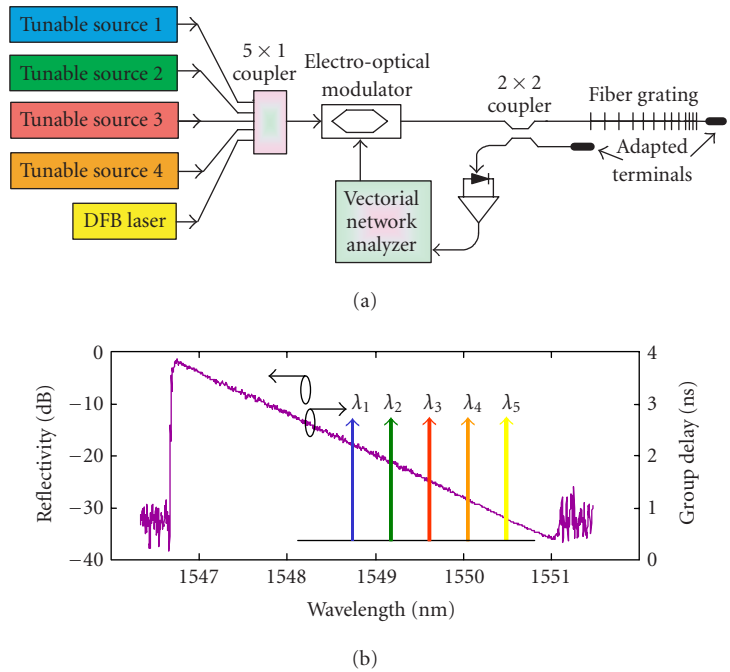


Figure 3: Architecture of a tunable and reconfigurable RF photonic filter using a laser array and a linearly chirped fiber Bragg grating.

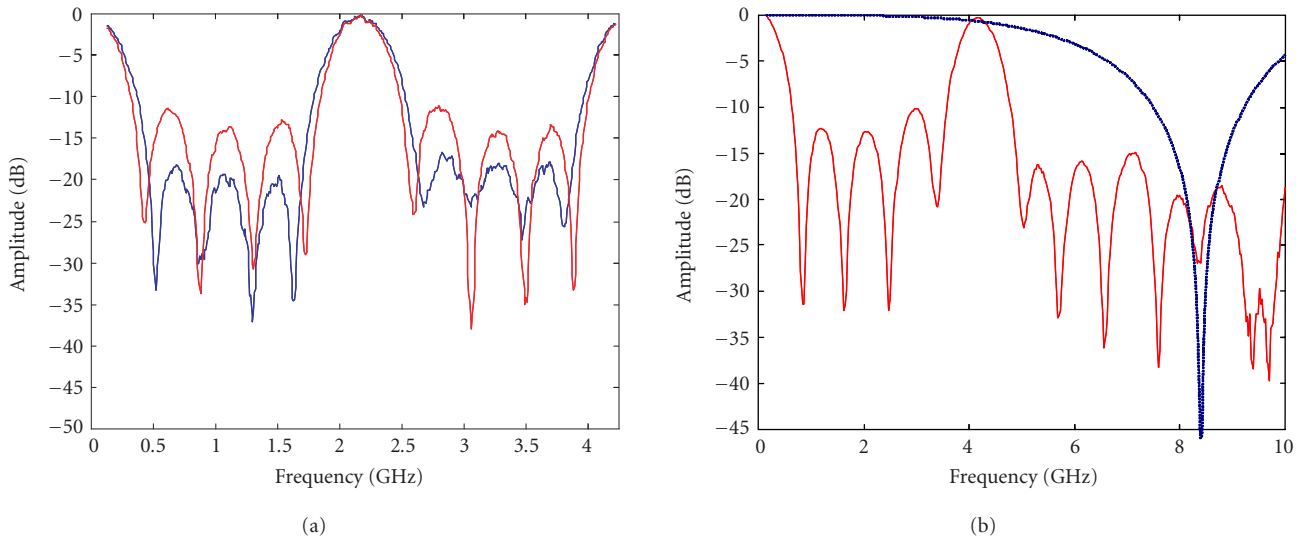


Figure 4: Results of the architecture using a laser array and an LCFBG. (a) The upper trace: uniform filter (i.e., unapodized) output powers from the lasers in the array are [1 1 1 1 1]. The intermediate trace: five-stage Gaussian windowed filter where the normalized output powers are [0.46 0.81 1 0.81 0.46]. (b) Resonance tunability, resonance position from approximately 2 GHz up to 4 GHz, and details of the CSE are demonstrated.

reducing in a factor of two the wavelength separation). In addition, this figure shows the carrier suppression effect (CSE) suffered by the second resonance in this specific case of dispersive media and wavelength spacing.

An additional advantage of employing laser arrays is the possibility of exploiting WDM techniques for parallel signal processing [60]. The possibility of implementing a bank of parallel transversal filters is feasible by extending the concept of a single fiber-optic RF transversal filter based on multiple linearly chirped fiber Bragg gratings and dispersive elements into the implementation of a bank of transversal filters, by means of utilizing wavelength division multiplexing techniques. This technique allows for the simultaneous processing of a single RF signal by various filters.

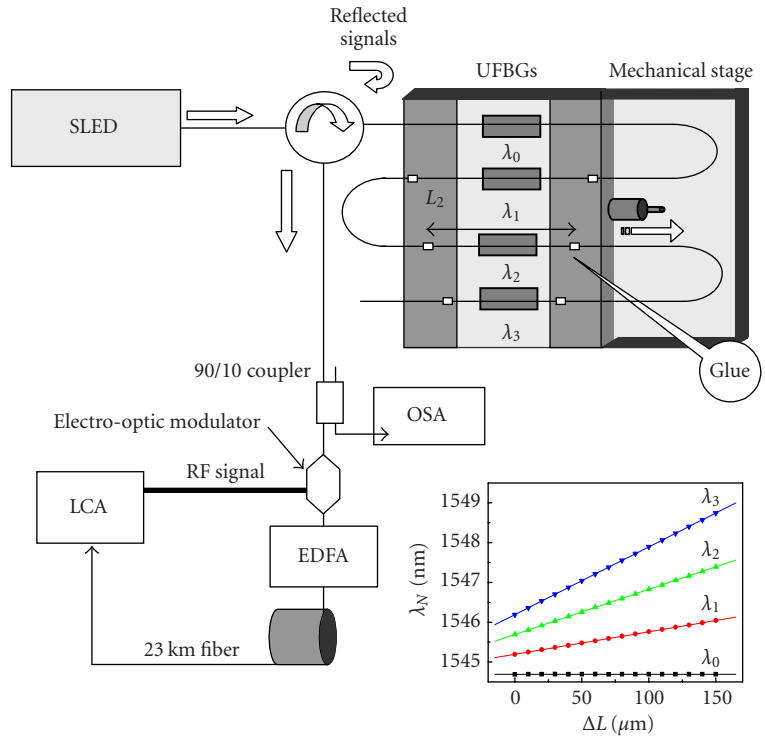


Figure 5: Setup of the flexible uniform FBG-based RF filter. Inset: spectral position of the reflectivity peaks (filter taps) when the fiber is stretched.

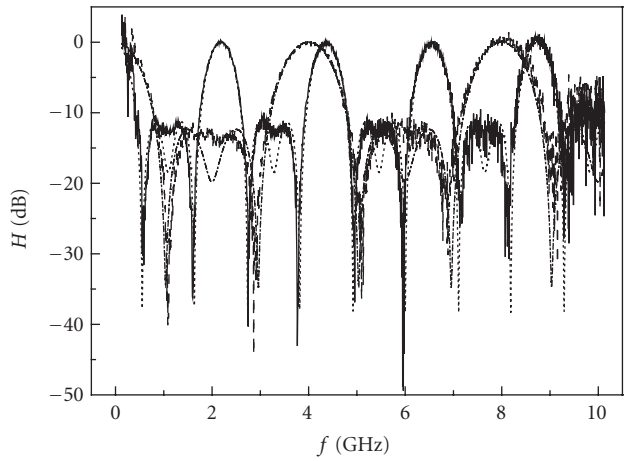


Figure 6: Tunability of the RF-filters. Experimental (solid line: filter 1, dashed line: filter 2) and calculated (dotted line: filter 1, dash-dotted line: filter 2) filter response versus RF signal frequency with different spectral spacing between taps.

A high-performance and continuously tunable positive coefficient RF filter with larger FSR tuning range, low-cost and simple tuning scheme was presented in [35]. This filter consists of a broadband optical source, that is, a superelectroluminescent diode, SLED, and uniform fiber Bragg gratings as filtering elements. This tunable approach was previously demonstrated to provide a simple tunable notch filter where the broadband optical source was sliced by means of only two FBGs, which can be tuned by means of a strain application stage [34]. The output light of the source is driven to the FBG through an optical circulator, and therefore, the reflected signal will be driven to the rest of the system. The uniform FBGs are 5 cm long, written in a series configuration as shown in Figure 5, and they will be stretched to tune the reflection bandwidth, initially centered at λ_{init}. Since the central optical frequency, ω_N, of different gratings must be equidistant [35], each grating must be stretched over a different fiber length, $L_N = L/N$, so the total device length is determined by the number of optical taps.

The device employs identical Bragg gratings whose initial responses, λ_{init_N}, have been tuned by tension before gluing the gratings on the mechanical stage. One of the gratings is not glued on the stage but the others are glued over a fiber length given by $N = 0$, not stretched, $L_1 = 21$ cm, $L_2 = 10.5$ cm, and $L_3 = 7$ cm. The inset of Figure 5 shows the wavelength tunability of all four optical taps corresponding to reflected signals by the gratings when different elongations are applied. The lowest wavelength is kept constant (the grating is not stretched, so $\lambda_0 = \lambda_{\text{init}_0}$) and the others show a linear behaviour, in such a way all of them are equidistant for different elongations.

Figure 6 shows the measured RF-transfer function of two filters corresponding to optical tap spectral spacing of 1.20 and 0.65 nm and a fiber length of 23 km as the dispersive element, together with the theoretical calculation. The filter FSRs were 2.19 and 4.05 GHz, respectively, although the FSR tuning range was demonstrated to be 1–6 GHz.

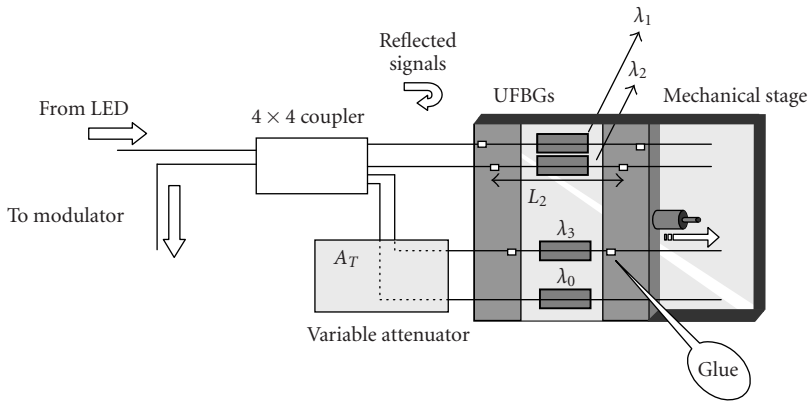


Figure 7: System configuration for reconfigurable sidelobe suppression.

A similar configuration for a 4-tap filter where the gratings are written in a parallel configuration to achieve large sidelobe suppression by weighting the taps was also proposed in [35]. Figure 7 shows the new configuration of the gratings, which will show large flexibility in the implementation of the filters with the only drawback of larger optical losses, which will be compensated by the optical amplifier. The gratings are written in different arms of a 4×4 optical coupler and will be glued onto the mechanical stage over different fiber lengths, as explained above. As known from filter theory [58], the shape of the transfer function of a discrete-time transversal filter can be changed or reconfigured by changing the optical power of the different taps according to an apodization function. Therefore, a decrease in the secondary sidelobes of the filter can be achieved. The optical signals corresponding to the side taps (in our case, $N = 0$ and $N = 3$) will go through a two-input variable attenuation, which will be varied according to the desired degree of MSLR of the RF filter.

Figure 8 shows the measured main to sidelobe ratios of implemented RF filters by introducing different attenuation values to the optical signal taps, together with the theoretical curve. As an example, the intensity of the four taps of two filters is shown in two different insets, exhibiting different apodization profiles. The uniform intensity pattern leads us to the theoretical (and measured) limitation of 11.3 dB (see Figure 6) and sidelobe reduction has been demonstrated in these filters up to 25 dB.

Another interesting option is to employ AWG devices with high port count in order to implement source slicing with a high number of samples [73, 74]. With this technique we have recently reported a twelve-sample transversal filter using a two-stage 1×40 AWG configuration shown in Figure 9. This structure has the advantage of allowing also filter reconfiguration if switches or variable attenuators are placed in between the demultiplexing and the multiplexing stages.

In the structure above, the combinations of the SLED and the erbium-doped fiber amplifier (EDFA) sources provide an almost flat optical spectrum to be sliced by the couple of AWGs. The proper channel-by-channel interconnection

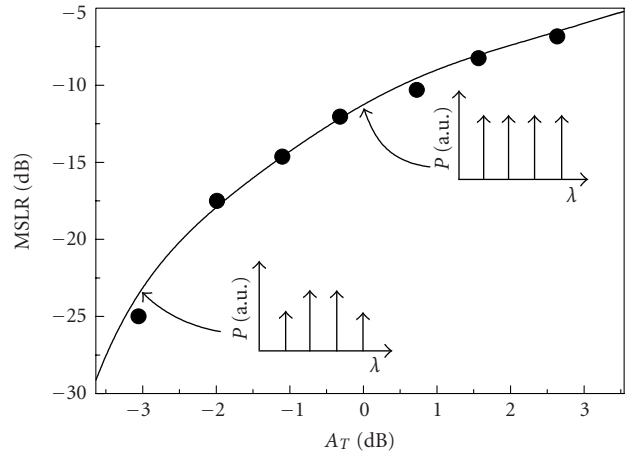


Figure 8: Calibration curve of sidelobe suppression versus attenuation tuning parameter. Insets: intensity of the taps in different filters.

between AWGs provides a certain degree of freedom to select the wavelength spacing between the slices. The employed AWGs were designed for DWDM applications with 0.8 nm spacing between adjacent channels and 0.4 nm of 3 dB optical bandwidth. Figure 10 shows two examples of channels interconnections to provide different RF lobe tuning and RF lobe 3 dB bandwidth. Proper optical attenuation of the interconnections provide the feasibility of taps apodization to reduce the sidelobe level. Special attention must be paid to the lowpass filtering effect arising from the combined use of broadband slices (0.4 nm) and large amount of dispersion, and on the other hand also the effect of the dispersion slope (S) and the large wavelength range covered must be considered. This phenomenon manifests as an amplitude reduction and an RF lobe bandwidth increasing for higher RF frequencies as it can be observed from results in Figure 10, and it was extensively discussed in [74].

Another recently reported slicing technique employs a fiber Bragg grating and an acoustic wave, which generates the optical taps when it is propagated along a uniform FBG

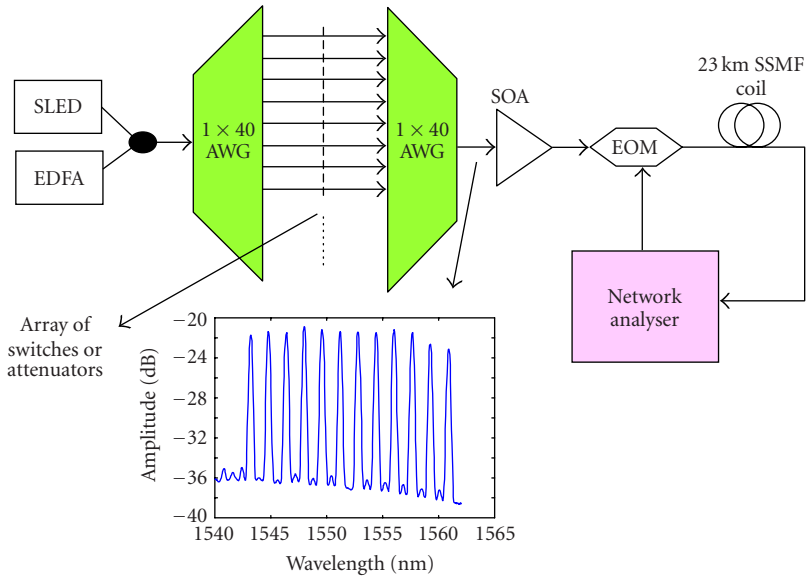


Figure 9: Spectral slicing technique employing AWGs. Inset: sliced spectrum employing twelve channels of the AWGs.

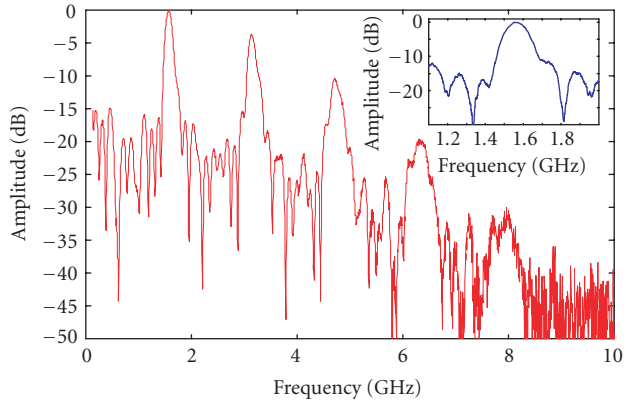
[75]. The acoustic wave creates a periodic strain perturbation that modulates periodically the period and the refractive index of the FBG. In this system, the spontaneous emission of an EDFA around 1540 nm is used as a broadband source, the FBG is written at the neck of a symmetric tapered fiber in order to increase the efficiency of the acoustic interaction and the longitudinal acoustic wave is generated by a piezoelectric transducer driven by an RF signal and launched into the fiber using a silica horn. It produces spectrally equispaced bands of reflection on both sides of the original Bragg grating as we can observe in Figure 11.

By using an SMF-28 fiber length of 46 km as a dispersive element, two transversal filters were implemented with the acoustic frequencies of 0.755 MHz and 1.444 MHz. The wavelength spacing of the optical taps were 0.11 nm and 0.22 nm, respectively, and the RF filter characteristics are shown in Figure 12. The former one showed an FSR of 6.25 GHz and a 3 dB bandwidth of 1.44 GHz, whereas the second one had an FSR of 11.5 GHz and a 3 dB bandwidth of 2.76 GHz. The reconfigurability of the filter can be obtained by applying different voltages to the piezoelectric transducer since different degrees of apodization of the optical taps intensities are achieved by controlling the acoustic power. A main to sidelobe ratio up to 20 dB has been demonstrated.

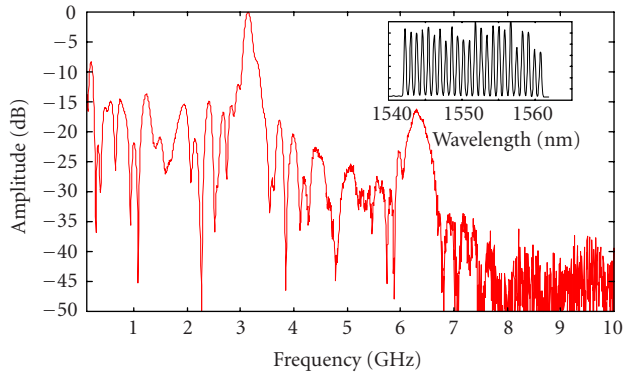
However, in RF applications when an optical signal is used to carry several RF signals providing different services, and photonic filters are used to select one of these services, as happens in next-generation optical access networks, there is a need for obtaining a single and very selective tunable radiofrequency band in sliced broadband optical source [76]. The presence of a periodic transversal filter response where different RF bands are selected by the filter implies a limitation in the number of services carried by the same optical carrier. Thus, a new approach to obtain single bandpass RF filters is extremely interesting for their implementation in optical access networks. It is based on a broadband optical source and a fiber Mach-Zehnder interferometer (MZI), as shown in Figure 13. When the source optical output is transmitted by the interferometer and launched into a fiber delay line, a tunable bandpass filter is achieved showing a single bandpass frequency response, large tunability without changing the bandwidth of the filter, and high attainable Q values.

The experimental arrangement is given by the Figure 13. The spontaneous emission of an EDFA is used as broadband optical source, with a 3 dB bandwidth of 5.4 nm and the MZI leads to obtain different periodicities $\Delta\omega$. In this system, a 46 km fiber length is used as dispersive element, and the RF filter response shows a bandpass characteristic centered at the frequency Ω_0, that can be tuned varying the periodicity $\Delta\omega$ of the interferometer Mach-Zehnder. As shown in Figure 14, periodical wavelength spacing in the interferometer ouput of 0.237 and 0.173 nm leads to bandpass filters at 5.83 and 7.88 GHz. A tuning range of several tens of GHz was achieved with an MSLR lower than 20 dB and the Q maximum value achieved was around 40.

The Q factor is plotted in Figure 15 for several RF filters implemented with different wavelength periodicity $\Delta\lambda$ of the signal at the electro-optic modulator (EOM) input, for two different dispersion values according to 23 km(■) and 46 km(•) length of fiber. The dashed curve plots the Q factor that can be achieved by a Gaussian optical source of 5.4 nm 3 dB bandwidth. As shown, Q value is improved by increasing the optical source bandwidth although the influence of the dispersion slope drives to a degradation of the radiofrequency response. Therefore, potential high Q values can be achieved by choosing the appropriated broadband source and reducing the dispersion slope that can be achieved by using two different fibers to compensate the dispersion slope. The dotted line shows the Q factor when the dispersion slope is compensated.



(a)



(b)

Figure 10: (a) Normalized RF amplitude for the sliced spectrum: slices from 1543.7–1561.3 nm up to a total number of twelve bands (samples), 1.6 nm apart. The RF response presents a band spacing of 1.56 GHz. Inset: precise measurement of the 1° passband. (b) Normalized RF response for 24 slices (1542.9–1561.3 nm), with 0.8 nm wavelength spacing. FSR = 3.1 GHz, MSLR = 14 dB, 3 dB bandwidth of 125 MHz, Q factor = 24.8.

3.1.2. Negative coefficient transversal filter

One of the main limitations of incoherent transversal filters is that only positive taps are possible since the intensity is a positive magnitude. Some solutions have been reported to overcome this limitation, from the optoelectronic approach that uses differential detection [21] to various configurations which use active elements to generate negative taps, that is, amplitude inversion due to gain saturation in the homogeneously broadened gain medium of a semiconductor optical amplifiers (SOA) [77], carrier depletion effect in a DFB laser diode [78], cross-intensity modulation of the longitudinal modes of an injection-locked Fabry-Perot laser diode [79]. Recently, microwave filters based on wavelength conversion employing cross-gain modulation of amplified spontaneous emission spectrum (ASE) of an SOA have been demonstrated [80]. Recently, a new low-cost approach based on passive elements has been proposed to generate negative taps and it is based on filtering a broadband source with the transmission of uniform FBGs [81].

The implemented filter is formed by a tunable laser (TL) and the signal transmitted by a uniform FBG, which is illuminated with the ASE of an EDFA. The broadband optical source has a 3 dB bandwidth of 5 nm around 1530 nm. The uniform FBG is 1 cm long with Bragg wavelength of 1530.96 nm and maximum reflectivity of 8 dB. The output light of the FBG and the TL are driven to a 90/10 optical coupler. The combined signal can be monitorised by an optical spectrum analyser, OSA, by using the 10% arm. The 90% arm signal is amplitude-modulated in the EOM whose RF-signal of frequency f is generated by a lightwave component analyser, LCA. A fiber length of 46 km will be the dispersive element in the filter, and finally, the transfer function of the filter is measured in the LCA (Figure 16).

Figure 17 gives the measured and theoretically predicted free spectral range of 2-tap RF filter versus different wavelength spacing between the central Bragg wavelength of the FBG and the TL output signal showing tunability in the 0.7–5.6 GHz.

In order to show the good performance of these filters when various taps are added, 5-tap RF filter has been measured using two FBGs and three lasers whose wavelength separation is 1.16 nm (Figure 18).

Therefore, the system shows a broad tuning range and a good performance of the transversal filter.

A different and promising technique also used to obtain RF filter with negative coefficients and consists of using laser arrays (like the first approach described in this section) has been recently reported in [82]. It relies in the counter-phase modulation on a Mach-Zehnder external modulator (MZM) device by means of employing the linear part of the transfer function with positive and negative slopes. The concept is illustrated in the upper part of Figure 19 with a simple RF modulating source.

The upper left part of the figure depicts the typical output versus input optical power sinusoidal transfer function of an MZM as a function of the applied bias voltage V_{BIAS}. Two linear modulation regions with opposite slopes can be observed centered at V_{BIAS}^{-} and V_{BIAS}^{+}, respectively. As shown in the right part of the figure, the same RF modulation signal applied to the modulator at each of the former bias points will produce an optical-modulated output signal with the same average power but where the modulating signals are 180° shifted or in counter-phase. In other words they can be considered of different signs. This principle can be employed to implement RF photonic filters with negative coefficients if the output wavelengths from a multi-wavelength source (either a laser array or a sliced broadband source) are applied to an MZM biased at either V_{BIAS}^{+} or V_{BIAS}^{-} depending on whether they are employed to implement positive or negative filter samples. The output from both modulators are combined and sent to a dispersive element (i.e., an LCFBG or a fiber coil) that implements the constant differential time delay between the filter samples.

The feasibility of this approach has been experimentally demonstrated with a six-sample uniform RF photonic filter with three positive and three negative coefficients using the laser array implementation described in [82].

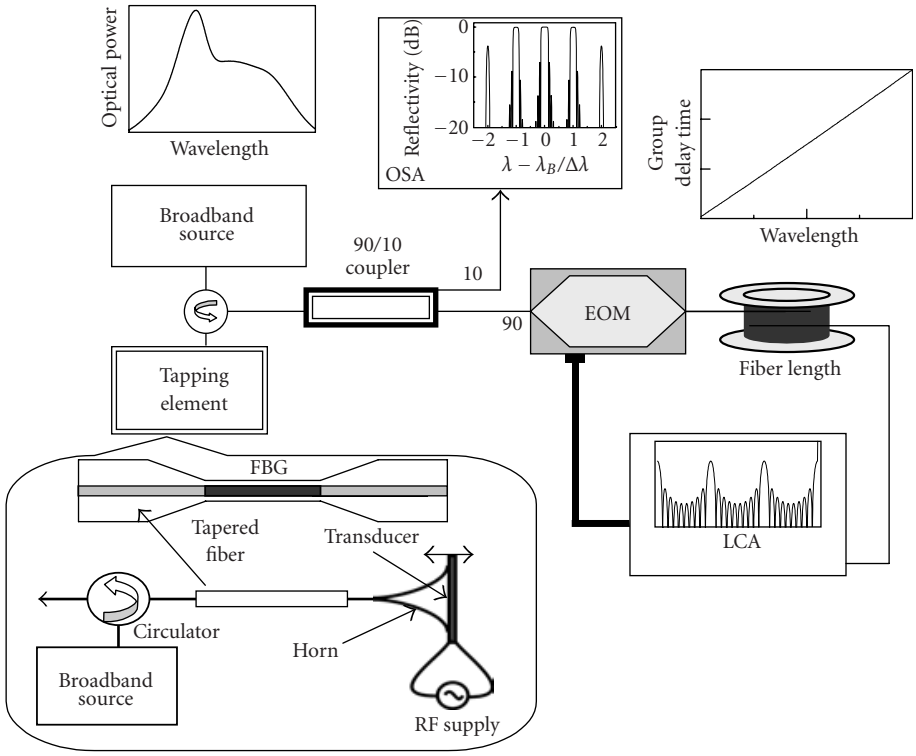


Figure 11: Experimental setup to implement transversal filters consisting of a Bragg grating-based acousto-optic superlattice modulator.

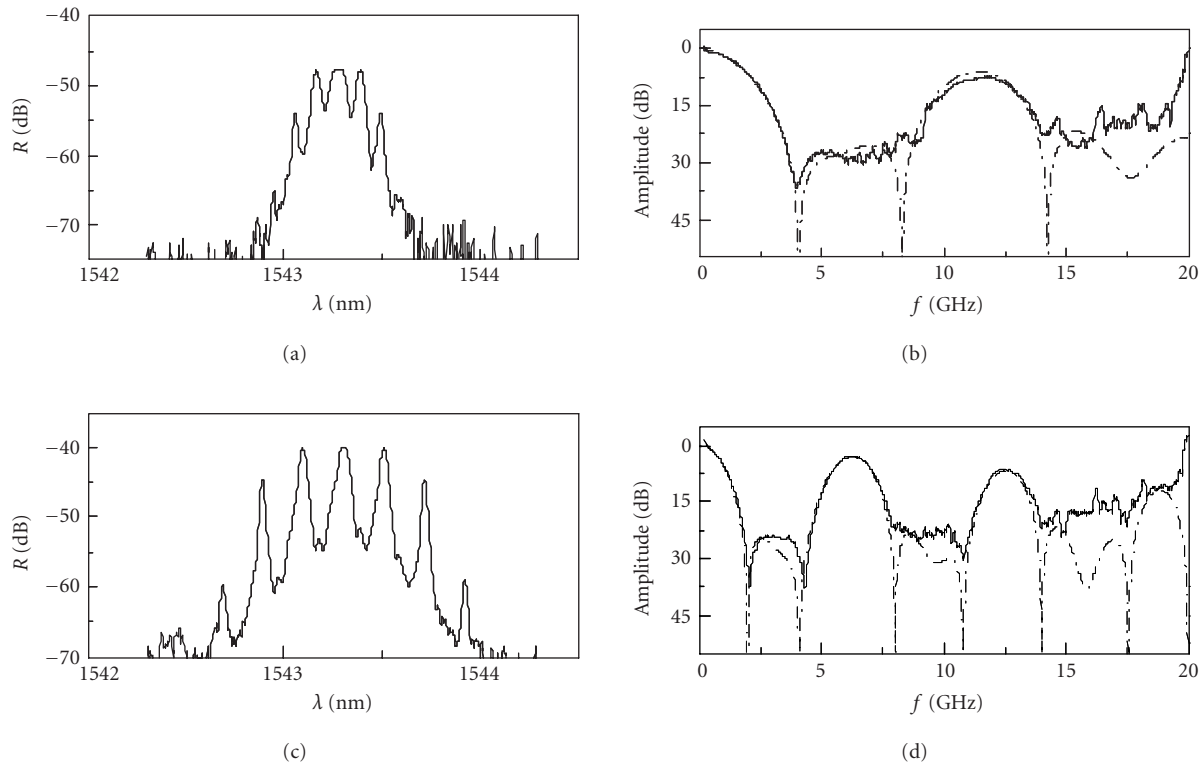


Figure 12: (a), (c) Optical spectra and their experimental (solid lines) and (b), (d) theoretical (dashed line) transfer functions for the frequencies (a), (b) 0.755 MHz and (c), (d) 1.444 MHz.

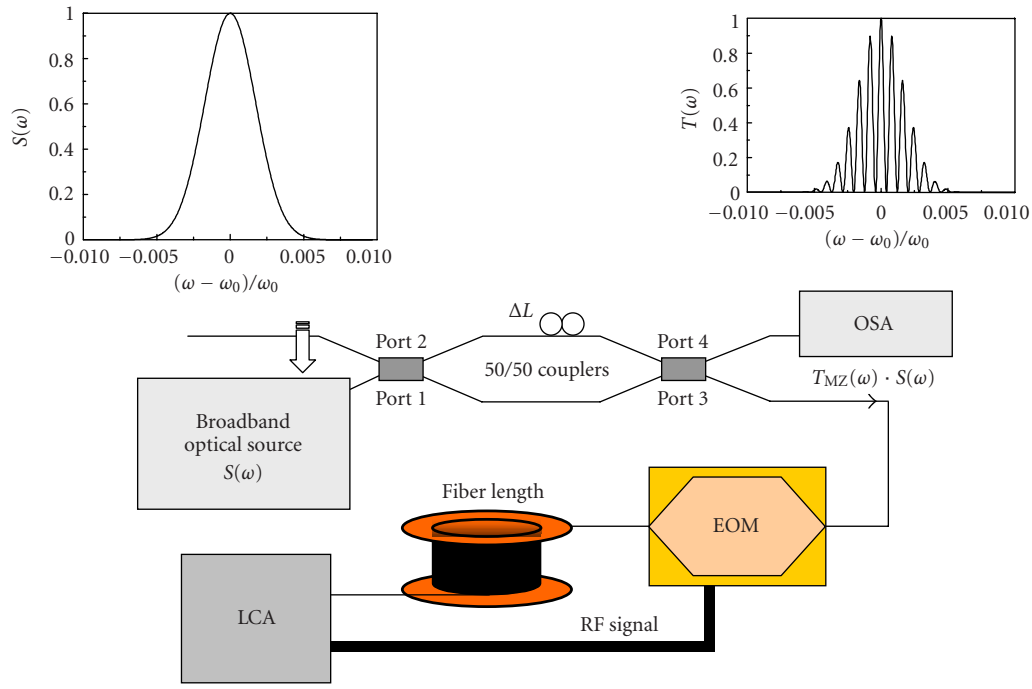


Figure 13: Schematic of the implementation of the RF single bandpass filter based on a broadband source and an MZI.

Figure 19 shows the experimental layout. An array of six tunable lasers emitting at $\lambda_1 = 1546.65$ nm, $\lambda_2 = 1548.43$ nm, $\lambda_3 = 1550.11$ nm, $\lambda_4 = 1551.86$ nm, $\lambda_5 = 1553.47$ nm, and $\lambda_6 = 1555.24$ nm was selectively fed to two MZMs biased at $V^+_{\text{BIAS}} = 0$ V and $V^-_{\text{BIAS}} = -3.9$ V, respectively. Wavelengths λ_1, λ_3, and λ_5 were fed to the MZM biased at V^+_{BIAS} (positive samples) , whereas wavelengths λ_2, λ_4, and λ_6 were fed to the MZM biased at V^-_{BIAS} (negative samples). Both EOMs were modulated by the same RF signal, a 5 dBm sinusoidal signal provided by an LCA. The frequency of the RF modulating signal was varied from 130 MHz to 5 GHz in order to measure the transfer function characteristic of the filter.

Figure 20 shows the measured modulus of the transfer function for a filter with six uniform coefficients. Both the experimental (solid line) and the theoretical (broken line) results are shown for reference and comparison. As expected, the filter resonance at baseband (typical of positive coefficient filters) has been eliminated, thus confirming the feasibility of the proposed scheme for the implementation of negative coefficients. Although in principle, as shown in Figure 19, two modulators are required in the transmitter, in practice this requirement can be reduced to only one modulator if this device is provided with two input ports. A main advantage of this configuration is that there is no need to duplicate the optical structure of the filter to implement positive and negative coefficients since the taps already carry their sign prior to being delayed. Another interesting feature is that the sign is decoupled from any sample weighting process.

As it was anticipated, in principle two MZM devices are required for the previous approach, but only one MZM is required to incorporate 2 inputs by replacing the input Y branch to the integrated modulator by a 2×2 integrated coupler. We have reported novel results [83] using a newly developed 2×1 MZM device fabricated by AMS (Alenia Marconi Systems) featuring the above desired characteristics. We demonstrate a filter structure that requires only this device and we employ it to implement a 9-tap (positive and negative) transversal RF photonic reconfigurable filter with square-type resonances.

Figure 21 shows the layout and the intensity transfer function of the newly developed device. The device is an $LiNbO_3$ dual drive MZM developed for this application within the framework of the EU-funded IST-LABELS project. As it can be observed, the input Y branch has been replaced by a 3 dB 2×2 integrated coupler. Figure 21 also shows the measured modulation curves for the two (input 1-output) (input 2-output) input/output configurations. As expected, the curves for the two input/output configurations show a clear 180° phase shift on the two RF-modulated outputs with maximum dynamic range at the quadrature point. Note as well that the 180° phase shift on the two RF-modulated outputs is maintained regardless of the value of the bias voltage although the output average output power changes. For instance, the figure shows two cases (case (i) and case (ii)) as particular examples. The 2×1 MZM previously described has been employed to demonstrate the operation concept in an experiment designed to implement, for the first time to our knowledge, a reconfigurable square-type

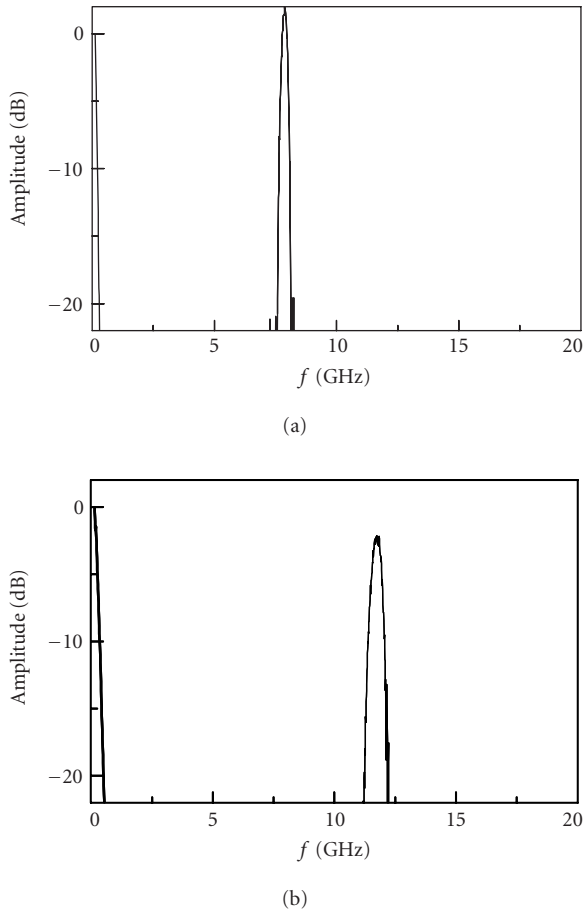


Figure 14: Response of the RF filters implemented by 46 km of fiber: (a) 0.237 nm and (b) 0.173 nm.

filter with negative coefficients. The filter layout was similar to that of Figure 19 [83] with the exception that the two standard MZMs have been replaced by the especially designed 2×1 MZM. In this case the dispersive element was implemented by a 23 km fiber and the filter samples were produced by means of tunable lasers with wavelength separation of 0.96 nm between adjacent carriers which resulted in a filter free spectral range of 2.63 GHz. The implemented RF filter was a square shape type with nine coefficients (7 different from 0 + 2 null ones) and with four negative taps. The values for the taps were [−0.16 0 0.27, −0.77 1 −0.77 0.27 0 −0.16]. Figure 22a shows the optical spectrum at the fiber coil output prior to the detector. Figure 22b demonstrates the measured transfer function (dotted trace), the theoretical expected trace for the square filter design (continuous trace), and also the theoretical trace for a filter with uniform coefficients (broken trace). The agreement between the theoretical and the experimental results is excellent inside the filter bandpass. Outside the filter bandpass the agreement is worse, but this is due mainly to the noise arising from delay nonuniformities due to inaccuracies in the central wavelengths of the lasers and also due to the feasible inaccuracy in the tap amplitude.

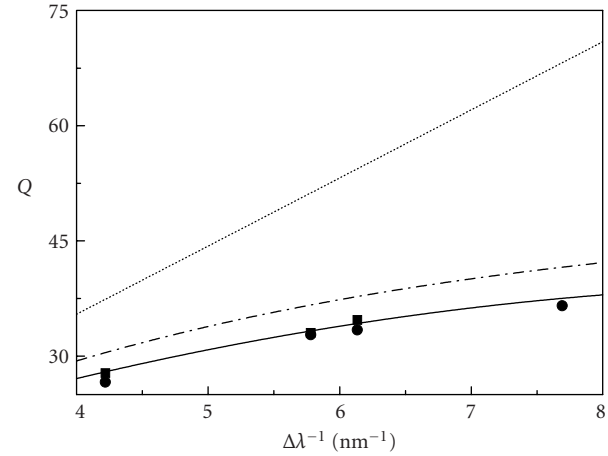


Figure 15: Dependence of the Q factor of an RF filter on the wavelength periodicity $\Delta\lambda$ of the signal at the EOM input, for two different dispersions L_1(■) and L_2(•). Solid line: theoretical fit of the experimental data. Dashed-dotted line: Q factor obtained when optical source (bandwidth of 5.4 nm) is Gaussian. Dotted line: Q factor obtained when dispersion slope is compensated.

3.2. Transversal filters based on tunable dispersive devices

Linearly chirped fiber Bragg gratings have been proposed to obtain tunable dispersion slope gratings showing suitable optical bandwidth for RF applications. By acting on them, it is possible to vary the time delay of each optical wavelength carrier. Temperature and strain gradients on the CFBG [3] or the use of piezoelectric transducer [4] are some of the most extended approaches. Recently, we demonstrated the dynamic chirp of an original uniform fiber Bragg grating based on a grating fixed to a magnetostrictive rod, which could be disturbed with a tapered magnetic circuit [5]. These magnetic systems show advantages such as good dynamic response and easy implementation.

A new device is based on tuning the phase response of a tapered fiber Bragg grating (TFBG) by using a magnetostrictive transducer and the magnetic field inside a simple coil. Tunable transversal filters can be implemented by using this device since the dispersion slope is tuned when applied current to the coil is varied.

The tunable dispersion device consists of a chirped grating with tapered transversal section held on a magnetostrictive material, which is subjected to the nonuniform magnetic field produced by the current flow through a finite solenoid.

A 5 cm long FBG with uniform period is written in a tapered fiber fabricated by fusion and elongation [6]. This TFBG is fixed on a magnetostrictive rod and placed inside of a 4 cm long magnetic coil (see Figure 23). Thus, the magnetostrictive material suffers a local lengthening, which is proportional to the intensity of the applied magnetic field. The TFBG is located at the axial region where the magnetic field variation is quasilinear [7]. Therefore, when an electrical current of a given intensity is injected to the

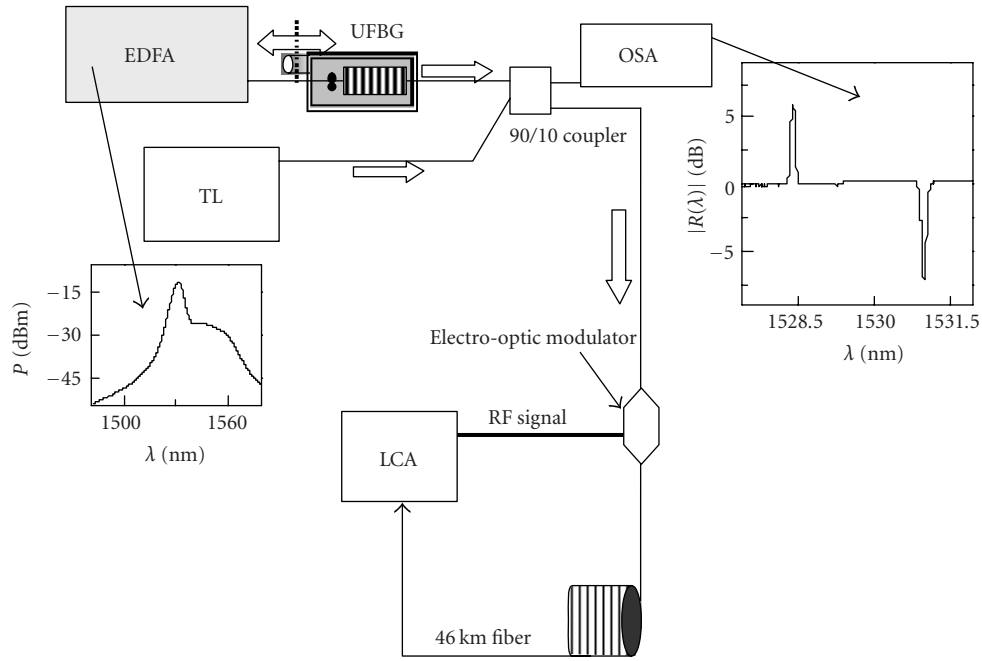


Figure 16: Setup of the RF negative-taps filter. Inset: input signal launched into EOM relative to EDFA power level.

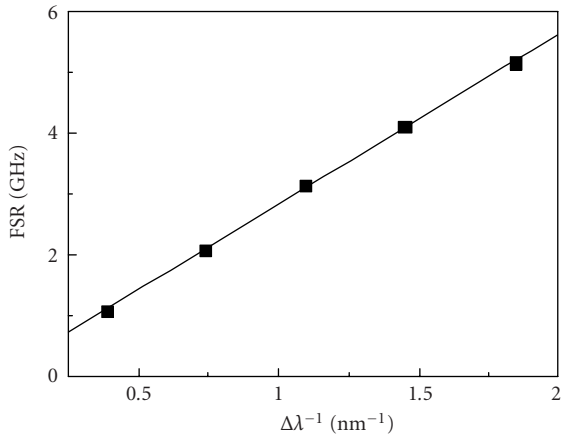


Figure 17: Free spectral range of the RF filters dependence on the reciprocal of the wavelength spacing between taps: theoretical calculation (solid line) and experimental results (dots).

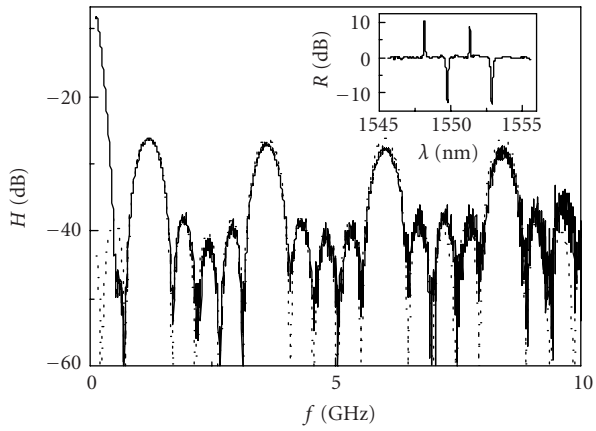


Figure 18: Filter response versus RF signal frequency with 1.16 nm equispaced wavelength separation. Theoretical calculation (dotted line) and experimental results (solid line). Inset: spectral position of the five taps.

solenoid, the magnetic field applied to the TFBG leads to different dispersion slopes depending on the intensity current. When no current is applied, the TFBG has a linear dispersion due to the design of the taper profile [6]. It has a flat reflectivity and a 3 dB bandwidth of 1.58 nm.

Inset of Figure 23 shows the dispersion slope when electric currents are 0 A, 3 A, and 5 A, with a 3 dB bandwidth of 1.58 nm, 2.05 nm, and 2.51 nm, respectively. Time delay slopes are achieved from 188–472 ps/nm, with a useful passband larger than 1 nm.

The implementation of RF filters requires N optical carriers, equidistant by $\Delta\lambda_0/(N-1)$, which are provided by a multi-wavelength tunable laser. They are amplitude-modulated by an EOM and launched into the tunable TFBG as shown in Figure 23 [84]. The dispersion of the TFBG, D, gives the differential delay between adjacent optical taps $\Delta\tau = D \cdot \Delta\lambda$. Because of the variation of the time delay slope of the TFBG when we apply different magnetic gradients, transversal notch filters with tunable free spectral range (FSR) are measured.

Figure 24 shows the range of FSR values that can be obtained by using our device (shaded region). Moreover, several filters have been implemented by changing the number of optical taps and the total optical bandwidth $\Delta\lambda_0$ of the filter

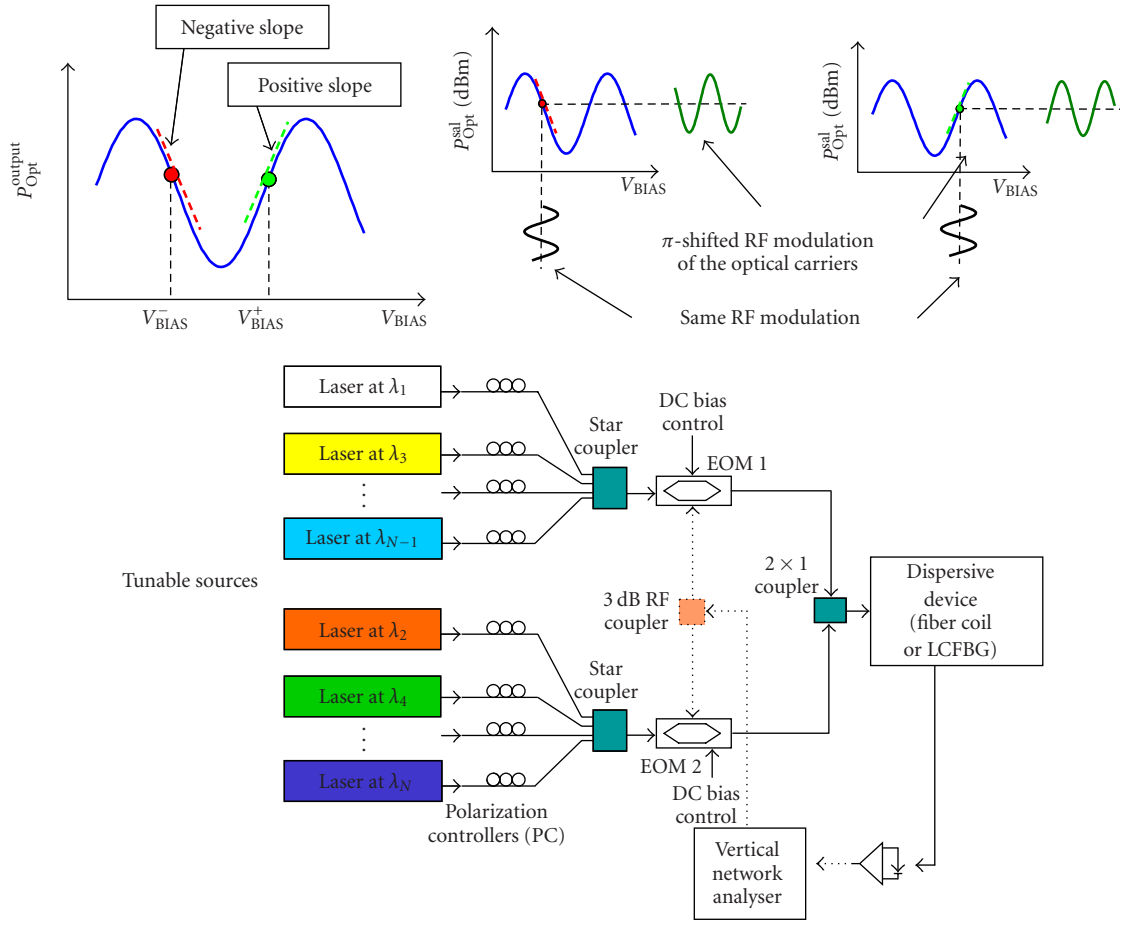


FIGURE 19: Experimental layout of the negative coefficient approach employing two EOMs and counter-phase modulation details at different biasing points of the EOM response.

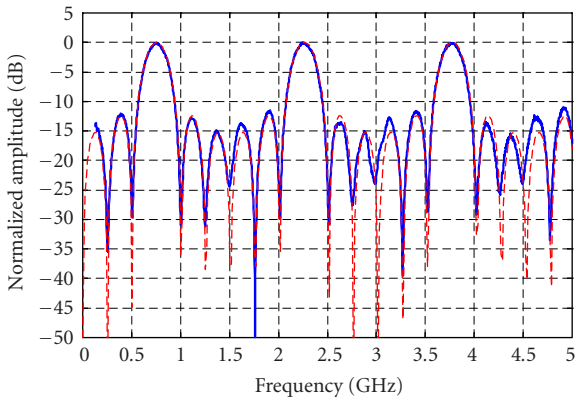


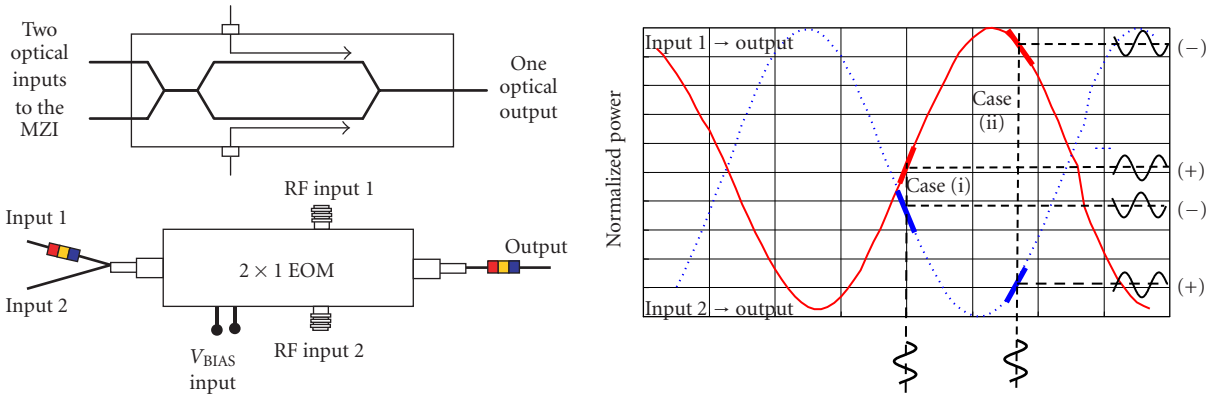
FIGURE 20: Measured modulus of the transfer function for a filter with six uniform coefficients: experimental (solid line) and theoretical (broken line).

for 0 A, 5 A, and −5 A. The inset shows an RF filter with $\Delta\lambda_0 = 1.0$ nm and two taps when $I = -5$ A ($D = 188$ ps/nm) and $I = 5$ A ($D = 472$ ps/nm) showing an FSR of 5.4 GHz and 2.3 GHz, respectively.

The tuning range of the previous approach can be enlarged by using a new system composed of the cascade of two switched tunable stages [85]. Each one includes a tapered fiber grating subjected to the nonuniform magnetic field created inside an electrical coil.

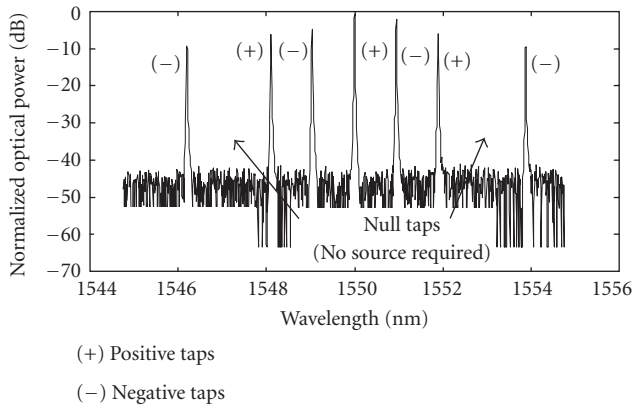
Figure 25 shows a scheme of this tunable dispersion system. A laser source is amplitude-modulated and launched into the first TFBG through a circulator. The optical signal reflected goes through a 50/50 coupler, and then, to an optical switch. When it is in the bar state (BS) the optical signal is launched into the second TFBG, and therefore, the response of the global system is given by the cascade of both subsystems. After reflection in the second grating, the signal is measured at one of the input ports of the coupler by using a lightwave components analyzer. Alternatively, when the optical switch is in the cross state (CS), the signal is driven to the LCA through the output port of the switch.

Measurements of the amplitude and time delay response of the whole system were performed with the optical switch in CS and BS. First state led to dispersion slopes from 230–351 ps/nm, whereas a range between 420 and 715 ps/nm was obtained in the second state.

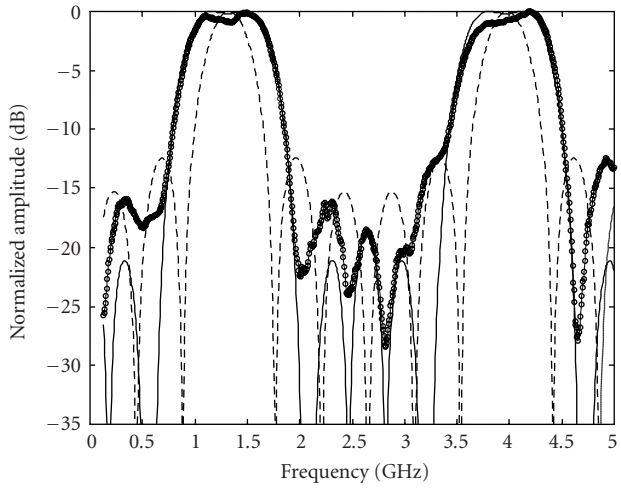


(a) (b)

FIGURE 21: Layout and the intensity transfer function of the newly developed device. (a) Two electrical travelling wave electrodes for dual drive operation. (b) Optical power versus bias voltage of 2 × 1 EOM.



(a)



(b)

FIGURE 22: (a) Optical spectrum at the detector, 9 taps (7 active and 2 null taps). (b) Measured transfer function (dotted trace), the theoretical expected trace (continuous trace), and the theoretical trace for a filter with uniform coefficients (broken trace).

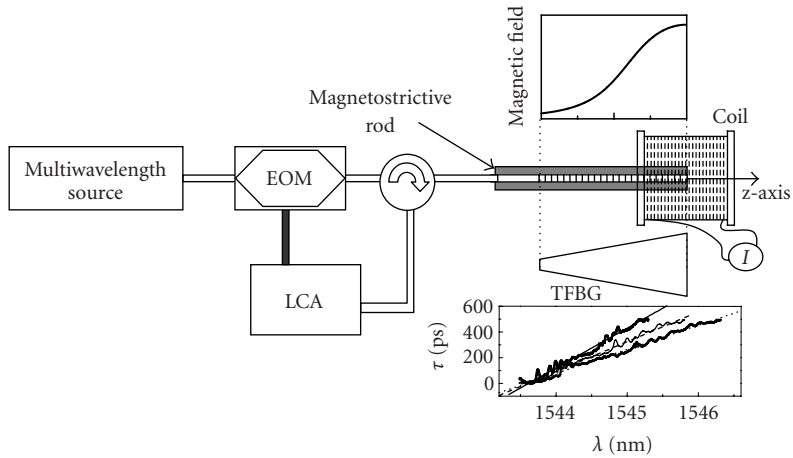


Figure 23: Tunable device for RF filtering implementation based on fiber Bragg gratings subjected to nonuniform magnetic fields. Inset: time delay response when different electrical currents are applied: $I = 0$ (solid line), 3 (dashed line), and 5 A (dotted line).

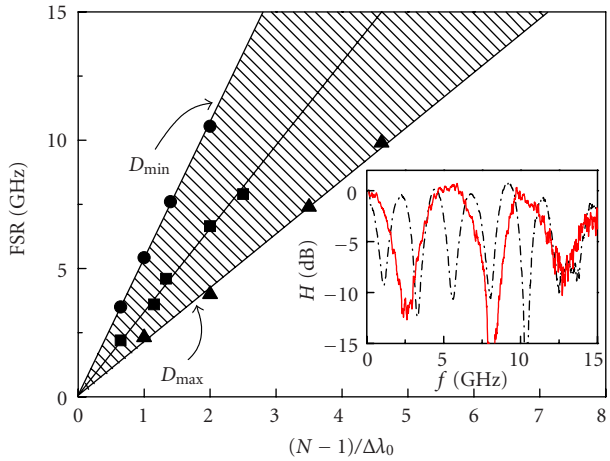


Figure 24: FSR versus number of optical taps and optical bandwidth of the filter for different time delay slopes. Inset: transfer function of a tuned RF filter (solid line: $D = 188$ ps/nm, dashed-dotted line: $D = 472$ ps/nm).

To show the performance of this system, 3-tap RF filters were implemented for CS and BS when different electrical currents are injected to the solenoid (0, 2, and 4 A). The multi-wavelength was set to emit light at three different optical wavelengths equispaced by 0.41 nm. Figure 26 shows the FSR of the measured RF filters for different electrical currents. CS led to achieve FSRs between 7 and 10.6 GHz and BS led to a tuning range from 3.5–5.8 GHz. The inset of Figure 26 shows the group time delay versus wavelength when no electrical current is applied to the coils. Note the increase of the dispersion when BS was set instead of CS.

Amongst the approaches based on the concept of dispersive adjusting and maintaining the wavelengths of the taps fixed to obtain RF lobe tuning we can include the following RF photonic filter, although in this case other important techniques as spectral slicing employing uniform FBGs and fiber delay lines have been combined to perform the filter specifications [8]. More specifically, we report a tunable photonic filter for noise suppression and channel interference mitigation in the front-end stage of a UMTS base station prior to the highly selective SAW filter. As it has been reported elsewhere [9], the inclusion of such a filter can increase the capacity of UMTS systems. It is possible to switch the designed filter along the twelve-channel UMTS.

The proposed configuration implements a classical FIR transversal filter. We also employ the spectral slicing of a high-power broadband optical source to obtain an "equivalent" multi-wavelength source. The slicing is performed by an array of fiber Bragg gratings which also introduces a fixed time delay between the reflected slices of the signal. Finally, we employ a reconfigurable chain of dispersive modules (SMF-28 fiber) to vary the time delay between the slices and thus introduce tuneability to the filter. Figure 27 shows the filter structure. The first block is the optical source and modulator, where an SLED is employed as the broadband optical source (power of 10 mW and with the 40 nm bandwidth at 1555 nm). The RF modulation of the optical signal is performed by means of an EOM directly over the entire SLED spectrum. The second block consists of an array of N fiber Bragg gratings written in series with certain uniform spacing between the gratings. This block accomplishes a double task: the slicing of the SLED spectrum and the introduction of the time delay ΔT between the signals reflected from the gratings. Finally, the spectral slices are fed to the third block: a reconfigurable chain of dispersive modules, where each module employs a standard nonshifted SMF-28 fiber (the use of a highly dispersive, e.g., dispersion compensating fiber, is also possible). The optical switches in the third block provide a stepwise tuning of the accumulated dispersion in the block. By varying the dispersion of the block, the time delay between the signals reflected from different gratings is changed and thus, tuneability of the filter RF response is achieved.

The UMTS channel filtering requires a high Q factor (about 400), since the required 3 dB passband of the filter should be less than 5 MHz and the operating frequency of the

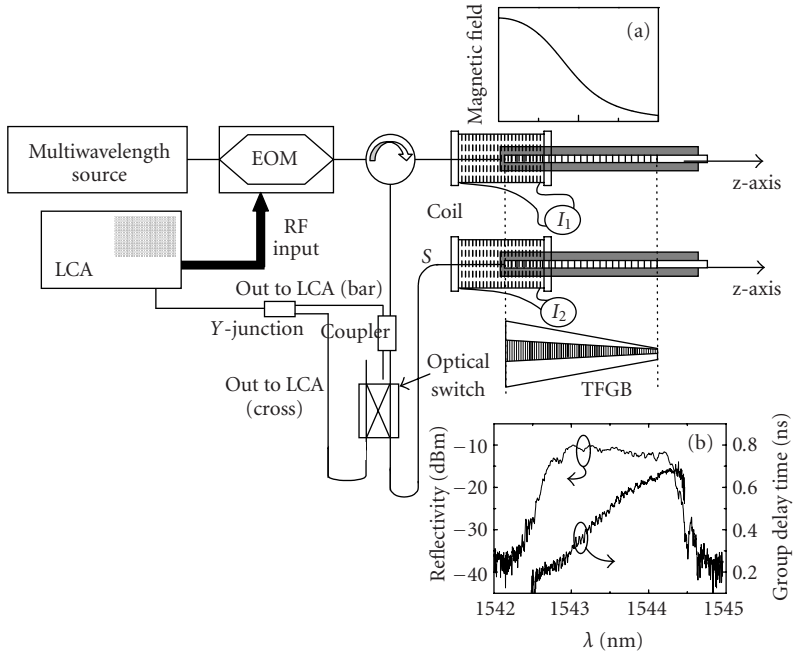


Figure 25: Schematic of the optically switched time delay line. Insets: (a) dependence of the magnetic field on the z-axis along the grating, and (b) reflectivity and time delay response of one of the chirped fiber gratings.

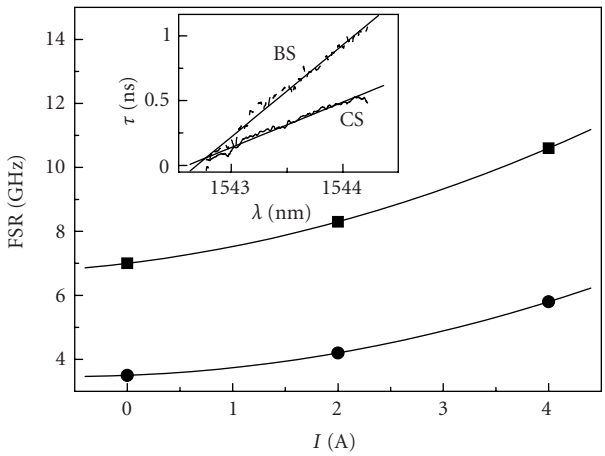


Figure 26: Free spectral range of the RF filters versus the current intensity for cross (■) and bar (•) states of the switch. Inset: group delay time for cross state (351 ps/nm; solid line) and bar state (712 ps/nm; dotted line) when no current is applied to the solenoids.

filter lies within 1920–1980 MHz. Furthermore, the UMTS channel filtering also requires the tuneability of the RF passband within the 12 channels allocated along the 60 MHz band (between 1920 and 1980 MHz). In order to achieve such a high Q factor, the present transversal filter operates at a higher order "resonance" of its periodic response. In this case, the FSR of the filter is an integer fraction of the UMTS operating frequency. The present filter has been designed to employ the resonance number 18. To keep the filter tuned to the upper UMTS channel at 1977 MHz when the dispersive module is "switched off," the FSR of the filter has been set to 109 MHz, and the corresponding spacing between the adjacent gratings has been set to 930 mm. The other design parameters are the total number of gratings N and the gratings reflectivity as a function of wavelength or grating number. The goal of the filter design was to achieve a 3 dB bandwidth within 5-6 MHz, a 1 dB bandwidth larger than 3 MHz, and the sidelobe rejection level larger than 40 dB. In order to meet the above rejection level, a normalized Gaussian apodization of the taps weights has been employed. The total number of required gratings has been determined from the indicated above target bandwidth being finally fixed to $N = 30$. The last filter parameter to be determined was the FBG wavelength spacing that was set to 1 nm which ensured that the SLED optical spectrum (40 nm) was used efficiently. The 1 nm spacing has also established the length of the SMF-28 fiber required for proper switching of the operating frequency between the UMTS channels. We determined that 5 MHz steps towards lower RF frequencies require an increase of 23 picoseconds in the time delay ΔT which corresponds to about 1.35 km of SMF-28 fiber with the dispersion of 17 ps/(nm km) at 1550 nm.

The fabricated gratings have been made 1 cm long, Hamming-apodized, and had the 3 dB bandwidth of about 44 pm. Each grating has been fabricated separately and then the gratings have been spliced to each other with the 930 mm center-to-center spacing. Figure 28 shows the view over the 17–19th and over the 18th resonance of the filter response. Each of the eight traces in either figure corresponds to a different amount of dispersion provided by the reconfigurable

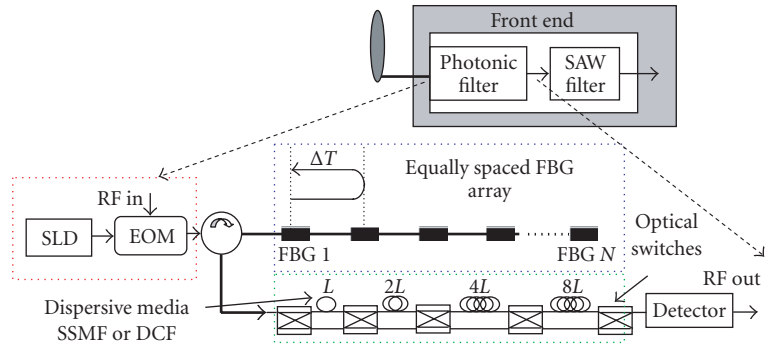
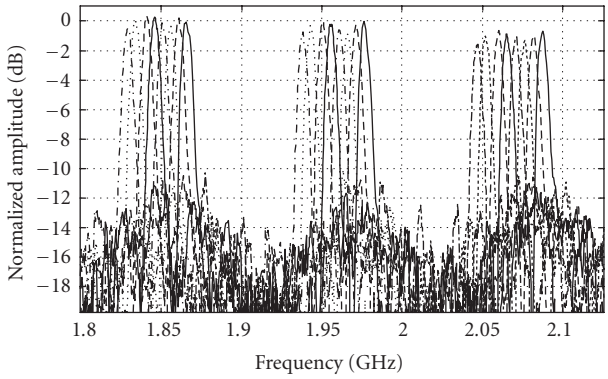
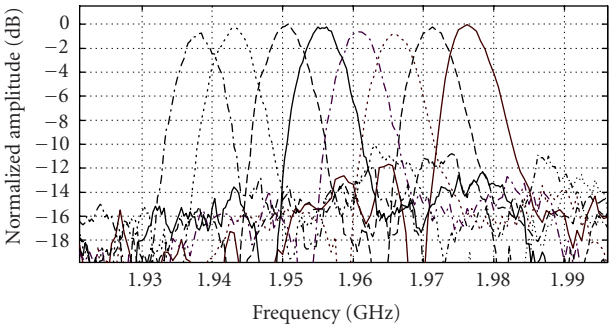


Figure 27: UMTS photonic filter setup.



(a)



(b)

Figure 28: (a) UMTS filter response over the 17–19th resonance and (b) detail over the 18th resonance. Each of the eight traces in either figure corresponds to a different amount of dispersion provided by the reconfigurable dispersive module in different steps from 0 km (higher-frequency trace) up to 10.7 km (lower-frequency trace).

dispersive module. The introduced optical insertion loss of 7 dB due to the 10.7 km long fiber and connectors (2 dB from the fiber and 5 dB from the connectors) has been compensated in the experiment by an adjustable EDFA prior to the detector. One can see that the operation frequency of the filter varies almost linearly versus the standard fiber length with the slope of 3.577 MHz/km (i.e., 1.39 km less or extra fiber is required for the shift of 5 MHz). The rejection levels have been obtained for each tuning position by measuring the range from the main lobe to the highest sidelobe level inside the rejected band, giving the 12, 13.4, 13.4, 12.8, 13, 11.4, 10.9, and 10.9 dB, respectively. The small MSLR is mainly due to the spacing errors between the gratings, which we can also conclude from the fact that it is frequency dependent (e.g., the measurements give MSLR > 24 dB at 500 MHz).

4. SUMMARY AND CONCLUSIONS

In this paper, we have presented the last advances in the implementation of photonic tunable transversal filters for RF signal processing. Following a distinction between filters based on wavelength tunable optical taps and others based on the tunability of the dispersive elements that provide the time delay between samples, last approaches have been discussed.

On one hand, interesting results related to the large degree of flexibility shown by the tunable laser array and chirped grating-based structure have been shown together with other cheaper alternatives based on the slicing of broadband optical sources either by using tunable fiber Bragg gratings or by using arrayed waveguide gratings. Other alternatives include the implementation of flexible and low-cost tunable transversal filters based on the generation of multiple taps by using an acousto-optic modulation on the fiber Bragg grating.

We also introduced structures that provide high-performance filters such as the bandpass filter implemented by using a Mach-Zehnder interferometer and a broadband optical source and discussed the implementation of filters with negative coefficients. These filters have been recently demonstrated by using two different approaches, which are also described in this paper: either by setting different bias voltage in an electro-optic modulator or by using fiber Bragg gratings in transmission; larger flexibility in the filter response given by the implementation of negative coefficients is demonstrated.

On the other hand, last proposals on tunable dispersive devices based on the tunability of the chirped fiber Bragg

gratings characteristics have also been addressed in a single stage or in a cascade by means of using an optical switch to increase the tunability range of these structures. Finally, a 30-tap transversal filter implemented by using a broadband optical source sliced by 30 uniform fiber gratings is designed for selecting channels in a UMTS application. In this structure, the physical spacing between the gratings provide the initial response and the filter tunability is achieved by means of a series of different switched fiber delay lines.

ACKNOWLEDGMENTS

The authors wish to acknowledge the EU-funded research projects IST-2001-37435 (LABELS) and IST-2001-32786 (NEFERTITI) and the MCYT support through the Ayudas a Parques Científicos and Infrestructuras FEDER programmes and projects TIC2002-04344 PROFECIA and TIC 2001-2969-c03-01 under which part of the work reported here is being carried.

REFERENCES

[1] A. J. Seeds, "Microwave photonics," *IEEE Trans. Microwave Theory Tech.*, vol. 50, no. 3, pp. 877–887, 2002.

[2] K. P. Jackson, S. A. Newton, B. Moslehi, et al., "Optical fiber delay-line signal processing," *IEEE Trans. Microwave Theory Tech.*, vol. 33, no. 3, pp. 193–210, 1985.

[3] K. Wilner and A. P. Van Den Heuvel, "Fiber-optic delay lines for microwave signal processing," *Proc. IEEE*, vol. 64, no. 5, pp. 805–807, 1976.

[4] C. Chang, J. A. Cassaboom, and H. F. Taylor, "Fibre optic delay line devices for RF signal processing," *Electronics Letters*, vol. 13, pp. 678–680, 1977.

[5] H. F. Taylor, "Fiber and integrated optical devices for signal processing," in *Proc. SPIE*, vol. 176, pp. 17–27, 1979.

[6] K. P. Jackson and H. J. Shaw, "Fiber-optic delay-line signal processors," in *Optical Signal Processing*, chapter 7, Academic Press, San Diego, Calif, USA, 1987.

[7] R. A. Minasian, K. E. Alameh, and E. H. W. Chan, "Photonics-based interference mitigation filters," *IEEE Trans. Microwave Theory Tech.*, vol. 49, no. 10, pp. 1894–1899, 2001.

[8] D. Pastor, B. Ortega, J. Capmany, P.-Y. Fonjallaz, and M. Popov, "Tunable microwave photonic filter for noise and interference suppression in UMTS base stations," *Electronics Letters*, vol. 40, no. 16, pp. 997–998, 2004.

[9] TSG-RAN Working Group 4, Rec. TSGR4#299048 Evaluation of up- and downlink adjacent channel performance, 3rd Generation Partnership Project 3GPP, www.3gpp.org, 1999.

[10] T. Sugiyama, M. Suzuki, and S. Kubota, "An integrated interference suppression scheme with adaptive equalizer for digital satellite communication systems," *IEICE Transactions on Communications*, vol. E79-B, pp. 191–196, 1996.

[11] M. I. Skolnik, *Introduction to Radar Systems*, McGraw-Hill, New York, NY, USA, 1980.

[12] B. Moslehi, "Fibre-optic filters employing optical amplifiers to provide design flexibility," *Electronics Letters*, vol. 28, no. 3, pp. 226–228, 1992.

[13] J. Capmany and J. Cascón, "Optical programmable transversal filters using fibre amplifiers," *Electronics Letters*, vol. 28, no. 13, pp. 1245–1246, 1992.

[14] V. Polo, F. Ramos, J. Martí, D. Moodie, and D. Wake, "Synthesis of photonic microwave filters based on external optical modulators and wide-band chirped fiber gratings," *J. Lightwave Technol.*, vol. 18, no. 2, pp. 213–220, 2000.

[15] B. Moslehi, K. Chau, and J. W. Goodman, "Optical amplifiers and liquid-crystal shutters applied to electrically reconfigurable fiber optic signal processors," *Optical Engineering*, vol. 32, pp. 974–981, 1993.

[16] J. Capmany and J. Cascón, "Direct form I fiber-optic discrete-time signal processors using optical amplifiers and embedded Mach-Zehnder structures," *IEEE Photon. Technol. Lett.*, vol. 5, no. 7, pp. 842–844, 1993.

[17] M. C. Vazquez, R. Civera, M. López-Amo, and M. A. Muriel, "Analysis of double-parallel amplified recirculating optical delay lines," *Applied Optics*, vol. 33, no. 6, pp. 1015–1021, 1994.

[18] A. Ho-Quoc and S. Tedjini, "Experimental investigation on the optical unbalanced Mach-Zehnder interferometers as microwave filters," *IEEE Microwave Guided Wave Lett.*, vol. 4, no. 6, pp. 183–185, 1994.

[19] D. Pastor, S. Sales, J. Capmany, J. Martí, and J. Cascón, "Amplified double-coupler fiber-optic delay line filter," *IEEE Photon. Technol. Lett.*, vol. 7, no. 1, pp. 75–77, 1995.

[20] J. Capmany, J. Cascón, J. L. Martín, and J. Martí, "Fibre-optic delay line filter synthesis using a modified PADE method," *Electronics Letters*, vol. 31, no. 6, pp. 479–480, 1995.

[21] S. Sales, J. Capmany, J. Martí, and D. Pastor, "Experimental demonstration of fibre-optic delay line filters with negative coefficients," *Electronics Letters*, vol. 31, no. 13, pp. 1095–1096, 1995.

[22] S. Sales, J. Capmany, D. Pastor, and J. Martí, "Fiber-optic delay line filters employing fiber loops: signal and noise analysis and experimental characterization," *Journal of the Optical Society of America* {*A*}, vol. 12, no. 10, pp. 2129–2135, 1995.

[23] J. Capmany, J. Cascón, J. L. Martín, S. Sales, D. Pastor, and J. Martí, "Synthesis of Fiber-optic delay line filters," *J. Lightwave Technol.*, vol. 13, no. 10, pp. 2003–2012, 1995.

[24] E. C. Heyde and R. A. Minasian, "A solution to the synthesis problem of recirculating optical delay line filters," *IEEE Photon. Technol. Lett.*, vol. 6, no. 7, pp. 833–835, 1994.

[25] S. Sales, J. Capmany, J. Martí, and D. Pastor, "Novel and significant results on the nonrecirculating delay line with a fibre loop," *IEEE Photon. Technol. Lett.*, vol. 7, no. 12, pp. 1439–1440, 1995.

[26] D. Norton, S. Johns, C. Keefer, and R. Soref, "Tunable microwave filtering using high dispersion fiber time delays," *IEEE Photon. Technol. Lett.*, vol. 6, no. 7, pp. 831–832, 1994.

[27] M. E. Frankel and R. D. Esman, "Fiber-optic tunable microwave transversal filter," *IEEE Photon. Technol. Lett.*, vol. 7, no. 2, pp. 191–193, 1995.

[28] S. Yegnanarayanan, P. D. Trinh, and B. Jalali, "Recirculating photonic filter: a wavelength-selective time delay for phased-array antennas and wavelength code-division multiple access," *Optics Letters*, vol. 21, no. 10, pp. 740–742, 1996.

[29] F. Coppinger, S. Yegnanarayanan, P. D. Trinh, B. Jalali, and I. L. Newberg, "Nonrecursive tunable photonic filter using wavelength-selective true time delay," *IEEE Photon. Technol. Lett.*, vol. 8, no. 9, pp. 1214–1216, 1996.

[30] F. Coppinger, S. Yegnanarayanan, P. D. Trinh, and B. Jalali, "Continuosly tunable photonic radio-frequency notch filter," *IEEE Photon. Technol. Lett.*, vol. 9, no. 3, pp. 339–341, 1997.

[31] N. You and R. A. Minasian, "A novel tunable microwave optical notch filter," *IEEE Trans. Microwave Theory Tech.*, vol. 49, no. 10, pp. 2002–2005, 2001.

[32] A. P. Foord, P. A. Davies, and P. A. Greenhalgh, "Synthesis of microwave and millimetre-wave filters using optical spectrum-slicing," *Electronics Letters*, vol. 32, no. 4, pp. 390–391, 1996.

[33] J. Capmany, D. Pastor, and B. Ortega, "Fibre-optic microwave and millimetre wave filter with high density sampling and very high sidelobe supression using subnanometre optical spectrum slicing," *Electronics Letters*, vol. 35, no. 6, pp. 494–496, 1999.

[34] D. Pastor, J. Capmany, and B. Ortega, "Broad-band tunable microwave transversal notch filter based on tunable uniform fiber Bragg gratings as slicing filters," *IEEE Photon. Technol. Lett.*, vol. 13, no. 7, pp. 726–728, 2001.

[35] J. Mora, B. Ortega, J. Capmany, et al., "Automatic tunable and reconfigurable fiberoptic microwave filters based on a broadband optical source sliced by uniform fiber Bragg gratings," *Optics Express*, vol. 10, no. 22, pp. 1291–1298, 2002.

[36] S. Yegnanarayanan, F. Coppinger, P. D. Trinh, and B. Jalali, "Microwave transversal filter based on spectral tapping of broadband light in an integrated waveguide prism, P CWF55," in *Conference on Lasers and Electro-Optics (CLEO '97)*, vol. 11, pp. 259–259, Baltimore, Md, USA, May 1997.

[37] D. Pastor, B. Ortega, J. Capmany, S. Sales, A. Martinez, and P. Muñoz, "Flexible and tunable microwave filters based on arrayed waveguide gratings," in *Proc. IEEE International Topical Meeting on Microwave Photonics (MWP '02)*, pp. 189–192, November 2002.

[38] T. A. Yost, P. R. Herczfeld, A. Rosen, and S. Singh, "Hybrid transversal filter utilizing MMIC and optical fiber delay lines," *IEEE Microwave Guided Wave Lett.*, vol. 5, no. 9, pp. 287–289, 1995.

[39] W. Zhang and J. A. R. Williams, "Fibre optic bandpass transversal filter employing fibre grating arrays," *Electronics Letters*, vol. 35, no. 12, pp. 1010–1011, 1999.

[40] D. Pastor, J. Capmany, S. Sales, P. Muñoz, and B. Ortega, "Reconfigurable fiber-optic-based RF filters using current injection in multimode lasers," *IEEE Photon. Technol. Lett.*, vol. 13, no. 11, pp. 1224–1226, 2001.

[41] G. A. Ball, W. H. Glenn, and W. W. Morey, "Programmable fiber optic delay line," *IEEE Photon. Technol. Lett.*, vol. 6, no. 6, pp. 741–743, 1994.

[42] D. B. Hunter, R. A. Minasian, and P. A. Krug, "Tunable optical transversal filter based on chirped gratings," *Electronics Letters*, vol. 31, no. 25, pp. 2205–2207, 1995.

[43] D. B. Hunter and R. A. Minasian, "Microwave optical filters using in-fiber Bragg grating arrays," *IEEE Microwave Guided Wave Lett.*, vol. 6, no. 2, pp. 103–105, 1996.

[44] W. Zhang and J. A. R. Williams, "Fibre optic bandpass transversal filter employing fibre grating arrays," *Electronics Letters*, vol. 35, no. 12, pp. 1010–1011, 1999.

[45] G. Yu, W. Zhang, and J. A. R. Williams, "High-performance microwave transversal filter using fiber Bragg grating arrays," *IEEE Photon. Technol. Lett.*, vol. 12, no. 9, pp. 1183–1185, 2000.

[46] W. Zhang, J. A. R. Williams, and I. Bennion, "Polarization synthesized optical transversal filter employing high birefringence fiber gratings," *IEEE Photon. Technol. Lett.*, vol. 13, no. 5, pp. 523–525, 2001.

[47] W. Zhang, G. Yu, and J. A. R. Williams, "Tap multiplexed fibre grating-based optical tranversal filter," *Electronics Letters*, vol. 36, no. 20, pp. 1708–1710, 2000.

[48] W. Zhang, J. A. R. Williams, and I. Bennion, "Optical fiber recirculating delay line incorporating a fiber grating array," *IEEE Microwave and Wireless Components Letters*, vol. 11, no. 5, pp. 217–219, 2001.

[49] D. B. Hunter and R. A. Minasian, "Microwave optical filters based on a fibre Bragg grating in a loop structure," in *Proc. International Topical Meeting on Microwave Photonics (MWP '96)*, pp. 273–276, Kyoto, Japan, December 1996.

[50] D. B. Hunter and R. A. Minasian, "Photonic signal processing of microwave signals using active-fiber Bragg-grating-pair structure," *IEEE Trans. Microwave Theory Tech.*, vol. 45, no. 8, pp. 1463–1466, 1997.

[51] N. You and R. A. Minasian, "A novel high-Q optical microwave processor using hybrid delay line filters," *IEEE Trans. Microwave Theory Tech.*, vol. 47, no. 7, pp. 1304–1308, 1999.

[52] R. A. Minasian, K. E. Alameh, and E. H. W. Chan, "Photonics-based interference mitigation filters," *IEEE Trans. Microwave Theory Tech.*, vol. 49, no. 10, pp. 1894–1899, 2001.

[53] E. H. W. Chan, K. E. Alameh, and R. A. Minasian, "Photonic bandpass filters with high skirt selectivity and stopband attenuation," *J. Lightwave Technol.*, vol. 20, no. 11, pp. 1962–1967, 2002.

[54] W. Zhang, J. A. R. Williams, L. A. Everall, and I. Bennion, "Fibre-optic radio frequency notch filter with linear and continous tuning by using a chirped fibre grating," *Electronics Letters*, vol. 34, no. 18, pp. 1770–1772, 1998.

[55] D. B. Hunter and R. A. Minasian, "Tunable microwave fiber-optic bandpass filters," *IEEE Photon. Technol. Lett.*, vol. 11, no. 7, pp. 874–876, 1999.

[56] D. Pastor and J. Capmany, "Fiber optic tunable transversal filter using laser array and linearly chirped fibre grating," *Electronics Letters*, vol. 34, no. 17, pp. 1684–1685, 1998.

[57] J. Capmany, D. Pastor, and B. Ortega, "Experimental demonstration of tunability and transfer function reconfiguration in fibre-optic microwave filters composed of linearly chirped fibre grating fed by a laser array," *Electronics Letters*, vol. 34, no. 23, pp. 2262–2264, 1998.

[58] J. Capmany, D. Pastor, and B. Ortega, "New and flexible fiber-optic delay-line filters using chirped Bragg gratings and laser arrays," *IEEE Trans. Microwave Theory Tech.*, vol. 47, no. 7, pp. 1321–1326, 1999.

[59] D. Pastor, J. Capmany, and B. Ortega, "Efficient sidelobe suppression by source power apodisation on fibre-optic microwave filters composed of linearly chirped fibre grating by laser array," *Electronics Letters*, vol. 35, no. 8, pp. 640–642, 1999.

[60] D. Pastor, J. Capmany, and B. Ortega, "Experimental demonstration of parallel fiber-optic-based RF filtering using WDM techniques," *IEEE Photon. Technol. Lett.*, vol. 12, no. 1, pp. 77–78, 2000.

[61] N. You and R. A. Minasian, "Synthesis of WDM grating-based optical microwave filter with arbitrary impulse response," in *Proc. International Topical Meeting on Microwave Photonics (MWP '99)*, vol. 1, pp. 223–226, Melbourne, Victoria, Australia, November 1999.

[62] J. Mora, B. Ortega, M. V. Andrés, et al., "Dynamic optical transversal filters based on a tunable dispersion fiber Bragg grating," in *Proc. International Topical Meeting on Microwave Photonics (MWP '01)*, pp. 203–206, Long Beach, Calif, USA, January 2001.

[63] A. V. Oppenheim, R. W. Schaffer, and J. R. Buck, *Discrete Time Signal Processing*, Prentice Hall, Englewood Cliffs, NJ, USA, 1996.

[64] W. Zhang, J. A. R. Williams, and I. Bennion, "Optical fibre delay line filter free of limitation imposed by optical coherence," *Electronics Letters*, vol. 35, no. 24, pp. 2133–2134, 1999.

[65] A. Ho-Quoc, S. Tedjini, and A. Hilt, "Optical polarization effect in discrete time fiber-optic structures for microwave signal processing," in *Proc. IEEE MTT-S International Microwave Symposium Digest*, vol. 2, pp. 907–910, San Francisco, Calif, USA, June 1996.

[66] F. Coppinger, S. Yegnanarayanan, P. D. Trinh, and B. Jalali, "All-optical incoherent negative taps for photonic signal processing," *Electronics Letters*, vol. 33, no. 11, pp. 973–975, 1997.

[67] M. Tur and B. Moslehi, "Laser phase noise effects in fiber-optic signal processors with recirculating loops," *Optics Letters*, vol. 8, no. 4, pp. 229–231, 1983.
[68] M. Tur, B. Moslehi, and J. W. Goodman, "Theory of laser phase noise in recirculating fiber optic delay lines," *J. Lightwave Technol.*, vol. 3, no. 1, pp. 20–30, 1985.
[69] M. Tur and A. Arie, "Phase induced intensity noise in concatenated fiber-optic delay lines," *J. Lightwave Technol.*, vol. 6, no. 1, pp. 120–130, 1988.
[70] B. Moslehi, "Analysis of optical phase noise in fiber-optic systems employing a laser source with arbitrary coherence time," *J. Lightwave Technol.*, vol. 4, no. 9, pp. 1334–1351, 1986.
[71] J. T. Kringlebotn and K. Blotekjaer, "Noise analysis of an amplified fiber-optic recirculating-ring delay line," *J. Lightwave Technol.*, vol. 12, no. 3, pp. 573–582, 1994.
[72] J. Capmany, "Investigation on phase-induced intensity noise in amplified fibre-optic recirculating delay line," *Electronics Letters*, vol. 29, no. 4, pp. 346–347, 1993.
[73] D. Pastor, B. Ortega, J. Capmany, S. Sales, A. Martinez, and P. Muñoz, "Flexible and tunable microwave filters based on arrayed waveguide gratings," in *Proc. IEEE International Topical Meeting on Microwave Photonics (MWP '02)*, pp. 189–192, Long Beach, Calif, USA, November 2002.
[74] D. Pastor, B. Ortega, J. Capmany, S. Sales, A. Martinez, and P. Muñoz, "Optical microwave filter based on spectral slicing by use of arrayed waveguide gratings," *Optics Letters*, vol. 28, no. 19, pp. 1802–1804, 2003.
[75] M. Delgado-Pinar, J. Mora, B. Ortega, A. Díez, M. V. Andrés, and J. Capmany, "Tunable and reconfigurable microwave filter by use of a Bragg-grating-based acousto-optic superlattice modulator," *Optics Letters*, vol. 30, no. 1, pp. 8–10, 2005.
[76] J. Mora, B. Ortega, M. V. Andrés, J. Capmany, A. Díez, and D. Pastor, "A single bandpass tunable photonic transversal filter based on a broadband optical source and a Mach-Zehnder interferometer," in *Proc. IEEE International Topical Meeting on Microwave Photonics (MWP '03)*, pp. 251–254, Budapest, Hungary, September 2003.
[77] F. Coppinger, S. Yegnanarayanan, P. D. Trinh, and B. Jalali, "All-optical RF filter using amplitude inversion in a semiconductor optical amplifier," *IEEE Trans. Microwave Theory Tech.*, vol. 45, no. 8, pp. 1473–1477, 1997.
[78] S. Li, K. S. Chiang, A. Gambling, Y. Liu, L. Zhang, and I. Bennion, "A novel tunable all-optical incoherent negative-tap fiber-optic transversal filter based on a DFB laser diode and fiber Bragg gratings," *IEEE Photon. Technol. Lett.*, vol. 12, no. 9, pp. 1207–1209, 2000.
[79] X. Wang and K. T. Chan, "Tunable all-optical incoherent bipolar delay-line filter using injection-locked Fabry-Perot laser and fibre Bragg gratings," *Electronics Letters*, vol. 36, no. 24, pp. 2001–2003, 2000.
[80] Y. Xiaoke, F. Wei, N. J. Hong, and L. Chao, "Tunable microwave filter design using wavelength conversion technique and high dispersion time delays," *IEEE Photon. Technol. Lett.*, vol. 13, no. 8, pp. 857–859, 2001.
[81] J. Mora, M. V. Andrés, J. L. Cruz, et al., "Tunable all-optical negative multi-tap microwave filters based on uniform fiber Bragg gratings," *Optics Letters*, vol. 28, no. 15, pp. 1308–1310, 2003.
[82] J. Capmany, D. Pastor, A. Martinez, B. Ortega, and S. Sales, "Microwave photonic filters with negative coefficients based on phase inversion in an electro-optic modulator," *Optics Letters*, vol. 28, no. 16, pp. 1415–1417, 2003.
[83] D. Pastor, J. Capmany, B. Ortega, A. Martinez, L. Pierno, and M. Varasi, "Reconfigurable RF photonic filter with negative coefficients and flat-top resonances using phase inversion in a newly designed 2×1 integrated Mach-Zehnder modulator," *IEEE Photon. Technol. Lett.*, vol. 16, no. 9, pp. 2126–2128, 2004.
[84] J. Mora, B. Ortega, M. V. Andrés, et al., "Tunable dispersion device based on a tapered fiber Bragg grating and nonuniform magnetic fields," *IEEE Photon. Technol. Lett.*, vol. 15, no. 7, pp. 951–953, 2003.
[85] J. Mora, B. Ortega, A. Díez, et al., "Highly tunable optically switched time delay line for transversal filtering," *Electronics Letters*, vol. 39, no. 25, pp. 1799–1800, 2003.

Beatriz Ortega was born in Valencia, Spain, in 1972. She received the M.S. degree in physics in 1995 from the Universidad de Valencia, and the Ph.D. degree in telecommunications engineering in 1999 from the Universidad Politécnica de Valencia. She joined the Departamento de Comunicaciones, Universidad Politécnica de Valencia, in 1996, where she was engaged with the Optical Communications Group and her research was mainly in the field of fiber gratings. From 1997 to 1998, she joined the Optoelectronics Research Centre, University of Southampton, United Kingdom, where she was involved in several projects developing new add-drop filters or twin-core fiber-based filters. She has published more than 60 papers and conference contributions in fiber Bragg gratings, microwave photonics, and fiber filters. Currently, she is an Associate Lecturer at the Telecommunications Engineering Faculty and her main interests include fiber gratings applications, optical delay lines, and optical networks.

Daniel Pastor was born in Elda, Spain, on November 5, 1969. He received the Ingeniero de Telecomunicacion degree in 1993 from the Universidad Politécnica de Valencia, and the Doctor Ingeniero de Telecomunicacion (Ph.D.) degree in 1996 from the Universidad Politécnica de Valencia, Spain. He joined the Departamento de Comunicaciones, Universidad Politécnica de Valencia, in 1993, where he was engaged with the Optical Communications Group. From 1994 to 1998 he was a Lecturer at the Telecommunications Engineering Faculty and he became an Associate Professor in 1999. He has published over 45 papers and conference contributions in the fields of optical delay-line filters, fiber Bragg gratings, microwave photonics, and WDM and SCM lightwave systems. His current technical interests include microwave photonics, fiber Bragg grating applications, and WDM networks.

José Mora was born in Torrent, Valencia, Spain, in 1976. He received the Licenciado en Física degree in 1999 from the Universidad de Valencia, Spain. From 1999 to 2004 he has served as Ph.D. student in the Departamento de Física Aplicada, Universidad de Valencia. Currently, he is a Researcher in the Optical Communications Group, Universidad Politécnica de Valencia. His research activities are focused on the use of fiber Bragg gratings in fields such as sensor applications and telecommunication systems.

José Capmany was born in Madrid, Spain, in 1962. He received the Ingeniero de Telecomunicacion degree from the Universidad Politécnica de Madrid in 1987. From 1988 to 1991, he worked as a Research Assistant at the Departamento de Tecnología Fotónica, Universidad Politécnica de Madrid, where he received the Ph.D. degree in 1991. In 1991 he moved to the Departamento de Comunicaciones, Universidad Politécnica de Valencia, where he started the activities on optical communications and photonics, founding the Optical Communications Group (www.gco.upv.es). He has been an Associate Professor from 1992 to 1996, and Full Professor in optical communications, systems, and networks since 1996. Since 2002, he has also been the Director of the IMCO2 Research Institute, Universidad Politécnica de Valencia. His research activities and interests cover a wide range of subjects related to optical communications including optical signal processing, fiber resonators, fiber gratings, RF filters, SCM, WDM, and CDMA transmission, wavelength conversion, and optical bistability. He has published over 200 papers in international refereed journals and conferences and has been a Member of the Technical Programme Committees of the European Conference on Optical Communications (ECOC), the Optical Fiber Conference (OFC), the Integrated Optics and Optical Communications Conference (IOOC), CLEO Europe.

Miguel V. Andrés was born in Valencia, Spain, in 1957. He received the Licenciado en Física degree in 1979, and the Doctor en Física (Ph.D.) degree in 1985, both from the Universidad de Valencia, Spain. From 1983 he has served successively as Assistant Professor and Lecturer in the Departamento de Física Aplicada, Universidad de Valencia. From 1984 to 1987 he was visiting for several periods the Department of Physics, University of Surrey, UK, as a Research Fellow. Until 1984 he was engaged in research on microwave surface waveguides. His current research interests include waveguide theory (inhomogeneous waveguides and microstructured optical fibres) and optical fiber devices and systems for microwave photonics and sensor applications (optical fiber interferometers, evanescent field devices based on optical fiber tapers, in-fiber Bragg gratings, and photonic crystal fibres). Dr. M. V. Andrés is a Member of IEEE, OSA, and IOP.

EURASIP Journal on Applied Signal Processing 2005:10, 1485–1497

DSP Approach to the Design of Nonlinear Optical Devices

Geeta Pasrija
Department of Electrical and Computer Engineering, University of Utah, Salt Lake City, UT 84112, USA
Email: pasrija@eng.utah.edu

Yan Chen
Department of Electrical and Computer Engineering, University of Utah, Salt Lake City, UT 84112, USA
Email: ychen@ece.utah.edu

Behrouz Farhang-Boroujeny
Department of Electrical and Computer Engineering, University of Utah, Salt Lake City, UT 84112, USA
Email: farhang@ece.utah.edu

Steve Blair
Department of Electrical and Computer Engineering, University of Utah, Salt Lake City, UT 84112, USA
Email: blair@ece.utah.edu

Received 5 April 2004; Revised 19 October 2004

Discrete-time signal processing (DSP) tools have been used to analyze numerous optical filter configurations in order to optimize their linear response. In this paper, we propose a DSP approach to design nonlinear optical devices by treating the desired nonlinear response in the weak perturbation limit as a discrete-time filter. Optimized discrete-time filters can be designed and then mapped onto a specific optical architecture to obtain the desired nonlinear response. This approach is systematic and intuitive for the design of nonlinear optical devices. We demonstrate this approach by designing autoregressive (AR) and autoregressive moving average (ARMA) lattice filters to obtain a nonlinear phase shift response.

Keywords and phrases: DSP tools, nonlinear optical devices, nonlinear phase shift.

1. INTRODUCTION

In order to satisfy the ever-increasing demand for high bit rates, next generation optical communication networks can be made all-optical to overcome the electronic bottleneck and more efficiently utilize the intrinsic broad bandwidth of optical fibers. Currently, there are two possible technologies for achieving high transmission rate: optical time division multiplexing (OTDM) and dense wavelength division multiplexing (DWDM). However, neither the full potential of OTDM nor that of DWDM technology has been realized due to lack of suitable nonlinear, all-optical devices that can perform signal regeneration, ultrafast switching, encoding/decoding, and/or wavelength conversion efficiently. There are a number of problems with current nonlinear optical materials and devices.

There are two types of nonlinear optical materials from which devices can be made: nonresonant and resonant. Nonresonant materials have a weak nonlinear response, but the passage of light occurs with very low loss and the response is broadband, typically exceeding 10 THz. However, because of the weak nonlinear response, these devices tend to be bulky and impose a long latency. Resonant materials have a very strong nonlinear response, but at the expense of reduced bandwidths and increased loss. Artificial resonances can be used in optical architectures to overcome the limitations of current nonlinear devices and materials [1]. In this paper, we design nonlinear optical devices that exhibit enhanced nonlinear phase shift response using microring resonators constructed from nonresonant nonlinear material.

The nonlinear optical response of many artificial resonant structures has been studied previously, but most of the

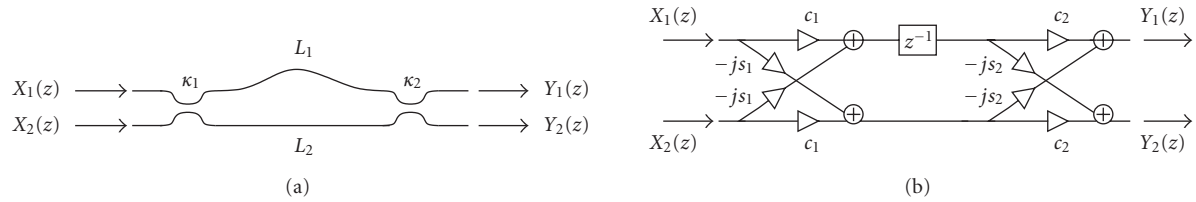


Figure 1: MZI device [2]. (a) Waveguide layout. (b) z-schematic.

studies have been limited to analyzing the nonlinear properties of specific architectures and do not provide a synthesis approach to device design that can produce a specific nonlinear response. Discrete-time signal processing (DSP) provides an easy to use mathematical framework, the z-transform, for the description of discrete-time filters. The z-transform has already been used to analyze numerous optical filter configurations in order to optimize their linear response [2]. We propose a similar approach to optimize the nonlinear response by treating the nonlinear response in the weak perturbation limit as a linear discrete-time filter. The field of discrete-time filter design has been extensively researched and various algorithms are available for designing and optimizing discrete-time filters. In this paper, we use existing discrete-time[1] filter design algorithms to design nonlinear optical devices.

This paper shows that the DSP approach is a systematic and intuitive way to design nonlinear optical devices. Six steps are involved in designing a nonlinear optical device using the DSP approach. First, a prototype linear frequency response (in the weak perturbation limit) is selected for the desired nonlinear optical device. Next, the optical architecture's unit cell is selected and the multistage optical architecture is analyzed using the z-transform. Then, an optimized discrete filter is designed to give the same frequency response as the prototype response desired from the optical architecture in the weak perturbation limit. Next, a mapping algorithm is derived to synthesize the parameters of the optical architecture from the discrete filter. The synthesized optical filter is then simulated using electromagnetic models and its linear response is verified to be the same as that of the discrete filter. Finally, the optical device is simulated to evaluate the desired nonlinear response and confirm the design.

This approach can be used to design optical devices to obtain various nonlinear responses, for example, all-optical switching [3, 4], nonlinear phase shift [5, 6, 7], second-harmonic generation [8], four-wave mixing [9, 10] (i.e., frequencies ν_m and ν_n mix to produce $2\nu_m - \nu_n$ and $2\nu_n - \nu_m$), solitons [11, 12, 13] (which is a carrier of digital information), bistability [14, 15, 16] (which results in two stable, switchable output states and can be used as a basis for logic operations and thresholding with restoration), and amplification (which can overcome loss). The nonlinear phase shift is a fundamental nonlinear process that enables many all-optical switching and logic devices, and is the process used to demonstrate our approach. Artificial resonant structures are used in the devices to overcome the aforementioned traditional drawbacks.

The rest of this paper is organized as follows. Section 2 provides some background on optical filters in relation to discrete-time filters. Section 3 explains the nonlinear phase shift process. Section 4 describes the prototype linear response desired for the nonlinear phase shift. Section 5 discusses the selection of optical architectures. Section 6 details the design procedure for AR and ARMA discrete filters. Sections 7 and 8 outline the mapping of discrete filters on to the optical architectures and their optical response, respectively. Sections 9 and 10 discuss an example and evaluation of AR lattice filters and ARMA lattice filters, respectively, followed by conclusions.

2. OPTICAL FILTERS AND z-TRANSFORMS

Discrete filters are designed and analyzed using z-transforms. In this section, we discuss the important aspects of optical filters in relation to discrete filters, and explain how z-transforms can be used to describe optical filters as well. This section borrows heavily from Madsen and Zhao's book on optical filters [2]. Like discrete filters, optical filters are completely described by their frequency response. Filters are broadly classified into two categories: finite impulse response (FIR) and infinite impulse response (IIR). FIR filters have no feedback paths between the output and input and their transfer function has only zeros. These are also referred to as moving average (MA) filters. IIR filters have feedback paths and their transfer functions have poles and may or may not have zeros. When zeros are not present or all the zeros occur at the origin, IIR filters are referred as autoregressive (AR) filters. When both poles and nonorigin zeros are present, they are referred to as autoregressive moving average (ARMA) filters.

Optical architectures can be of restricted type or general type. With restricted architectures, we cannot obtain arbitrary frequency response, while general architectures, like discrete filters, allow arbitrary frequency response to be approximated over a frequency range of interest. To approximate any arbitrary function in discrete-time signal processing, a set of sinusoidal functions whose weighted sum yields a Fourier series approximation is used. The optical analog is found in interferometers. Interferometers come in two general classes: (1) Mach-Zehnder interferometer (MZI), and (2) Fabry-Perot interferometer (FPI). MZI is shown in Figure 1a and has finite number of delays and no recirculating (or feedback) delay paths. Therefore, these are MA filters. FPI consists of a cavity surrounded by two partial

[1] Henceforth, *discrete-time filters* will be referred to as *discrete filters*.

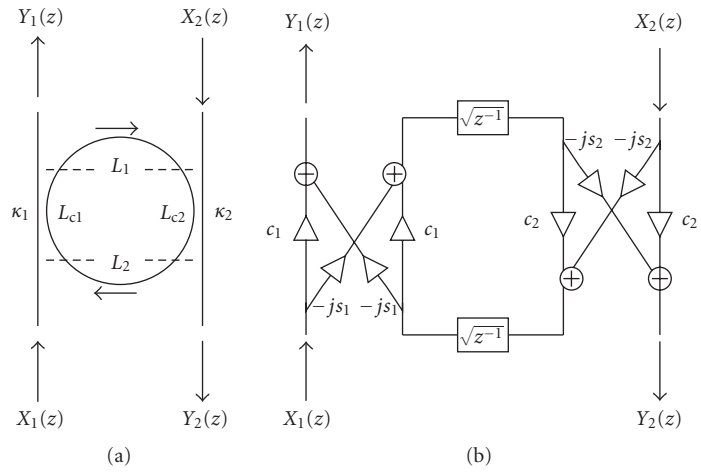


FIGURE 2: Ring resonator. After [2]. (a) Waveguide layout. (b) z-schematic.

reflectors that are parallel to each other. The waveguide analog of the FPI is the ring resonator shown in Figure 2a. The output is the sum of delayed versions of the input signal weighted by the roundtrip cavity transmission. The transmission response is of AR type while the reflection response is of ARMA type. The ring resonator is an example of an artificial resonator.

The z-transform schematics for the MZI and FPI device are shown in Figures 1b and 2b, respectively. κ is the power coupling ratio for each directional coupler, $c = \sqrt{1-\kappa}$ is the through-port transmission term, and $-js = -j\sqrt{\kappa}$ is the cross-port transmission term. Also, $z = e^{j\Omega T}$, and $\Omega T = \beta Ł_u$, where L_u is the smallest path length called the unit delay length, T is the unit delay and is equal to $L_u n/c$, β is the propagation constant and is equal to $2\pi n/\lambda$, n is the refractive index of the material, c is the speed of light in vaccum, and λ is the wavelength of light. Propagation loss of a delay line is accounted for by multiplying z^{-1} by $\gamma = 10^{-\alpha L/20}$, where α is the average loss per unit length in dB, and L is the delay path length. Because delays are discrete values of the unit delay, the frequency response is periodic. One period is defined as the free spectral range (FSR) and is given by FSR $= 1/T$. The normalized frequency, $f = \omega/2\pi$, is related to the optical frequency by $f = (\nu - \nu_c)T$, or $f = (\Omega - \Omega_c)T/2\pi$. The center frequency $\nu_c = c/\lambda_c$ is defined so that the product of refractive index and unit length is equal to an integer number of wavelengths, that is, $m\lambda_c = nL_u$, where m is an integer.

To analyze the frequency response of the MZI, the unit delay is set equal to the difference in path lengths, $L_u = L_1 - L_2$. The overall transfer function matrix of the MZI is the product of the matrices:

$$\begin{aligned}\mathbf{\Phi}_{MZI} &= \mathbf{\Phi}_{cplr}(\kappa_2)\mathbf{\Phi}_{delay}\mathbf{\Phi}_{cplr}(\kappa_1)\\ &= \begin{bmatrix} c_2 & -js_2 \\ -js_2 & c_2 \end{bmatrix}\begin{bmatrix} z^{-1} & 0 \\ 0 & -1 \end{bmatrix}\begin{bmatrix} c_1 & -js_1 \\ -js_1 & c_1 \end{bmatrix}.\end{aligned} \tag{1}$$

For the ring resonator, the unit delay is equal to $L_u = L_1 + L_2 + L_{c1} + L_{c2}$, where L_{c1} and L_{c2} are the coupling region lengths for each coupler. The sum of all-optical paths is given by

$$Y_2(z) = -s_1 s_2\sqrt{\gamma z^{-1}}\left(1 + c_1 c_2 \gamma z^{-1} + c_1^2 c_2^2 \gamma^2 z^{-2} + \cdots\right)X_1(z). \tag{2}$$

The infinite sum simplifies to the following expression for the ring's transfer function:

$$H_{21}(z) = \frac{Y_2(z)}{X_1(z)} = \frac{-\sqrt{\kappa_1\kappa_2\gamma z^{-1}}}{1 - c_1 c_2 \gamma z^{-1}}. \tag{3}$$

Other responses for the ring resonator can similarly be obtained. Hence we see that optical resonances are represented by poles in a filter transfer function. Therefore the filters built using artificial resonances are IIR filters.

We have used the MZI and microring resonator as the building blocks to design the nonlinear optical devices for obtaining nonlinear phase shift in this paper. Detailed description of using z-transforms for analyzing single-stage and multistage optical filters is provided in [2].

3. NONLINEAR OPTICAL PROCESSES

Nonlinear optics is the study of phenomena that occur as a consequence of the modification of the optical properties of a material under intense illumination. Typically, only laser light is sufficiently intense to modify the optical properties of a material. Nonlinear optical phenomena are *nonlinear* in the sense that the induced material polarization is nonlinear in the electric field:

$$\mathbf{P} = \underbrace{\epsilon_o\mathbf{E} + \epsilon_o\overline{\overline{\chi}}^{(1)} : \mathbf{E}}_{\text{linear } \mathbf{P}^L} + \underbrace{\epsilon_o\overline{\overline{\overline{\chi}}}^{(2)} :: \mathbf{E}\cdot\mathbf{E} + \epsilon_o\overline{\overline{\overline{\chi}}}^{(3)} ::: \mathbf{E}\cdot\mathbf{E}\cdot\mathbf{E} + \cdots}_{\text{nonlinear } \mathbf{P}^{NL}}, \tag{4}$$

where dielectric dispersion is ignored. The optical Kerr effect (i.e., nonlinear refraction index) results from the third-order nonlinear susceptibility $\overline{\overline{\overline{\chi}}}^{(3)}$, which is a fourth-rank tensor.

An optical wave is a real quantity and is usually expressed as

$$\mathbf{E}(t) = \mathrm{Re}\{\mathbf{E}\exp j(\mathbf{k}\cdot\mathbf{r} + \omega t)\}, \tag{5}$$

or similarly as

$$\mathbf{E}(t) = \frac{1}{2}\mathbf{E}\exp j(\mathbf{k}\cdot\mathbf{r}+\omega t) + \text{cc}, \tag{6}$$

where cc represents the complex conjugate of the preceding term. Thus, an x-polarized optical wave, propagating in z-direction in an isotropic medium, is represented mathematically as

$$\mathbf{E}(t) = \frac{1}{2}E_x\hat{x}\exp j(kz+\omega t) + \text{cc}. \tag{7}$$

3.1. Nonlinear phase shift

The third-order polarization (mediated by $\chi^{(3)}$) in a material leads to a nonlinear intensity dependent contribution to its refractive index, that is, the refractive index of the material changes as the incident intensity on the material changes. The susceptibility tensors in isotropic material can be further simplified as $\overline{\overline{\chi}}^{(1)} = \chi^{(1)}$, being a scalar quantity, and $\overline{\overline{\chi}}^{(2)} = 0$, due to inversion symmetry. The third-order nonlinear susceptibility will only have one contributing term χ_{xxxx} since the light is x-polarized and there are no means for sourcing additional polarization components. The linear- and nonlinear-induced polarizations are

$$\begin{aligned} P^{\text{L}} &= \epsilon_o(1+\chi^{(1)})E, \\ P^{\text{NL}} &= P^{(3)} \\ &= \epsilon_o\chi_{xxxx}(\omega;-\omega,\omega,\omega)E^*EE \\ &\quad + \epsilon_o\chi_{xxxx}(\omega;\omega,-\omega,\omega)EE^*E \\ &\quad + \epsilon_o\chi_{xxxx}(\omega;\omega,\omega,-\omega)EEE^* \\ &= 3\epsilon_o\chi_{xxxx}|E|^2E \\ &= \frac{3}{4}\epsilon_o\chi_{xxxx}\left|E_x\right|^2E, \end{aligned} \tag{8}$$

respectively. Hence,

$$P = P^{\text{L}} + P^{\text{NL}} = \epsilon_o\left(1+\chi^{(1)}+\frac{3}{4}\epsilon_o\chi_{xxxx}\left|E_x\right|^2\right)E. \tag{9}$$

The total dielectric constant is

$$\epsilon_r^{\text{tot}} = \epsilon_r + \Delta\epsilon_r. \tag{10}$$

Comparing with the expression for P, we obtain $\epsilon_r = 1+\chi^{(1)} = n_o^2$ and $\Delta\epsilon = (3/4)\chi_{xxxx}|E_x|^2$. The refractive index is related to the dielectric constant as

$$n = \sqrt{\epsilon_r+\Delta\epsilon_r} \approx \sqrt{\epsilon_r} + \frac{\Delta\epsilon_r}{2\sqrt{\epsilon_r}} = n_o + \frac{3\chi_{xxxx}}{8n_o}\left|E_x\right|^2. \tag{11}$$

The intensity dependent refractive index for a nonlinear material is given by

$$n = n_o + n_2|E|^2. \tag{12}$$

Comparing (11) and (12), the nonlinear refractive index is directly determined by the third-order susceptibility as

$$n_2 = \frac{3\chi_{xxxx}}{8n_o} = \frac{3\chi^{(3)}}{8n_o}, \tag{13}$$

which characterizes the strength of the optical nonlinearity. The intensity I of an optical wave is proportional to $|E|^2$ as $I = (1/2\eta)|E|^2$ where η is the impedance of the medium. When comparing the optical response in the same medium, $I = |E|^2$ is taken for simplification.

This intensity dependent refractive index, in turn, results in various processes, one of which is the nonlinear phase shift. For a material with positive n_2, increasing the intensity results in a red shift of the frequency response of an optical filter. This can be explained using the equation $nL_{\text{u}} = m\lambda_{\text{c}} \Rightarrow (n_o+n_2|E|^2)L_{\text{u}} = m\lambda_{\text{c}}$, where m is an integer. The product nL_{u} is called the optical path length. Increasing intensity I results in the increase of optical path length and wavelength λ_{c}, and hence a decrease in the center frequency ν_{c} causing a red shift of the frequency response. When optical path length is increased by varying L_{u} and keeping n constant, the red shift will be perfect and the shape of the frequency response curve will not change. In nonlinear materials, the refractive index n as well as the loss in the material changes with changing intensity and hence the red shift is not perfect.

As discussed, current nonlinear optical materials and devices either have weak nonlinear response (nonresonant materials) or have high loss (resonant materials). Using artificial resonances, for example, microring resonators made of nonresonant nonlinear material, we can obtain strong nonlinear response with low loss [1]. Light circulates within the resonator and coherent interference of multiple beams occurs, resulting in intracavity intensity build-up and group delay enhancement which in turn enhances the nonlinear response.

4. PROTOTYPE RESPONSE FOR NONLINEAR PHASE SHIFT

The nonlinear phase shift is a fundamental nonlinear process that enables many all-optical switching and logic devices [5] that can be used in the next generation optical communication systems. An ideal nonlinear phase shifting element has constant intensity transmission up to at least a π radian phase shift upon increasing the incident intensity. The lesser the intensity required to obtain a π phase shift, the better the nonlinear performance.

The first step in the design approach is to select a linear frequency response for the desired device. Figure 3 illustrates the notion of producing a nonlinear phase shift response through the nonlinear detuning of a periodic (discrete) filter. To act as an ideal nonlinear phase shifter, in the weak perturbation limit, a flat magnitude response and steep linear phase are desired within the passband.

Light incident on the filter (at a frequency ν_m, e.g.) will be transmitted with efficiency given by the magnitude response, but will also experience a phase change due to the

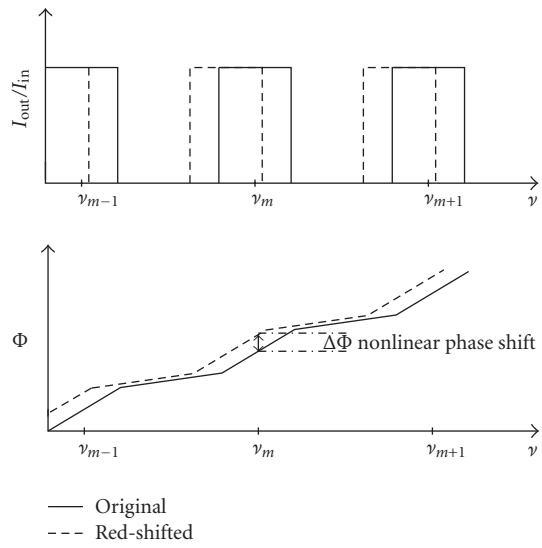


Figure 3: Prototype linear response for nonlinear phase shift.

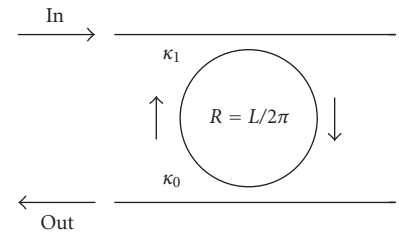


Figure 4: Single-pole structure.

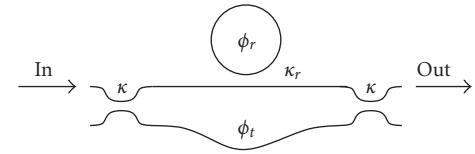


Figure 5: Independent pole-zero structure.

phase response. As the light intensity increases, the overall filter response will red shift due to intensity-induced changes in the filter components, which are themselves constructed from (weakly) nonlinear materials. Ideally, under weak detuning, the transmitted intensity fraction will not change (and hence the desire for a flat-topped magnitude response), but the phase at the output will change due to a steep linear phase response within the filter passband. The slope of the phase determines the group delay. Ripples in group delay may result in bistability in the nonlinear response, and therefore, linear phase is desired in the passband to have constant group delay. In effect, what this approach does is to amplify the intrinsic nonlinearity of a material, where the efficiency of the process improves with increasing the filter group delay. However, strong detuning in multiresonator systems can result in distortions of the filter response.

The red-shifted response is shown by the dotted curve in Figure 3. It can be seen that the transmitted output does not change (in the weak perturbation limit) and a nonlinear phase shift is obtained because of the shifted phase response. An increase in the input intensity I_{in} results in greater red shift and hence more nonlinear phase shift. The input intensity at which a π phase shift is obtained is denoted as I_π. The nonlinear phase shift response should be such that a phase shift of π can be obtained at a lower input intensity, I_π, than that required for the bulk material. The lower the I_π, the better the filter. Also, the transmission ratio at the intensity at which π phase shift is obtained should be at least 0.5, for maximum of 3 dB insertion loss.

5. OPTICAL ARCHITECTURES FOR NONLINEAR PHASE SHIFTER

The second step is to select the optical architecture's unit cell and analyze it using the z-transform. Artificial resonances produced by ring resonators can be used to enhance the nonlinear phase shift response of an optical device [1, 7]. The presence of a ring resonator in the architecture implies the presence of a pole in the filter's transfer function. To select the optical architecture for obtaining a nonlinear phase shift response, we analyze two ring resonator configurations (1) single pole (2) single pole-zero with the pole and zero independent of each other.

(i) Single-pole design. Figure 4 shows a single-pole architecture with a zero at the origin. The transfer function for this architecture in the z-domain is given by

$$\frac{E_{out}(z)}{E_{in}(z)} = \frac{\sqrt{\kappa_0\kappa_1}\sqrt{\gamma e^{-j\phi}z^{-1}}}{1 - c_0c_1\gamma e^{-j\phi}z^{-1}}. \tag{14}$$

The total phase change in the fundamental range $-\pi \le \omega \le \pi$ for this unit cell is equal to π. By cascading N such unit cells, we can obtain a total phase change of $N\pi$ in the fundamental range.

(ii) Single pole-zero design with independent pole and zero. Figure 5 shows a single pole-zero architecture with the pole and zero independent of each other. The transfer function for this architecture in the z-domain is given by

$$\frac{E_{out}(z)}{E_{in}(z)} = \frac{\left(c^2c_r - s^2e^{-j\phi_t}\right) - \left(c^2e^{-j\phi_r} - s^2c_re^{-j(\phi_r+\phi_t)}\right)z^{-1}}{1 - c_re^{-j\phi_r}z^{-1}}. \tag{15}$$

The total phase change in the fundamental range $-\pi \le \omega \le \pi$ for this unit cell is equal to 2π if the filter is maximum phase, and 0 if it is minimum phase. We are interested in lowpass maximum phase systems ($|\text{zero}| > 1/|\text{pole}|$) since they have the maximum net phase change and most of the phase change lies within the passband. The architecture shown in Figure 5 can be designed to be a lowpass maximum phase system since the poles and zeros are independent of each other. By cascading N such unit cells, we can obtain a total phase change of $2N\pi$ in the fundamental range.

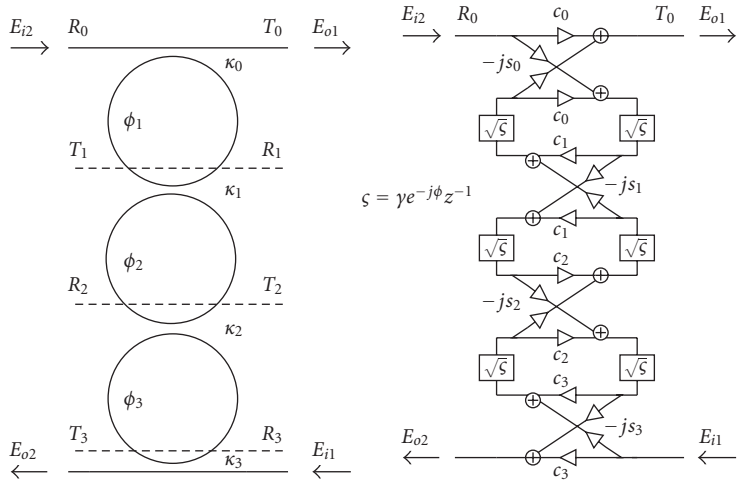


Figure 6: AR lattice filter [2].

A third possible configuration is a ring resonator with a single coupler. However, this is a pole-zero architecture with dependent pole and zero and is always highpass for a maximum phase system. The total phase change is equal to 2π but most of the phase change is present in the stopband and hence, we cannot obtain the prototype response of Figure 3 using this unit cell. Therefore, we decided to use the first and second configurations as the unit cells for our designs. Joining the first configuration unit cell in a lattice structure gives us an AR lattice filter architecture shown in Figure 6 and joining the second configuration unit cell in a lattice structure gives us an ARMA lattice filter architecture shown in Figure 7. Lattice structures are chosen since they have low passband loss and can operate at significantly higher component variations as compared to transversal or cascade structures.

The next step is to obtain a z-transform description of the multistage architecture obtained by joining the unit cells. First, a DSP schematic is drawn for the architecture and then it is analyzed to obtain a transfer function matrix. The AR and ARMA lattice architecture's DSP schematics and transfer functions are given below. The detailed derivations are presented in [2].

(i) AR lattice filter. Figure 6 shows the waveguide layout and DSP schematic of an AR lattice architecture. The transfer matrix for this architecture is [2]

$$\begin{bmatrix} T_{n+1}(z) \\ R_{n+1}(z) \end{bmatrix} = \mathbf{\Phi}_N \mathbf{\Phi}_{N-1} \cdots \mathbf{\Phi}_1 \mathbf{\Phi}_0 \begin{bmatrix} T_0(z) \\ R_0(z) \end{bmatrix}, \tag{16}$$

where

$$\mathbf{\Phi}_n = \frac{1}{-js_n\sqrt{\gamma e^{-j\phi_{n+1}} z^{-1}}} \begin{bmatrix} 1 & -c_n \\ c_n \gamma e^{-j\phi_{n+1}} z^{-1} & -\gamma e^{-j\phi_{n+1}} z^{-1} \end{bmatrix}. \tag{17}$$

(ii) ARMA lattice filter. Figure 7 shows the waveguide layout and DSP schematic of an ARMA lattice architecture. The transfer matrix for this architecture is [2]

$$\begin{bmatrix} X_n(z) \\ Y_n(z) \end{bmatrix} = \mathbf{\Phi}_N \mathbf{\Phi}_{N-1} \cdots \mathbf{\Phi}_1 \mathbf{\Phi}_0 \begin{bmatrix} X_0(z) \\ Y_0(z) \end{bmatrix}, \tag{18}$$

where

$$\mathbf{\Phi}_n = \frac{1}{A_n(z)} \begin{bmatrix} -c_{nt} A_n^R(z) e^{-j\phi_{nr}} & -js_{nt} A_n(z) e^{-j\phi_{nt}} \\ js_{nt} A_n^R(z) e^{-j\phi_{nr}} & c_{nt} A_n(z) e^{-j\phi_{nt}} \end{bmatrix}, \tag{19}$$

$$A_n = 1 - c_{nr} e^{-j\phi_{nr}} z^{-1}, \qquad A_n^R = -c_{nr} + e^{-j\phi_{nr}} z^{-1}. \tag{20}$$

6. DESIGN OF ARMA AND AR DISCRETE FILTERS

The next step is to design discrete filters to be mapped onto AR and ARMA lattice architectures with the response as shown in Figure 3 (where the number of stages, i.e., poles and zeros are given). For mapping onto the AR lattice architecture having N rings (unit cells), an Nth-order discrete AR filter (N poles, no zeros) is designed. Similarly, for mapping onto the ARMA lattice architecture having N stages, an Nth-order discrete ARMA filter (N poles and N zeros) is designed. The discrete filter design procedure for designing AR and ARMA filters is described below. The design needs to meet the constraints of linear phase within the passband with as high group delay as possible, and flat magnitude response with as large bandwidth as possible.

6.1. Design of AR discrete filters

Each stage of the AR optical architecture represents a pole in the transfer function. Therefore, the discrete filter designed to be mapped on this architecture should have only poles. To obtain the nonlinear phase shift, the AR discrete filter should be designed to obtain the prototype response

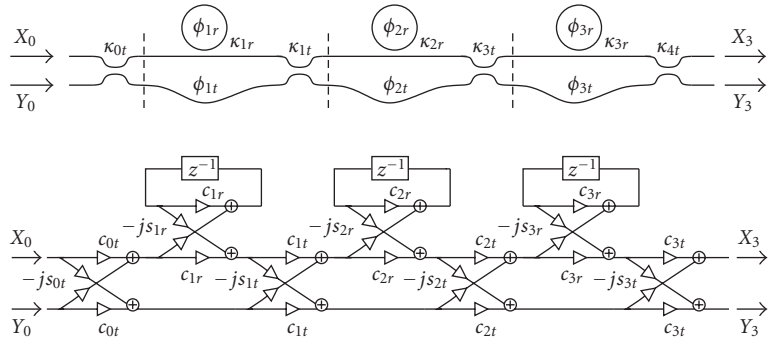


Figure 7: ARMA lattice filter [2].

of Figure 3. The prototype response requires a flat passband and linear phase within the passband. If $H(z)$ is the transfer function of the discrete filter, the condition to obtain linear phase is $H(z^{-1}) = z^{-\Delta}H(z)$, where Δ is a delay. In the case of IIR filters, since all poles are inside the unit circle, satisfying the above condition requires that there are mirror image poles outside the unit circle thereby making the filter unstable. Therefore, stable IIR filters can only approximate a linear phase response.

In the next subsection, we formulate the problem of ARMA discrete filter design as a least squares minimization problem. Since the case of AR filters can be thought as a special case of ARMA filters with all zeros at origin, the least squares formulation of ARMA filter design can be easily adopted to AR filters as well. However, unfortunately, numerical examples reveal that this approach results in either unstable IIR filters or, if the poles of the filter are constrained to the stable region, $|z| < 1$, the group delay of the resulting filter will be unsatisfactory. Therefore, other methods of filter design have to be adopted. Selesnick and Burrus [17] have proposed a generalized Butterworth discrete filter design procedure that allows arbitrary constraints to be imposed on the number of poles and nontrivial zeros, that is, zeros other than those at the origin. Hence, it can be adopted for designing AR filters. The designs satisfy the condition of maximally flat magnitude response at the center of passband, the Butterworth condition. This fulfills the required flat passband response. The filter's group delay shows some variation over the passband. However, it remains relatively flat over a good portion of the passband, which, to some extent, satisfies the constant group delay condition.

The generalized Butterworth filter design uses the mapping $x = (1/2)(1 - \cos(\omega))$ and provides formulas for two real and nonnegative polynomials $P(x)$ and $Q(x)$ where $P(x)/Q(x)$ resembles a lowpass response, over the range $x \in [0, 1]$ (equivalent to $\omega \in [0, \pi]$). A stable IIR filter $B(z)/A(z)$ that satisfies

$$|H(e^{j\omega})|^2 = \frac{P(1/2 - (1/2)\cos\omega)}{Q(1/2 - (1/2)\cos\omega)} \tag{21}$$

is then obtained. To this end, the spectral factorizations $P(1/2 - (1/2)\cos\omega) = B(e^{j\omega})B(e^{-j\omega})$ and $Q(1/2 - (1/2)\cos\omega) = A(e^{j\omega})A(e^{-j\omega})$, from which the transfer functions $B(z)$ and $A(z)$ could be extracted, are performed. Note that the latter factorizations are possible since $P(x)$ and $Q(x)$, for $x \in [0, 1]$, are real and nonnegative [18].

Reference [17] details the design process and provides the closed form expressions for obtaining $B(z)$ and $A(z)$. The routine *maxflat* provided in the Matlab's signal processing toolbox is an implementation of the generalized Butterworth filter design procedure. We use this routine of Matlab to design the AR filters whose response matches the prototype response. The number of poles and the bandwidth are given as parameters to the routine which delivers the desired transfer function.

6.2. Design of ARMA discrete filters

The generalized Butterworth filter design procedure that was considered above for the design of AR filters could also be adopted for the design of ARMA filters. However, our experiments have shown that better designs could be obtained by adopting a least squares method. The idea is to find the coefficients of an IIR transfer function

$$H(z) = \frac{B(z)}{A(z)} = \frac{b_0 + b_1 z^{-1} + \cdots + b_N z^{-N}}{1 + a_1 z^{-1} + \cdots + a_N z^{-N}} \tag{22}$$

such that its frequency response resembles that of a desired response. Two approaches are commonly adopted [19]: (i) the output error method, and (ii) the equation error method. In the output error method, the coefficients of $A(z)$ and $B(z)$ are chosen by minimizing the cost function

$$\xi_{oe} = \frac{1}{2\pi}\int_0^{2\pi} W(\omega)\left|\frac{B(e^{j\omega})}{A(e^{j\omega})} - H_o(e^{j\omega})\right|^2 d\omega, \tag{23}$$

where $W(\omega)$ is a weighting function and $H_o(e^{j\omega})$ is the desired (prototype filter) response. In the equation error method, on the other hand, the coefficients of $A(z)$ and $B(z)$ are chosen by minimizing the cost function

$$\xi_{ee} = \frac{1}{2\pi}\int_0^{2\pi} W(\omega)\left|B(e^{j\omega}) - A(e^{j\omega})H_o(e^{j\omega})\right|^2 d\omega. \tag{24}$$

In this paper, we choose the equation error method as it leads to a closed form solution for the filter coefficients. The output error method leads to a nonlinear optimization procedure. It is thus much harder to solve. Moreover, any solution that could be obtained from the output error method may also be obtained from the equation error method by an appropriate selection of the weighting function $W(\omega)$.

The common approach of optimizing $B(e^{j\omega})$ and $A(e^{j\omega})$ in (24) is to first replace the integral (24) by the weighted sum

$$J_{\text{ee}} = \sum_{i=1}^{K} w_i \left| B(e^{j\omega_i}) - A(e^{j\omega_i}) h_{o,i} \right|^2, \tag{25}$$

where ω_i is a grid of dense frequencies over the range $0 \le \omega \le 2\pi$ and w_i is the short-hand notation for $W(\omega_i)$. Defining the column vectors

$$\begin{aligned} \mathbf{e}_i = \big[1 \; e^{j\omega_i} \; e^{j2\omega_i} \; \cdots \; e^{jN\omega_i} \; - h_{o,i} e^{j\omega_i} \; - h_{o,i} e^{j2\omega_i} \\ \cdots \; - h_{o,i} e^{jN\omega_i}\big]^H, \end{aligned} \tag{26}$$

$\mathbf{b} = [b_0 \;\; b_1 \;\; b_2 \;\; \cdots \;\; b_N]^H$, $\mathbf{a} = [a_1 \;\; a_2 \;\; \cdots \;\; a_N]^H$, where the superscript H denotes Hermitian, and $\mathbf{c} = \left[\begin{smallmatrix} \mathbf{b} \\ \mathbf{a} \end{smallmatrix}\right]$, (25) can be rearranged as

$$J_{\text{ee}} = \mathbf{c}^H \boldsymbol{\Psi} \mathbf{c} - \boldsymbol{\theta}^H \mathbf{c} - \mathbf{c}^H \boldsymbol{\theta} + \eta, \tag{27}$$

where

$$\begin{aligned} \boldsymbol{\Psi} &= \sum_{i=1}^{K} w_i \mathbf{e}_i \mathbf{e}_i^H, \\ \boldsymbol{\theta} &= \sum_{i=1}^{K} w_i h_{o,i} \mathbf{e}_i, \end{aligned} \tag{28}$$

and $\eta = \sum_{i=1}^{K} w_i |h_{o,i}|^2$.

The cost function (27) has a quadratic form whose solution is well known to be [19]

$$\mathbf{c} = \boldsymbol{\Psi}^{-1} \boldsymbol{\theta}. \tag{29}$$

Once $\mathbf{c}$ is obtained, one can easily extract the coefficients b_i and a_i from it. This procedure was originally developed in [20].

The routine *invfreqz* in Matlab signal processing tool box can be used to find the coefficients $A(z)$ and $B(z)$ according to the above procedure.

7. MAPPING DISCRETE FILTERS ONTO OPTICAL ARCHITECTURES

The optical architectures were analyzed using the z-transform and their transfer functions were derived in Section 5. The discrete filter's transfer functions obtained in the previous step are now set equal to the corresponding optical filter's transfer function. Backward relations are derived to calculate the optical architecture's parameters for each stage. Thus, the optical filter is synthesized from the discrete filter using a mapping algorithm. The AR discrete filter designed in the previous section is mapped onto the AR lattice optical architecture using the recursion-based algorithm developed by Madsen and Zhao [21]. The ARMA discrete filter designed in the previous section is mapped onto the ARMA lattice optical architecture using the recursion-based algorithm developed by Jinguji [22]. These algorithms return the coupling ratios and phase solutions for each stage of the lattice architectures.

8. FROM DISCRETE RESPONSE TO THE OPTICAL RESPONSE

The optical filter designed using the above steps is now simulated for its linear response [23] using electromagnetic models. Also, the linear optical response is compared with the discrete filter's response. Both should have exactly the same shape (different scales) since the optical filter was synthesized from the discrete filter.

The discrete frequency response curve can be converted to an optical frequency response curve once we know the optical parameters such as unit length and center frequency. We had previously defined $z = e^{j\Omega T}$ with $\Omega = 2\pi\nu$, and $T = L_{\text{u}} n/c$ where ν is the optical frequency, L_{u} is the unit length, n is the refractive index, and c is the speed of light. Also the FSR was defined to be equal to $1/T$.

The discrete frequency response plotted over the fundamental range $-\pi \le \omega \le \pi$ or $-1/2 \le f \le 1/2$ which is normalized to $-1 \le f_{\text{norm}} \le 1$ by Matlab's *freqz* routine is equal to one optical FSR. The normalized frequency $f_{\text{norm}} = \omega_{\text{norm}}/2\pi$ is related to the optical frequency by $f_{\text{norm}} = (\nu - \nu_{\text{c}})T$ or $f_{\text{norm}} = (\Omega - \Omega_{\text{c}})T/2\pi$. To plot the optical frequency response over one FSR directly using freqz, the sampling frequency F_s can be set equal to the FSR and the frequency response can be plotted from $-F_s/2$ to $F_s/2$.

Since FSR $= 1/T = c/nL_{\text{u}}$, we need to know the unit length to know FSR. The unit length is chosen such that the product of refractive index and unit length is equal to an integer number of wavelengths, that is, $m\lambda_{\text{c}} = nL_{\text{u}}$ where m is an integer and λ_{c} is the desired central wavelength. The center frequency is then defined as $\nu_{\text{c}} = c/\lambda_{\text{c}}$. It is the frequency at which resonance occurs.

Once the linear response of the optical architecture is verified to be the same as that of the discrete filter, the optical filter is simulated to obtain the nonlinear phase shift response [23].

9. EXAMPLE AND EVALUATION OF AR LATTICE FILTERS

9.1. Design and synthesis example

In this section, we design an optical AR lattice filter and simulate it to obtain the nonlinear phase shift response. The filter is synthesized by designing discrete filters according to the

description in Section 6.1 and then using the mapping algorithm derived by Madsen and Zhao [21]. The circumference of each microring in the AR lattice architecture is chosen as the unit delay length and is equal to 50 μm. The center frequency corresponds to a wavelength of 500 nm.

A generalized digital Butterworth filter with five poles is designed using the procedure discussed in Section 6.1. Filter bandwidth is set to be 0.16π in the fundamental range $-\pi \leq \omega \leq \pi$. Assuming the loss in the material to be 1cm^{-1}, the obtained filter transfer function is

$$\frac{N(z)}{D(z)} = \frac{6.5941 \times 10^{-4}}{1.0000 - 4.1912z^{-1} + 7.0824z^{-2} - 6.0254z^{-3} + 2.5789z^{-4} - 0.4439z^{-5}}. \tag{30}$$

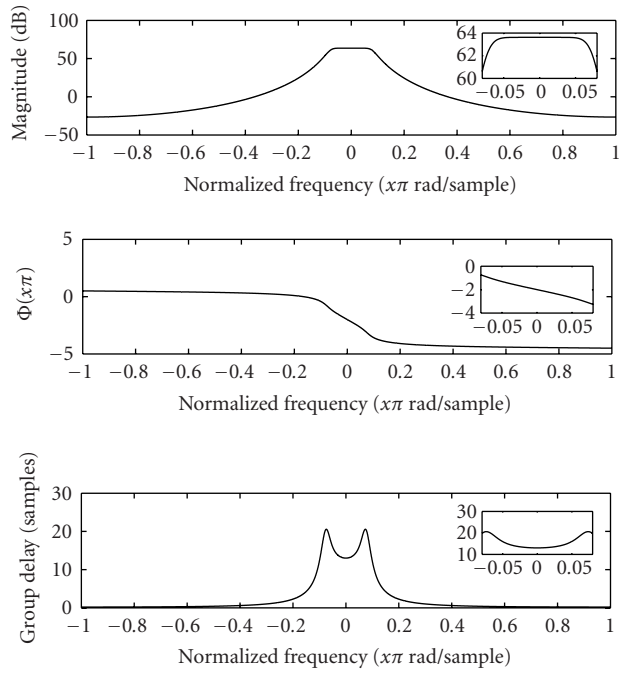


Figure 8: Frequency response and group delay characteristic of 5th-order AR filter.

Table 1: Design values for a 5th-order AR lattice filter.

	$n = 0$	1	2	3	4	5
κ_n	0.7336	0.1416	0.0357	0.0198	0.0232	0.2488
ϕ_n	—	0	0	0	0	0

The frequency response and the group delay characteristic of this filter are presented in Figure 8 showing that the designed filter's response matches with the ideal prototype response of Figure 3 for nonlinear phase shift. The magnitude response is maximally flat as desired. Also, even though most of the group delay is pushed towards the passband edges, the group delay and magnitude response does not have ripples and hence bistability is largely avoided.

This discrete filter is then mapped onto the optical AR lattice architecture of Figure 6. Table 1 shows the coupling ratios and phase values thus obtained for each stage of the optical filter.

The linear response of the synthesized optical filter is the same as that of the discrete filter for low input intensity. The nonlinear phase shift response of the AR filter is shown in Figure 9 as a function of the normalized input intensity $n_2 I_{\text{in}}$, where n_2 is the nonlinear coefficient of the underlying material and I_{in} is the input intensity. As can be seen from the figure, a π radian phase change is obtained at $n_2 I_\pi = 9.0 \times 10^{-5}$ and the transmission ratio at this input intensity is 0.66. The nonlinear response is also plotted for incident frequencies at $\nu_m \pm \delta\nu/4$ where ν_m is the center frequency. Because of the flat magnitude response in the filter's linear response, the nonlinear phase response (up to a π phase shift) is weakly sensitive to frequency within the passband of the filter, as shown, allowing for a broadband nonlinearity. Also plotted for comparison is the phase shift produced by the underlying material of length $L = k_{\text{gd}} c/n \sim 0.65$ mm, which gives the same group delay as that of the AR lattice architecture. The nonlinear phase shift produced by the designed AR filter is 5 times better than that of the bulk material.

The allowable amount of parameter error is an important information for fabrication. Random errors were added to each of the design parameters, that is, the coupling ratios and the phase values, and the nonlinear response was obtained to determine the parameter sensitivity. The allowable errors below which the nonlinear response is within 10% of the original value are $\pm 0.001\pi$ for κ_{rn}, and $\pm 0.003\pi$ for ϕ_{rn}. A detailed sensitivity analysis is presented in [24].

9.2. Improving the nonlinear phase shift response

The nonlinear phase shift response improves upon increasing the group delay. This is because high group delay implies steeper phase response which results in greater nonlinear phase shift as the frequency response red shifts upon increasing input intensity. For a maximum phase discrete filter with no poles at the origin, the total phase change across the FSR is expressed by $\Phi_{\text{ob}} + \Phi_{\text{ib}} = 2\pi N_z$, where Φ_{ob} is the out-of-band phase, Φ_{ib} is the in-band phase, and N_z is the number of zeros in the discrete filter. This simple analysis shows

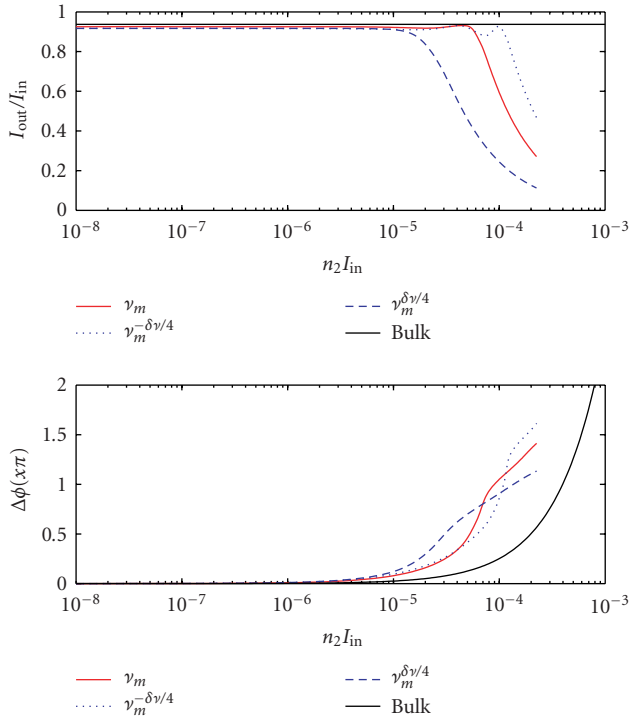


Figure 9: Nonlinear response vsersus incident intensity $n_2 I_{in}$.

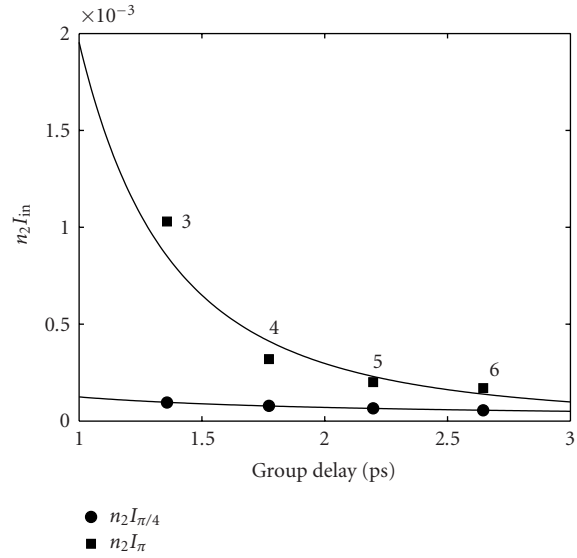


Figure 10: Improving nonlinear response by increasing the number of stages and keeping BW = 0.12 FSR.

Table 2: Improving nonlinear response by increasing the AR filter order with BW = 0.12 FSR.

Filter order	Group delay (ps)	$n_2 I_{\pi}$	$n_2 I_{\pi/4}$
3	1.36	1.03×10^{-3}	9.59×10^{-5}
4	1.77	3.20×10^{-4}	7.89×10^{-5}
5	2.19	2.01×10^{-4}	6.56×10^{-5}
6	2.64	1.70×10^{-4}	5.54×10^{-5}

that there are two means to increase the group delay (and hence, the nonlinear response) within the passband:

(1) increase the in-band phase change Φ_{ib}, and/or
(2) increase the filter order.

In general, the bandwidth, $\delta\nu$ (along with the FSR) should be a quantity chosen at the outset to match a specific application. For example, if the desired application were to produce a phase shift on a single channel of a DWDM system, then $\delta\nu \sim \delta\nu_{ch}$ and FSR $\sim N_{ch}\delta\nu_{ch}$, where $\delta\nu_{ch}$ is the channel spacing and N_{ch} is the number of channels.

Since AR filters are designed using the generalized Butterworth filter design, we do not have control over the in-band phase to increase the group delay. We increase the group delay by increasing the filter order, that is, the number of stages in the architecture, which in turn increases the total phase as well as the in-band phase. Figure 10 plots $n_2 I_{\pi}$ as a function of the group delay where the group delay is increased by increasing the filter order while keeping the bandwidth constant. The quantity $n_2 I_{\pi}$ scales as $1/k_{gd}^{2.72}$ and is given by $n_2 I_{\pi} = 19.55 \times 10^{-4} k_{gd}^{-2.72}$. The scaling of $n_2 I_{\pi}$ with group delay is not an accurate representation of the initial design of the filter because by the time a π radian nonlinear phase shift is obtained, the filter characteristics change (i.e., the new filter function is no longer just a shifted version of the initial function as assumed in the weak perturbation limit) because of increasing input intensity. Hence $n_2 I_{\pi/4}$ is plotted as a function of group delay and is shown in Figure 10. The quantity $n_2 I_{\pi/4}$ scales as $1/k_{gd}^{0.82}$ and is given by $n_2 I_{\pi/4} = 12.46 \times 10^{-5} k_{gd}^{-0.82}$. This implies that in principle, the nonlinear response can be improved while maintaining constant bandwidth by using higher-order filters. The filter order, group delay, $n_2 I_{\pi}$, and $n_2 I_{\pi/4}$ are shown in Table 2 for a bandwidth of 0.12FSR.

10. EXAMPLE AND EVALUATION OF ARMA LATTICE FILTERS

10.1. Design and synthesis example

In this section, we design an optical ARMA lattice filter and simulate it to obtain the nonlinear phase shift response. The filter is synthesized by designing discrete filters according to the description in Section 6.2 and then using the mapping algorithm derived by Jinguji [22]. The circumference of each microring in the ARMA lattice architecture is chosen as the unit delay length and is equal to 50 μm. The center frequency corresponds to a wavelength of 500 nm.

A maximum phase ARMA filter with four zeros and four poles is designed using the procedure discussed in Section 6.2. The filter bandwidth is set to be 0.05π in the fundamental range $-\pi \le \omega \le \pi$. 4π out of the total 8π phase change is allocated to the out-of-band phase change to maintain flat magnitude and linear phase response. Passband ripple is less than 0.1 dB and the stop-band magnitude is 18 dB.

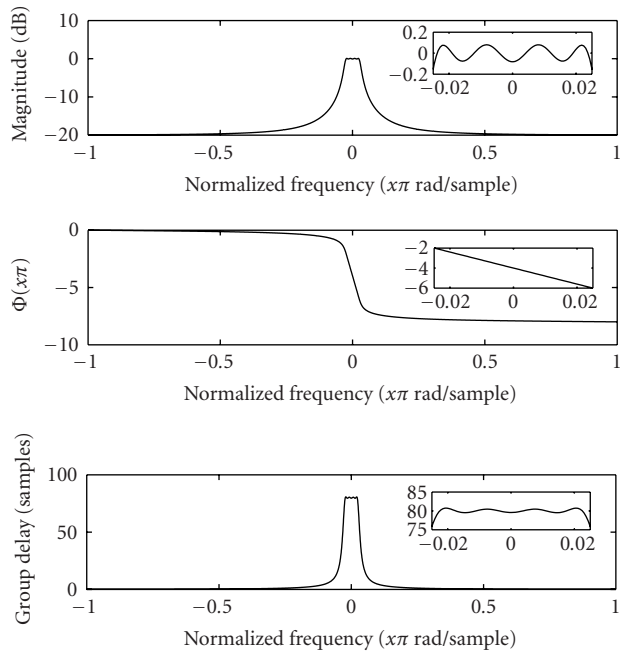


Figure 11: Frequency response and group delay characteristic of a 4th-order real ARMA filter.

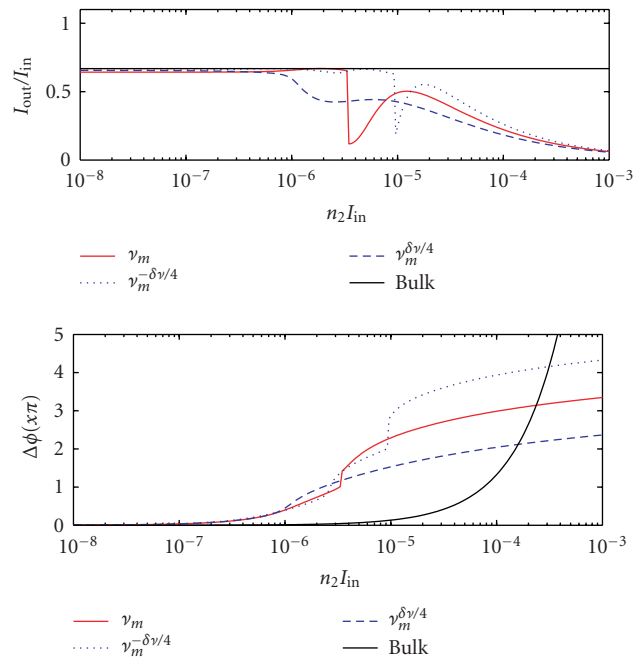


Figure 12: Nonlinear response versus incident intensity $n_2 I_{in}$.

Table 3: Design values for a 4th-order real ARMA lattice filter.

	$n = 0$	1	2	3	4
k_{tn}	0.1555	0.5513	0.5289	0.1733	0.9418
ϕ_{tn}	—	2.9754	-1.4868	1.3928	2.0702
k_{rn}	—	0.0594	0.0594	0.0784	0.0784
ϕ_{rn}	—	0.0771	-0.0771	0.0267	-0.0267

Assuming the loss in the material to be 1cm^{-1}, the obtained filter transfer function is

$$\frac{N(z)}{D(z)} = \frac{0.0656 - 0.3176z^{-1} + 0.5661z^{-2} - 0.4424z^{-3} + 0.1283z^{-4}}{1.0000 - 3.8531z^{-1} + 5.5736z^{-2} - 3.5872z^{-3} + 0.8667z^{-4}}. \quad (31)$$

The frequency response and the group delay characteristic of this filter are shown in Figure 11 showing that the designed filter's response matches with the ideal prototype response of Figure 3 for the nonlinear phase shift.

This discrete filter is then mapped onto the optical ARMA lattice architecture of Figure 7. Table 3 shows the coupling ratios and phase values thus obtained for each stage of the optical filter.

The linear response of the synthesized optical filter is the same as that of the discrete filter for low input intensity. The nonlinear phase shift response of the ARMA filter is shown in Figure 12 as a function of the normalized input intensity $n_2 I_{in}$, where n_2 is the nonlinear coefficient of the underlying material and I_{in} is the input intensity. A π radian phase change is obtained at $n_2 I_\pi = 3.3 \times 10^{-6}$ and the transmission ratio at this input intensity is 0.65. The nonlinear response is also plotted for incident frequencies at $\nu_m \pm \delta\nu/4$ where ν_m is the center frequency. As in the case of AR filter, the flat magnitude response in the filter's linear response allows for a broadband nonlinearity. Also plotted for comparison is the phase shift produced by the underlying material of length $L = k_{gd}c/n \sim 4\,\text{mm}$, which gives the same group delay as that of the ARMA lattice architecture. The nonlinear phase shift produced by the filter is 19 times better than the bulk material [25]. The nonlinear phase shift enhancement over bulk material is larger in the case of ARMA filters because of two reasons. (1) The total phase change in the case of ARMA filters is twice that of AR filters for equal number of stages. This results in higher group delay in the case of ARMA filters and hence better nonlinear phase shift response. (2) The group delay in case of AR filters is pushed towards the passband edges and hence, lower group delay at center frequency results in lower nonlinear phase-shift enhancement.

As in the case of AR filters, random errors were added to each of the design parameters, and the nonlinear response was obtained to determine the parameter sensitivity. The allowable errors below which the nonlinear response is within 10% of the original value are $\pm 0.001\pi$ for κ_{rn}, $\pm 0.001\pi$ for ϕ_{rn}, $\pm 0.01\pi$ for κ_{tn}, and $\pm 0.01\pi$ for ϕ_{tn}. A detailed sensitivity analysis is presented in [24].

10.2. Improving the nonlinear phase shift response

Similar to AR filters, the nonlinear phase shift response improves upon increasing the group delay and two means to increase the group delay (aside from decreasing passband width) are to either increase the in-band phase change Φ_{ib}, and/or increase the filter order.

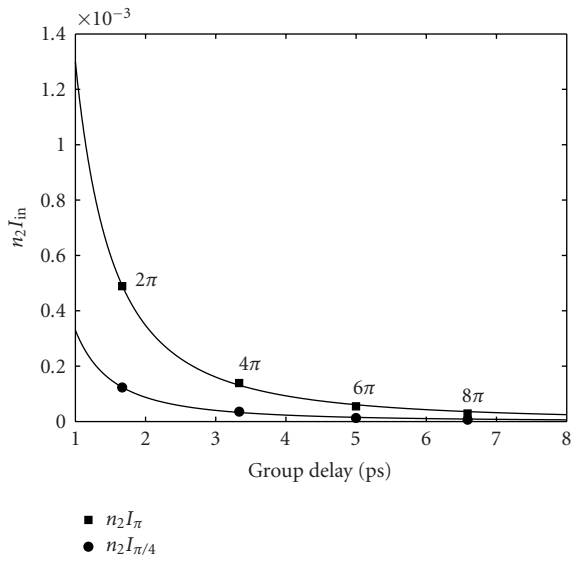


Figure 13: Improving nonlinear response by increasing the in-band phase for a 6th-order ARMA filter with BW = 0.15 FSR.

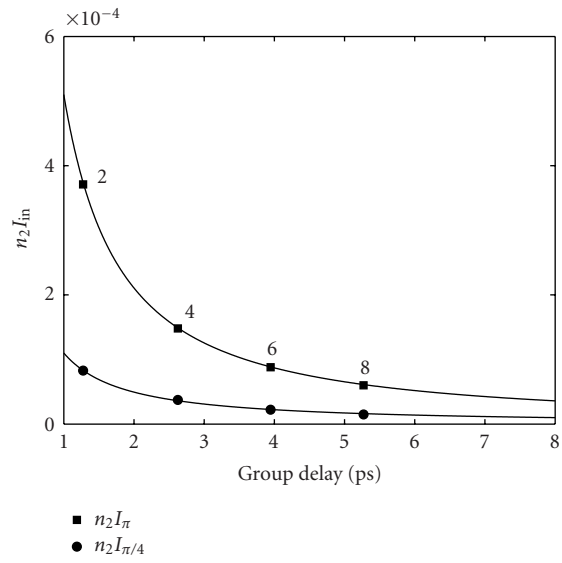


Figure 14: Improving nonlinear response by increasing the number of stages and keeping BW = 0.19 FSR, $\Phi_{ib}/\Phi_{ob} = 0.5$.

Table 4: Improving nonlinear response by increasing the in-band phase for a 6th-order ARMA filter with BW = 0.15 FSR.

In-band phase Φ_{ib}	Group delay (ps)	n_2I_π	$n_2I_{\pi/4}$
2π	1.67	4.89×10^{-4}	1.23×10^{-4}
4π	3.34	1.39×10^{-4}	3.55×10^{-5}
6π	4.99	5.49×10^{-5}	1.30×10^{-5}
8π	6.59	2.90×10^{-5}	6.51×10^{-6}

Table 5: Improving nonlinear response by increasing the ARMA filter order with BW = 0.19 FSR, in-band to out-band phase ratio = 0.5.

Filter order	Group delay (ps)	n_2I_π	$n_2I_{\pi/4}$
2	1.27	3.71×10^{-4}	8.27×10^{-5}
4	2.62	1.48×10^{-4}	3.73×10^{-5}
6	3.95	8.80×10^{-5}	2.23×10^{-5}
8	5.27	6.01×10^{-5}	1.50×10^{-5}

For a chosen bandwidth and fixed filter order, the first approach results in a trade-off between retaining the full phase within the band and in-band ripple (there is also a trade-off between Φ_{ib} and rejection ratio, but, unlike for true bandpass filters, here we are not concerned with having high rejection). Therefore, a certain amount of the total phase change needs to be allocated to Φ_{ob} in order to reduce ripple. Figure 13 plots n_2I_{in} as a function of the group delay where group delay is increased by increasing the in-band phase in a 6th-order ARMA lattice filter while keeping a constant bandwidth of 0.15FSR. The quantity n_2I_π scales as $1/k_{gd}^{1.90}$ and is given by $n_2I_\pi = 1.30 \times 10^{-3} k_{gd}^{-1.90}$. The quantity $n_2I_{\pi/4}$ scales as $1/k_{gd}^{1.92}$ and is given by $n_2I_{\pi/4} = 3.30 \times 10^{-4} k_{gd}^{-1.92}$. The in-band phase, group delay, n_2I_π, and $n_2I_{\pi/4}$ are shown in Table 4.

The second approach increases the group delay by increasing the filter order, that is, the number of stages in the architecture, which in turn increase the total phase as well as the in-band phase. Figure 14 plots n_2I_{in} as a function of the group delay where the group delay is increased by increasing the filter order while keeping the bandwidth and the Φ_{ib}/Φ_{ob} ratio constant. The quantity n_2I_π scales as $1/k_{gd}^{1.28}$ and is given by $n_2I_\pi = 5.10 \times 10^{-4} k_{gd}^{-1.28}$. The quantity $n_2I_{\pi/4}$ scales as $1/k_{gd}^{1.15}$ and is given by $n_2I_{\pi/4} = 1.10 \times 10^{-4} k_{gd}^{-1.15}$. This implies that in principle, the nonlinear response can be improved while maintaining constant bandwidth by using higher-order filters. The filter order, group delay, n_2I_π, and $n_2I_{\pi/4}$ are shown in Table 5 for a bandwidth of 0.19 FSR and the in-band to out-band phase ratio of 0.5.

11. CONCLUSIONS

In this paper, we have proposed that a discrete-time signal processing approach can be used to design nonlinear optical devices by treating the desired nonlinear response in the weak perturbation limit as a linear discrete filter. This provides a systematic and intuitive method for the design of nonlinear optical devices. We have demonstrated this approach by designing AR and ARMA filters to obtain a nonlinear phase shift response. This approach can be used for designing optical devices for various other nonlinear processes as well.

REFERENCES

[1] S. Blair, J. E. Heebner, and R. W. Boyd, "Beyond the absorption-limited nonlinear phase shift with microring resonators," *Optics Letters*, vol. 27, no. 5, pp. 357–359, 2002.

[2] C. K. Madsen and J. H. Zhao, *Optical Filter Design and Analysis: A Signal Processing Approach*, Wiley, New York, NY, USA, 1999.

[3] T. A. Ibrahim, V. Van, and P.-T. Ho, "All-optical time-division demultiplexing and spatial pulse routing with a GaAs AlGaAs microring resonator," *Optics Letters*, vol. 27, no. 10, pp. 803–805, 2002.

[4] M. Soljacic, C. Luo, J. D. Joannopoulos, and S. Fan, "Nonlinear photonic crystal microcavities for optical integration," *Optics Letters*, vol. 28, no. 8, pp. 637–639, 2003.

[5] S. Blair, *Optical soliton-based logic gates*, Ph.D. dissertation, University of Colorado, Boulder, Colo, USA, 1998.

[6] M. Soljacic, S. G. Johnson, S. Fan, et al., "Photonic-crystal slow-light enhancement of nonlinear phase sensitivity," *Journal of the Optical Society of America B*, vol. 19, no. 9, pp. 2052–2059, 2002.

[7] Y. Chen and S. Blair, "Nonlinear phase shift of cascaded microring resonators," *Journal of Optical Society of America B*, vol. 20, no. 10, pp. 2125–2132, 2003.

[8] W. Nakagawa, R.-C. Tyan, and Y. Fainman, "Analysis of enhanced second-harmonic generation in periodic nanostructures using modified rigorous coupled-wave analysis in the undepleted-pump approximation," *Journal of the Optical Society of America A*, vol. 19, no. 9, pp. 1919–1928, 2002.

[9] P. P. Absil, J. V. Hryniewicz, B. E. Little, et al., "Wavelength conversion in GaAs micro-ring resonators," *Optics Letters*, vol. 25, no. 8, pp. 554–556, 2000.

[10] A. Melloni, F. Morichetti, and M. Martinelli, "Optical slow wave structures," *Optics and Photonics News*, vol. 14, no. 11, pp. 44–48, 2003.

[11] A. Hasegawa and Y. Kodama, "Signal transmission by optical solitons in monomode fiber," *Proc. IEEE*, vol. 69, no. 9, pp. 1145–1150, 1981.

[12] A. Melloni, F. Morichetti, and M. Martinelli, "Linear and nonlinear pulse propagation in coupled resonator slow-wave optical structures," *Optical and Quantum Electronics*, vol. 35, no. 4-5, pp. 365–379, 2003.

[13] J. E. Heebner and R. W. Boyd, "SCISSOR solitons and other novel propagation effects in microresonator-modified waveguides," *Journal of the Optical Society of America B*, vol. 19, no. 4, pp. 722–731, 2002.

[14] D. A. B. Miller, S. D. Smith, and A. Johnston, "Optical bistability and signal amplification in a semiconductor crystal: applications of new low-power nonlinear effects in InSb," *Applied Physics Letters*, vol. 35, no. 9, pp. 658–660, 1979.

[15] H. M. Gibbs, *Optical Bistability: Controlling Light with Light*, Academic Press, Orlando, Fla, USA, 1985.

[16] M. Soljacic, M. Ibanescu, S. G. Johnson, Y. Fink, and J. D. Joannopoulos, "Optimal bistable switching in nonlinear photonic crystals," *Physical Review E*, vol. 66, no. 5, pp. 055601-1–055601-4, 2002.

[17] I. W. Selesnick and C. S. Burrus, "Generalized digital butterworth filter design," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP '99)*, vol. 3, pp. 90–95, Phoenix, Ariz, USA, May 1999.

[18] P. P. Vaidyanathan, *Multirate Systems and Filter Banks*, Prentice Hall, Englewood Cliffs, NJ, USA, 1993.

[19] B. Farhang-Boroujeny, *Adaptive Filters: Theory and Applications*, John Wiley & Sons, Chichester, UK, 1998.

[20] E. C. Levi, "Complex-curve fitting," *IRE Transactions on Automatic Control*, vol. 4, pp. 37–44, 1959.

[21] C. K. Madsen and J. H. Zhao, "A general planar waveguide autoregressive optical filter," *J. Lightwave Technol.*, vol. 14, no. 3, pp. 437–447, 1996.

[22] K. Jinguji, "Synthesis of coherent two-port optical delay line circuits with ring waveguides," *J. Lightwave Technol.*, vol. 14, no. 8, pp. 1882–1898, 1996.

[23] Y. Chen, *Nonlinear optical process enhancement by artificial resonant structures*, Ph.D. dissertation, University of Utah, Salt Lake City, Utah, USA, 2004.

[24] G. Pasrija, "Discrete-time signal processing approach to the design of nonlinear optical devices," M.S. thesis, University of Utah, Salt Lake City, Utah, USA, 2004.

[25] Y. Chen, G. Pasrija, B. Farhang-Boroujeny, and S. Blair, "Engineering the nonlinear phase shift," *Optics Letters*, vol. 28, no. 20, pp. 1945–1947, 2003.

Geeta Pasrija received her M.S. degree in electrical engineering from the University of Utah in 2004. Currently, she is working at SR Technologies Inc., and is involved in the design of a software-defined radio for satellite communications.

Yan Chen received her Ph.D. degree in electrical engineering from the University of Utah in 2004. From 1997 to 1999, she worked at Optical Memory National Engineering Research Center China, as a Research Assistant and Lecturer. Currently, she is working in the area of semiconductor lithography software development for Timbre Technologies Inc.

Behrouz Farhang-Boroujeny received his Ph.D. degree from Imperial College, University of London, UK, in 1981. From 1981 to 1989 he was at Isfahan University of Technology, Isfahan, Iran. From 1989 to 2000 he was at the National University of Singapore. Since August 2000, he has been with the University of Utah. He is an expert in the general area of signal processing. His current scientific interests are in adaptive filters, multicarrier communications, detection techniques for space-time coded systems, and signal processing applications to optical devices.

Steve Blair received his Ph.D. degree from the University of Colorado at Boulder in 1998. He has been an Assistant Professor in the Electrical and Computer Engineering Department at the University of Utah since then. His research interests include slow-light nonlinear optics, nanoscale photonics, and real-time molecular detection arrays.

EURASIP Journal on Applied Signal Processing 2005:10, 1498–1519

Fractional Transforms in Optical Information Processing

Tatiana Alieva
Facultad de Ciencias Físicas, Universidad Complutense de Madrid, Ciudad Universitaria, 28040 Madrid, Spain
Email: talieva@fis.ucm.es

Martin J. Bastiaans
Faculteit Elektrotechniek, Technische Universiteit Eindhoven, Postbus 513, 5600 MB Eindhoven, The Netherlands
Email: m.j.bastiaans@tue.nl

Maria Luisa Calvo
Facultad de Ciencias Físicas, Universidad Complutense de Madrid, Ciudad Universitaria, 28040 Madrid, Spain
Email: mlcalvo@fis.ucm.es

Received 31 March 2004

We review the progress achieved in optical information processing during the last decade by applying fractional linear integral transforms. The fractional Fourier transform and its applications for phase retrieval, beam characterization, space-variant pattern recognition, adaptive filter design, encryption, watermarking, and so forth is discussed in detail. A general algorithm for the fractionalization of linear cyclic integral transforms is introduced and it is shown that they can be fractionalized in an infinite number of ways. Basic properties of fractional cyclic transforms are considered. The implementation of some fractional transforms in optics, such as fractional Hankel, sine, cosine, Hartley, and Hilbert transforms, is discussed. New horizons of the application of fractional transforms for optical information processing are underlined.

Keywords and phrases: fractional Fourier transform, fractional convolution, fractional cyclic transforms, fractional optics.

1. INTRODUCTION

During the last decades, optics is playing an increasingly important role in computing technology: data storage (CD-ROM) and data communication (optical fibres). In the area of information processing, optics also has certain advantages with respect to electronic computing, thanks to its massive parallelism, operating with continuous data, and so forth [1, 2, 3]. Moreover, the modern trend from binary logic to fuzzy logic, which is now used in several areas of science and technology such as control and security systems, robotic vision, industrial inspection, and so forth, opens up new perspectives for optical information processing. Indeed, typical optical phenomena such as diffraction and interference inherit fuzziness and therefore permit an optical implementation of fuzzy logic [4].

The first and highly successful configuration for optical data processing—the optical correlator—was introduced by Van der Lugt more than 30 years ago [5]. It is based on the ability of a thin lens to produce the two-dimensional Fourier transform (FT) of an image in its back focal plane. This invention led to further creation of a great variety of optical and optoelectronic processors such as joint correlators, adaptive filters, optical differentiators, and so forth [6]. More sophisticated tools such as wavelet transforms [7] and bilinear distributions [8, 9, 10, 11, 12, 13, 14], which are actively used in digital data processing, have been implemented in optics.

Nowadays, fractional transforms play an important role in information processing [15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31], and the obvious question is: why do we need fractional transformations if we successfully apply the ordinary ones? First, because they naturally arise under the consideration of different problems, for example, in optics and quantum mechanics, and secondly, because fractionalization gives us a new degree of freedom (the fractional order) which can be used for more complete characterization of an object (a signal, in general) or as an additional encoding parameter. The canonical fractional FT, for instance, is used for phase retrieval [32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42], signal characterization [43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56],

space-variant filtering [29, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77], encryption [78, 79, 80, 81, 82, 83, 84, 85], watermarking [86, 87], creation of neural networks [88, 89, 90, 91, 92, 93], and so forth, while the fractional Hilbert transform was found to be very promising for selective edge detection [94, 95, 96]. Several fractional transforms can be performed by simple optical configurations.

In this paper, we review the progress achieved in optical information processing during the last decade by application of fractional transforms. We will start from the definition of a fractional transformation in Section 2. Then we consider, in Section 3, the fractionalization in paraxial optics described by the canonical integral transformation. Two fractional canonical transforms, the Fresnel transform and the fractional FT, are commonly used in optical information processing. The fractional FT, which is a generalization of the ordinary FT with an additional parameter α that can be interpreted as a rotational angle in phase space, is considered in more detail.

Since the convolution operation is fundamental in information processing, there were several proposals to generalize it to the fractional case. In Section 4, we define the generalized fractional convolution, and in Sections 5–8, we consider its application for information processing: phase retrieval, signal characterization, filtering, noise reduction, encryption, and watermarking.

The second part of the paper will be devoted to the fractionalization procedure of other important transforms. We will restrict ourselves to the consideration of cyclic transforms, which produce the identity transform when they act an integer number of times N. In Sections 9–11, we will show that there are different ways for the construction of a fractional transform for a given cyclic transform. In Section 12, we briefly mention the common properties of fractional cyclic transforms.

The fractional Hankel, Hartley, sine and cosine, and Hilbert transforms, which can all be implemented in optics, will be considered in Section 13. Finally, we discuss the main lines of future development of fractional optics in Section 14 and make some conclusions.

2. FRACTIONAL TRANSFORM: A GENERAL DEFINITION

The word "fraction" is nowadays very popular in different fields of science. We recall fractional derivatives in mathematics, fractal dimension in geometry, fractal noise, fractional transformations in signal processing, and so forth. In general, "fractional" means that some parameter has no longer an integer value.

To define the fractional version of a given linear integral transform, we consider the operator $\mathcal{R}$ of such a transform, acting on a function $f(x)$,

$$\mathcal{R}[f(x)](u) = \int_{-\infty}^{\infty} K(x,u) f(x) dx, \tag{1}$$

with $K(x,u)$ the operator kernel. As an example we mention the Fourier transformation, for which the kernel reads $K(x,u) = \exp(-i2\pi ux)$. The fractional transform operator is denoted by $\mathcal{R}^p$, where p is the parameter of fractionalization:

$$\mathcal{R}^p[f(x)](u) = \int_{-\infty}^{\infty} K(p,x,u) f(x) dx. \tag{2}$$

We will formulate some desirable properties of this fractional transform first.

The fractional transform has to be continuous for any real value of the parameter p, and additive with respect to this parameter: $\mathcal{R}^{p_1+p_2} = \mathcal{R}^{p_2}\mathcal{R}^{p_1}$. Moreover it has to reproduce the ordinary transform and powers of it for integer values of p. In particular, for $p = 1$ we should get the ordinary transform $\mathcal{R}^1 = \mathcal{R}$, and for $p = 0$ the identity transform $\mathcal{R}^0 = I$. From the additivity property it follows that $\int_{-\infty}^{\infty} K(p_1,x,u)K(p_2,u,y)du = K(p_1+p_2,x,y)$. Note that the parameter p, as we will see further, may be given by a matrix, and the additivity property is then formulated easily as the product of the corresponding matrices.

As we have mentioned in the introduction, some fractional transforms arise under consideration of different problems: description of paraxial diffraction in free space and in a quadratic refractive index medium, resolution of the nonstationary Schrödinger equation in quantum mechanics, phase retrieval, and so forth. Other fractional transforms can be constructed for their own sake, even if their direct application may not be obvious yet. In particular, in Section 9 we consider a general algorithm for the fractionalization of a given linear cyclic integral transform. The application of a particular fractional transform for optical information processing then depends on its properties and on the possibility of its experimental realization in optics.

3. FRACTIONALIZATION IN PARAXIAL OPTICS: THE CANONICAL INTEGRAL TRANSFORM

Analog optical signal processing systems are often described in the framework of paraxial scalar diffraction theory. A typical subset of such a system is displayed in Figure 1 and contains a thin lens with focal distance f, preceded and followed by two sections of free space with distances z_1 and z_2, respectively. Note that the conventional Van der Lugt correlator [5, 6], mentioned in the introduction, is constructed by a cascade of two such subsets, with each subset forming an FT system ($z_1 = z_2 = f$) and with a filter mask inserted between them. A monochromatic optical field in a transversal plane (x,y) is then described either by a complex field amplitude $f(x,y)$ for the coherent case, or by the two-point correlation function $\Gamma(x_1,x_2;y_1,y_2) = \langle f(x_1,y_1)f^*(x_2,y_2)\rangle$ for the partially coherent case, where the asterisk denotes complex conjugation and $\langle\cdot\rangle$ indicates ensemble averaging; note that these cases correspond to a deterministic or a stochastic signal description in signal theory, respectively.

Under the paraxial approximation of scalar diffraction theory, the complex amplitude $f(x_{\text{in}},y_{\text{in}})$ of a monochromatic coherent optical field at the input plane of the

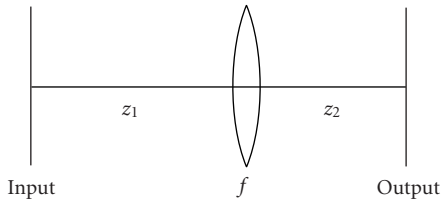


Figure 1: A typical optical information processing system.

setup depicted in Figure 1 and the complex amplitude $F_M(x_{\text{out}}, y_{\text{out}})$ at the output plane of it are related by the input-output relationship [97]

$$\begin{aligned} F_M(x_{\text{out}}, y_{\text{out}}) &= \mathcal{R}^M[f(x_{\text{in}}, y_{\text{in}})](x_{\text{out}}, y_{\text{out}}) \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} K_{M_x}(x_{\text{in}}, x_{\text{out}}) K_{M_y}(y_{\text{in}}, y_{\text{out}}) \\ &\qquad \times f(x_{\text{in}}, y_{\text{in}})\, dx_{\text{in}}\, dy_{\text{in}}, \end{aligned} \tag{3}$$

where the kernel $K_{M_x}(x_{\text{in}}, x_{\text{out}})$ takes the form

$$\begin{aligned} &K_{M_x}(x_{\text{in}}, x_{\text{out}}) \\ &\quad = \begin{cases} \dfrac{1}{\sqrt{ib_x}} \exp\left(i\pi \dfrac{a_x x_{\text{in}}^2 + d_x x_{\text{out}}^2 - 2x_{\text{out}}x_{\text{in}}}{b_x}\right), & b_x \neq 0, \\ \dfrac{1}{\sqrt{|a_x|}} \exp\left(i\pi \dfrac{c_x x_{\text{out}}^2}{a_x}\right) \delta\left(x_{\text{in}} - \dfrac{x_{\text{out}}}{a_x}\right), & b_x = 0, \end{cases} \end{aligned} \tag{4}$$

with

$$\begin{aligned} M_x &= \begin{pmatrix} a_x & b_x \\ c_x & d_x \end{pmatrix} \\ &= \begin{pmatrix} 1 - \dfrac{z_2}{f_x} & \lambda\left(z_1 + z_2 - \dfrac{z_1 z_2}{f_x}\right) \\ -\dfrac{1}{\lambda f_x} & 1 - \dfrac{z_1}{f_x} \end{pmatrix} \end{aligned} \tag{5}$$

and λ the optical wavelength, and where similar expressions, with x replaced by y, hold for the kernel $K_{M_y}(y_{\text{in}}, y_{\text{out}})$ and the matrix M_y. Note that the optical wavelength λ enters the expressions for b and c as a mere scaling factor; very often, we like to work with reduced, dimensionless coordinates, in which case b and c take a form that would also be achieved by assigning an appropriate value to λ. We remark that the application of cylindrical lenses, $f_x \neq f_y$, permits to perform anamorphic transformations.

The coefficients a_x, b_x, c_x, and d_x that arise in the kernel (4), are entries of the general, symplectic ray transformation matrix [98] that relates the position (x, y) and direction (ξ, η) of an optical ray in the input and the output plane of a so-called first-order optical system, and we have

$$\begin{pmatrix} x_{\text{out}} \\ \xi_{\text{out}} \end{pmatrix} = \begin{pmatrix} a_x & b_x \\ c_x & d_x \end{pmatrix} \begin{pmatrix} x_{\text{in}} \\ \xi_{\text{in}} \end{pmatrix} = M_x \begin{pmatrix} x_{\text{in}} \\ \xi_{\text{in}} \end{pmatrix} \tag{6}$$

and a similar relation for the other dimension, with x and ξ replaced by y and η, respectively. For separable systems, to which we restrict ourselves throughout, symplecticity reads simply $a_x d_x - b_x c_x = 1$ and $a_y d_y - b_y c_y = 1$. The transform described by (3) is known by such names as canonical integral transform and generalized Fresnel transform [97, 98, 99, 100].

Special cases of canonical integral transform systems include

(i) an imaging system ($1/z_1 + 1/z_2 = 1/f$, and hence $ad = 1$ and $b = 0$);

(ii) a simple lens ($z_1 = z_2 = 0$, and hence $a = d = 1$ and $b = 0$);

(iii) a section of free space ($f \to \infty$, and hence $a = d = 1$ and $c = 0$), which is also known as a parabolic system [97] and which in the paraxial approximation performs a Fresnel transformation;

(iv) an FT system ($z_1 = z_2 = f$, and hence $a = d = 0$ and $bc = -1$), and more generally, a fractional FT system [15, 16, 17, 18] ($z_1 = z_2 = 2f \sin^2(\alpha/2)$ [22], and hence $a = d = \cos\alpha$ and $bc = -\sin^2\alpha$), which is also known as an elliptic system [97]; the common case for which $b = -c = \sin\alpha$ follows when we normalize x/ξ with respect to $\lambda f \sin\alpha$, and can also be achieved by formally choosing $\lambda f \sin\alpha = 1$;

(v) a hyperbolic system [97], with $a = d = \cosh\alpha$ and $bc = \sinh^2\alpha$.

To treat the propagation of partially coherent light through first-order optical systems, it is advantageous to describe such light not by its two-point correlation function $\Gamma(x_1, x_2; y_1, y_2)$ as mentioned before, but by the related Wigner distribution (WD) [101, Chapter 12]. Of course, the coherent case considered in (3) is just a special case of this more general, partially coherent case. The Wigner distribution of partially coherent light is defined in terms of the two-point correlation function by

$$\begin{aligned} &W(x, \xi; y, \eta) \\ &\quad = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \Gamma\left(x + \frac{x'}{2}, x - \frac{x'}{2}; y + \frac{y'}{2}, y - \frac{y'}{2}\right) \\ &\qquad \times \exp\left[-i2\pi(\xi x' + \eta y')\right] dx'\, dy'. \end{aligned} \tag{7}$$

A distribution function according to definition (7) was first introduced in optics by Walther [8, 9], who called it the generalized radiance. The WD $W(x, \xi; y, \eta)$ represents partially coherent light in a combined space/spatial-frequency domain, the so-called phase space, where ξ, η are the spatial-frequency variables associated to the positions x, y, respectively.

The WD is closely related to another bilinear distribution, the ambiguity function (AF) [101, Chapter 12], which was also applied to the description of optical fields [10] and which is related to the WD by a combined FT/inverse FT.

Note that the introduction of the WD and the AF in optics [8, 9, 10, 11, 12, 13, 14] has allowed to describe—through the same function—both coherent and partially coherent optical fields, and to unify approaches for optical and digital information processing.

It is well known that the input-output relationship between the WDs $W_{\text{in}}(x,\xi;y,\eta)$ and $W_{\text{out}}(x,\xi;y,\eta)$ at the input and the output plane of a separable first-order optical system, respectively, reads [12, 13, 14]

$$W_{\text{out}}(x,\xi;y,\eta) = W_{\text{in}}(d_x x - b_x\xi, -c_x x + a_x\xi; d_y y - b_y\eta, -c_y y + a_y\eta), \tag{8}$$

which elegant expression can be considered as the counterpart of the canonical integral transform (3) in phase space, valid for partially coherent and completely coherent light. A similar relation holds for the AF [10].

Every separable, first-order optical system is described by a set of 2×2 matrices M, one for each transversal coordinate, whose entries are real-valued and whose determinants are equal 1, and we have the important symmetry property $K_M^*(x_{\text{in}},x_{\text{out}}) = K_{M^{-1}}(x_{\text{out}},x_{\text{in}})$. The cascade of two such systems is characterized by the matrix product $M_3 = M_2M_1$, which expresses the additivity of first-order optical systems. We might say that each separate subsystem performs a separate fraction of the total canonical integral transform that corresponds to the system as a whole. We may demand that in distributing the total canonical transform over the separate subsystems, certain rules of the dividing procedure should hold, for example, that all fractional subsets should be identical and be defined by the same matrix [102]. It is often possible to separate the original setup into equal subsets characterized by a one-parameter matrix; this is in particular the case for one-parameter systems like the parabolic, the elliptic, and the hyperbolic system.

It is easy to see from (4) that two canonical systems whose parameters are related as $b_1/a_1 = b_2/a_2$ produce the same transformation of the complex amplitude of the input field, and differ only in a scaling (determined by b_2/b_1) and an additional quadratic phase shift [51, 103]:

$$\mathcal{R}^{M_1}[f(x_{\text{in}})](x_{\text{out}}) = \frac{b_2}{b_1}\exp\left[\frac{ix^2}{2b_1^2}(d_1b_1 - d_2b_2)\right]\mathcal{R}^{M_2}[f(x_{\text{in}})]\left(\frac{b_2}{b_1}x_{\text{out}}\right). \tag{9}$$

In this sense, the elliptic (fractional FT), parabolic (Fresnel transform), and hyperbolic systems with the same b/a, determined by the angle α or the propagation distance z, behave similarly.

The fractional FT and the Fresnel transform are usually applied in optical information processing due to their simple analog realizations. Since both of them belong to the class of canonical integral transforms, we summarize the main theorems for the canonical transform in Table 1. For simplicity, we consider only the one-dimensional case, and we will do the same in the rest of the paper if the generalization to the two-dimensional case is straightforward. The eigenfunctions of the linear canonical transform were considered in [99, 104].

Table 1: Canonical integral transform: main theorems.

(1) Linearity:

$$\mathcal{R}^M\left[\sum_j \mu_j f_j(x)\right](u) = \sum_j \mu_j \mathcal{R}^M[f_j(x)](u)$$

(2) Parseval's equality:

$$\int_{-\infty}^{\infty} f(x)g^*(x)dx = \int_{-\infty}^{\infty} F_M(u)G_M^*(u)du$$

(3) Shifting:

$$\mathcal{R}^M[f(x-x_\circ)](u) = \exp\left[i\pi(2ux_\circ - ax_\circ^2)c\right]\mathcal{R}^M[f(x)](u-ax_\circ)$$

(4) Scaling:

$$\mathcal{R}^M[f(\mu x)](u) = \left(\frac{1}{\mu}\right)\mathcal{R}^{M_\mu}[f(x)](u)$$

with $M_\mu = \begin{pmatrix} a & b \\ c & d\end{pmatrix}\begin{pmatrix} 1/\mu & 0 \\ 0 & \mu\end{pmatrix}$

(5) Differentiation:

$$\mathcal{R}^M\left[\frac{d^n f(x)}{dx^n}\right](u) = (2\pi i)^n\left[-cu + \frac{a}{2\pi i}\frac{d}{du}\right]^n \mathcal{R}^M[f(x)](u)$$

4. FRACTIONAL FOURIER TRANSFORM AND GENERALIZED FRACTIONAL CONVOLUTION

Since the FT plays an important role in data processing, its generalization—the fractional FT—was probably the most intensively studied among all fractional transforms. Although the FT can be divided into fractions in different ways, the canonical fractional FT certainly has advantages for application in optical information processing. First, because this fractional FT can easily be realized experimentally by using simple optical setups [22], and secondly, because it produces a mere rotation of the two fundamental phase-space distributions: the WD and the AF.

The canonical fractional FT was introduced more than 60 years ago in the mathematical literature [19]; after that, it was reinvented for applications in quantum mechanics [20, 21], optics [15, 16, 18], and signal processing [23]. After the main properties of the fractional FT were established, the perspectives for its implementations in filter design, signal analysis, phase retrieval, watermarking, and so forth became clear. Moreover, the use of refractive optics for analog realizations of the fractional FT opened a way for fractional Fourier optical information processing. In this section, we will point out the basic properties of the fractional FT and its applications in optics.

In the one-dimensional case, we define the fractional FT of a signal $f(x)$ as

$$F_\alpha(u) = \mathcal{R}^\alpha[f(x)](u) = \int_{-\infty}^{\infty} K(\alpha, x, u) f(x) dx, \tag{10}$$

where the kernel $K(\alpha, x, u)$ is given by

$$K(\alpha, x, u) = \frac{\exp(i\alpha/2)}{\sqrt{i \sin\alpha}} \exp\left[i\pi \frac{(x^2 + u^2)\cos\alpha - 2ux}{\sin\alpha}\right]. \tag{11}$$

Here we use reduced, dimensionless variables x and u. Note the slight change in notation in comparison to Section 2; it will soon be clear that in the case of the fractional FT, we prefer to use the fractional angle $\alpha = p(\pi/2)$. The fractional FT is a particular case of the canonical integral transform (4), except for the constant factor $\exp(i\alpha/2)$.

The fractional FT can be considered as a generalization of the ordinary FT for the parameter α, which may be interpreted as a rotation angle in phase space [22]. This can easily be seen by considering the WD (or the AF) and by noting that a fractional FT system is a special case of a first-order optical system with $a = d = \cos\alpha$ and $b = -c = \sin\alpha$. If $f_{\text{out}}(u) = \mathcal{R}^\alpha[f_{\text{in}}(x)](u)$ is the fractional FT of $f_{\text{in}}(x)$, then the WD $W_{\text{in}}(x, \xi)$ of $f_{\text{in}}(x)$ and the WD $W_{\text{out}}(u, v)$ of $f_{\text{out}}(u)$ are related as $W_{\text{in}}(x, \xi) = W_{\text{out}}(u, v)$, see (8), where x and ξ are related to u and v by the rotation operation

$$\begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} \cos\alpha & \sin\alpha \\ -\sin\alpha & \cos\alpha \end{pmatrix} \begin{pmatrix} x \\ \xi \end{pmatrix}. \tag{12}$$

A detailed analysis of the fractional FT can be found in [24, 25, 29, 30, 31]. From its properties we mention that for $\alpha = \pm\pi/2$, we have the normal FT and its inverse (and also $F_{\alpha+\pi}(u) = F_\alpha(-u)$), while for $\alpha \to 0$, we have the identity transformation $F_0(x) = f(x)$. Note also the symmetry properties $K(\alpha, x, u) = K(\alpha, u, x)$ and $K^*(\alpha, x, u) = K(-\alpha, u, x)$, and the reversion property $\mathcal{R}^\alpha[f(-x)](u) = \mathcal{R}^\alpha[f(x)](-u)$. The analysis and synthesis of eigenfunctions of the fractional FT for a given angle were discussed in [105, 106, 107, 108, 109].

Besides the optical realization of a fractional FT system mentioned before in Section 3, other optical schemes have been proposed [22, 110, 111, 112, 113]. In particular, the complex amplitudes at two spherical surfaces of given curvature and spacing are related by a fractional FT, where the angle is proportional to the Gouy phase shift between the two surfaces [111, 112, 113]. This relationship can be helpful for the analysis of quasi-confocal resonators and data transmission between a spherical emitter and receiver.

In the sequel, optical systems performing a fractional FT will be called fractional FT systems. As we have mentioned before, the use of cylindrical refractive index media allows to perform a separable, two-dimensional fractional FT for different angles in the two dimensions [114, 115].

One of the most important properties of the FT is related to the convolution operation on two signals $f(x)$ and $g(x)$,

$$h_{f,g}(x) = \int_{-\infty}^{\infty} f(x') g(x - x') dx', \tag{13}$$

which in the spectral domain takes the form

$$\mathcal{R}^{\pi/2}[h_{f,g}(x)] = \{\mathcal{R}^{\pi/2}[f(x)]\}\{\mathcal{R}^{\pi/2}[g(x)]\}. \tag{14}$$

After the introduction of the fractional FT, several kinds of fractional convolution and correlation operations were proposed [57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70]. These operations can be expressed in the form of a generalized fractional convolution (GFC) $H_{f,g}(x, \alpha, \beta, \gamma)$, defined by [66]

$$\mathcal{R}^\alpha[H_{f,g}(x, \alpha, \beta, \gamma)] = \{\mathcal{R}^\beta[f(x)]\}\{\mathcal{R}^\gamma[g(x)]\} \tag{15}$$

(cf. (14)), or equivalently by

$$\mathcal{R}^{\alpha-\pi/2}[H_{f,g}(x, \alpha, \beta, \gamma)](u) = \int_{-\infty}^{\infty} F_{\beta-\pi/2}(u') G_{\gamma-\pi/2}(u - u') du' \tag{16}$$

(cf. (13)).

It is easy to see that the GFC includes as particular cases almost all definitions of the fractional convolution and correlation operations proposed before [57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70]. Also the expressions for the cross-WD and cross-AF can easily be given in terms of the GFC; for the cross-WD and cross-AF expressed in polar coordinates [34],

$$W_{f,g}(r, \phi) = 2\int_{-\infty}^{\infty} F_{\phi+\pi/2}(u) G^*_{\phi+\pi/2}(-u) \exp[i2\pi u(2r)] du, \tag{17}$$

$$A_{f,g}(r, \phi) = \int_{-\infty}^{\infty} F_{\phi+\pi/2}(u) G^*_{\phi+\pi/2}(u) \exp(i2\pi u r) du, \tag{18}$$

we thus have

$$W_{f,g}(r, \phi) = 2H_{f,g^*}\left(2r, \frac{\pi}{2}, \phi + \frac{\pi}{2}, -\phi + \frac{\pi}{2}\right), \tag{19}$$

$$A_{f,g}(r, \phi) = H_{f,g^*}\left(r, \frac{\pi}{2}, \phi + \frac{\pi}{2}, -\phi - \frac{\pi}{2}\right), \tag{20}$$

respectively. The GFC system is represented schematically in Figure 2, indicating a general procedure to obtain the GFC.

In view of the canonical integral transform, a further generalization of the convolution operation $H_{f,g}(x, M_1, M_2, M_3)$ can be proposed as [69]

$$\mathcal{R}^{M_1}[H_{f,g}(x, M_1, M_2, M_3)] = \{\mathcal{R}^{M_2}[f(x)]\}\{\mathcal{R}^{M_3}[g(x)]\}, \tag{21}$$

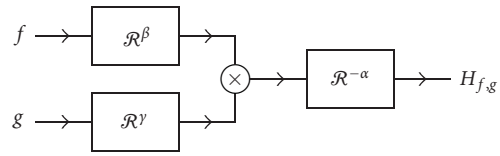


Figure 2: Schematic representation of the generalized fractional convolution system.

where the kernels of the three canonical integral transforms are parameterized by a matrix M (see (6)). This definition corresponds to the nonconventional convolution that is used in real optical systems under the paraxial approximation of the scalar diffraction theory, where the image and filter planes are shifted from their conventional positions [68, 71]. As particular cases, the GFC and the Fresnel convolution can thus be realized. The introduction of the canonical convolution operation permits to find features similar to the ones of the fractional Fourier correlators and the Fresnel correlator, proposed several years ago in [71], and to treat easily the fractional correlator based on the modified fractional FT [68].

Note that the GFC of a one-dimensional signal is a function of four variables: x, α, β, and γ. The angle variables are often considered as parameters, and the function becomes one-dimensional. As we will see below in Sections 5 and 6, optical signal processing allows to treat the GFC as a two-dimensional function, where one of the parameters is considered as the second coordinate. The choice of the parameters and the number of variables of the GFC depends on the particular application. In Sections 5–8, we will consider the applications of the GFC for phase retrieval, signal characterization, pattern recognition, and filtering tasks, respectively.

5. FRACTIONAL POWER SPECTRA FOR PHASE RETRIEVAL

Phase retrieval from intensity information is an important problem in many areas of science, including optics, quantum mechanics, X-ray radiation, and so forth. In particular, noninterferometric techniques have attracted considerable attention recently. In this section, we consider the application of fractional FT systems for the phase retrieval problem.

The squared moduli of the fractional FT, also called fractional power spectra, correspond to the projection of the WD upon the direction at an angle α in phase space. Note also that the fractional power spectrum is the particular case of the GFC

$$\left|F_{\alpha}(u)\right|^{2} = H_{f,f^{*}}(u,0,\alpha,-\alpha). \tag{22}$$

Fractional power spectra play an important role in fractional optics: they are related to the intensity distributions at the output plane of a fractional FT system and therefore can be easily measured in optics. The set of fractional power spectra for $\alpha \in [0,\pi]$ is called the Radon-Wigner transform [116], because it defines the Radon transform of the WD. The WD can be obtained from the Radon-Wigner transform by applying the inverse Radon transform [101, Chapter 8]. This is a basis for phase-space tomography [32], a method for experimental determination of the complex field amplitude in the coherent case or the two-point correlation function for partially coherent fields, from the measurements of only intensity distributions. Application of cylindrical lenses allows the reconstruction of two-dimensional optical fields.

In the case of coherent optical signals, other methods for phase retrieval based on the measurements of fractional power spectra have been proposed. One of them is related to the estimation of the instantaneous spatial frequency $\Xi(x)$ from two close fractional power spectra. It was shown that the instantaneous frequency is related to the convolution of the angular derivative of the fractional power spectrum and the signum function [33],

$$\begin{aligned}\Xi_{F_{\beta}}(x) &= \frac{\int_{-\infty}^{\infty} \xi W_f(x\cos\beta - \xi\sin\beta, x\sin\beta + \xi\cos\beta)d\xi}{\int_{-\infty}^{\infty} W_f(x\cos\beta - \xi\sin\beta, x\sin\beta + \xi\cos\beta)d\xi} \\ &= \frac{1}{2\left|F_{\beta}(x)\right|^{2}} \int_{-\infty}^{\infty} \frac{\partial\left|F_{\alpha}(x')\right|^{2}}{\partial\alpha}\bigg|_{\alpha=\beta} \operatorname{sgn}(x-x')dx',\end{aligned} \tag{23}$$

where $\operatorname{sgn}(x) = \pm 1$ for $x \gtrless 0$. Moreover, since the instantaneous frequency is the phase derivative of the fractional FT of a signal,

$$2\pi\Xi_{F_{\beta}}(x) = \frac{d\varphi_{\beta}(x)}{dx}, \tag{24}$$

where $\varphi_{\beta}(x) = \arg F_{\beta}(x)$, the complex field amplitude up to a constant phase factor can be reconstructed from only two close fractional power spectra [33, 34, 35]. This method has been demonstrated on different examples of multicomponent and noisy signals and exhibits high quality of phase reconstruction [35]. Note that a similar method of phase retrieval can be applied for any one-parameter canonical transform [36]. Thus, in the case of the Fresnel transform, we can mention a noniterative approach for phase retrieval in free space, based on the so-called transport-of-intensity equation in optics, proposed by Teague [37] and then further developed by others.

In the case that two fractional power spectra are known for angles which are not close to each other, iterative methods of phase retrieval can be applied [38, 39, 40]. These methods are a generalization of the iterative Gerchberg-Saxton algorithm, designed for the recovery of a complex signal from its intensity distribution and power spectrum.

Another method for phase retrieval is based on a signal decomposition as a series of orthogonal Hermite-Gauss modes [41]. It has been shown that if a coherent optical signal contains only a finite number of Hermite-Gauss modes N, then it can be reconstructed from the knowledge of its $2N$ fractional power spectra—associated with the intensity distribution in a fractional FT system—at only two transversal points. Note that this method can be generalized to the case of other fractional optical systems to be discussed below, such as, for example, the fractional Hankel one.

A further method for phase retrieval is based on filtering of the optical field in fractional Fourier domains [42]. Indeed, the phase derivative $d\varphi/dx$, and therefore the phase $\varphi(x)$ up to a constant term, can be reconstructed from the knowledge of the intensity $|f(x)|^2$ and the intensity distributions at the output of two fractional FT filters with mask u:

$$\frac{d\varphi(x)}{dx} = \pi \frac{\left|\mathcal{R}^{-\alpha}[F_\alpha(u)u](x)\right|^2 - \left|\mathcal{R}^{\alpha}[F_{-\alpha}(u)u](x)\right|^2}{x\left|f(x)\right|^2 \sin 2\alpha}. \tag{25}$$

The efficiency of this approach has been demonstrated by numerical simulations. A simple optical configuration for the experimental realization of the method was discussed in [42].

6. FRACTIONAL POWER SPECTRA FOR OPTICAL BEAM CHARACTERIZATION

Since the AF, the WD, and other bilinear distributions of two-dimensional optical signals are functions of four variables, their direct application for the analysis and characterization is limited. Mostly the moments of these distributions are used for beam characterization. The normalized moments μ_{pqrs} of the WD are defined by

$$\mu_{pqrs}E = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} W(x,\xi;y,\eta) \times x^p \xi^q y^r \eta^s \, dx\, d\xi\, dy\, d\eta \quad (p,q,r,s \geq 0), \tag{26}$$

where the normalization is with respect to the total energy E of the signal (and hence $\mu_{0000} = 1$). Note that in a first-order optical system, with a symplectic ray transformation matrix, the total energy E is invariant. The low-order moments represent the global features of the optical signal such as total energy, width, principal axes, and so forth. Thus the second-order moments of the WD ($p + q + r + s = 2$) are used as a basis of an International Organization for Standardization standard of beam quality. The combination of the second-order moments $(\mu_{1001} - \mu_{0110})E$, for instance, describes the orbital angular momentum of the optical beam, which is actively used for the description of vortex beams [117]. The moments of higher order are related to finer details of the optical signal.

Note that for $q = s = 0$ and for $p = r = 0$, we have the position and frequency moments, which can easily be obtained from measurements of the intensities in the signal and the Fourier domain, respectively:

$$\mu_{p0r0}E = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x^p y^r \left|F_0(x,y)\right|^2 dx\, dy, \tag{27}$$

$$\mu_{0q0s}E = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \xi^q \eta^s \left|F_{\pi/2}(\xi,\eta)\right|^2 d\xi\, d\eta. \tag{28}$$

Since in optics only intensity distributions can be measured directly, it was proposed in [43] to apply fractional FT systems in order to calculate other moments from the intensity moments. It was shown that the moments at the output plane of a separable fractional FT system, with fractional angles α and β in the x- and the y-direction, respectively, are related to the input ones as

$$\mu^{\text{out}}_{pqrs} = \sum_{k=0}^{p}\sum_{l=0}^{q}\sum_{m=0}^{r}\sum_{n=0}^{s}\binom{p}{k}\binom{q}{l}\binom{r}{m}\binom{s}{n} \times (-1)^{l+n}(\cos\alpha)^{p-k+q-l}(\sin\alpha)^{k+l}(\cos\beta)^{r-m+s-n} \times (\sin\beta)^{m+n}\mu^{\text{in}}_{p-k+l,q-l+k,r-m+n,s-n+m}, \tag{29}$$

and for the intensity moments in particular, we have

$$\mu^{\text{out}}_{p0r0} = \sum_{k=0}^{p}\sum_{m=0}^{r}\binom{p}{k}\binom{r}{m}(\cos\alpha)^{p-k}(\sin\alpha)^{k} \times (\cos\beta)^{r-m}(\sin\beta)^{m}\mu^{\text{in}}_{p-k,k,r-m,m}. \tag{30}$$

From (30) a set of fractional FT systems can be found for which the input moments can be derived from knowledge of the intensity moments in the output, that is, from fractional power spectra for selected angles α and β. It was demonstrated [43] that in order to find all nth order moments—and we have $(n+1)(n+2)(n+3)/6$ of such moments—we need N fractional power spectra, where $N = (n+2)^2/4$ for even n and $N = (n+1)(n+3)/4$ for odd n. Moreover, $N - (n+1)$ spectra have to be anamorphic, that is, spectra with nonequal fractional order for the two transversal coordinates ($\alpha \neq \beta$). In particular, we need 2 fractional spectra to find the 4 first-order moments, 4 fractional spectra (one of which has to be anamorphic) to find the 10 second-order moments, 6 fractional spectra (with 2 anamorphic ones) to find the 20 third-order moments, and so forth.

Regarding the evolution of the second-order moments in a fractional FT system, we can find the fractional domain where the signal has the best concentration or where it is the most widely spread, by calculating the zeros of the angular derivatives of the central moments $\mu_{p0r0}(\alpha,\beta)$. This analysis [33, 34] is helpful, for example, in search for an appropriate fractional domain to perform filtering operations [45]. Smoothing interferograms in the optimal fractional domain leads to a weighted WD with significantly reduced interference terms of multicomponent signals, while the auto terms remain almost the same as in the WD. In general, based on this approach, optimal signal-adaptive distributions can be constructed with low cost [46].

The way to determine the moments from measurements of intensity distributions as described by (30) has been generalized to the case of arbitrary separable first-order optical systems [44]. Using an equation similar to (29), one can easily determine the evolution of these moments during propagation of the beam in any first-order optical system; in particular, this was applied to the analysis of optical vortices [47].

In signal processing, the fractional FT spectra were primarily developed for detection and classification of multicomponent linear FM in noise [48, 49].

It was shown [50, 51, 52, 53, 54, 55, 56] that the fractional FT spectra as well as the Fresnel spectra are also useful for the analysis of fractal signals. Thus the hierarchical structure of the fractal fields and its main characteristics such as fractal dimension, Hurst exponent, scaling parameters, fractal level, and so forth, can be obtained from the analysis of the fractional spectra for the angular region from 0 to $\pi/2$ [50, 51, 52, 53]. Since in this region the fractional FT spectra and the Fresnel transform spectra differ only by a scaling parameter, the Fresnel diffraction is applied for this task [51, 52, 55]. Recently the experimental fractal tree of triadic Cantor bars has been constructed from the observation of the evolution of diffraction patterns in free space [54]. The general properties of the Fresnel diffraction by structures constructed through the multiplicative iterative procedure have been studied in [56].

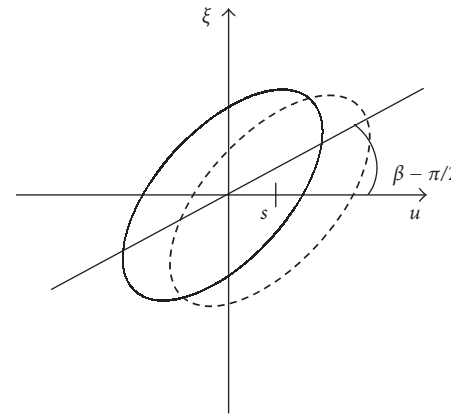


Figure 3: Schematic representation of the cross-AF of two signals, before (solid line) and after (dashed line) shifting of one of the signals.

7. GENERALIZED FRACTIONAL CONVOLUTION FOR PATTERN RECOGNITION

A great part of the proposed applications of the GFC is related to pattern recognition tasks [57, 60, 66, 67, 68, 69, 70, 71, 72, 73, 74]. It was shown [66, 67] that for this purpose, the following relation between the angular parameters has to hold:

$$\cot\alpha = \cot\beta + \cot\gamma. \tag{31}$$

Then the amplitude of the GFC is expressed in the form [66]

$$\begin{aligned} &|H_{f,g^*}(x,\alpha,\beta,\gamma)| \\ &\quad = C\left| \int_{-\infty}^{\infty} f\left[\frac{\sin\beta}{\sin\gamma}\left(x\frac{\sin\gamma}{\sin\alpha} - y\right)\right] g^*(y) \right. \\ &\qquad \left. \times \exp\left[i\pi y^2 \frac{\cot\alpha(1+\cot\gamma\cot\beta)}{1+\cot^2\beta} \right.\right. \\ &\qquad \left.\left. - i\pi yx\frac{\sin 2\beta}{\sin\alpha\sin\gamma}\right] dy \right|, \end{aligned} \tag{32}$$

where C is a constant for fixed α, β, and γ. The quadratic phase factor under the integral vanishes—which brings the integral in the form of a windowed FT—if $\cot\alpha(1 + \cot\gamma\cot\beta) = 0$. In the case $\cot\alpha = 0$ (and hence $\alpha = \pi/2$ and $\gamma = -\beta$) which is usually considered, $H_{f,g^*}(r, \pi/2, \beta, -\beta)$ corresponds to radial slices $A_{f,g}(r, \beta - \pi/2)$ of the cross-AF of the signals $f(x)$ and $g(x)$ (cf. (20)).

If the position and the size of the object is known, then the correlation operation $H_{f,g^*}(x,\alpha,\beta,-\beta)$ for pattern recognition can be performed in any fractional domain β, since the auto-AF has a maximum at the coordinate origin $r = 0$. Nevertheless, in spite of the fact that the magnitude of the correlation maximum is the same in any fractional domain, the forms of the correlation peaks are different. It was shown [70] on the example of a rectangular function that the narrowest correlation peak is observed in the fractional domain with fractional angle $\beta = 0$. Note also that the object is usually corrupted by noise, or is blurred. The characteristics of the noise (except for white noise) in different fractional domains depend on the fractional angle [75]. The fractional correlation offers the flexibility to choose the domain where the effect of noise on the correlation operation is minimized. Moreover, for the recognition of complex or highly degraded objects, several fractional correlation operations for different angles can be performed in order to make the right decision.

On the other hand, if the position of the object is unknown, the choice of the fractional domain is related to the tolerance to a shift variance of the correlation operation. A shift of the signal leads to a shift and a modulation of the cross-AF:

$$A_{f(y-s),g(y)}(x,\xi) = A_{f(y),g(y)}(x-s,\xi)\exp(-i\pi s\xi). \tag{33}$$

Then the form of the AF radial slices of a shifted signal is changing except for the angle corresponding to the ordinary correlation (see Figure 3).

Therefore fractional correlations are shift variant for $\beta \neq \pi/2 + n\pi$. Thus if in the conventional correlator a shift of the object results in a shift with opposite sign of the correlation peak at the output plane, the shape of the peak is also changed in the fractional correlator. This effect increases with decreasing parameter β from $\pi/2$ down to 0. For large β, the fractional correlator is almost shift invariant, whereas for small β, it becomes strongly shift variant. Note that there are applications, such as cryptography or image coding, where the location of the object can be as important as its form. In these cases fractional correlators with fractional parameter β, $0 < \beta < \pi/2$, must be used.

The shift tolerance condition is usually written in the form [29, 59, 60] $\pi s\sigma\cot\beta \ll 1$, where s is the signal shift and σ the signal width. More precisely, the shift variance depends on the fractional order, the signal size, and also the form of the AF.

The tasks of pattern detection and recognition in optics are mostly related to two-dimensional signals (images). It is also possible to choose different fractional orders for the two

orthogonal coordinates and thus to better control the shift variance. In order to recognize a letter on a certain line of the text, for example, one can choose the parameter $\beta_x = \pi/2$ and $\beta_y < \pi/2$ while the filter corresponds to the inverse fractional FT with parameters β_x, β_y of a letter situated on a given line. The exciting results demonstrating the efficiency of shift-variant pattern recognition in the fractional domain can be found in [72, 73, 74].

The fractional correlation operation can be performed in optics by a fractional Van der Lugt correlator [72, 73, 74] or by a nonconventional joint-transform correlator [118].

In order to maximize the Horner efficiency of the correlation operation, phase-only filters are often used. It was shown in [76] that in general, the phase of the fractional FT for $\alpha \neq n\pi$ contains more information about the signal/image than the amplitude. Therefore the phase-only filters can also be applied in the fractional Fourier domain. The development of liquid crystal spatial light modulators allows their relatively simple implementation in optics.

Another particular case of GFC which can be applied for recognition tasks is related to the fractional FT of the ordinary correlation operation [23] $H_{f,g^*}(x, \alpha, \pi/2, -\pi/2)$. We believe that this type of operation can be useful for angles α at the region near $\pi/2$ in order to improve the performance of the conventional correlation operation. Thus it was shown [77] that for α slightly different from $\pi/2$, the performance of the joint-transform correlator improves and higher correlation peaks are observed. Efficient use of the light source and a larger joint-transform spectrum were achieved. Moreover, for these angles α, the correlator still remains shift invariant. Nevertheless using angles α far from $\pi/2$ leads to confusing results for interpreting the correlation peaks. Indeed, if the conventional correlation operation does not produce a clear local maximum and is almost constant, then a sharp peak in fractional correlation $H_{f,g^*}(x, \alpha \approx 0, \pi/2, -\pi/2)$ can appear.

8. GENERALIZED FRACTIONAL CONVOLUTION FOR FILTERING AND DATA PROTECTION

We consider now the filtering operation in the fractional domain. The parameters of the GFC in this case depend on the particular application of filtering. If the filter is used for improvement of image quality or for manipulation of the image f in order to extract its features (e.g., for edge detection or image deblurring), then we have to choose $\beta = \alpha$ in order to represent the result of filtering in the position domain. Since we are free to assign an arbitrary fractional domain for the filter function g, we can as well put $\gamma = \alpha$. Thus the complete operation leads to the $H_{f,g}(x, \alpha, \alpha, \alpha)$. The useful properties of this type of GFC,

$$\mathcal{R}^{\beta}[H_{f,g}(x, \alpha, \alpha, \alpha)](u) = H_{F_\beta, G_\beta}(u, \alpha - \beta, \alpha - \beta, \alpha - \beta),$$
$$H_{f,g}(x, \alpha, \alpha, \alpha) = H_{f,g}(x, \alpha + \pi, \alpha + \pi, \alpha + \pi), \tag{34}$$

were proved in [62]. Moreover, this type of convolution operation is associative for a fixed parameter α.

The GFC $H_{f,g}(x, \alpha, \alpha, \alpha)$ has been found very powerful for noise reduction, if the noise is separable from the signal or very well concentrated in some fractional domain [57]. It was shown that in particular for chirp-like noise, the performance of filtering in a fractional domain is more relevant [24, 29]. Since the fractional FT of a chirp becomes proportional to a Dirac-delta function in an appropriate fractional domain, it can be detected as a local maximum on the Radon-Wigner transform map and then easily removed by a notch filter, which minimizes the signal information loss.

Several applications of fractional FT filtering systems for industrial devices have been proposed recently.

Chirp detection, localization, and estimation via the fractional FT formalism are applied now in different areas of science. Appropriate filtering in fractional domains, which allows to extract linear chirps out of a multicomponent and noisy signal, is used to analyze the propagation of acoustic waves in a dispersive medium [119]. In particular, the nonlinear effects due to the Helmholtz resonators are considered.

A new spatial filtering technique for partially coherent light in the fractional Fourier domain [120] was proposed to improve image contrast and depth of focus in projection photolithography. Unlike the currently applied pupil method of filtering in the Fourier domain, the fractional filter can be placed at any location along the projection optical path other than the pupil plane. On the examples of designed phase filters for contact hole and line-space patterns, it was demonstrated that the fractional FT filtering technique can significantly improve image fidelity, reduce the optical proximity effect, and increase the depth of focus.

Optical technologies play an increasing role in securing information [121]. Also the GFC found its way into security protection: encryption and watermarking techniques originally proposed for the Fourier domain were generalized to the fractional domain.

Optical image encryption by random-phase filtering in the fractional Fourier domain was proposed in [78, 79]. It can be described by the GFC $H_{f,g}(x, \alpha, \beta, \beta)$, where the phase mask G_β and the parameters α and β are the encryption codes. This procedure was further generalized by application of the cascaded fractional FT with random-phase filtering [80]. In order to encode the image, the fractional transform is performed and random-phase is introduced by means of a spatial light modulator. After repeating this procedure several times, the encrypted image is obtained. In order to decode it, not only the information about the used random-phase masks has to be known, but also the parameters and the types of the fractional transforms. It was demonstrated that it is impossible to reconstruct the image using the correct masks but with the wrong fractional orders. Without increasing the complexity of the hardware, the fractional-Fourier optical image encryption system has additional keys provided by the fractional order of the fractional convolution operation. Due to the double domain properties of the fractional FT, the algorithm demonstrates the robustness to the blind deconvolution.

Recently, some modifications of the optical encryption procedures in the fractional Fourier domain were proposed.

Thus in [81] the combination of a jigsaw transform and a localized fractional FT was applied. The image to be encrypted is divided into independent nonoverlapping segments, and each segment is encrypted using different fractional parameters and two statistically independent random-phase codes. The random-phase codes, the set of fractional orders, and the jigsaw transform index are the keys to the encrypted data. The encryption by juxtaposition of sections of the image in fractional Fourier domains without random-phase screen keys was proposed in [82].

Another encryption technique discussed in [83] is based on a method of phase retrieval using the fractional FT. The encrypted image consists of two intensity distributions, obtained in the output of two fractional FT systems of different fractional orders, where the input of each system is formed by the 2D complex signal multiplied by a random-phase mask. The two statistically independent random-phase masks and the fractional orders form the encryption key. Decryption is based on the correlation property of the fractional FT, which allows to recover the signal recursively.

The implementation of a fully phase encryption system, using a fractional FT to encrypt and decrypt a 2D phase image obtained from an amplitude image, was reported in [85]. A comparative analysis of the encryption techniques based on the implementation of the fractional FT has been done in [84].

Watermarking is another widely applied data protection operation. A watermarking technique in the fractional domain was proposed in [86, 87]. In this case, the GFC $H_{f,g}(x,\alpha,\alpha,\alpha)$ is commonly used. In order to include the watermark, the α-fractional FT of the image is performed. The signature has to be a function that is spread in the image domain and well localized in the fractional domain α. Usually a chirp signal, which becomes a δ-function in a certain fractional domain and should be spread in the image domain, is used. Introducing the watermark and performing the inverse fractional FT, finally we obtain the protected image. Usually several watermarks in different fractional FT domains are introduced. Only the owner of the image who knows the all fractional domains will be able to remove them. This watermarking technique is robust to translation, rotation, cropping, and filtering [86, 87].

9. GENERAL ALGORITHM FOR THE FRACTIONALIZATION OF CYCLIC TRANSFORMS

We have considered the properties and application of the fractional FT. Now the following key questions arise.

(i) Is this fractional FT unique? Or is it possible to generate other fractional FTs?

(ii) How can we generate the fractional version of other transformations, for example, Hilbert, sine, cosine?

(iii) Do fractional transforms have some common properties?

In order to answer these questions, we will consider the procedure of fractionalization of a given transform [27, 28]. Similar approaches for fractionalization of the integral transform, and the FT in particular, were reported in [122] and [123], respectively. We will restrict ourselves to the consideration of cyclic transforms. There is a long list of linear transforms, actively used in optics and signal/image processing, which belong to this class of cyclic transforms. Thus, if $\mathcal{R}$ is an operator of a linear integral transform, see (1), this transform is a cyclic one, if it produces the identity transform when it acts an integer number of times N:

$$\mathcal{R}^N[f(x)](u) = f(u). \tag{35}$$

For example, the Fourier and Hilbert transforms are cyclic with a period $N = 4$, and the Hankel and Hartley transforms have a period $N = 2$. Cyclic canonical transforms of period N with kernel $K(x,u) = K_M(x,u)$ (cf. (4)),

$$K(x,u) = \frac{1}{\sqrt{ib}} \exp\left(i\pi \frac{ax^2 + du^2 - 2ux}{b}\right), \tag{36}$$

where $a + d = 2\cos(2\pi m/N)$ and m and N are integers, were mentioned in [124].

All cyclic transforms have some common properties. In particular, the eigenvalues of cyclic transforms can be represented as $A = \exp(i2\pi L/N)$, where L is an integer. Indeed, let $\Phi(x)$ be an eigenfunction of $\mathcal{R}$ with eigenvalue $A = |A|\exp(i\varphi)$; from (35) one gets that $A^N = 1$, and hence $|A| = 1$ and $\varphi = 2\pi L/N$.

In Section 2, we have formulated the requirements for the fractional $\mathcal{R}$-transform $\mathcal{R}^p$, where p is the parameter of the fractionalization: continuity of $\mathcal{R}^p$ for any real value p; additivity of $\mathcal{R}^p$ with respect to the parameter p; reproducibility of the ordinary transform for integer values of p: $\mathcal{R}^1 = \mathcal{R}$ and $\mathcal{R}^0 = I$. In the case of cyclic transforms, we obviously demand that $\mathcal{R}^N = I$.

Let us analyze the structure of the kernel $K(p,x,u)$ of a fractional $\mathcal{R}$-transform with period N. Due to its periodicity with respect to the parameter p, one can represent $K(p,x,u)$ in the form

$$K(p,x,u) = \sum_{n=-\infty}^{\infty} k_n(x,u) \exp\left(\frac{i2\pi pn}{N}\right), \tag{37}$$

where the coefficients $k_n(x,u)$ have to satisfy the system of N equations [27]

$$K(l,x,u) = \sum_{n=-\infty}^{\infty} k_n(x,u) \exp\left(\frac{i2\pi ln}{N}\right) \tag{38}$$

with $l = 0,\ldots,N-1$. From the additivity property for the fractional transform, it follows that the coefficients have to be orthonormal to each other [27, 28]:

$$\int_{-\infty}^{\infty} k_n(x,u)k_m(u,y)du = \delta_{n,m}k_n(x,y), \tag{39}$$

where $\delta_{n,m}$ denotes the Kronecker delta.

Note that all coefficients $k_{n+mN}(x,u)$ for fixed n and an arbitrary integer m have the same exponent factor in the system of (38). Therefore, we can rewrite (38) as

$$K(l,x,u) = \sum_{n=0}^{N-1} \exp\left(\frac{i2\pi ln}{N}\right) \sum_{m=-\infty}^{\infty} k_{n+mN}(x,u). \tag{40}$$

If we introduce the new variables $C_n(x,u)$, which are the partial sums of the coefficients in the Fourier expansions (37) and (38),

$$C_n(x,u) = \sum_{m=-\infty}^{\infty} k_{n+mN}(x,u), \tag{41}$$

equation (40) reduces to a system of N linear equations with N variables. This system has a unique solution [27]

$$C_n(x,u) = \frac{1}{N} \sum_{l=0}^{N-1} \exp\left(-\frac{i2\pi ln}{N}\right) K(l,x,u). \tag{42}$$

It is easy to see that the variables C_n satisfy a condition similar to (39):

$$\int_{-\infty}^{\infty} C_n(x,u) C_m(u,y)\,du = \delta_{n,m} C_n(x,y). \tag{43}$$

Note that some partial sums for certain transforms may be equal to zero. As we will see further on, this is the case for the Hilbert transform, for instance.

So, if we find the coefficients $k_n(x,u)$ that satisfy the condition (39) and whose partial sums are given by (42), we can construct the fractional transform. In general, there are a number of sets $\{k_n(x,u)\}$ that generate fractional transforms of a given $\mathcal{R}$-transform.

10. N-PERIODIC FRACTIONAL TRANSFORM KERNELS WITH N HARMONICS

We first construct the fractional transform kernel with N harmonics, where N is the period of the cyclic transform. Then every sum $C_n(x,u)$ ($n \in [0, N-1]$) contains only one element $k_{n+\varphi_n}(x,u) = C_n(x,u)$ from the decomposition (37), where $\varphi_n = mN$ and m is an arbitrary integer. Therefore, in the general case, the kernel of the fractional $\mathcal{R}$-transform with N harmonics can be written as

$$\begin{aligned} K(p,x,u) &= \sum_{n=0}^{N-1} k_{n+\varphi_n}(x,u) \exp\left[\frac{i2\pi p(n+\varphi_n)}{N}\right] \\ &= \frac{1}{N} \sum_{l=0}^{N-1} K(l,x,u) \sum_{n=0}^{N-1} \exp\left(-\frac{i2\pi ln}{N}\right) \\ &\quad \times \exp\left[\frac{i2\pi p(n+\varphi_n)}{N}\right]. \end{aligned} \tag{44}$$

This equation provides a formula for recovering the continuous periodic function $K(p,x,u)$ from its N samples $K(l,x,u)$, under the assumption that the spectrum of $K(p,x,u)$ contains only N harmonics at the frequencies $\{\varphi_0, 1+\varphi_1, \ldots, n+\varphi_n, \ldots, N-1+\varphi_{N-1}\}$.

If we put $\varphi_n = 0$ ($n = 0, 1, \ldots, N-1$), we obtain the fractional transform with the kernel

$$\begin{aligned} K(p,x,u) &= \frac{1}{N} \sum_{l=0}^{N-1} \exp\left[\frac{i\pi(N-1)(p-l)}{N}\right] \\ &\quad \times \frac{\sin[\pi(p-l)]}{\sin[\pi(p-l)/N]} K(l,x,u) \end{aligned} \tag{45}$$

proposed by Shih in [125]. In particular, this formula is used as the definition of a kind of fractional FT (for the continuous as well as the discrete case) [125, 126].

With N an odd integer and choosing N nonzero coefficients in the decomposition (37) with indices $j = -(N-1)/2, \ldots, 0, \ldots, (N-1)/2$ (corresponding to the indices $n + mN$ for $m = 0$ and $n = 0, 1, \ldots, (N-1)/2$, and $m = -1$ and $n = (N-1)/2+1, \ldots, N-1$), we obtain the kernel

$$K(p,x,u) = \frac{1}{N} \sum_{l=0}^{N-1} \frac{\sin[\pi(p-l)]}{\sin[\pi(p-l)/N]} K(l,x,u). \tag{46}$$

This equation corresponds to the recovering procedure of a band-limited periodic function from its values on equidistant sampling points [127]. In particular, if $K(l,x,u)$ is real for integer $l = 0, 1, \ldots, N-1$, then the kernel of the fractional transform determined by (46) is real, too. It also means that the Fourier spectrum of $K(p,x,u)$ with respect to the parameter p is symmetric: $|k_j| = |k_{-j}|$.

As an example, we consider the general expression (44) for the kernel of the fractional $\mathcal{R}$-transform with period 4 (which is the case for the Fourier and Hilbert transforms):

$$K(p,x,u) = \frac{1}{4} \sum_{l=0}^{3} K(l,x,u) S(l) \tag{47}$$

with $S(l) = \sum_{n=0}^{3} \exp(-inl\pi/2) \exp[i(n+\varphi_n)p\pi/2]$.

Note that for the Hilbert transform, the number of harmonics reduces to two, because $C_0(x,u) = C_2(x,u) = 0$, which follows from $K(0,x,u) = -K(2,x,u)$ and $K(1,x,u) = -K(3,x,u)$. From (44) we then conclude that the fractional Hilbert transform kernel can be written as

$$\begin{aligned} K(p,x,u) &= \exp[i(m_1+m_3+1)p\pi] \\ &\quad \times \left\{K(0,x,u) \cos\left[\left(m_3 - m_1 + \frac{1}{2}\right)p\pi\right]\right. \\ &\quad \left. - K(1,x,u) \sin\left[\left(m_3 - m_1 + \frac{1}{2}\right)p\pi\right]\right\}, \end{aligned} \tag{48}$$

where m_1 and m_3 are integers. In particular, for the case $m_1 = m_3 = 0$ ($k_n = 0$ if $n \neq 1, 3$), one gets

$$K(p,x,u) = \exp(ip\pi) \times \left[K(0,x,u)\cos\left(\frac{p\pi}{2}\right) - K(1,x,u)\sin\left(\frac{p\pi}{2}\right)\right], \tag{49}$$

while for the case $m_1 = 0$ and $m_3 = -1$ ($k_n = 0$ if $n \neq -1, 1$), the common form for the fractional Hilbert transform [94] with a real kernel is obtained:

$$K(p,x,u) = K(0,x,u)\cos\left(\frac{p\pi}{2}\right) + K(1,x,u)\sin\left(\frac{p\pi}{2}\right). \tag{50}$$

Therefore, even for the same number of harmonics, there are several ways for the fractionalization of cyclic transforms.

11. FRACTIONAL TRANSFORM KERNELS CONSTRUCTION USING EIGENFUNCTIONS OF CYCLIC TRANSFORMS

In the case that the set of orthonormal eigenfunctions of the cyclic transform exists, one can construct fractional kernels with a number of harmonics $M > N$, where N is the period of the cyclic transform [27, 28].

Suppose that there is a complete set of orthonormal eigenfunctions $\{\Phi_n\}$ of the operator $\mathcal{R}$ with eigenvalues $\{A_n = \exp(i2\pi L_n/N)\}$, $n = 0, 1, \ldots$ (see Section 9). Then we can represent a kernel of the $\mathcal{R}$-transform of the integer power q as

$$\begin{aligned} K(q,x,u) &= \sum_{n=0}^{\infty} \Phi_n(x) A_n^q \Phi_n^*(u) \\ &= \sum_{n=0}^{\infty} \Phi_n(x) \exp\left(\frac{i2\pi q L_n}{N}\right) \Phi_n^*(u). \end{aligned} \tag{51}$$

One of the possible series of kernels for the fractional $\mathcal{R}$-transform can then be written in the form

$$K(p,x,u) = \sum_{n=0}^{\infty} \Phi_n(x) \exp\left[i2\pi\left(\frac{L_n}{N} + l_n\right)p\right] \Phi_n^*(u), \tag{52}$$

where l_n is an integer and indicates the location of the harmonics. This kernel satisfies the additivity condition due to the orthonormality of the eigenfunctions $\Phi_n(x)$.

Note that not all cyclic operators have a complete set of orthonormal eigenfunctions, as it is the case, for example, for the Hilbert operator, whose eigenfunctions $\Phi(x)$ are self-orthogonal. Nevertheless, the majority of cyclic transforms of interest in optics, such as Fourier, Hartley, Hankel, and so forth, have this set. For the Fourier and Hartley transforms, $\Phi_n(x)$ are the Hermite-Gauss modes [15, 16]

$$\Phi_n(x) = 2^{1/4}(2^n n!)^{-1/2} H_n(x\sqrt{2\pi}) \exp(-\pi x^2), \tag{53}$$

where $H_n(x)$ are the Hermite polynomials; for the Hankel transform of different orders, $\Phi_n(x)$ are the normalized Laguerre-Gauss functions [128, 129].

The canonical fractional FT kernel, discussed in the previous sections, can be obtained from (52) as a particular case: $L_n = -n$ and $l_n = 0$,

$$\begin{aligned} K_F(p,x,u) &= \sum_{n=0}^{\infty} \Phi_n(x) \exp\left(-\frac{inp\pi}{2}\right) \Phi_n^*(u) \\ &= \frac{\exp(inp\pi/4)}{\sqrt{i\sin(p\pi/2)}} \exp\left[i\pi \frac{(x^2+u^2)\cos(p\pi/2) - 2ux}{\sin(p\pi/2)}\right] \end{aligned} \tag{54}$$

(cf. (11)). The fractional Hankel transform, defined by (52) for $L_n = -n$ and $l_n = 0$ and $\Phi_n(x)$ being the normalized Laguerre-Gauss functions, describes the propagation of rotationally symmetric optical beams through a medium with a quadratic refractive index [128, 129]. The kernels of these transforms contain an infinite number of harmonics.

We rewrite (52) in the form

$$K(p,x,u) = \sum_{n=-\infty}^{\infty} z_n(x,u) \exp\left(\frac{i2\pi np}{N}\right). \tag{55}$$

Here $z_n(x,u)$ is a sum of the elements $\Phi_j(x)\Phi_j^*(u)$ over j, where $\Phi_j(x)$ is the eigenfunction of the $\mathcal{R}$-transform with eigenvalue $\exp(i2\pi n/N)$. Thus for the case of the canonical fractional FT,

$$\begin{aligned} K_F(p,x,u) &= \sum_{n=0}^{\infty} \Phi_n(x) \exp\left(-\frac{inp\pi}{2}\right) \Phi_n^*(u) \\ &= \sum_{n=-\infty}^{0} z_n(x,u) \exp\left(\frac{inp\pi}{2}\right), \end{aligned} \tag{56}$$

the coefficients $z_n(x,u)$ vanish for positive n and $z_n(x,u) = \Phi_n(x)\Phi_n^*(u)$ for $n \leq 0$. As we will see below, the fractional Hartley transform [27] can be represented in the form

$$\begin{aligned} K(p,x,u) &= \sum_{n=0}^{\infty} \exp(-i\pi np) z_{-n}(x,u), \\ z_{-n}(x,u) &= \Phi_{2n}(x)\Phi_{2n}(u) + \Phi_{2n+1}(x)\Phi_{2n+1}(u). \end{aligned} \tag{57}$$

It is easy to see from (55) that we can generate another kernel series with M harmonics,

$$K(p,x,u) = \sum_{n=0}^{M-1} \exp\left(\frac{i2\pi np}{M}\right) \sum_{m=-\infty}^{\infty} z_{n+mM}(x,u), \tag{58}$$

which satisfy the requirements for the fractional transforms. Here the sums of the elements $z_j(x,u)$,

$$k_n(M,x,u) = \sum_{m=-\infty}^{\infty} z_{n+mM}(x,u), \tag{59}$$

are used as the coefficients $k_n(x,u)$ in (37). Note that the relationship (39) holds for the coefficients $k_n(M,x,u)$ and $k_m(M,x,u)$ because they are constructed from the disjoint series of orthonormal elements.

One can prove that the kernel (58) for $p = 1$ reduces to (51). In particular, if $\{\Phi_n\}$ is the Hermite-Gauss mode set and $z_{-n}(x,u) = \Phi_n(x)\Phi_n^*(u)$ for $n = 0, 1, \ldots$ and $z_{-n}(x,u) = 0$ for negative n, then (58) corresponds to the series of the M-harmonic fractional FTs proposed in [130],

$$\begin{aligned} K(p,x,u) &= \sum_{n=0}^{M-1} \exp\left[-\frac{i2\pi np(1-M)}{M}\right] \\ &\quad\times \sum_{m=0}^{\infty} \Phi_{n+mM}(x)\Phi_{n+mM}^*(u) \\ &= \frac{1}{M}\sum_{n=0}^{M-1} \exp\left[\frac{i\pi(M-1)(pl-n)}{M}\right] \\ &\quad\times \frac{\sin[\pi(pl-n)]}{\sin[\pi(pl-n)/M]} K_F\left(\frac{n}{l},x,u\right), \end{aligned} \tag{60}$$

where $K_F(n/l,x,u)$ is the kernel of the canonical fractional FT. Application of such types of fractional FTs for image encryption was reported in [80]. If $M = N$ ($l = 1$), we obtain that the kernel of the Shih fractional transform defined by (45) can also be represented as

$$\begin{aligned} K(p,x,u) &= \sum_{n=0}^{N-1} \exp\left[-\frac{i2\pi np(1-N)}{N}\right] \\ &\quad\times \sum_{m=0}^{\infty} \Phi_{n+mN}(x)\Phi_{n+mN}^*(u). \end{aligned} \tag{61}$$

Finally we can conclude that if a complete orthonormal set of eigenfunctions for a given cyclic transform exists, then an infinite number of fractional transform kernels with an arbitrary number of harmonics can be constructed using the procedure (52). Some examples of fractional FTs whose kernels contain different numbers of harmonics were considered in [27].

12. SOME PROPERTIES OF FRACTIONAL CYCLIC TRANSFORMS

Although there is a variety of schemes for the construction of fractional transforms, all of them have some common properties.

If the coefficients $k_n(x,u)$ in the decomposition (37) are real, then the following relationship holds:

$$\{\mathcal{R}^p[f^*(x)](u)\}^* = \mathcal{R}^{-p}[f(x)](u). \tag{62}$$

This is the case for the canonical fractional FT, the related fractional sine, cosine, and Hartley transforms, and the canonical fractional Hankel transform.

Eigenfunctions of fractional transforms

By analogy with the analysis of the fractional FT eigenfunctions, made in [106, 107], the eigenfunction $\Psi_{1/M}(x)$ for the fractional transform $\mathcal{R}^p$ for $p = 1/M$ with eigenvalue $A = \exp(i2\pi L/M)$ can be constructed from the arbitrary generator function $g(u)$ by the following procedure:

$$\Psi_{1/M}(x) = \frac{1}{M}\sum_{n=0}^{M-1} \exp\left(-\frac{i2\pi nL}{M}\right)\mathcal{R}^{n/M}[g(u)](x). \tag{63}$$

In the limiting case $M \to \infty$, one gets the eigenfunction for any value p with eigenvalue $\exp(i2\pi pL)$:

$$\Psi_p^L(x) = \frac{1}{N}\int_0^N \exp(-i2\pi pL)\mathcal{R}^p[g(u)](x)dp. \tag{64}$$

In particular, for fractional transforms generated by (52) (as it was shown by the example for the fractional FT [107]), the functions $\Psi_p^L(x)$ correspond to the elements of the orthogonal set $\{a_L\Phi_L\}$, where the constant factors depend on the generator function.

Complex and real fractional transform kernels

We have seen in the previous section that if there exists a complete orthonormal set of eigenfunctions $\{\Phi_n\}$ for the $\mathcal{R}$-transform, then any coefficient in the harmonic decomposition of the fractional kernel $k_n(x,u)$ (37) can be expressed as a linear composition of the elements $\Phi_j(x)\Phi_j^*(u)$. For the kernel of the fractional transform to be real, the Fourier spectrum of the fractional kernel with respect to the parameter p has to be symmetric; this means that $|k_{-n}(x,u)| = |k_n(x,u)|$. Since the coefficients $k_n(x,u)$ with different indices n contain disjoint series of the orthogonal elements, their amplitudes cannot be equal. In the case that there exists a complete orthonormal set of eigenfunctions $\{\Phi_n\}$ for the $\mathcal{R}$-transform, the fractional kernel of the $\mathcal{R}^p$-transform cannot be real, even if the $\mathcal{R}$-transform kernel is real.

As we have seen above, the fractional Hilbert kernel can be real because there is no complete orthonormal set of eigenfunctions for the Hilbert transform.

13. FRACTIONAL CYCLIC TRANSFORMS IMPLEMENTED IN OPTICS

Besides the canonical fractional FT discussed in Sections 4 and 11, other fractional cyclic transforms can be performed by optical setups. Thus the fractional FTs described in Sections 10 and 11 and represented as a sum of the weighted canonical fractional FTs for the corresponding parameters $\{\alpha_n\}$ (see for instance (45) and (60)) can be obtained as an interference of optical beams at the output of the related canonical fractional FT optical systems. In general, most fractional cyclic transforms proposed for optical implementation are closely connected to the canonical fractional FT.

The two-dimensional fractional FT of a rotationally symmetric function leads to the fractional Hankel transform,

analogous to the fact that its two-dimensional FT produces the Hankel transform [128, 129]. The fractional Hankel transform of a function $f(r)$ is defined as

$$\mathcal{R}^{\alpha}[f(r)](u) = H_{\alpha}(u) = \int_0^{\infty} K(\alpha, r, u) f(r) r\, dr, \tag{65}$$

where the kernel $K(\alpha, r, u)$ is given by

$$K(\alpha, r, u) = \frac{\exp(i\alpha)}{i \sin\alpha} \exp\left[i\pi (r^2 + u^2) \cot\alpha\right] \times J_0\left(\frac{2\pi r u}{\sin\alpha}\right) \tag{66}$$

with J_0 the first-type, zero-order Bessel function. One can represent the fractional Hankel kernel in the form (52), where $L_n = -n$, $l_n = 0$, and $\Phi_n(x)$ are the Laguerre-Gauss functions, which are the eigenfunctions of the fractional Hankel transform.

The fractional Hankel transform inherits the main properties of the fractional FT [103, 128] and can be performed by the fractional FT setups described in Sections 3 and 4, if the input optical field is rotationally symmetric. The fractional Hankel transform can substitute the fractional FT in many optical signal processing tasks where rotationally symmetric beams are used.

Since the FT is closely related to sine, cosine, and Hartley transforms, which are cyclic ones with period $N = 2$, several attempts to introduce the fractional sine, cosine, and Hartley transforms were made in [25, 26], where the authors supposed that the kernels of these transforms are the real part of the kernel of the optical fractional FT, the imaginary part of this kernel, or the sum of these two parts, respectively. Nevertheless, they have mentioned that the transforms defined in such a manner are not angle additive, and therefore, in our view, cannot be interpreted as fractional transforms. The kernels K_S, K_C, and K_H of the fractional sine, cosine, and Hartley transforms (ST, CT, HT) [27, 28, 131], respectively, which are closely related to the canonical fractional FT with kernel K_F and which are indeed angle additive, are defined as

$$\begin{aligned} ie^{-i\alpha} K_S(\alpha, x, u) &= 2k_{\alpha}(x, u) \sin\left(\frac{2\pi u x}{\sin\alpha}\right), \\ K_C(\alpha, x, u) &= 2k_{\alpha}(x, u) \cos\left(\frac{2\pi u x}{\sin\alpha}\right), \\ K_H(\alpha, x, u) &= k_{\alpha}(x, u) \operatorname{cas}\left(\frac{2\pi u x}{\sin\alpha}\right), \\ K_F(\alpha, x, u) &= k_{\alpha}(x, u) \exp\left(-\frac{i2\pi u x}{\sin\alpha}\right), \end{aligned} \tag{67}$$

where

$$k_{\alpha}(x, u) = \frac{\exp(i\alpha/2)}{\sqrt{i \sin\alpha}} \exp\left[i\pi (x^2 + u^2) \cot\alpha\right], \tag{68}$$

where, on the analogy of $\exp(i\varphi) = \cos\varphi + i \sin\varphi$, we have introduced $\operatorname{cas}\varphi = \cos\varphi + \sin\varphi$, and where, for easy reference, we have repeated the expression of the canonical fractional FT kernel K_F.

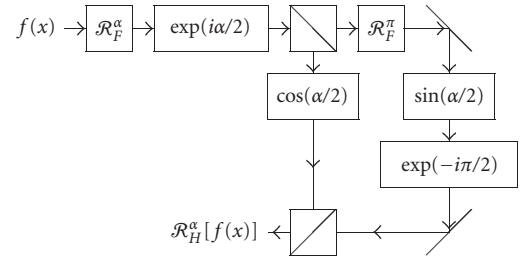


Figure 4: Schematic representation of a fractional Hartley transformer. The setup consists of two fractional FTs $\mathcal{R}_F^{\alpha}$ and $\mathcal{R}_F^{\pi}$, two beam splitters, two mirrors, two absorbing plates $\cos(\alpha/2)$ and $\sin(\alpha/2)$, and two phase plates $\exp(i\alpha/2)$ and $\exp(-i\pi/2)$.

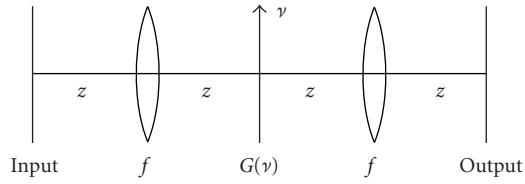


Figure 5: Schematic representation of a (fractional) Hilbert transformer: $z = f$, $G(\nu) = \exp[i\alpha \operatorname{sgn}(\nu)]$.

Since the fractional ST, CT, and HT can easily be expressed in terms of the fractional FT, and since optical realizations of the fractional FT [25] are well known, optical realizations of the fractional ST, CT, and HT can easily be constructed. One of the possible schemes for the fractional HT, based on [27]

$$\mathcal{R}_H^{\alpha} = \exp\left(\frac{i\alpha}{2}\right) \mathcal{R}_F^{\alpha}\left[\cos\left(\frac{\alpha}{2}\right) - i \sin\left(\frac{\alpha}{2}\right) \mathcal{R}_F^{\pi}\right], \tag{69}$$

is given in Figure 4.

As the ST, CT, and HT are widely used in signal processing, the application of their fractional versions in signal/image processing is very promising.

Since, as we have seen in Section 10, the kernel of the fractional Hilbert transform has only two harmonics, the number of possible fractionalization procedures is significantly reduced. The real kernel of the fractional Hilbert transform introduced in [94, 95] and described by (50) is commonly used. Optical setups performing this transform were proposed in [94, 95]. As the fractional Hilbert transform is a weighted mixture of the optical field $f(u)$ itself and its Hilbert transform $H(u)$,

$$\mathcal{R}^{\alpha}[f(x)](u) = f(u) \cos\alpha + H(u) \sin\alpha, \tag{70}$$

an optical scheme performing the ordinary Hilbert transform (see Figure 5, with $G(\nu) = i \operatorname{sgn}(\nu)$) can easily be adapted to perform a fractional Hilbert transform, by having the filter function $G(\nu)$ now taking the more general form $\exp[i\alpha \operatorname{sgn}(\nu)] = \cos\alpha + i \operatorname{sgn}(\nu) \sin\alpha$.

The Hilbert transform can be considered as a convolution of a function with a step function, which is a model

for a perfect edge. Therefore the Hilbert transform produces edge enhancement. It was shown that the fractional Hilbert transform stresses the right-hand and the left-hand slopes unequally [94, 95, 96] and that variation of the fractional order changes the nature of the edge enhancement. Thus, for $\alpha \approx \pi/4, \pi/2, 3\pi/4$, the right-hand edges, both edges, and the left-hand edges of the input object are emphasized, respectively. In general, we can conclude that the fractional Hilbert transform produces an output image that is selectively edge enhanced. This property of the fractional Hilbert transform makes it a perspective tool for image processing and pattern recognition.

14. NEW HORIZONS OF FRACTIONAL OPTICS

Fractional optics is a rapidly developing research area. Novel applications of the fractional transforms for motion detection and analysis, holographic data storage, optical neural networking, and optical security (see Section 8) have been proposed recently. In this section, we give a short overview of the main directions of development of fractional optics.

Fractional Fourier transformers

Significant work has been done to improve fractional transformers.

The effect of the spherical aberration of a lens on the performance of the fractional Fourier transformation in the optical systems proposed by Lohmann in [22] was analyzed in [132]. It was shown that the effect of spherical aberration on the output intensity distribution of the fractional FT system depends on the sign and the absolute value of the aberration coefficient. Moreover, Lohmann's two types of optical setups for implementing the fractional FT are no longer equivalent if the lenses suffer from spherical aberration.

In the optical systems proposed in [22], the fractional order is fixed by the ratio between the focal length of the lens and the distance of free space preceding and following the lens. This fact introduces a difficulty in the design of fractional Fourier transformers with a variable order. Fractional FT systems with a fixed optical setup, but with different fractional orders, can be obtained by the implementation of programmable lenses, written onto a liquid-crystal spatial light modulator [133].

A one-dimensional, variable fractional Fourier transformer, based on the application of a reconfigurable electro-optical waveguide, was proposed in [134]. In general, this device produces a variable canonical transformation, with a ray transformation matrix for which $a = d$ and for which the matrix entry b is controlled by the amplitude of an electric field.

A quantum circuit for the calculation of a fractional FT whose kernel contains four harmonics was proposed in [135].

Propagation through a fractional FT system

The evolution during propagation through fractional FT systems of different types of beams frequently used in modern optics, such as flattened Gaussian [136, 137], elliptical Gaussian [138], and partially coherent and partially polarized Gaussian-Shell beams [139], has been studied. In particular, it was shown that the intensity distribution and polarization properties in the fractional FT plane are closely related to the fractional order of the fractional FT system and the initial coherence of the partially coherent beam [139, 140, 141].

Several devices for manipulation of optical beams based on the fractional FT have been proposed recently.

The fractional FT is applied in the $\pi/2$ converter, which is used to obtain focused Laguerre-Gauss beams from Hermite-Gauss radiation modes [142].

The design of a diffractive optical element for beam smoothing in the fractional Fourier domain was described in [143].

An iterative method for the reconstruction of a wave field or a beam profile from measurements obtained using low-resolution amplitude and phase sensors in several fractional Fourier domains was proposed in [144].

Motion analysis

Several applications of the fractional FT for motion analysis have been proposed.

A method for the independent estimation of both surface tilting and translational motion using the speckle photographic technique by capturing consecutive images in two different fractional Fourier domains, has been proposed in [145].

In [146] the fractional FT is applied to airborne, synthetic aperture radar, slow-moving target detection. Since the echo from a ground moving target can be approximated as a chirp signal, the fractional FT is used to concentrate its energy. An iterative detection of strong moving targets and weak ones, based on filtering in the fractional Fourier domain, has been proposed.

The application of fractional FT correlators to control movements in a specific range has been considered in [147]. Based on the controllable shift variance of fractional correlations, only the movements limited to a specific range are determined. Fractional FT correlators operating with a log-polar representation of two dimensional images (fractional Mellin-based correlator) allow to control the similarity of objects under rotation and scale transformations. Optically implemented fractional FT and Mellin correlators, providing correlation images directly at image acquisition time, have been proposed to be used in detecting or controlling a specific range of movements in navigation tasks.

Beamforming is another application of the fractional FT indirectly related to motion analysis. Beamforming is widely used in sensor arrays, signal processing for signal enhancement, direction of arrival and velocity estimation, and so forth. The conventional minimum mean-squared error beamforming in the frequency domain or the spatial domain has been generalized to the fractional Fourier domain case [148]. It is especially useful for radar problems where chirp signals are encountered. Note that acceleration of the sinusoidal signal source yields that, due to the Doppler effect,

a chirp signal arrives at the sensor. Such a chirp signal is often transmitted in active radar systems.

Fractional FT implemented as neural networks

Several neural network schemes have been proposed recently, in which the canonical fractional FT was implemented.

An optical neural network based on the fractional correlation realized by a Van der Lugt correlator that employs fractional FTs was proposed in [88]. The error back-propagation algorithm was used to provide the learning rule by which the filter values are changed iteratively to minimize the error function.

The replacement of the mean square error with the log-likelihood and the introduction of parallelism to this network significantly improve its learning convergence and recall rate [89].

It was demonstrated in [90] that, due to the shift variance of the fractional convolution, the fractional Van der Lugt correlator is more suitable than the conventional one for classification tasks. For a phase modulation filter, the optimal learning rate to improve the learning convergence and the classification performance can quickly be found by Newton's method.

Besides these static networks with fixed weights and the learning based on the adjustment of the filter coefficients, another type of neural networks using the fractional FT has been proposed [91]. In this scheme, the fractional FT is used for preprocessing of input signals to neural networks. Adjusting the fractional order of the fractional FT of the input signal leads to an overall improvement of the neural network performance, as has been demonstrated on the example of recognition and position estimation of different objects from their sonar returns. In [92], a comparative analysis has been made of different approaches of target differentiation and localization, including the target differentiation algorithm, Dempster-Shafer evidential reasoning, different kinds of voting schemes, statistical pattern recognition techniques (the k-nearest neighbor classifier, the kernel estimator, the parameterized density estimator, linear discriminant analysis, and the fuzzy c-means clustering algorithm), as well as artificial neural networks, trained with different input signal representations obtained using preprocessing techniques such as discrete ordinary and fractional Fourier, Hartley and wavelet transforms, and Kohonen's self-organizing feature map. It has been shown that the use of neural networks trained by the back-propagation algorithm with fractional FT preprocessing results in near-perfect differentiation, around 85% correct range estimation, and around 95% correct azimuth estimation.

The potential application of a spatially varying, fractional correlation in implementing parallel fuzzy association has been explored in [93].

Fresnel and fractional FT holograms

Holographic recording/reconstruction techniques are very well established for image, Fourier, and Fresnel holograms. In particular, since the fractional FT and the Fresnel transform belong to the class of canonical integral transforms, see (3), one can analyze the feasibility of fractional Fourier holograms in relation to Fresnel holograms properties. The fundamentals of Fresnel holograms have been known for about four decades. In 1965, Armstrong [149] published a general contribution on the basic formulation for describing the recording and reconstruction of two- and three-dimensional Fresnel holograms using paraxial approximation. The total complex amplitude at recording step was formulated as a Fresnel integral. Assuming linearity conditions on the hologram development, the final field for the reconstructed image was formulated. This general analytical expression contains the basic contributions to the final field, namely, illumination conditions of the object, hologram diffraction properties, and illumination conditions of the hologram. To this respect, one of the major keys for Fresnel holograms with high reconstruction fidelity is resolution, which is determined by the hologram aperture. If the effect of a finite aperture on the complex amplitude field is not taken into account, the hologram diffraction modulation turns out to be represented by a Dirac-delta function. Assuming that under a fractional FT regime, we are dealing with a particular scaling defined by the order of the transformation, one can assert that resolution plays a similar role as in a standard Fresnel hologram. Taking into account the finite aperture implies that an intrinsic dependence of the hologram quality (fidelity) on the aberrated reconstructed wavefront has also to be considered.

Recently, several contributions on these subjects have been published. An algorithm for digital holography based on the so-called Fresnelets, which arise when the Fresnel transform is applied to a wavelet basis, has been developed in [150]. An experimental digital holography setup was shown, as well as results for Fresnelet holograms. An interesting result of this paper is related to the uncertainty relation for the Fresnel transform as a condition for signal localization. The condition suggests that Gaussian and Gabor-like functions, modulated with a Fresnel kernel, optimize the processing and reconstruction of Fresnel holograms. This is related to the illumination conditions of the object and the behavior of the hologram aperture as an apodizing optical pupil.

Other experiments for implementing Fresnel holograms on LCDs have been reported [151]. The hologram is obtained by back-propagating the object function applying an inverse Fresnel transform. Results indicate the presence of noise in the reconstruction due to the limited number of amplitude levels of the signal (8 bits).

Other authors [152] claim the implementation of multiple fractional FT holograms. Nevertheless, the lack of information related to critical parameters as real illumination conditions, object size, hologram size, type of holographic material, and minimum distance between objects for avoiding aliasing makes it difficult to arrive to a precise conclusion on the actual proposed technique.

From the above-mentioned results, we can assert that Fresnel and fractional FT holograms show great practical interest for signal localization, data coding and decoding, and optical security systems.

Novel fractional transforms

Another direction for further research is the investigation of novel types of fractional transforms: their properties, applications, and possible implementations in optics.

The fractional cosine, sine, and Hartley transforms and their digital implementations were discussed in [27, 28, 131, 153]. It was shown that the fractional cosine and sine transforms are useful for processing one-sided signals, that is, the independent variable is an element of $[0, \infty)$. From our point of view, the different types of fractional FT, ST, CT, and HT, constructed by the general fractionalization algorithm (see Sections 10 and 12), may be suitable for signal/image encryption and watermarking. Thus an image watermarking scheme based on different types of fractional discrete Fourier, Hartley, cosine, and sine transforms was proposed in [154]. To remove the watermarks in this case, the type of the used fractional transform and the orders of the fractional domains where signatures were introduced have to be known.

Several other fractional transforms have been introduced recently.

The fractional FT of log-polar representation of a two-dimensional image generates the fractional Mellin transform [122, 147].

Another fractional transform, the complex fractional FT, closely related to the canonical fractional FT has been introduced in [155]. With $\xi = \xi_1 + i\xi_2$ and $\eta = \eta_1 + i\eta_2$, the kernel of the complex fractional FT takes the form (cf. (11))

$$\frac{\exp(i\alpha)}{i\sin\alpha}\exp\left[i\pi\frac{(|\xi|^2+|\eta|^2)\cos\alpha+\xi\eta^*-\xi^*\eta}{2\sin\alpha}\right]. \quad (71)$$

Based on the approach of eigenfunction kernel decomposition similar to [27, 28], some new fractional integral transforms, including the fractional Mellin transform, a fractional transform associated with the Jacobi polynomials, a Riemann-Liouville fractional derivative operator, and a fractional Riemann-Liouville integral, have been proposed in [122]. In the analogy with canonical fractional Fourier and Hankel transforms, the fractional Laplace and Barut-Girardello transforms have been introduced in [156].

The applications of these transforms in science and engineering are still subject of research.

15. CONCLUSIONS

We have reviewed the fractional transformations implemented in paraxial optics and their applications for optical information processing: phase retrieval, signal/image characterization, optical beam manipulation, pattern recognition and classification, adaptive filter design, encryption, watermarking, motion detection, holography, and so forth. A general algorithm of fractionalization, which allows to construct various fractional transforms related to a given cyclic transform, has been discussed. The usefulness of a specific fractional transform is related to its optical feasibility, as well as to its possible application in signal/image processing. The analysis of the harmonic contents for various types of fractional transforms offers a procedure for their experimental realization. It seems that the fractional sine, cosine, Hartley, and Hankel transforms, discussed in Sections 13 and 14, due to their similarity to the canonical fractional FT, may act as a substitute for it in many tasks. The usage of the fractional Hilbert transform for selective edge enhancement produces very promising results. The exploration of other recently proposed fractional transforms is expected in near future.

Beside the theoretical and numerical simulation works demonstrating an important impact of the optical implementation of the fractional FT, the experimental realization of the corresponding devices and techniques takes up a significant place of research.

We believe that fractional optics significantly increases the importance of analog optical information processing. The design of new devices based on fractional optics will lead to unified approaches of signal/image processing used in optics and electrical engineering, which will significantly enrich the fields of optoelectronics, optical security technology, and optical computing.

ACKNOWLEDGMENTS

This work has been financially supported partially by the project TIC 2002-01846 from the Spanish Ministry of Education and Science and by the project IST-2001-34168 "Two-dimensional optical storage for high density and high data rate" from the European Commission. T. Alieva thanks the Spanish Ministry of Science and Technology ("Ramon y Cajal" grant).

REFERENCES

[1] H. J. Caulfield, "Perspectives in optical computing," *IEEE Computer*, vol. 31, no. 2, pp. 22–25, 1998.
[2] D. L. Flannery and J. L. Horner, "Fourier optical signal processors," *Proc. IEEE*, vol. 77, no. 10, pp. 1511–1527, 1989.
[3] B. Javidi and J. L. Horner, Eds., *Real-Time Optical Information Processing*, Academic Press, New York, NY, USA, 1994.
[4] A. V. Pavlov, "Mathematical models of optical methods in data processing," *Journal of Computer and Systems Sciences International*, vol. 39, no. 3, pp. 441–447, 2000.
[5] A. Van der Lugt, Ed., *Optical Signal Processing*, Wiley Series in Pure and Applied Optics, John Wiley, New York, NY, USA, 1992.
[6] G. O. Reynolds, J. B. DeVelis, B. G. Parrent, and B. J. Thompson, *Physical Optical Notebook: Tutorials in Fourier Optics*, SPIE Optical Engineering Press, New York, NY, USA, 1989.
[7] Y. Li, H. H. Szu, Y. Sheng, and H. J. Caulfield, "Wavelet processing and optics," *Proc. IEEE*, vol. 84, no. 5, pp. 720–732, 1996.
[8] A. Walther, "Radiometry and coherence," *Journal of the Optical Society of America*, vol. 58, no. 9, pp. 1256–1259, 1968.
[9] A. Walther, "Propagation of the generalized radiance through lenses," *Journal of the Optical Society of America*, vol. 68, no. 11, pp. 1606–1610, 1978.
[10] A. Papoulis, "Ambiguity function in Fourier optics," *Journal of the Optical Society of America*, vol. 64, no. 6, pp. 779–788, 1974.
[11] M. J. Bastiaans, "The Wigner distribution function applied to optical signals and systems," *Optics Communication*, vol. 25, no. 1, pp. 26–30, 1978.

[12] M. J. Bastiaans, "Wigner distribution function and its application to first-order optics," *Journal of the Optical Society of America*, vol. 69, no. 12, pp. 1710–1716, 1979.
[13] M. J. Bastiaans, "Application of the Wigner distribution function to partially coherent light," *Journal of the Optical Society of America A*, vol. 3, no. 8, pp. 1227–1238, 1986.
[14] W. Mecklenbräuker and F. Hlawatsch, Eds., *The Wigner Distribution—Theory and Applications in Signal Processing*, Elsevier, Amsterdam, The Netherlands, 1997.
[15] D. Mendlovic and H. M. Ozaktas, "Fractional Fourier transforms and their optical implementation: I," *Journal of the Optical Society of America A*, vol. 10, no. 9, pp. 1875–1881, 1993.
[16] H. M. Ozaktas and D. Mendlovic, "Fractional Fourier transforms and their optical implementation. II," *Journal of the Optical Society of America A*, vol. 10, no. 12, pp. 2522–2531, 1993.
[17] H. M. Ozaktas and D. Mendlovic, "Fourier transforms of fractional order and their optical interpretation," *Optics Communications*, vol. 101, no. 3–4, pp. 163–169, 1993.
[18] T. Alieva, V. Lopez, F. Agullo-Lopez, and L. B. Almeida , "The fractional Fourier transform in optical propagation problems," *Journal of Modern Optics*, vol. 41, no. 5, pp. 1037–1044, 1994.
[19] H. Kober, "Wurzeln aus der Hankel-, Fourier- und aus anderen stetigen Transformationen," *Quart. J. Math. Oxford. Ser.*, vol. 10, pp. 45–59, 1939.
[20] V. Namias, "The fractional order Fourier transform and its applications to quantum mechanics," *Journal of the Institute of Mathematics and Its Applications*, vol. 25, pp. 241–265, 1980.
[21] A. C. McBride and F. H. Kerr, "On Namias' fractional Fourier transforms," *IMA Journal of Applied Mathematics*, vol. 39, no. 2, pp. 159–175, 1987.
[22] A. W. Lohmann, "Image rotation, Wigner rotation, and the fractional Fourier transform," *Journal of the Optical Society of America A*, vol. 10, no. 10, pp. 2181–2186, 1993.
[23] L. B. Almeida, "The fractional Fourier transform and time-frequency representations," *IEEE Trans. Signal Processing*, vol. 42, no. 11, pp. 3084–3091, 1994.
[24] H. M. Ozaktas, M. A. Kutay, and D. Mendlovic, "Introduction to the fractional Fourier transform and its applications," in *Advances in Imaging and Electron Physics*, P. W. Hawkes, Ed., vol. 106, pp. 239–291, Academic Press, London, UK, 1999.
[25] A. W. Lohmann, D. Mendlovic, and Z. Zalevsky, "Fractional transformation in optics," in *Progress in Optics*, E. Wolf, Ed., vol. 38, pp. 263–342, Elsevier, Amsterdam, The Netherlands, 1998.
[26] A. W. Lohmann, D. Mendlovic, Z. Zalevsky, and R. G. Dorsch, "Some important fractional transformations for signal processing," *Optics Communications*, vol. 125, no. 1–3, pp. 18–20, 1996.
[27] T. Alieva and M. L. Calvo, "Fractionalization of the linear cyclic transforms ," *Journal of the Optical Society of America A*, vol. 17, no. 12, pp. 2330–2338, 2000.
[28] T. Alieva, M. J. Bastiaans, and M. L. Calvo, "Fractional cyclic transforms in optics: theory and applications," in *Recent Research and Developments in Optics*, S. G. Pandalai, Ed., vol. 1, pp. 105–122, Research Signpost, Trivandrum, India, 2001.
[29] H. M. Ozaktas, Z. Zalevsky, and M. A. Kutay, *The Fractional Fourier Transform with Applications in Optics and Signal Processing*, Wiley, New York, NY, USA, 2001.
[30] T. Alieva and M. J. Bastiaans, "Wigner distribution and fractional Fourier transform," in *Time-Frequency Signal Analysis and Processing: A Comprehensive Reference*, B. Boashash, Ed., pp. 145–152, Elsevier, Oxford, UK, 2003.
[31] A. Torre, "The fractional Fourier transform and some of its applications to optics," in *Progress in Optics*, E. Wolf, Ed., vol. 43, pp. 531–596, North-Holland, Amsterdam, The Netherlands, 2002.
[32] M. G. Raymer, M. Beck, and D. F. McAlister, "Complex wave-field reconstruction using phase-space tomography," *Physical Review Letters*, vol. 72, no. 8, pp. 1137–1140, 1994.
[33] T. Alieva and M. J. Bastiaans, "On fractional Fourier transform moments," *IEEE Signal Processing Lett.*, vol. 7, no. 11, pp. 320–323, 2000.
[34] T. Alieva and M. J. Bastiaans, "Phase-space distributions in quasi-polar coordinates and the fractional Fourier transform," *Journal of the Optical Society of America A*, vol. 17, no. 12, pp. 2324–2329, 2000.
[35] T. Alieva, M. J. Bastiaans, and L. Stanković, "Signal reconstruction from two close fractional Fourier power spectra," *IEEE Trans. Signal Processing*, vol. 51, no. 1, pp. 112–123, 2003.
[36] M. J. Bastiaans and K. B. Wolf, "Phase reconstruction from intensity measurements in linear systems," *Journal of the Optical Society of America A*, vol. 20, no. 6, pp. 1046–1049, 2003.
[37] R. M. Teague, "Deterministic phase retrieval: a Green's function solution," *Journal of the Optical Society of America*, vol. 73, no. 11, pp. 1434–1441, 1983.
[38] Z. Zalevsky, D. Mendlovic, and R. G. Dorsch, "Gerchberg-Saxton algorithm applied in the fractional Fourier or the Fresnel domain," *Optics Letters*, vol. 21, no. 12, pp. 842–844, 1996.
[39] B.-Z. Dong, Y. Zhang, B.-Y. Gu, and G.-Z. Yang, "Numerical investigation of phase retrieval in a fractional Fourier transform," *Journal of the Optical Society of America A*, vol. 14, no. 10, pp. 2709–2714, 1997.
[40] W.-X. Cong, N.-X. Chen, and B.-Y. Gu, "Recursive algorithm for phase retrieval in the fractional Fourier transform domain," *Applied Optics*, vol. 37, no. 29, pp. 6906–6910, 1998.
[41] T. Alieva and M. J. Bastiaans, "Finite-mode analysis by means of intensity information in fractional optical systems," *Journal of the Optical Society of America A*, vol. 19, no. 3, pp. 481–484, 2002.
[42] T. Alieva, M. L. Calvo, and M. J. Bastiaans, "Power filtering of *n*th order in the fractional Fourier domain," *Journal of Physics A: Mathematical and General*, vol. 35, no. 36, pp. 7779–7785, 2002.
[43] M. J. Bastiaans and T. Alieva, "Wigner distribution moments in fractional Fourier transform systems," *Journal of the Optical Society of America A*, vol. 19, no. 9, pp. 1763–1773, 2002.
[44] M. J. Bastiaans and T. Alieva, "Wigner distribution moments measured as intensity moments in separable first-order optical systems," *EURASIP Journal on Applied Signal Processing*, vol. 2005, no. 10, pp. 1535–1540, 2005.
[45] L. Stanković, T. Alieva, and M. J. Bastiaans, "Time-frequency signal analysis based on the windowed fractional Fourier transform," *Signal Processing*, vol. 83, no. 11, pp. 2459–2468, 2003.
[46] M. J. Bastiaans, T. Alieva, and L. Stanković, "On rotated time-frequency kernels," *IEEE Signal Processing Lett.*, vol. 9, no. 11, pp. 378–381, 2002.
[47] T. Alieva and M. J. Bastiaans, "Evolution of the vortex and the asymmetrical parts of orbital angular momentum in separable first-order optical systems," *Optics Letters*, vol. 29, no. 14, pp. 1587–1589, 2004.

[48] J. C. Wood and D. T. Barry, "Tomographic time-frequency analysis and its application toward time-varying filtering and adaptive kernel design for multicomponent linear-FM signals," *IEEE Trans. Signal Processing*, vol. 42, no. 8, pp. 2094–2104, 1994.

[49] J. C. Wood and D. T. Barry, "Linear signal synthesis using the Radon-Wigner transform," *IEEE Trans. Signal Processing*, vol. 42, no. 8, pp. 2105–2111, 1994.

[50] T. Alieva, "Fractional Fourier transform as a tool for investigation of fractal objects," *Journal of the Optical Society of America A*, vol. 13, no. 6, pp. 1189–1192, 1996.

[51] T. Alieva and F. Agullo-Lopez, "Optical wave propagation of fractal fields," *Optics Communications*, vol. 125, no. 4–6, pp. 267–274, 1996.

[52] T. Alieva and F. Agullo-Lopez, "Diffraction analysis of random fractal fields," *Journal of the Optical Society of America A*, vol. 15, no. 3, pp. 669–674, 1998.

[53] T. Alieva and A. Barbé, "Fractional Fourier analysis of objects with scaling symmetry," in *Fractals in Engineering*, J. L. Vehel, E. Lutton, and C. Tricot, Eds., pp. 252–265, Springer-Verlag, London, UK, 1997.

[54] D. Rodriguez Merlo, J. A. Rodrigo Martín-Romo, T. Alieva, and M. L. Calvo, "Fresnel diffraction by deterministic fractal gratings: an experimental study," *Optics and Spectroscopy*, vol. 95, no. 1, pp. 131–133, 2003.

[55] J. Uozumi and T. Asakura, "Fractal optics," in *Current Trends in Optics*, J. C. Dainty, Ed., pp. 83–94, Academic Press, London, UK, 1994.

[56] T. Alieva and M. L. Calvo, "Paraxial diffraction on structures generated by multiplicative iterative procedures," *Journal of Optics A: Pure and Applied Optics*, vol. 5, no. 5, pp. S324–S328, 2003.

[57] H. M. Ozaktas, B. Barshan, D. Mendlovic, and L. Onural, "Convolution, filtering, and multiplexing in fractional Fourier domains and their relation to chirp and wavelet transforms," *Journal of the Optical Society of America A*, vol. 11, no. 2, pp. 547–559, 1994.

[58] L. B. Almeida, "Product and convolution theorems for the fractional Fourier transform," *IEEE Signal Processing Lett.*, vol. 4, no. 1, pp. 15–17, 1997.

[59] D. Mendlovic, H. M. Ozaktas, and A. W. Lohmann, "Fractional correlation," *Applied Optics*, vol. 34, no. 2, pp. 303–309, 1995.

[60] D. Mendlovic, Y. Bitran, R. G. Dorsch, and A. W. Lohmann, "Optical fractional correlation: experimental results," *Journal of the Optical Society of America A*, vol. 12, no. 8, pp. 1665–1670, 1995.

[61] Z. Zalevsky and D. Mendlovic, "Fractional Wiener filter," *Applied Optics*, vol. 35, no. 20, pp. 3930–3936, 1996.

[62] D. J. Mustard, "Fractional convolution," *Journal of the Australian Mathematical Society. Series B: Applied Mathematics*, vol. 40, pp. 257–265, 1998, Part 2.

[63] B. Ruiz and H. Rabal, "Fractional Fourier transform description with use of differential operators," *Journal of the Optical Society of America A*, vol. 14, no. 11, pp. 2905–2913, 1997.

[64] A. I. Zayed, "A convolution and product theorem for the fractional Fourier transform," *IEEE Signal Processing Lett.*, vol. 5, no. 4, pp. 101–103, 1998.

[65] I. Raveh and D. Mendlovic, "New properties of the Radon transform of the cross Wigner/ambiguity distribution function," *IEEE Trans. Signal Processing*, vol. 47, no. 7, pp. 2077–2080, 1999.

[66] A. W. Lohmann, Z. Zalevsky, and D. Mendlovic, "Synthesis of pattern recognition filters for fractional Fourier processing," *Optics Communications*, vol. 128, no. 4–6, pp. 199–204, 1996.

[67] S. Granieri, R. Arizaga, and E. E. Sicre, "Optical correlation based on the fractional Fourier transform," *Applied Optics*, vol. 36, no. 26, pp. 6636–6645, 1997.

[68] A. M. Almanasreh and M. A. G. Abushagur, "Fractional correlations based on the modified fractional order Fourier transform," *Optical Engineering*, vol. 37, no. 1, pp. 175–184, 1998.

[69] T. Alieva and M. L. Calvo, "Generalized fractional convolution," in *Perspectives in Modern Optics and Optical Instrumentation*, J. Joseph, A. Sharma, and V. K. Rastogi, Eds., pp. 163–173, Anita Publications, New Delhi, India, 2002.

[70] O. Akay and G. F. Boudreaux-Bartels, "Fractional convolution and correlation via operator methods and an application to detection of linear FM signals," *IEEE Trans. Signal Processing*, vol. 49, no. 5, pp. 979–993, 2001.

[71] J. A. Davis, D. M. Cottrell, N. Nestorovic, and S. M. Highnote, "Space-variant Fresnel transform optical correlator," *Applied Optics*, vol. 31, no. 32, pp. 6889–6893, 1992.

[72] D. Mendlovic, Y. Bitran, R. G. Dorsch, C. Ferreira, J. García, and H. M. Ozaktas, "Anamorphic fractional Fourier transform: optical implementation and applications," *Applied Optics*, vol. 34, no. 32, pp. 7451–7456, 1995.

[73] J. García, D. Mendlovic, Z. Zalevsky, and A. W. Lohmann, "Space-variant simultaneous detection of several objects by the use of multiple anamorphic fractional-Fourier-transform filters," *Applied Optics*, vol. 35, no. 20, pp. 3945–3952, 1996.

[74] J. García, R. G. Dorsch, A. W. Lohmann, C. Ferreira, and Z. Zalevsky, "Flexible optical implementation of fractional Fourier transform processors. Applications to correlation and filtering," *Optics Communications*, vol. 133, no. 1–6, pp. 393–400, 1997.

[75] Y. Bitran, Z. Zalevsky, D. Mendlovic, and R. G. Dorsch, "Fractional correlation operation: performance analysis," *Applied Optics*, vol. 35, no. 2, pp. 297–303, 1996.

[76] T. Alieva and M. L. Calvo, "Importance of the phase and amplitude in the fractional Fourier domain," *Journal of the Optical Society of America A*, vol. 20, no. 3, pp. 533–541, 2003.

[77] F. T. S. Yu, C. Zhang, Y. Jin, and D. A. Gregory, "Nonconventional joint-transform correlator," *Optics Letters*, vol. 14, no. 17, pp. 922–924, 1989.

[78] G. Unnikrishnan and K. Singh, "Optical encryption using quadratic phase systems," *Optics Communications*, vol. 193, no. 1–6, pp. 51–67, 2001.

[79] B. Zhu and S. Liu, "Optical image encryption based on the generalized fractional convolution operation," *Optics Communications*, vol. 195, no. 5–6, pp. 371–381, 2001.

[80] B. Zhu, S. Liu, and Q. Ran, "Optical image encryption based on multifractional Fourier transforms," *Optics Letters*, vol. 25, no. 16, pp. 1159–1161, 2000.

[81] N. K. Nishchal, G. Unnikrishnan, J. Joseph, and K. Singh, "Optical encryption using a localized fractional Fourier transform," *Optical Engineering*, vol. 42, no. 12, pp. 3566–3571, 2003.

[82] B. Hennelly and J. T. Sheridan, "Optical image encryption by random shifting in fractional Fourier domains," *Optics Letters*, vol. 28, no. 4, pp. 269–271, 2003.

[83] B. Hennelly and J. T. Sheridan, "Fractional Fourier transform-based image encryption: phase retrieval algorithm," *Optics Communications*, vol. 226, no. 1–6, pp. 61–80, 2003.

[84] B. M. Hennelly and J. T. Sheridan, "Image encryption and the fractional Fourier transform," *Optik*, vol. 114, no. 6, pp. 251–265, 2003.

[85] N. K. Nishchal, J. Joseph, and K. Singh, "Fully phase encryption using fractional Fourier transform," *Optical Engineering*, vol. 42, no. 6, pp. 1583–1588, 2003.
[86] S. Stankovic, I. Djurovic, and I. Pitas, "Watermarking in the space/spatial-frequency domain using two-dimensional Radon-Wigner distribution," *IEEE Trans. Image Processing*, vol. 10, no. 4, pp. 650–658, 2001.
[87] I. Djurovic, S. Stankovic, and I. Pitas, "Digital watermarking in the fractional Fourier transformation domain," *Journal of Network and Computer Applications*, vol. 24, no. 2, pp. 167–173, 2001.
[88] S.-Y. Lee and H. H. Szu, "Fractional Fourier transforms, wavelet transforms, and adaptive neural networks," *Optical Engineering*, vol. 33, no. 7, pp. 2326–2330, 1994.
[89] S.-G. Shin, S.-I. Jin, S.-Y. Shin, and S.-Y. Lee, "Optical neural network using fractional Fourier transform, log-likelihood, and parallelism," *Optics Communications*, vol. 153, no. 4–6, pp. 218–222, 1998.
[90] S.-I. Jin, Y.-S. Bae, and S.-Y. Lee, "Generalized Vander Lugt correlator as an optical pattern classifier and its optimal learning rate," *Optics Communications*, vol. 206, no. 1–3, pp. 19–25, 2002.
[91] B. Barshan and B. Ayrulu, "Fractional Fourier transform pre-processing for neural networks and its application to object recognition," *Neural Networks*, vol. 15, no. 1, pp. 131–140, 2002.
[92] B. Barshan and B. Ayrulu, "Comparative analysis of different approaches to target differentiation and localization with sonar," *Pattern Recognition*, vol. 36, no. 5, pp. 1213–1231, 2003.
[93] S. Q. Zhang and M. A. Karim, "Fractional correlation filter for fuzzy associative memories," *Optical Engineering*, vol. 41, no. 1, pp. 126–129, 2002.
[94] A. W. Lohmann, D. Mendlovic, and Z. Zalevsky, "Fractional Hilbert transform," *Optics Letters*, vol. 21, no. 4, pp. 281–283, 1996.
[95] A. W. Lohmann, E. Tepichin, and J. G. Ramirez, "Optical implementation of the fractional Hilbert transform for two-dimensional objects," *Applied Optics*, vol. 36, no. 26, pp. 6620–6626, 1997.
[96] J. A. Davis, D. E. McNamara, and D. M. Cottrell, "Analysis of the fractional Hilbert transform," *Applied Optics*, vol. 37, no. 29, pp. 6911–6913, 1998.
[97] K. B. Wolf, *Integral Transforms in Science and Engineering*, chapter 9, Plenum Press, New York, NY, USA, 1979.
[98] R. K. Luneburg, *Mathematical Theory of Optics*, University of California Press, Berkeley and Los Angeles, Calif, USA, 1966.
[99] D. F. V. James and G. S. Agarwal, "The generalized Fresnel transform and its applications to optics," *Optics Communications*, vol. 126, no. 4–6, pp. 207–212, 1996.
[100] S. Abe and J. T. Sheridan, "Almost-Fourier and almost-Fresnel transformations," *Optics Communications*, vol. 113, no. 4–6, pp. 385–388, 1995.
[101] A. D. Poularikas, Ed., *The Transforms and Applications Handbook*, CRC Press, Boca Raton, Fla, USA, 1996, chapter 8, S. R. Deans, "Radon and Abel transforms," and chapter 12, G. F. Boudreaux-Bartels, "Mixed time-frequency signal transformations".
[102] J. Shamir and N. Cohen, "Root and power transformations in optics," *Journal of the Optical Society of America A*, vol. 12, no. 11, pp. 2415–2423, 1995.
[103] C. J. R. Sheppard and K. G. Larkin, "Similarity theorems for fractional Fourier transforms and fractional Hankel transforms," *Optics Communications*, vol. 154, no. 4, pp. 173–178, 1998.
[104] S.-C. Pei and J.-J. Ding, "Eigenfunctions of linear canonical transform," *IEEE Trans. Signal Processing*, vol. 50, no. 1, pp. 11–26, 2002.
[105] D. Mendlovic, H. M. Ozaktas, and A. W. Lohmann, "Self Fourier functions and fractional Fourier transforms," *Optics Communications*, vol. 105, no. 1–2, pp. 36–38, 1994.
[106] T. Alieva, "On the self-fractional Fourier functions," *Journal of Physics A: Mathematical and General*, vol. 29, no. 15, pp. L377–L379, 1996.
[107] T. Alieva and A. M. Barbé, "Self-fractional Fourier images," *Journal of Modern Optics*, vol. 46, no. 1, pp. 83–99, 1999.
[108] T. Alieva and A. M. Barbé, "Self-fractional Fourier functions and selection of modes," *Journal of Physics A: Mathematical and General*, vol. 30, no. 8, pp. L211–L215, 1997.
[109] T. Alieva and M. J. Bastiaans, "Mode analysis in optics through fractional transforms," *Optics Letters*, vol. 24, no. 17, pp. 1206–1208, 1999.
[110] L. M. Bernardo and O. D. D. Soares, "Fractional Fourier transforms and optical systems," *Optics Communications*, vol. 110, no. 5–6, pp. 517–522, 1994.
[111] P. Pellat-Finet, "Fresnel diffraction and the fractional-order Fourier transform," *Optics Letters*, vol. 19, no. 18, pp. 1388–1390, 1994.
[112] P. Pellat-Finet and G. Bonnet, "Fractional order Fourier transform and Fourier optics," *Optics Communications*, vol. 111, no. 1–2, pp. 141–154, 1994.
[113] H. M. Ozaktas and D. Mendlovic, "Fractional Fourier transform as a tool for analyzing beam propagation and spherical mirror resonators," *Optics Letters*, vol. 19, no. 21, pp. 1678–1680, 1994.
[114] Y. B. Karasik, "Expression of the kernel of a fractional Fourier transform in elementary functions," *Optics Letters*, vol. 19, no. 11, pp. 769–770, 1994.
[115] A. Sahin, H. M. Ozaktas, and D. Mendlovic, "Optical implementation of the two-dimensional fractional Fourier transform with different orders in the two dimensions," *Optics Communications*, vol. 120, no. 3–4, pp. 134–138, 1995.
[116] A. W. Lohmann and B. H. Soffer, "Relationships between the Radon-Wigner and fractional Fourier transforms," *Journal of the Optical Society of America A*, vol. 11, no. 6, pp. 1798–1811, 1994.
[117] A. Ya. Bekshaev, M. S. Soskin, and M. V. Vasnetsov, "Optical vortex symmetry breakdown and decomposition of the orbital angular momentum of light beams," *Journal of the Optical Society of America A*, vol. 20, no. 8, pp. 1635–1643, 2003.
[118] S.-I. Jin and S.-Y. Lee, "Joint transform correlator with fractional Fourier transform," *Optics Communications*, vol. 207, no. 1–6, pp. 161–168, 2002.
[119] G. Gonon, O. Richoux, and C. Depollier, "Acoustic wave propagation in a 1-D lattice: analysis of nonlinear effects by the fractional Fourier transform method," *Signal Processing*, vol. 83, no. 11, pp. 2469–2480, 2003.
[120] J. Du, Z. Cui, Y. Zhang, C. Du, J. Yang, and Y. Guo, "Enhancement of photolithography resolution by fractional Fourier domain filtering," *Microelectronic Engineering*, vol. 67–68, pp. 31–38, 2003.
[121] B. Javidi, "Securing information with optical technologies," *Physics Today*, vol. 50, no. 3, pp. 27–32, 1997.
[122] A. I. Zayed, "A class of fractional integral transforms: A generalization of the fractional Fourier transform," *IEEE Trans. Signal Processing*, vol. 50, no. 3, pp. 619–627, 2002.
[123] D. S. Yeung, Q. Ran, E. C. C. Tsang, and K. L. Teo, "Complete way to fractionalize Fourier transform," *Optics Communications*, vol. 230, no. 1–3, pp. 55–57, 2004.

[124] T. Alieva and M. J. Bastiaans, "Powers of transfer matrices determined by means of eigenfunctions," *Journal of the Optical Society of America A*, vol. 16, no. 10, pp. 2413–2418, 1999.
[125] C.-C. Shih, "Fractionalization of Fourier transform," *Optics Communications*, vol. 118, no. 5–6, pp. 495–498, 1995.
[126] B. W. Dickinson and K. Steiglitz, "Eigenvectors and functions of the discrete Fourier transform," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 30, no. 1, pp. 25–31, 1982.
[127] D. A. Linden, "A discussion of sampling theorems," *Proc. IRE*, vol. 47, no. 7, pp. 1219–1226, 1959.
[128] L. Yu, Y. Lu, X. Zeng, et al., "Deriving the integral representation of a fractional Hankel transform froma fractional Fourier transform," *Optics Letters*, vol. 23, no. 15, pp. 1158–1160, 1998.
[129] L. Yu, W. Huang, M. Huang, Z. Zhu, X. Zeng, and W. Ji, "The Laguerre-Gaussian series representation of two-dimensional fractional Fourier transform," *Journal of Physics A: Mathematical and General*, vol. 31, no. 46, pp. 9353–9357, 1998.
[130] S. Liu, J. Jiang, Y. Zhang, and J. Zhang, "Generalized fractional Fourier transforms," *Journal of Physics A: Mathematical and General*, vol. 30, no. 3, pp. 973–981, 1997.
[131] T. Alieva and M. J. Bastiaans, "Fractional cosine and sine transforms," *Asian Journal of Physics*, vol. 12, no. 1, pp. 25–32, 2003.
[132] D. Zhao, H. Mao, H. Liu, X. Wei, F. Jing, and Q. Zhu, "Effect of spherically aberrated lens on the fractional Fourier transformation systems," *Optics Communications*, vol. 227, no. 4–6, pp. 213–220, 2003.
[133] I. Moreno, J. A. Davis, and K. Crabtree, "Fractional Fourier transform optical system with programmable diffractive lenses," *Applied Optics*, vol. 42, no. 32, pp. 6544–6548, 2003.
[134] D. Dragoman and M. Dragoman, "Reconfigurable electro-optical waveguide for optical processing," *Applied Optics*, vol. 42, no. 32, pp. 6439–6444, 2003.
[135] A. Klappenecker and M. Rötteler, "Engineering functional quantum algorithms," *Physical Review A*, vol. 67, no. 1, 2003, art. no. 010302.
[136] Y. Cai and Q. Lin, "Properties of a flattened Gaussian beam in the fractional Fourier transform plane," *Journal of Optics A: Pure and Applied Optics*, vol. 5, no. 3, pp. 272–275, 2003.
[137] D. Zhao, H. Mao, M. Shen, et al., "Propagation of flattened Gaussian beams in apertured fractional Fourier transforming systems," *Journal of Optics A: Pure and Applied Optics*, vol. 6, no. 1, pp. 148–154, 2004.
[138] Y. Cai and Q. Lin, "Fractional Fourier transform for elliptical Gaussian beams," *Optics Communications*, vol. 217, no. 1–6, pp. 7–13, 2003.
[139] Y. Cai, D. Ge, and Q. Lin, "Fractional Fourier transform for partially coherent and partially polarized Gaussian-Schell model beams," *Journal of Optics A: Pure and Applied Optics*, vol. 5, no. 5, pp. 453–459, 2003.
[140] Y. Cai and Q. Lin, "Transformation and spectrum properties of partially coherent beams in the fractional Fourier transform plane," *Journal of the Optical Society of America A*, vol. 20, no. 8, pp. 1528–1536, 2003.
[141] Q. Lin and Y. Cai, "Fractional Fourier transform for partially coherent Gaussian Schell model beams," *Optics Letters*, vol. 27, no. 19, pp. 1672–1674, 2002.
[142] A. A. Malyutin, "Simple scheme for the astigmatic transformation of laser modes," *Quantum Electronics*, vol. 33, no. 11, pp. 1015–1018, 2003.
[143] Q. Tan, Y. Yan, G. Jin, and D. Xu, "True beam smoothing in the fractional Fourier transform domain," *Journal of Modern Optics*, vol. 50, no. 14, pp. 2147–2153, 2003.
[144] A. E. Cetin, H. Ozaktas, and H. M. Ozaktas, "Resolution enhancement of low resolution wavefields with POCS algorithm," *Electronics Letters*, vol. 39, no. 25, pp. 1808–1810, 2003.
[145] J. T. Sheridan, B. Hennelly, and D. Kelly, "Motion detection, the Wigner distribution function, and the optical fractional Fourier transform," *Optics Letters*, vol. 28, no. 11, pp. 884–886, 2003.
[146] H.-B. Sun, G.-S. Liu, H. Gu, and W.-M. Su, "Application of the fractional Fourier transform to moving target detection in airborne SAR," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 38, no. 4, pp. 1416–1424, 2002.
[147] D. Sazbon, Z. Zalevsky, E. Rivlin, and D. Mendlovic, "Using Fourier/Mellin-based correlators and their fractional versions in navigational tasks," *Pattern Recognition*, vol. 35, no. 12, pp. 2993–2999, 2002.
[148] I. S. Yetik and A. Nehorai, "Beamforming using the fractional Fourier transform," *IEEE Trans. Signal Processing*, vol. 51, no. 6, pp. 1663–1668, 2003.
[149] J. A. Armstrong, "Fresnel holograms: their imaging properties and aberrations," *IBM Journal of Research and Development*, vol. 9, no. 3, pp. 171–178, 1965.
[150] M. Liebling, T. Blu, and M. Unser, "Fresnelets: new multiresolution wavelet bases for digital holography," *IEEE Trans. Image Processing*, vol. 12, no. 1, pp. 29–43, 2003.
[151] R. Tudela, E. Martin-Badosa, I. Labastida, S. Vallmitjana, I. Juvells, and A. Carnicer, "Full complex Fresnel holograms displayed on liquid crystal devices," *Journal of Optics A: Pure and Applied Optics*, vol. 5, no. 5, pp. S189–S194, 2003.
[152] Y. Zeng, Y. Guo, F. Gao, and J. Zhu, "Principle and application of multiple fractional Fourier transform holography," *Optics Communications*, vol. 215, no. 1–3, pp. 53–59, 2003.
[153] S.-C. Pei and J.-J. Ding, "Fractional cosine, sine, and Hartley transforms," *IEEE Trans. Signal Processing*, vol. 50, no. 7, pp. 1661–1680, 2002.
[154] C.-C. Tseng, "Eigenvalues and eigenvectors of generalized DFT, generalized DHT, DCT-IV and DST-IV matrices," *IEEE Trans. Signal Processing*, vol. 50, no. 4, pp. 866–877, 2002.
[155] H.-Y. Fan and H.-L. Lu, "Eigenmodes of fractional Hankel transform derived by the entangled-state method," *Optics Letters*, vol. 28, no. 9, pp. 680–682, 2003.
[156] A. Torre, "Linear and radial canonical transforms of fractional order," *Journal of Computational and Applied Mathematics*, vol. 153, no. 1–2, pp. 477–486, 2003.

Tatiana Alieva was born in Moscow, Russia. In 1983 and 1996, she received the M.S. and Ph.D. degrees (with honours) in physics from M. V. Lomonosov Moscow State University and University Autonoma of Madrid, respectively. From 1983 till 1999, she worked for A. L. Mintz Radiotechnical Institute, Academy of Science, Moscow. She spent several years as a Postdoctoral Fellow at the Departments of Electrical Engineering of Catholic University of Leuven, Belgium, and Eindhoven University of Technology, The Netherlands. She is presently a researcher at the Faculty of Physics, Complutense University of Madrid. She has worked in diverse fields of physics, such as the theory of high energetic beams channelling in crystals, and radio waves scattering on the anisotropic formations in the ionosphere. Her current research interests are in optical information processing, fractal signal analysis, theory of fractional integral transforms, phase retrieval, and characterization of optical fields by means of phase-space distributions. She is the author of more than eighty scientific publications.

Martin J. Bastiaans was born in Helmond, The Netherlands, in 1947. He received the M.S. degree in electrical engineering (with honours) and the Ph.D. degree in technical sciences from the Technische Universiteit Eindhoven (Eindhoven University of Technology), Eindhoven, The Netherlands, in 1969 and 1983, respectively. In 1969 he became an Assistant Professor and since 1985, he has been an Associate Professor in the Department of Electrical Engineering, Technische Universiteit Eindhoven, in the Signal Processing Systems Group, where he teaches electrical circuit analysis, signal theory, digital signal processing, and Fourier optics and holography. His main current research interest is in describing signals by means of a local frequency spectrum (for instance, the Wigner distribution function, the windowed Fourier transform, Gabor's signal expansion, etc.). He is the author and coauthor of more than 140 papers in international scientific journals and proceedings of scientific conferences. Dr. Bastiaans is a Fellow of the Optical Society of America and a Senior Member of the Institute of Electrical and Electronics Engineers.

Maria Luisa Calvo graduated in physics in 1969 from the Complutense University of Madrid (UCM), Spain, and received the Ph.D. degree (with honours) in 1977. In 1972 she joined the Optics Department of UCM as an Assistant Professor, became an Associate Professor in 1981 and Chair of Optics in 1999. She has visited as a researcher the following institutions: Philips, Eindhoven, The Netherlands, in 1976; Institute of Optics, Paris, France, in 1968–1972; Bremen University, Germany, in 1983; School of Optometry, University of Berkeley, USA, in 1985–1986; School of Optometry, University of Missouri, St. Louis, USA, in 1994 and 2000. Professor Calvo's main areas of interest are electromagnetic theory of optical waveguides, holography, optical information processing, and bio-optics. She is the author of more than one hundred scientific publications on the mentioned subjects. In 1993, she established the Interdisciplinary Group for Bio-Optics Research (GIBO-UCM), now Group for Optical Computing (GICO-UCM). She has been the President of the Image Technical Committee of the Spanish Optical Society (SEDO) (1992–1997). She was a Travelling Lecturer of the International Commission for Optics (ICO) in 1998, elected Vice-president at the ICO Bureau (1999–2002), and Secretary-General of ICO for the term in 2002–2005. She collaborates with the General Directorate for Research and Technology of the European Union in Brussels. She is a Fellow of OSA and a Member of SPIE and EOS.

EURASIP Journal on Applied Signal Processing 2005:10, 1520–1534

Applications of the Wigner Distribution Function in Signal Processing

Daniela Dragoman
Faculty of Physics, University of Bucharest, P.O. Box MG-11, 077125 Bucharest-Magurele, Romania
Email: daniela@solid.fizica.unibuc.ro

Received 1 April 2004; Revised 23 August 2004

We present a review of the applications of the Wigner distribution function in various areas of signal processing: amplitude and phase retrieval, signal recognition, characterization of arbitrary signals, optical systems and devices, and coupling coefficient estimation in phase space. Although reference is made to specific signals and systems, the mathematical formulation is general and can be applied to either spatial, temporal, or spatio-temporal phase spaces, to coherent, partially coherent, or discrete signals. The universal and intuitive character of the Wigner distribution approach to signal characterization and processing and its simplicity in solving many issues are evidenced throughout the paper.

Keywords and phrases: signal processing, phase space, Wigner distribution function.

1. INTRODUCTION

Phase space methods become increasingly exploited in signal processing applications due to their intuitive character, universal validity, and last but not least simplicity in an increasing number of practical situations. These methods refer to any description of a signal or an optical system through a function that depends jointly on the canonical conjugate phase space variables, which are the transverse position vector $\mathbf{r}$ and the angular (spatial frequency) vector $\mathbf{k}$ for light beams, the time variable t, and frequency ω for optical pulses or, more generally, all four coordinates $\mathbf{r}, \mathbf{k}, t$, and ω. There are many phase space distribution functions that can be employed in signal processing applications; the spectrogram, the ambiguity function, or the Wigner distribution function (WDF) are just a few examples. These phase space distribution functions are actually closely related to one another, the use of one or another depend to a great extent on the specific application and/or on the type of experiments that are carried on.

This paper focuses on the applications of the WDF in signal processing; this phase space distribution function is often referred to as Wigner-Ville distribution, but the WDF term is used in a larger range of applications and therefore the term WDF will be used throughout this paper. The intention is not to present an exhaustive list of WDF properties (including its relation to other phase space distributions) or problems that can be tackled with this distribution function; there are already a large number of review papers and even books dedicated to these subjects [1, 2, 3, 4, 5, 6, 7, 8, 9]. Rather, it will be shown how to make use of the specific properties of the WDF and why it is desirable to use this phase space distribution in addressing some relevant issues in signal processing. Based on the available mathematical and experimental results, the paper offers a glimpse on the beauty, simplicity, and potentialities of the WDF approach to this research field; therefore, many details will be left aside and only the most significant results will be discussed in greater detail. The responsibility of doing justice to the so many briefly mentioned issues is entrusted to the references.

2. THE WIGNER DISTRIBUTION FUNCTION AND SOME OF ITS PROPERTIES

The WDF has been originally introduced in quantum mechanics [10], but found itself rapidly in hosts of applications in the analysis and processing of both light beams and pulses [1, 2, 3, 11, 12]. The WDF of coherent signals that propagate along the z direction and that are characterized by a scalar field distribution $\varphi(\mathbf{u}, z)$ is defined at a $z = \text{const.}$ plane as

$$\begin{aligned} W_{\varphi}(\mathbf{u}, \mathbf{v}; z) &= (2\pi)^{-n/2} \int \varphi\left(\mathbf{u} + \frac{\mathbf{u}'}{2}, z\right) \varphi^{*}\left(\mathbf{u} - \frac{\mathbf{u}'}{2}, z\right) \exp\left(i\mathbf{u}'\mathbf{v}^{T}\right) d\mathbf{u}' \\ &= (2\pi)^{-n/2} \int \bar{\varphi}\left(\mathbf{v} + \frac{\mathbf{v}'}{2}, z\right) \bar{\varphi}^{*}\left(\mathbf{v} - \frac{\mathbf{v}'}{2}, z\right) \exp\left(-i\mathbf{u}\mathbf{v}'^{T}\right) d\mathbf{v}', \end{aligned} \tag{1}$$

where $*$ and T denote the operations of complex conjugation and transposition, respectively, and

$$\tilde{\varphi}(\mathbf{v}, z) = (2\pi)^{-n/2} \int \varphi(\mathbf{u}, z) \exp\left(i\mathbf{u}\mathbf{v}^T\right) d\mathbf{u} \tag{2}$$

is the Fourier transform of $\varphi(\mathbf{u}, z)$. (For vectorial field distributions, a WDF matrix can be introduced, as shown in [13, 14].) In the expressions above, $\mathbf{u}$ represents the transverse coordinate vector $\mathbf{r} = (x, y)$ in the spatial phase space appropriate for describing stationary and monochromatic light beams, symbolizes in the temporal phase space the time coordinate t in a reference frame that travels with the pulse, or designates the ensemble $(\mathbf{r}, t)$ for a general spatio-temporal phase space. The coordinate $\mathbf{v}$ is its canonically conjugate variable equal, respectively, to the projection $\mathbf{k} = (k_x, k_y)$ of the wavevector on the transverse coordinates (known also as spatial frequency), to the frequency ω measured from the central frequency ω_0 of a light pulse, or to $(\mathbf{k}, \omega)$ in general; n denotes the dimensionality of $\mathbf{u}$ and $\mathbf{v}$ (the phase space is $2n$-dimensional). The parameter z in the WDF definition can be removed for simplicity if no ambiguity arises. It should be mentioned that in literature slightly different forms of the WDF are sometimes encountered, defined either with an opposite sign in the Fourier integral or even with scaled spatial frequency or frequency coordinates; due account of these differences must be taken when comparing the formulas in this paper with results derived with a different WDF definition.

In the paraxial approximation $\mathbf{k} = (k\phi_x, k\phi_y)$, where $k = 2\pi/\lambda$ is the wavenumber of the monochromatic field with wavelength λ, and ϕ_x, ϕ_y are the angles between the wavevector direction and the z-axis; thus, for monochromatic fields, $\mathbf{k}$ has the significance of a scaled angular coordinate. The WDF characterization of fields in the temporal phase space is sometimes referred to as the chronocyclic representation of light pulses [15]; it can adequately describe the propagation of signals when their spatial extent is of no concern. The use of spatio-temporal phase spaces is justified when optical systems, such as prisms or diffraction gratings couple the space and time coordinates through a frequency-dependent refraction phenomenon, due to which different frequency components follow different paths; this situation is common for ultrashort light pulses characterized by a broad spectrum.

Extremely important for a large number of applications is that the WDF can be defined also for partially coherent light distributions such as [16]

$$W_\varphi(\mathbf{u}, \mathbf{v}) = (2\pi)^{-n/2} \int \Gamma\left(\mathbf{u} + \frac{\mathbf{u}'}{2}, \mathbf{u} - \frac{\mathbf{u}'}{2}\right) \exp\left(i\mathbf{u}'\mathbf{v}^T\right) d\mathbf{u}', \tag{3}$$

where the correlation function $\Gamma(\mathbf{u}_1, \mathbf{u}_2) = \langle \varphi(\mathbf{u}_1)\varphi^*(\mathbf{u}_2)\rangle$ is the ensemble average of $\varphi(\mathbf{u}_1)\varphi^*(\mathbf{u}_2)$; similarly to (1), the Fourier transform of the correlation function can also be used to define the WDF for partially coherent fields.

Digital signal processing applications can benefit from a discrete form of the WDF, which is defined in the temporal phase space, for example, as [17]

$$W_\varphi(n\tau, \omega) = 2\tau \sum_{l=-\infty}^{\infty} \varphi(n\tau + l\tau)\varphi^*(n\tau - l\tau) \exp(i2l\tau\omega), \tag{4}$$

where $\varphi(t)$ is a complex-valued function that is sampled with a sampling interval τ, n, and l are discrete and ω is continuous. Note that the discrete-time WDF (4) is periodic in the frequency coordinate, with a period equal to one-half the sampling frequency $\Delta\omega = 2\pi/\tau$, that is, $W_\varphi(n\tau, \omega) = W_\varphi(n\tau, \omega + \Delta\omega/2)$. The practical considerations regarding the sampling rate that is needed to avoid aliasing are detailed in [17], which also investigates an efficient algorithm to implement in real time the temporal WDF. In an analogous manner a discrete-frequency WDF can be defined [18] through a formula similar to (4) but referring to the Fourier transform of the signal; a change in sign in the exponential term is also required. A WDF discrete in both n and θ variables, with properties similar to that of the continuous WDF in (1), has been introduced by Brenner [19]. The discrete-time/frequency WDF is defined as

$$W_\varphi(n, m) = \left(\frac{\tau}{2N}\right) \sum_{l=-\infty}^{\infty} \varphi(l\tau)\varphi^*(n\tau - l\tau) \exp\left[\frac{im(2l - n)}{2N}\right], \tag{5}$$

where it is assumed that the signal can be approximated by a limited number N of samples, the values of the WDF being evaluated for $t = n\tau/2$ and $\omega = m/N\tau$ (the product between the sample rates in time and frequency is equal to N^{-1}). The discrete-time/frequency WDF is periodic in both variables with a period $2N$, its properties and its relation to the continuous WDF being detailed in [18].

Irrespective of its definition, the WDF satisfies some properties that are tremendously important in practical applications. For the case of coherent light distributions these properties are as follows.

(i) The WDF is real, but can take negative values. Only Gaussian fields have WDFs that are everywhere positive. For a forward propagating field the negative WDF values were associated with small regions of backward flux, the local flux of the monochromatic field being defined as $J(\mathbf{r}) = k^{-1} \int W(\mathbf{r}, \mathbf{k})\mathbf{k}\, d\mathbf{k}$ [20]. According to another interpretation [21], the negative values of the WDF originate from the phase space interference between neighboring minimum-value phase space areas that correspond to the Gaussian beams into which an arbitrary field distribution can be decomposed.

(ii) The WDF is limited to the $\mathbf{u}$ and $\mathbf{v}$ intervals that limit the field distribution and its Fourier transform, respectively.

(iii) A shift of the field distribution in $\mathbf{u}$ or a shift of its Fourier transform in $\mathbf{v}$ leads to the same shift for the WDF.

(iv) The partial integral of the WDF over $\mathbf{v}$ is proportional to the field intensity:

$$\int W_\varphi(\mathbf{u},\mathbf{v})d\mathbf{v} = (2\pi)^{n/2}\,|\,\varphi(\mathbf{u})\,|^2. \tag{6}$$

(v) The WDF integral over $\mathbf{u}$ is proportional to the intensity of the Fourier transform of the field distribution (the spectrum of the field):

$$\int W_\varphi(\mathbf{u},\mathbf{v})d\mathbf{u} = (2\pi)^{n/2}\,|\,\bar{\varphi}(\mathbf{v})\,|^2. \tag{7}$$

(vi) The field distribution and its Fourier transforms can be reconstructed from the WDF up to constant phase factors, according to

$$\varphi(\mathbf{u})\varphi^*(\mathbf{0}) = (2\pi)^{-n/2}\int W_\varphi\left(\frac{\mathbf{u}}{2},\mathbf{v}\right)\exp\left(-i\mathbf{u}\mathbf{v}^T\right)d\mathbf{v}, \tag{8}$$

$$\bar{\varphi}(\mathbf{v})\bar{\varphi}^*(\mathbf{0}) = (2\pi)^{-n/2}\int W_\varphi\left(\mathbf{u},\frac{\mathbf{v}}{\mathbf{2}}\right)\exp\left(\mathbf{i}\mathbf{u}\mathbf{v}^{\mathbf{T}}\right)\mathbf{d}\mathbf{u} \tag{9}$$

with $\mathbf{0}$ the null vector, since the modulus of $\varphi^*(\mathbf{0})$ and $\bar{\varphi}^*(\mathbf{0})$ can be found from (8) and (9), respectively, for $\mathbf{u} = 0$ and $\mathbf{v} = 0$.

(vii) The WDF satisfies the so-called overlap or Moyal relation

$$\begin{aligned}&\int W_{\varphi_1}(\mathbf{u},\mathbf{v})W_{\varphi_2}(\mathbf{u},\mathbf{v})d\mathbf{u}\,d\mathbf{v}\\ &\quad= \left|\int \varphi_1(\mathbf{u})\varphi_2^*(\mathbf{u})d\mathbf{u}\right|^2 = \left|\int \bar{\varphi}_1(\mathbf{v})\bar{\varphi}_2^*(\mathbf{v})d\mathbf{v}\right|^2.\end{aligned} \tag{10}$$

For partially coherent sources with correlation functions $\Gamma_1(\mathbf{u},\mathbf{u}')$, $\Gamma_2(\mathbf{u},\mathbf{u}')$, the Moyal relation becomes

$$\int W_{\varphi_1}(\mathbf{u},\mathbf{v})W_{\varphi_2}(\mathbf{u},\mathbf{v})d\mathbf{u}\,d\mathbf{v} = \int \Gamma_1(\mathbf{u},\mathbf{u}')\Gamma_2(\mathbf{u}',\mathbf{u})d\mathbf{u}\,d\mathbf{u}', \tag{11}$$

which reduces to (10) when $\Gamma_i(\mathbf{u},\mathbf{u}') = \varphi_i(\mathbf{u})\varphi_i^*(\mathbf{u}')$, $i = 1, 2$.

(viii) The WDF is bilinear in the field distribution, which means that the WDF of a superposition of signals $\varphi(\mathbf{u}) = \sum_m a_m\varphi_m(\mathbf{u})$ contains not only the WDF of the individual signals (the auto-terms), $W_{\varphi_m}(\mathbf{u},\mathbf{v})$ but also complex cross-terms $W_{\varphi_m\varphi_n}(\mathbf{u},\mathbf{v})$ that describe phase space interference effects:

$$W_\varphi(\mathbf{u},\mathbf{v}) = \sum_m |a_m|^2 W_{\varphi_m}(\mathbf{u},\mathbf{v}) + \sum_{m,\,n\neq m} a_m a_n^* W_{\varphi_m\varphi_n}(\mathbf{u},\mathbf{v}), \tag{12}$$

where

$$\begin{aligned}&W_{\varphi_m\varphi_n}(\mathbf{u},\mathbf{v})\\ &\quad= (2\pi)^{-n/2}\int \varphi_m\left(\mathbf{u}+\frac{\mathbf{u}'}{2}\right)\varphi_n^*\left(\mathbf{u}-\frac{\mathbf{u}'}{2}\right)\exp\left(i\mathbf{u}'\mathbf{v}'^T\right)d\mathbf{u}'.\end{aligned} \tag{13}$$

The interference terms in the WDF have an oscillatory behavior if the field distributions are coherent, and show no oscillations when the superposing beams are incoherent [22]. The oscillations indicate the eventuality of interference between the superposing states in the $\mathbf{u}$ and/or $\mathbf{v}$ domains, interferences in the $\mathbf{u}$ and $\mathbf{v}$ spaces, respectively, taking place only when the superposing field distributions have common projections along these axes. Note that the cross-terms in the WDF appear even when the individual WDFs do not overlap. The presence of the interference terms in the WDF of a multicomponent signal hampers in many situations the interpretation of the WDF and the reconstruction or synthesis of the field distribution. The geometry of the interference terms of the WDF and related signal representations in the temporal phase space is analyzed in [23]. Other properties of the WDF for coherent distributions can be found in [24], while the corresponding properties of the WDF for partially coherent distributions and for the discrete WDF are detailed in [16, 19], respectively.

An attractive property of the WDF is that its transformation law through first-order optical systems characterized in the geometrical optical approximation by a real symplectic matrix

$$S = \begin{pmatrix} A & B \\ C & D \end{pmatrix} \tag{14}$$

relating the ray vectors $(\mathbf{u}\mathbf{v})^T$ at the output $z = $ const. plane of an optical system to those at the input $z = 0$, is very simple [24, 25, 26]:

$$W_\varphi(\mathbf{u},\mathbf{v};z) = W_\varphi(\mathbf{u}D - \mathbf{v}B, -\mathbf{u}C + \mathbf{v}A; 0). \tag{15}$$

A matrix is symplectic if it satisfies the relation $S^TJS = J$ with

$$J = \begin{pmatrix} 0 & -I \\ I & 0 \end{pmatrix}, \tag{16}$$

where 0 and I are the n-dimensional null and identity matrices, respectively; for $n = 1$ the symplecticity condition reduces to $\det S = 1$.

The WDF embodies the link between geometrical optics and wave optics: it remains constant along geometrical ray paths, but incorporates wave optical effects when considering diffraction and interference phenomena since the WDF is defined in terms of the field distribution. Note that (15) is valid not only for the better known case of optical systems for monochromatic beams but also for optical systems for light pulses; in particular the propagation of the WDF of pulses through temporal lenses [27] and dispersive media (the analogs of free spaces for monochromatic beams [28]) is also accounted for by (15). The influence of ray-pulse matrices [29] on the WDF is also described by (15). A similar law can be derived for each temporal frequency component of a polychromatic-pulsed paraxial beam propagating through nondispersive and achromatic optical systems [30] (see also

[31] for the definition of the WDF for polychromatic paraxial fields), or for light propagation in misaligned optical systems [32]. Dragoman [33] has demonstrated that the *ABCD* transformation law for the WDF (15) holds also for particular optical systems characterized by complex S matrices when illuminated with partially coherent Gaussian-Schell beam; in this case, however, the A, B, C, D elements of the S matrix in (15) depend also on the incident field. The WDF transformation law in homogeneous, inhomogeneous, and weakly dispersive media has been derived by Bastiaans [34, 35], the corresponding transformation in nonstationary and inhomogeneous optical media can be found in [36], while the evolution of the WDF through optical systems characterized by complex matrices is given in [37]. It is also possible to derive the propagation law of the temporal WDF of pulsed plane-wave signals in continuous random media [38]. All these propagation laws, although very useful in some applications, are overcome in popularity by the *ABCD* transformation law of the WDF expressed in (15).

From (6), (7), and (15), it follows that the integral of the WDF over both $\mathbf{u}$ and $\mathbf{v}$ coordinates is proportional to the total energy content of the field distribution and that the energy is conserved at propagation through a first-order optical system characterized by a real and symplectic matrix. This conclusion no longer holds for propagation through optical systems characterized by a symplectic matrix with complex elements [39]; symplecticity of optical systems is not equivalent to energy conservation, with the notable exception of real matrices.

In the derivation of the elements of the symplectic matrix, S the reflection at various interfaces is neglected. An explicit accommodation of these reflections for layered structures is known to introduce a matrix that relates not the phase space coordinates but the field distribution and its derivative along the stratification direction [40]. A matrix-like relation exists even in this case between the vector WDF at the input and output plane, the elements of the WDF vector being the WDF of field, of its derivative, and the cross-terms [41]; the WDF moments at the output plane are related to those at the input plane through a more complicated relation than (15).

The simplicity of the WDF transformation law through first-order systems reflects itself in the simplicity of the transformation laws of the WDF moments. For a two-dimensional stationary and monochromatic light distribution $\varphi(x, y)$, for example, the moment of order $l + m + n + p$ is defined as

$$\overline{x^l y^m k_x^n k_y^p} = \frac{\int x^l y^m k_x^n k_y^p W_\varphi(x, y, k_x, k_y)\,dx\,dy\,dk_x\,dk_y}{\int W_\varphi(x, y, k_x, k_y)\,dx\,dy\,dk_x\,dk_y}, \tag{17}$$

and transforms through first-order systems according to [42]

$$M_j(z) = \underbrace{(S \otimes S \otimes \cdots)}_{[j/2]\text{times}} M_j(0) \underbrace{(S \otimes S \otimes \cdots)^T}_{(j-[j/2])\text{times}}. \tag{18}$$

In the expression above

$$M_j(z) = \underbrace{\overline{\begin{pmatrix}\mathbf{r}^T\\ \mathbf{k}^T\end{pmatrix} \otimes (\mathbf{rk}) \otimes \begin{pmatrix}\mathbf{r}^T\\ \mathbf{k}^T\end{pmatrix} \otimes \cdots}}_{j\text{ times}} \tag{19}$$

is the matrix that contains all moments of order j calculated at a $z = \text{const.}$ plane, with $\otimes$ denoting the direct product of matrices and with the bar on top indicating that each matrix element is averaged in the sense of (17); $M_j(0)$ has the same significance but with the moments calculated at the input $z = 0$ plane of the optical system. Numerous moment invariants of several orders can be derived starting from (18) (see [6] and the references therein). In particular, $(\det M_2)^{1/2}$, which is invariant at propagation through first-order optical systems, can be identified with the phase space area occupied by the system. Its invariance can be viewed as a selection rule for various optical systems, such as the astigmatic mode converter consisting of a pair of properly oriented cylindrical lenses that transforms a Hermite-Gauss mode with rectangular symmetry into a Laguerre-Gauss mode with cylindrical symmetry [43].

Note that, due to (6), the moments of the spatial coordinates in (17) (those with $n = p = 0$), calculated with the WDF as the weighting function, are identical to the moments of spatial coordinates calculated with the field intensity as the weighting function, while the moments of the angular coordinates (with $l = m = 0$) in (17) are identical to the moments calculated with the field spectrum as weighting function, in agreement with (7). The moments defined in (17), with the WDF as weighting function, allow however a simplified calculation procedure for mixed, that is, spatial and angular, coordinates.

The propagation law (18) for WDF-based moments can be extended in certain cases even to optical systems that are not of the first-order type. For example, a modified definition of second-order moments for hard-edge diffracted fields leads for $j = 2$ to a propagation law similar to (18) even if hard-edge apertures are not first-order systems [44]. Analogously, it was demonstrated that pure phase transmittances can be represented as 4×4 matrices and a law similar to (18) has been derived for the second-order moments of the WDF [45].

3. APPLICATIONS TO SIGNAL RETRIEVAL

The complex field distribution can be determined, according to (8), from its WDF; the recovery of both amplitude and phase distributions of a signal is one of the most important tasks in signal processing since many detectors are only sensitive to the incident intensity. The WDF of one-dimensional field distributions can be easily displayed on a two-dimensional screen using common optical elements. Optical set-ups that generate the spatial WDF of real signals, complex transparencies (holograms), or arbitrary complex one-dimensional fields are described in [46]; the product between the two displaced replicas of the field distribution

that enter the definition of the WDF in (1) is implemented using rotated transparencies or holograms through which the light passes twice (but with a transverse coordinate inversion between passages) or using a rotated object illuminated by a tilted light beam, the subsequent Fourier transform being generated with the help of lenses and free space propagation regions. More complex set-ups involving joint transform correlator architectures [47] or incorporating a degenerate phase conjugation device for the implementation of the product between the field and its complex conjugate (and shifted) replica [48] can also be imagined. The latter set-up can be employed as well to generate samples of the WDF of two-dimensional complex signals. A white-light implementation of the WDF of one-dimensional field distributions containing an achromatic processor that consists of linear blazed zone plates and an achromatic cylindrical objective was proposed in [49], while acousto-optic processors for real-time implementation of temporal WDFs are described in [50]. Note that, since the WDF is real, a measurement of the intensity at the output plane of a WDF-generating set-up yields the desired WDF values up to their sign, which can however be easily deduced from the (eventually) passages through zero of the intensity distribution if the WDF sign is known in one point; this is a trivial task for symmetric field distributions and a not-impossible task in general.

Except the rotationally symmetric beam for which the four-dimensional WDF can be recovered from the two-dimensional WDF of the field distribution along a line that passes through the center [51], only samples of the four-dimensional WDF of an arbitrary two-dimensional light source can be produced. Several methods have been proposed to generate these samples, and they differ through the manner of implementing the product between the field distribution and its shifted and complex conjugate replica: a dual-channel display technique [52], two-tilted [52] or properly shifted and rotated transparencies [52, 53], or a single transparency illuminated twice [53] can be used to generate desired samples of the WDF of real-valued fields, while a simultaneous optical production of several sections, $W_n(\mathbf{r}_n/2, \mathbf{k}/2)$, of the WDF can be implemented using a fiber grating to shift multiple copies of the input object with $\mathbf{r}_n$ [54].

Apart from the possibility of direct generation, the WDF function can be recovered from measurements of other distribution functions such as the spectrogram or the Radon or Hartley transforms (see the references in [6]). Alternatively, a powerful technique has been devised for the computation of the WDF, and hence for the retrieval of both amplitude and phase of the original field distribution, from measurements of light intensities at different planes or after passing through optical systems that rotate the original WDF in phase space. This method is known as tomography [55], and has been extensively used for the reconstruction of the WDF of one- [56] or two-dimensional optical beams [57], coherent and low-coherent light sources [58, 59], and optical pulses [60, 61]. As expected, the two methods of complex field recovery, that is, the direct generation of the WDF and the tomography, provide comparable results [62].

For two-dimensional monochromatic point-symmetric fields, that is, for fields for which $\varphi(-\mathbf{r}) = \pm\varphi(\mathbf{r})$, the WDF is the inverse Fourier transform of its scaled fractional Fourier transform spectrum, and hence can be determined from intensity measurements of the fractional Fourier transform of the incident field in a suitable number of points [63].

Closely related to the problem of signal retrieval is that of signal synthesis. A signal synthesis algorithm that involves the discrete-time WDF defined in (4) has been developed in [64]; it consists in finding the digital sequence $\varphi(n\tau)$ whose discrete-time WDF best approximates (in a least-squares sense) a desired function. A more detailed account on time-varying signal processing using WDF synthesis techniques can be found in [65].

4. APPLICATIONS TO IMAGE RECOGNITION

In applications involving pattern recognition and image analysis, the moments of the field distribution as well as the invariants that can be constructed with these moments are essential [66]. For example, the first-order moments of the spatial coordinates calculated with the monochromatic field intensity as the weighting function locate the centroid of the intensity distribution, while the second-order moments of the spatial coordinates characterize the orientation and the size of the image. Although the information contained in the original image can be recovered only if the infinite set of all moments is known, in most applications only a finite set of higher-order moments and moment invariants (in particular Zernike moment invariants [66]) are sufficient for image recognition and/or image classification purposes. Teague [67] showed that it is possible to calculate (with limited accuracy) image moments by optical means, that is, by using a Fourier transforming lens to perform the spatial integration. Why do we need then a WDF-based definition for moments? A powerful justification of the usefulness of WDF-based moments has been provided by Freeman and Saleh [68]. They showed that for many image recognition purposes it is better to use a set composed of moment invariants from both spatial and angular (spatial frequency) domains instead of the same number of moment invariants from the spatial or spatial frequency domain alone. This result, welcomed by experimentalists who know that noise hamper the accurate determination of moments to a degree that increases with the moment's order, validates the use of the mixed WDF-based moments in order to achieve the image recognition task with lower-order moments and moment invariants compared to the case when spatial or spatial frequency moments alone are used. Not to mention that the simple propagation laws of WDF-based moments through real and symplectic first-order systems offer the easiest way to find moment invariants.

A last but not least reason for the employment of moments defined with the WDF as a weighting function is that they can, for arbitrary orders, be determined from the measurement of intensity moments of an appropriate number of fractional power spectra of the incident field distribution, obtained by measuring the intensity at the output plane

of generally anamorphic fractional Fourier transform systems [69]. Other set-ups for measuring the ten second-order WDF-based moments of two-dimensional beams, which involve the measurements of second-order spatial moments at the output of astigmatic optical systems, have been put forward by Nemes and Siegman [70], and by Eppich et al. [71], whereas the phase space beam analyzer [72] measures simultaneously the beam radius and far-field divergence of one-dimensional field distributions.

5. APPLICATIONS TO THE CHARACTERIZATION OF SIGNALS

The spatial and angular extent of optical field distributions is determined by its second-order moments. For stationary monochromatic and centered two-dimensional field distributions (for which $\bar{x} = \bar{y} = \overline{k_x} = \overline{k_y} = 0$) the squared width of the beam along x, y and the associated squared far-field divergence angles are determined, respectively, by $\overline{x^2}$, $\overline{y^2}$ and $\overline{k_x^2}, \overline{k_y^2}, \overline{(x-\bar{x})^2}, \overline{(y-\bar{y})^2}$ and $\overline{(k_x-\overline{k_x})^2}, \overline{(k_y-\overline{k_y})^2}$ for non-centered light beams) while an overall characterization of the quality of the beam is provided by [73]

$$Q = \overline{x^2k_x^2} - \left(\overline{xk_x}\right)^2 + \overline{y^2k_y^2} - \left(\overline{yk_y}\right)^2 - 2\left(\overline{xk_y}\,\overline{yk_x} - \overline{xy}\,\overline{k_xk_y}\right). \tag{20}$$

The second-order moments and the beam quality parameter are tools that allow a quantitative comparison of the spatial and spatial frequency extent of beams that can otherwise differ substantially with respect to their shape or wavefront profile. Moreover, the propagation law (18) of the WDF-based moments through first-order systems is a useful tool for designing optical systems that can optimize the extent or the shape of the field distribution required in some applications.

First-order optical systems leave invariant the beam quality parameter Q and modify the WDF-based moments according to (18). The beam quality factor Q takes a minimum value for a Gaussian field distribution. The square root of the ratio between the Q value of an arbitrary beam and that of a Gaussian field is known as the beam-propagation factor and is denoted by M^2; M^2 describes the far-field spreading of an arbitrary beam relative to a Gaussian field with the same waist. A high-quality beam, that is, a beam with a small Q value is simultaneously localized in spatial and angular coordinates. The minimum value of Q is 1/4 for one-dimensional monochromatic beams, and 1/2 for two-dimensional field distributions. The occurrence of a minimum value for Q corresponds to the existence of a classical counterpart of the uncertainty relations in quantum mechanics; uncertainty relations can also be defined for nonparaxial wave fields [74]. For multimode beams in an optical resonator, M^2 determines the number of transverse modes in one dimension [75].

An effective radius of curvature matrix R^{-1} for an arbitrary quasimonochromatic but centered optical beam can be introduced in terms of second-order WDF-based moments as $R^{-1} = (1/k)\overline{\mathbf{u}^T \otimes \mathbf{v}}/\overline{\mathbf{u}^T \otimes \mathbf{u}}$; R^{-1} has in the spatio-temporal phase space the meaning of a spatial curvature and a temporal chirp [30]. The matrix division in the definition of R^{-1} should be understood as multiplication of the matrix in the numerator with the inverse of the matrix in the denominator. A similar definition of an effective radius of curvature matrix can be introduced for each frequency component of a polychromatic pulse that propagates through nondispersive and achromatic optical systems [30]. A complex radius of curvature matrix q^{-1} can also be introduced for arbitrary beams as $q^{-1} = R^{-1} - i\lambda M^2/(4\pi\overline{\mathbf{u}^T \otimes \mathbf{u}})$, the arbitrary beam satisfying the same propagation rule through first-order optical systems as a Gaussian beam with the same size if the wavelength λ is replaced with λM^2 [76, 77, 78]. In particular the second-order spatial moment (the beam width) obeys a parabolic law of propagation in free space, the minimum beam width occurring at the waist plane irrespective of the shape and wavefront profile of the beam. This remarkable result, which allows the characterization of arbitrary beams with concepts derived for Gaussian field distributions, is only possible when second-order moments, and in particular WDF-based second-order moments, are employed for beam characterization.

The twist parameter of partially coherent Gaussian field distributions [79], as well as the angular rotation of arbitrary fields under free space propagation can also be expressed in terms of second-order moments [80]. More precisely, for a two-dimensional monochromatic light beam, the angle between the principal axes, defined as those for which $\overline{xy} = 0$, and the absolute axes, that is, those for which $\overline{k_xk_y} = 0$, is given by $\cot 2\theta = (\overline{x^2} - \overline{y^2})/(2\overline{xy})$, the angular rotation offering a classification criterion of light beams. It is also worth mentioning that the Stokes matrices associated with partially polarized fields can be also expressed in terms of the second-order moments of the WDF matrix of the vectorial electric field [14].

The symmetry of one-dimensional fields is described by the third-order moment $\overline{x^3}$, the moments $\overline{x^2k_x}$ and $\overline{x^3k_x}$ measure, respectively, the spatial range of the beam's symmetry and sharpness, $\overline{x^2k_x^2}/(\overline{x^2}\,\overline{k_x^2})$ expresses the degree of similarity of an arbitrary beam with a quasihomogeneous field distribution [81], while the kurtosis parameter $K = \overline{x^4}/(\overline{x^2})^2$ provides a quantitative measure for the classification of beams with respect to their sharpness [82, 83]. K is useful to estimate the capabilities of laser beams in material processing applications.

Special relations between moments of the WDF occur for light beams with certain symmetries. For example, for one-dimensional self-Fourier field distributions $\varphi(x)$, which are characterized by $\bar{\varphi}(k_x) = a_1\varphi(a_2k_x)$ with constant a_1, a_2 parameters such that $|a_1^2/a_2| = 1$, the WDF satisfies the symmetry condition $W_\varphi(x, k_x) = W_\varphi(a_2k_x, -x/a_2)$, while its moments are related through $\overline{x^mk_x^n} = (-1)^m a_2^{m-n}\overline{x^nk_x^m}$ [84]. The symmetry of moments is thus an indication of the symmetry of the field. Field distributions with such a special symmetry can preserve it only at propagation through certain optical systems, characterized by a symplectic matrix that can be determined from (15). Bright solitons are examples of self-Fourier functions.

Not only the moments defined by (17), which are called global moments, have physical significance, but also local moments can be associated with parameters of the signal. For example, $\mathbf{K}(\mathbf{r}) = \int \mathbf{k} W(\mathbf{r},\mathbf{k}) d\mathbf{k} / \int W(\mathbf{r},\mathbf{k}) d\mathbf{k}$ can be interpreted as the scaled (with k) average propagation angle at position $\mathbf{r}$, $\mathbf{R}(\mathbf{k}) = \int \mathbf{r} W(\mathbf{r},\mathbf{k}) d\mathbf{r} / \int W(\mathbf{r},\mathbf{k}) d\mathbf{r}$ can be identified with the average beam position at the transverse wavevector k, $T(\omega) = \int t W(t,\omega) dt / \int W(t,\omega) dt$ can be seen as the group delay of the optical pulse, and $\Omega(t) = \int \omega W(t,\omega) d\omega / \int W(t,\omega) d\omega$ can be defined as the instantaneous frequency of a pulse [1].

Although the WDF is mainly used for characterizing optical beams and pulses that propagate through linear media, some studies have also focused on nonlinear propagation. A thorough analysis of the propagation of the WDF and of its moments through active media, for example, can be performed under the thin sheet condition, which restricts the diffraction along the amplifier and limits the magnitude and variation of the refractive index and gain coefficient of the active medium. Under these assumptions the field amplitude in the paraxial approximation can be expressed as a sum of two terms, one of them being approximately proportional to the z coordinate along the propagation direction. The WDF of this sum contains both auto-terms and a cross-term, each of them satisfying the free space propagation law in (15) with (in one dimension) $A = D = 1$, $C = 0$, and $B = z$ [85]. As a result, the WDF does not propagate along straight lines in active media and the radiant intensity, that is, the average of the WDF along the spatial coordinate, as well as the beam quality parameter do not remain invariant at propagation. Moreover, an incident aligned field distribution, with $\bar{x} = \overline{k_x} = 0$ at the input plane does not preserve this property at the output plane, in contrast to the behavior in homogeneous and passive media. Numerical simulations have shown that at propagation of Gaussian beams through active media under the thin sheet approximation, in both homogeneously and inhomogeneously broadened cases, the beam size increases compared to the propagation through passive media and the kurtosis parameter lowers (the beam becomes flatter) [86]. The form of the WDF also changes at propagation through nonlinear media. For example, at self-phase modulation the WDF of an incident Gaussian pulse develops an asymmetric pronounced dip and can even take negative values [87].

Studies performed using the WDF as a mathematical tool have shown that the degree of coherence of the beam has a determinant role in propagation. For example, partial incoherence weakens the nonlinearity, a higher intensity being required to achieve the same effect as that obtained with a coherent beam. In particular, partial incoherence tends to suppress the coherent instabilities, acting as a Landau-like damping effect on modulational instabilities, or tends to prevent the self-focusing collapse instability of bidimensional field distributions [88]. These effects have been demonstrated employing the WDF of the envelope function of partially coherent fields together with the Klimontovich statistical average.

The evolution of the beam quality factor in a nonlinear Kerr medium with a quadratic refractive index profile has been derived in [89]; an invariant beam quality parameter can be introduced for nonlinear Kerr-like media for a properly defined effective radius of curvature. Moreover, Nasalski [90] has demonstrated that the first-order formalism can be preserved at propagation of a light pulse or beam through a nonlinear Kerr-like medium if the phase space coordinates are properly scaled: the low-power (linear) coordinates $\mathbf{u}, \mathbf{v}$ should be replaced with $\mathbf{U} = \mathbf{u}/w(\zeta)$, $\mathbf{V} = \mathbf{v}w(\zeta)$, where w is the half-width pulse waist and ζ is the (temporal and spatial) chirp parameter, which vanishes at the actual waist plane. ζ replaces the z coordinate as propagation parameter and differs from it because of the self-shortening effect and because of the waist shift with respect to the linear propagation situation. Apart from these scaling operations of the phase space coordinates, an on-axis phase correction must be introduced in the expressions of both the field amplitude and its Fourier transform, and an additional scaling with $w^{n/2}(\zeta)$ and $w^{-n/2}(\zeta)$ of the field amplitude and its Fourier transform, respectively, must be taken into account, where n is the dimension of the nonlinear Schrödinger equation satisfied by the quasimonochromatic pulse.

The analytic expression of the WDF for bright solitons and for the two-soliton solution of the nonlinear Schrödinger equation, which is identical to the equation satisfied by optical solitons propagating in a dispersive and nonlinear medium, can be found in [91]. Bright solitons are characterized by a diamond-shaped WDF, while the WDF of the two-soliton solution evolves in time oscillating between profiles similar to the dark soliton and profiles with a large momentum distribution, which correspond to the most squeezed state in spatial coordinates. Most interesting, however, is the appearance of a distinct interference cross-term in the WDF of the two-soliton solution when the amplitudes of the two solitons are comparable. The WDF of black and gray solitons in the temporal phase space as well as the WDF moments of all envelope solitons (bright, black, gray) also have analytical expressions [92]. These expressions have provided test methods to decide if an optical pulse is or not a soliton; such phase space tests were applied to a bright soliton pulse source used for 270 terabit km/s transmission system [92] and for bright and dark magnetostatic solitons [93].

The transport equation of the WDF for field distributions that satisfy the nonlinear Schrödinger equation, the Korteweg-de Vries equation, as well as the Burgers equation was derived by Kamp [94], who separated the operator $G(\mathbf{r}, \partial/\partial\mathbf{r}, \varphi(\mathbf{r},t))$ acting upon the field distribution into a linear and a nonlinear part: $G(\mathbf{r}, \partial/\partial\mathbf{r}, \varphi(\mathbf{r},t)) = F(\mathbf{r}, \partial/\partial\mathbf{r}) + L[\varphi(\mathbf{r},t)]$. Note that the WDF evolution law under a linear operation that acts on the field distribution according to

$$\frac{\partial \varphi}{\partial t} = -F\left(\mathbf{r}, \frac{\partial}{\partial \mathbf{r}}\right)\varphi \tag{21}$$

is given by

$$\frac{\partial W_\varphi}{\partial t} = -2\mathrm{Re}\left[F\left(\mathbf{r} + \frac{i}{2}\frac{\partial}{\partial \mathbf{k}}, \frac{1}{2}\frac{\partial}{\partial \mathbf{r}} + i\mathbf{k}\right)\right] W_\varphi. \tag{22}$$

(See also [34, 35, 36, 37].) The equation of motion of the WDF for an underlying governing equation that is a linear ordinary or partial differential equation has been derived in [95].

Starting from (22) the transformation laws for the WDF, the radiant intensity, the radiant emittance, the first- and second-order moments of the WDF, and the kurtosis parameter of a field that propagates through an inhomogeneous Kerr-type medium have been derived [96]. Quite surprisingly, it was found that the WDF transforms according to a law similar to (15), where the elements of the real symplectic matrix depend on the field distribution. This fact leads, however, to a variation of the beam quality factor with the propagation distance, in contrast to the case when the elements of the real symplectic matrix do not depend on the field distribution.

The soliton solutions of the nonlinear Schrödinger equation refer to envelope solitary waves in $\chi^{(3)}$ materials, but stationary wave solutions of soliton type exist also in $\chi^{(2)}$ materials. These are related to a variety of phenomena that include degenerate and nondegenerate parametric amplification and the coupling of intense electromagnetic fields to polaritons. The solitons in $\chi^{(2)}$ media are solutions of coupled field equations in which the nonlinearity is second order in the field, and exist in both anomalously and normally dispersive regions. In absorptive media, the widths of $\chi^{(2)}$ solitons are maintained over long distances while their amplitudes decay exponentially. The WDF of the bright and dark $\chi^{(2)}$ solitons and their second-order moments are given in [97]; the obtained expressions are different from the corresponding ones in $\chi^{(3)}$ solitons, providing a signature of the soliton type, and can be used to test different pulses.

6. APPLICATIONS TO THE CHARACTERIZATION OF OPTICAL SYSTEMS AND DEVICES

A measurement of the WDF after it passes through an optical system is a valuable means to characterize the system if the incident field distribution is known. Due to the *ABCD* transformation law and due to the fact that the WDF can be defined for either coherent or partially coherent field distributions, the WDF incorporates information about the coherence of illumination and the optical system through which it passes. Therefore it can be used to evaluate the performance and tolerances of optical processors [98] and, in particular, to estimate the parameters of various optical systems. For example, according to (15), the WDF of a collimated, monochromatic, and one-dimensional field, $W(x,k_x) = \delta(k_x)$, becomes $W(x,k_x) = \delta(k_x + k_x x/f)$ after passing through a thin lens characterized by a transmission function $\exp(ikx^2/2f)$, so that the tilt angle of the WDF after the lens is a measure of its focal length. (The free space produces a shear transformation of the WDF along x.) Analogously, the passage of the same field distribution through a pure phase object with a transmission function $\exp[i\phi(x)]$ that cannot be in general characterized through a real symplectic matrix yields an output WDF of the form $W(x,k_x) = \delta(k_x + d\phi/dx)$, from which the profile of the pure phase object can be recovered. An example of the phase profile recovery from the WDF measurement can be found in [99].

One example of special interest is the case of spherically aberrated lenses. The type of aberration, as well as the value of the aberration coefficient can be determined from the form of the WDF of the output field distribution [100]. The presence of aberrations manifests itself in the change of shape of the output WDF, in addition to the coordinate transformation predicted by the *ABCD* law in (15). As demonstrated in [101], the output WDF of an aberrated optical system is obtained by applying an exponential differential operator to the WDF of the same optical system in the absence of aberrations, the first term of the operator predicting the coordinate transformation of the unaberrated WDF while the second term predicts the shape distortion.

In many applications the overall effect of aberrations can be quantified by the changes in Q. For a Gaussian incident field, for example, the effect of spherical aberrations can be calculated analytically [102]; in the two-dimensional case, spherical aberrations always deteriorate the beam quality if the incident field distribution is real. On the contrary, a Gaussian aperture always improves the beam quality parameter for one-dimensional partially coherent Gaussian-Schell beams since it reduces the beamwidth (the far-field divergence can be enlarged or decreased depending on the position of the aperture with respect to the waist plane) [103].

The WDF can also be used to express the intensity distribution along different paths in the image space of an optical system [104], and can hence be used for an efficient analysis of the performance of the imaging system in the presence of spherical aberrations. Moreover, the WDF is also useful for the design of pupil filters that would generate specific axial responses in the presence of spherical aberrations [105]. The WDF of the desired pupil filters is obtained through tomography from the desired axial irradiance distributions, and the pupil filter is subsequently determined from its WDF.

Due to the simplicity of the *ABCD* transformation law of the WDF through first-order optical system, it is easier to study optical systems using the WDF approach than to compute the field evolution through such systems. And this approach can also be used for polychromatic light beams, as long as each frequency component is treated separately. This is the reason why it is desirable to employ the WDF approach to follow the propagation of polychromatic coherent fields through a system designed to achromatize Fresnel diffraction patterns, for example, [106]. The residual chromatic aberrations can be obtained as well using this approach to achromatic white-light self-imaging.

The propagation of partially coherent fields through apertures is another subject that can be linked to the WDF. More precisely, the angular spectrum of the cross-spectral density of the partially coherent field at an observation plane is a WDF that depends on the path vectors between the aperture and the observation plane [107]. The complex degree of spatial coherence acts as a weighting factor for the contribution of aperture radiator pairs to the WDF at each position in the observation plane.

When pulses are launched into single-mode fibers, $\bar{t}$ and $\bar{\omega}$ represent the mean arrival time and the mean frequency of the pulse, respectively, the squared temporal and spectral pulse widths being determined by $\overline{(t-\bar{t})^2}$ and $\overline{(\omega-\bar{\omega})^2}$, respectively, [108]. Moreover, the transmission capacity of a fiber at a z = const. plane can be defined as the inverse of the pulse width at that plane, that is, as $N(z) = (\overline{(t-\bar{t})^2}|_z)^{-1/2}$. At propagation through dispersive fibers with a second-order dispersion coefficient, the beam quality parameter $Q = \overline{(t-\bar{t})^2}\,\overline{(\omega-\bar{\omega})^2} - (\overline{(t-\bar{t})(\omega-\bar{\omega})})^2$ remains constant and the moments of the time and frequency coordinates transform according to (18), whereas for fibers with third-order dispersion terms, Q is a second-order polynomial in the fiber's length z; the evolution of the transmission capacities of the fibers in the two cases can be calculated analytically [108].

The study of active devices, in particular lasers, can also benefit from the computation of moments of the field distribution and of the beam quality parameter. For example, the beam quality parameter and the second-order moments for the fields emitted by heterojunction monomode diode lasers with very thin active layers can be calculated analytically and can be related to physical constants that characterize the laser device [109]. These parameters of the emitted field can then be used to design first-order optical systems that can transform the field emitted by monomode diode lasers into a circular field distribution at a desired plane. (This transformation can only be achieved at some planes since the emitted field has different degrees of global coherence on each transverse axis.) An efficient coupling of such laser beams into circularly symmetric optical fibers, for example, require thus the introduction of precisely designed first-order optical systems between the laser and the fiber's end.

The WDF was also applied by Gase [110] to the design of multimode laser resonators consisting of two mirrors and a laser rod between them, the laser rod acting as a convergence lens due to the temperature gradient. The *ABCD* transformation law of the WDF and of its moments was used to match the radius of the laser beam, considered as an isotropic Gaussian-Schell beam, to the free-rod diameter, this condition describing the stability of the resonator; note that the radius of the laser beam depends on the beam quality parameter. The focusing performances of multimode laser beams as well as the disturbance effect of an aperture diaphragm were also treated with the WDF formalism.

Holograms can also be related to the WDF. A hologram is an interference pattern recorded by the superposition of the signal $\varphi_s(\mathbf{r})$ and the reference beam $\varphi_r(\mathbf{r})$, the signal wavefront being recovered by reading the hologram, that is, by illuminating it with a replica of the reference signal. It can be demonstrated that the interference pattern of a hologram is identical to the partial integral over $\mathbf{k}$ of the interference term (or cross-term) $W_{\varphi_s\varphi_r}(\mathbf{r},\mathbf{k})$ of the field distribution $\varphi_s(\mathbf{r}) + \varphi_r(\mathbf{r})$ obtained by superposing the signal and reference beams [111]. This cross-term in the WDF has a strongly oscillating behavior and therefore has been called "the smile function" of the superposition state. Note that once a hologram is recorded it is possible to read the information carried by it by measuring the WDF of the hologram without the need to employ a replica of the reference beam in order to recover the signal wavefront. This fact has been used successfully to extract the location of three-dimensional objects from inline holograms without reconstructing the three-dimensional field [112]. The procedure works if the concentration of particles in inline holograms is not too high, since otherwise the cross-terms in the WDF may interfere with the auto-terms and the interpretation of the WDF becomes extremely difficult.

Another effect that can be analyzed with the WDF is the moiré pattern, which is an undesired feature of pixelated displays that originates from undersampling and which is encountered in image processing, topography, Talbot interferometry, and submicrometer alignment methods. The multiplicative moiré patterns with well-defined local frequencies, resulting from the superposition of nonperiodic masks, for example, can be analyzed disregarding the cross-terms in the WDF of the superposed masks [113], but the cross-terms must be accounted for in interferometric applications. The WDF interpretation of moiré patterns provides an intuitive insight into the information stored in the phase modulation of nonperiodic moiré fringes.

Optical set-ups that shape ultrashort pulses, for example, can only be characterized by a WDF defined on a spatio-temporal phase space. A typical set-up of this kind consists of a grating that spatially separates the different frequency components of an ultrashort pulse, followed by a collimating lens, a mask that modifies in a desired manner each spectral component, and another lens that focuses the collimated beam onto a subsequent grating, which superimposes spatially the diverse spectral components. Each of these devices, with the exception of the mask, can be described by a real symplectic S matrix such that the evolution of the WDF through the set-up can be described in a much simpler way than the evolution of the spatially- and frequency-dependent field distribution. As mentioned before, a spatio-temporal phase space is needed since the grating couples the spatial and temporal features of the pulse. The effect of the mask with a transmission function $t(\mathbf{u})$ on the incident field distribution $\varphi_{\text{in}}(\mathbf{u})$, described as $\varphi_{\text{out}}(\mathbf{u}) = t(\mathbf{u})\varphi_{\text{in}}(\mathbf{u})$, corresponds to a change of the WDF given by (see also [12])

$$W_{\varphi_{\text{out}}}(\mathbf{u},\mathbf{v}) = (2\pi)^{-n/2}\int W_{\varphi_{\text{in}}}(\mathbf{u},\mathbf{v}')\,W_t(\mathbf{u},\mathbf{v}-\mathbf{v}')\,d\mathbf{v}'. \quad (23)$$

The convolution in the $\mathbf{v}$ coordinate describes the action of any filtering device in $\mathbf{u}$, a symmetrical effect, that is, a convolution of the WDF in $\mathbf{u}$, occurring for any filtering action in the $\mathbf{v}$ domain. The WDF treatment of the pulse shaper is detailed in [114].

Another optical system that has been studied with the spatio-temporal WDF is a Lukosz multiplexer with moving gratings, which sends a one-dimensional signal through a finite-width aperture by segmenting it into parts separated by small differences in wavelengths, and then reconstructs the signal [115]. More precisely, the multiplexer consists of

a moving grating, a Fourier transforming system, the finite-width aperture, another Fourier transforming system, and a counterpropagating grating. Since a Fourier-transforming system formed from a lens and free spaces is described by an *ABCD* matrix and the action of the finite-width aperture can be seen as that of a filtering device, a WDF approach to the analysis of the multiplexer would be justified if the moving grating could be described easily in phase space. (Note that, unlike in the beam shaper the grating is illuminated with an extended object.) Fortunately, such a simple description exists: the WDF $W_{\text{in}}(x, k_x; \lambda)$ of a one-dimensional input signal of wavelength λ, $\varphi(x, \lambda)$ is multiplied by the grating and transformed into

$$W_{\text{out}}(x, k_x; \lambda) = \sum_n W_n(x) W_{\text{in}}\left(x, k_x - \frac{n\pi}{\Lambda}; \lambda\right), \qquad (24)$$

where Λ is the grating period and $W_n(x)$ is the intensity of each copy, which is directly related to the shape of the grating. A moving grating viewed with optical sensors that integrate over time should be represented in phase space by the same expression as in (24), but the coefficients $W_n(x)$ become now time dependent and hence should be averaged over time. Other optical set-ups analyzed with spatio-temporal WDFs include a white-light spatial Fourier transformer, a rotated zone plate [116], and a grating compressor for optical pulses consisting of two parallel diffraction gratings separated by a free space region [117].

One of the most recently emerged applications of temporal phase space distribution functions is the synthesis [118] or reconstruction [119] of periodic grating profiles, in particular of fiber Bragg gratings. Such devices can function, for example, as wavelength-selective components, dispersion compensators, or mode converters. The problem is to reconstruct or synthesize the refractive index profile from the measured or desired values of the complex field reflection coefficient $\tilde{r}(\omega)$ or the reflection impulse response $r(t)$; these two functions are related by a Fourier transform. Once one of these is known, it is easy to compute the WDF but the next step, that is, the finding of the refractive index profile, is generally difficult due to the presence of cross-terms in the WDF of a multicomponent signal. These cross-terms, which reflect the correlations between pairs of auto-terms, have larger amplitudes than the auto-terms and are situated in the middle (in both temporal and frequency coordinates) between the corresponding auto-terms. It is even possible that cross-terms that connect pairs of auto-terms overlap with other auto-terms, the resulting WDF becoming indeed very difficult to interpret. So, for fiber grating reconstruction or synthesis, the WDF is not the best choice: the spectrogram, for example, or the Choi-Williams distribution [4] are better suited since they are not plagued by the presence of cross-terms. On the other hand, the resolution limitation of the spectrogram technique imposed by the choice of an appropriate window function is always worse than that achieved with the WDF approach; a suitable compromise is then to use the cross-term-free spectrogram as a mask over the WDF and so to reduce the unwanted cross-terms of the latter. Our point here is that the WDF is not always the best suited phase space distribution, although it satisfies a number of attractive properties that are not shared by other distribution functions (see, e.g., [1]).

Not only the refractive index profile of fiber gratings can be recovered using the WDF, but also this approach works well for the recovery of spatial nonperiodic profiles of the refractive index of single- or multimoded guided structures. The refractive index profile reconstruction can be achieved by solving a linear system of equations that involve moments of the gradients along $\mathbf{r}$ and $\mathbf{k}$ of the measured or desired WDF. The procedure works for either longitudinally constant [120] or longitudinally variant but continuous refractive index profiles [121], and for different propagating modes in the optical structure. If the WDF gradients $\nabla_{\mathbf{r}} W$ and $\nabla_{\mathbf{k}} W$ are approximated with $[W(\mathbf{r}+d\mathbf{r}, \mathbf{k}) - W(\mathbf{r}, \mathbf{k})]/d\mathbf{r}$ and $[W(\mathbf{r}, \mathbf{k} + d\mathbf{k}) - W(\mathbf{r}, \mathbf{k})]/d\mathbf{k}$, respectively, the recovery precision depends on the steps in $\mathbf{r}$ and $\mathbf{k}$ of the measured or simulated WDF.

It is important to note that the usefulness of the WDF for the characterization of optical systems is not limited to the case when these are of the first-order type, with an eventual inclusion of aberration effects. The effect of complex integrated structures on light propagation is usually estimated with the beam-propagation method, which implies the division of the guiding structure into a number of steps such that in each step the refractive index profile can be decomposed in longitudinally- and transverse-varying parts. The computation of the output field is carried out by applying in each step different operators to the field distribution and to its Fourier transform, which implies that at each step a computation of the direct and the inverse Fourier transform of the field is required [122]. The succession of these operations can be replaced with a much simpler WDF formulation of the beam-propagation method, which has the advantage that it depends on both spatial and spatial frequency coordinates and thus the operators at each step can be applied directly on the WDF, the successive time-consuming direct and indirect Fourier transforms being no longer needed [123]. The operators that must be applied on the WDF are simple differential operators; it is even possible to obtain in the first-order approximation an analytic formula for the WDF transformation through some integrated structures, such as a linear taper with a transverse parabolic refractive index profile.

7. APPLICATIONS TO THE CALCULATION OF THE COUPLING COEFFICIENT

The overlap or Moyal relation of the WDF (property (vii) of the WDF) can be used to express the coupling efficiency of light distributions to a region of space with different refractive or dispersive properties. The optimization of the coupling efficiency between different optical systems is an issue of enormous practical importance in any signal processing and signal transmission application.

A definition of the coupling coefficient between a light source and a waveguide, for example, in terms of spatial phase space distribution functions is more appropriate for the analysis of the coupling problem than a definition in terms of spatial or spatial frequency distributions since the numerical aperture (an angular parameter) is as important as the dimensions of the waveguide core (a spatial parameter) in the characterization of waveguide field propagation. The expression of the coupling efficiency between a coherent source, with field distribution $\varphi_s(\mathbf{r})$ and WDF $W_s(\mathbf{r},\mathbf{k})$, and a waveguide mode $\varphi_w(\mathbf{r})$ with a WDF $W_w(\mathbf{r},\mathbf{k})$ is given by [124]

$$\eta = (2\pi)^n \frac{\int W_s(\mathbf{r},\mathbf{k}) W_w(\mathbf{r},\mathbf{k}) d\mathbf{r}\, d\mathbf{k}}{\int W_s(\mathbf{r},\mathbf{k}) d\mathbf{r}\, d\mathbf{k} \int W_w(\mathbf{r},\mathbf{k}) d\mathbf{r}\, d\mathbf{k}}, \tag{25}$$

where the integrals extend over the illumination plane. The behavior of the coupling coefficient in (25), which is a real parameter, with spatial and angular misalignments between the source and the waveguide mode can be intuitively grasped from the graphical representations of $W_s(\mathbf{r},\mathbf{k})$ and $W_w(\mathbf{r},\mathbf{k})$ (see [124, 125] for more details). Note that in phase space, the spatial and angular misalignments correspond to displacements of the respective WDF along the $\mathbf{r}$ and $\mathbf{k}$ axes, respectively. The connection between the value of the coupling coefficient and the graphic representations of the WDF of a quantum-well modulator and the first four modes of a planar optical waveguide, for example, has been discussed in [126]. According to (25), coupling is achieved if the WDFs (or the fields) of the source and waveguide overlap.

For partially coherent light distributions, the expression of the coupling coefficient in (25) becomes [127]

$$\eta = \frac{\left| \int W_{sw}(\mathbf{r},\mathbf{k}) d\mathbf{r}\, d\mathbf{k} \right|^2}{\int W_s(\mathbf{r},\mathbf{k}) d\mathbf{r}\, d\mathbf{k} \int W_w(\mathbf{r},\mathbf{k}) d\mathbf{r}\, d\mathbf{k}}, \tag{26}$$

where

$$\begin{aligned} W_{sw}(\mathbf{r},\mathbf{k}) &= (2\pi)^{-n/2} \int \left\langle \varphi_s\left(\mathbf{r}+\frac{\mathbf{r}'}{2}\right) \varphi_w^*\left(\mathbf{r}-\frac{\mathbf{r}'}{2}\right) \right\rangle \exp\left(i\mathbf{r}'\mathbf{k}^T\right) d\mathbf{r}' \\ &= (2\pi)^{-n/2} \int \Gamma_{sw}\left(\mathbf{r}+\frac{\mathbf{r}'}{2}, \mathbf{r}-\frac{\mathbf{r}'}{2}\right) \exp\left(i\mathbf{r}'\mathbf{k}^T\right) d\mathbf{r}' \end{aligned} \tag{27}$$

while the coupling coefficient between a coherent multimode light source described by the field distribution $\varphi_s(\mathbf{r}) = \sum_n a_n \varphi_s^n(\mathbf{r})$ and a superposition $\varphi_w(\mathbf{r}) = \sum_m b_m \varphi_w^m(\mathbf{r})$ of waveguides modes $\varphi_w^m(\mathbf{r})$ can be expressed as [127]

$$\eta = \sum a_l a_k^* b_n b_m^* \eta^{lknm}. \tag{28}$$

Here

$$\eta^{lknm} = (2\pi)^n \frac{\int W_s^{lk}(\mathbf{r},\mathbf{k}) W_w^{nm}(\mathbf{r},\mathbf{k}) d\mathbf{r}\, d\mathbf{k}}{\int W_s(\mathbf{r},\mathbf{k}) d\mathbf{r}\, d\mathbf{k} \int W_w(\mathbf{r},\mathbf{k}) d\mathbf{r}\, d\mathbf{k}} \tag{29}$$

with

$$W_i^{lk}(\mathbf{r},\mathbf{k}) = (2\pi)^{-n/2} \int \varphi_i^l\left(\mathbf{r}+\frac{\mathbf{r}'}{2}\right) \varphi_i^{k*}\left(\mathbf{r}-\frac{\mathbf{r}'}{2}\right) \exp\left(i\mathbf{r}'\mathbf{k}^T\right) d\mathbf{r}', \tag{30}$$

$i = s, w$, $W_s(\mathbf{r},\mathbf{k})$ and $W_w(\mathbf{r},\mathbf{k})$ having the same significance as in (25).

8. CONCLUSIONS

The paper introduces the WDF and its most important properties as a mathematical tool in several areas of signal processing that include signal retrieval, signal recognition, characterization of signals and optical systems, and coupling coefficient estimation in phase space. The mathematical formalism can be applied to either spatial, temporal, or spatiotemporal phase spaces, to coherent, partially coherent or digital signals, offering a unified view for the analysis of field propagation through various optical systems. In many cases the WDF approach to the study of field propagation and coupling is very much simplified due to its desirable properties and due to the extremely simple propagation law (see (15)) through first-order optical systems. This propagation law can be employed to characterize optical systems described by symplectic matrices, active devices such as lasers, or to estimate the value of aberration effects. The moments of the WDF, subject also to a very simple evolution law expressed by (18), are related to spatial, temporal, angular, and/or frequency characteristics of the field distribution and can be either directly measurable or can be calculated from the experimentally obtained WDF. These moments and/or their invariant combinations (such as the beam quality factor) are quantitative parameters that allow the comparison of beams with totally different shapes or wavefronts. The WDF is thus a universal tool, in the sense that it can be applied to (almost) any field distribution and an intuitive and simple representation of field evolution (due to its *ABCD* law of transformation) and coupling (because it characterizes simultaneously the fields' overlap in both spatial and spatial frequency domains). The continuous growth of the number of applications of the WDF is hence predictable, desirable, and enjoyable.

REFERENCES

[1] T. A. C. M. Claasen and W. F. G. Mecklenbräuker, "The Wigner distribution—a tool for time-frequency signal analysis. Part I. Continuous-time signals," *Philips Journal of Research*, vol. 35, no. 3, pp. 217–250, 1980.

[2] T. A. C. M. Claasen and W. F. G. Mecklenbräuker, "The Wigner distribution—a tool for time-frequency signal analysis. Part II: Discrete-time signals," *Philips Journal of Research*, vol. 35, pp. 276–300, 1980.

[3] T. A. C. M. Claasen and W. F. G. Mecklenbräuker, "The Wigner distribution—a tool for time-frequency signal analysis. Part III: Relations with other time-frequency transformations," *Philips Journal of Research*, vol. 35, pp. 372–389, 1980.

[4] L. Cohen, "Time-frequency distributions—a review," *Proc. IEEE*, vol. 77, no. 7, pp. 941–981, 1989.
[5] P. Flandrin, *Temps-Fréquence*, Hermès, Paris, France, 1993.
[6] D. Dragoman, "The Wigner distribution function in optics and optoelectronics," *Progress in Optics*, vol. 37, pp. 1–56, 1997.
[7] G. Matz and F. Hlawatsch, "Wigner distributions (nearly) everywhere: time-frequency analysis of signals, systems, random processes, signal spaces, and frames," *Signal Processing*, vol. 83, no. 7, pp. 1355–1378, 2003.
[8] B. Boashash, Ed., *Time Frequency Signal Analysis and Processing—A Comprehensive Review*, Elsevier, Amsterdam, The Netherlands, 2003.
[9] F. Hlawatsch and G. F. Boudreaux-Bartels, "Linear and quadratic time-frequency signal representations," *IEEE Signal Processing Mag.*, vol. 9, no. 2, pp. 21–67, 1992.
[10] E. Wigner, "On the quantum correction for thermodynamic equilibrium," *Physical Review*, vol. 40, pp. 749–759, 1932.
[11] A. Walther, "Radiometry and coherence," *Journal of the Optical Society of America*, vol. 58, pp. 1256–1259, 1968.
[12] M. J. Bastiaans, "The Wigner distribution function applied to optical signals and systems," *Optics Communications*, vol. 25, pp. 26–30, 1978.
[13] D. S. Bugnolo and H. Bremmer, "The Wigner distribution matrix for the electric field in a stochastic dielectric with computer simulation," *Advances in Electronics and Electron Physics*, vol. 61, pp. 299–389, 1983.
[14] R. Martínez-Herrero, P. M. Mejías, and J. M. Movilla, "Spatial characterization of general partially polarized beams," *Optics Letters*, vol. 22, no. 4, pp. 206–208, 1997.
[15] J. Paye, "The chronocyclic representation of ultrashort light pulses," *IEEE J. Quantum Electron.*, vol. 28, no. 10, pp. 2262–2273, 1992.
[16] M. J. Bastiaans, "Application of the Wigner distribution function to partially coherent light," *Journal of the Optical Society of America* {*A*}, vol. 3, pp. 1227–1237 , 1986.
[17] B. Boashash and P. J. Black, "An efficient real-time implementation of the Wigner-Ville distribution," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 35, no. 11, pp. 1611–1618, 1987.
[18] F. Peyrin and R. Prost, "A unified definition for the discrete-time, discrete-frequency, and discrete-time/frequency Wigner distributions," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 34, no. 4, pp. 858–867, 1986.
[19] K.-H. Brenner, "A discrete version of the Wigner distribution function," in *Signal Processing II: Theories and Applications*, H. W. Schüssler, Ed., Elsevier (North-Holland), Amsterdam, The Netherlands, 1983.
[20] K. B. Wolf, M. A. Alonso, and G. W. Forbes, "Wigner functions for Helmholtz wave fields," *Journal of the Optical Society of America* {*A*}, vol. 16, no. 10, pp. 2476–2487, 1999.
[21] D. Dragoman, "Phase space interference as the source of negative values of the Wigner distribution function," *Journal of the Optical Society of America* {*A*}, vol. 17, no. 12, pp. 2481–2485, 2000.
[22] D. Dragoman and M. Dragoman, "Quantum coherent versus classical coherent light," *Optical and Quantum Electronics*, vol. 33, no. 3, pp. 239–252, 2001.
[23] F. Hlawatsch and P. Flandrin, "The interference structure of the Wigner distribution and related time-frequency signal representations," in *The Wigner Distribution—Theory and Applications in Signal Processing*, W. Mecklenbräuker and F. Hlawatsch, Eds., pp. 59–132, Elsevier, Amsterdam, The Netherlands, 1997.
[24] M. J. Bastiaans, "Wigner distribution function and its application to first-order optics," *Journal of the Optical Society of America*, vol. 69, pp. 1710–1716, 1979.
[25] M. J. Bastiaans, "The Wigner distribution function and Hamilton's characteristics of a geometric-optical system," *Optics Communications*, vol. 30, pp. 321–326, 1979.
[26] R. K. Luneburg, *Mathematical Theory of Optics*, University of California Press, Berkeley, Calif, USA, 1966.
[27] A. A. Godil, B. A. Auld, and D. M. Bloom, "Picosecond time-lenses," *IEEE J. Quantum Electron.*, vol. 30, no. 3, pp. 827–837, 1994.
[28] B. H. Kolner, "Space-time duality and the theory of temporal imaging," *IEEE J. Quantum Electron.*, vol. 30, no. 8, pp. 1951–1963, 1994.
[29] A. G. Kostenbauder, "Ray-pulse matrices: a rational treatment for dispersive optical systems," *IEEE J. Quantum Electron.*, vol. 26, no. 6, pp. 1148–1157, 1990.
[30] J. Yang and D. Fan, "Intensity-moments characterization of general pulsed paraxial beams with the Wigner distribution function," *Journal of the Optical Society of America* {*A*}, vol. 16, no. 10, pp. 2488–2493, 1999.
[31] K. B. Wolf, "Wigner distribution function for paraxial polychromatic optics," *Optics Communications*, vol. 132, no. 3–4, pp. 343–352, 1996.
[32] D. Onciul, "Characterization of the propagation of light beams through misaligned linear optical systems," *Optik*, vol. 93, pp. 23–26, 1993.
[33] D. Dragoman, "Wigner distribution function for Gaussian-Schell beams in complex matrix optical systems," *Applied Optics*, vol. 34, pp. 3352–3357, 1995.
[34] M. J. Bastiaans, "Transport equations for the Wigner distribution function," *Optica Acta*, vol. 26, pp. 1265–1272, 1979.
[35] M. J. Bastiaans, "Transport equations for the Wigner distribution function in an inhomogeneous and dispersive medium," *Optica Acta*, vol. 26, pp. 1333–1344, 1979.
[36] J. T. Mendonça and N. L. Tsintsadze, "Analog of the Wigner-Moyal equation for the electromagnetic field," *Physical Review E*, vol. 62, no. 3, pp. 4276–4282, 2000.
[37] D. Dragoman, "Wigner distribution function for a complex matrix optical system," *Optik*, vol. 100, pp. 137–139, 1995.
[38] J. Oz and E. Heyman, "Wigner-Ville distribution for pulse propagation in random media: pulsed plane wave," *Waves in Random Media*, vol. 8, no. 2, pp. 175–191, 1998.
[39] D. Dragoman, "The Wigner distribution function and the energy conservation of a light beam," *Journal of Modern Optics*, vol. 43, pp. 1127–1134, 1996.
[40] M. Born and E. Wolf, *Principles of Optics*, Pergamon, Oxford, UK, 1970.
[41] D. Dragoman, "Wigner-transform approach to paraxial light propagation in stratified media," *Journal of the Optical Society of America* {*A*}, vol. 14, no. 5, pp. 1103–1109, 1997.
[42] D. Dragoman, "Higher-order moments of the Wigner distribution function in first-order optical systems," *Journal of the Optical Society of America* {*A*}, vol. 11, no. 10, pp. 2643–2646, 1994.
[43] J. B. Murphy, "Phase space conservation and selection rules for astigmatic mode converters," *Optics Communications*, vol. 165, no. 1–3, pp. 11–18, 1999.
[44] R. Martínez-Herrero and P. M. Mejías, "Second-order spatial characterization of hard-edge diffracted beams," *Optics Letters*, vol. 18, no. 19, pp. 1669–1671, 1993.
[45] G. Piquero, P. M. Mejías, and R. Martínez-Herrero, "Propagation of laser beam parameters through pure phase transmittances," *Optics Communications*, vol. 129, no. 3–4, pp. 161–166, 1996.

[46] K.-H. Brenner and A. W. Lohmann, "Wigner distribution function display of complex 1D signals," *Optics Communications*, vol. 42, no. 5, pp. 310–314, 1982.
[47] G. Shabtay, D. Mendlovic, and Z. Zalevsky, "Proposal for optical implementation of the Wigner distribution function," *Applied Optics*, vol. 37, pp. 2142–2144, 1998.
[48] Y. Li, G. Eichmann, and M. Conner, "Optical Wigner distribution and ambiguity function for complex signals and images," *Optics Communications*, vol. 67, no. 3, pp. 177–179, 1988.
[49] J. Lancis, E. E. Sicre, E. Tajahuerce, and P. Andrés, "White-light implementation of the Wigner-distribution function with an achromatic processor," *Applied Optics*, vol. 34, pp. 8209–8212, 1995.
[50] R. A. Athale, J. N. Lee, E. L. Robinson, and H. H. Szu, "Acousto-optic processors for real-time generation of time-frequency representations," *Optics Letters*, vol. 8, no. 3, pp. 166–168, 1983.
[51] D. Dragoman, "Can the Wigner transform of a two-dimensional rotationally symmetric beam be fully recovered from the Wigner transform of its one-dimensional approximation?" *Optics Letters*, vol. 25, no. 5, pp. 281–283, 2000.
[52] M. Conner and Y. Li, "Optical generation of the Wigner distribution of 2-D real signals," *Applied Optics*, vol. 24, pp. 3825–3829, 1985.
[53] R. Bamler and H. Glünder, "The Wigner distribution function of two-dimensional signals. Coherent-optical generation and display," *Optica Acta*, vol. 30, pp. 1789–1803, 1983.
[54] T. Iwai, A. K. Gupta, and T. Asakura, "Simultaneous optical production of the sectional Wigner distribution function for a two-dimensional object," *Optics Communications*, vol. 58, no. 1, pp. 15–19, 1986.
[55] G. T. Hermann, *Image Reconstruction from Projections: The Fundamentals of Computerized Tomography*, Academic Press, New York, NY, USA, 1980.
[56] D. F. McAlister, M. Beck, L. Clarke, A. Mayer, and M. G. Raymer, "Optical phase retrieval by phase-space tomography and fractional-order Fourier transforms," *Optics Letters*, vol. 20, no. 10, pp. 1181–1183, 1995.
[57] M. G. Raymer, M. Beck, and D. F. McAlister, "Complex wave-field reconstruction using phase-space tomography," *Physical Review Letters*, vol. 72, no. 8, pp. 1137–1140, 1994.
[58] K. F. Lee, F. Reil, S. Bali, A. Wax, and J. E. Thomas, "Heterodyne measurement of Wigner distributions for classical optical fields," *Optics Letters*, vol. 24, no. 19, pp. 1370–1372, 1999.
[59] A. Wax, S. Bali, and J. E. Thomas, "Optical phase-space distributions for low-coherence light," *Optics Letters*, vol. 24, no. 17, pp. 1188–1190, 1999.
[60] M. Beck, M. G. Raymer, I. A. Walmsley, and V. Wong, "Chronocyclic tomography for measuring the amplitude and phase structure of optical pulses," *Optics Letters*, vol. 18, no. 23, pp. 2041–2043, 1993.
[61] J. W. Wu, P. K. Lam, M. B. Gray, and H. A. Bachor, "Optical homodyne tomography of information carrying laser beams," *Optics Express*, vol. 3, no. 4, pp. 154–161, 1998.
[62] D. Dragoman, M. Dragoman, and K.-H. Brenner, "Amplitude and phase recovery of rotationally symmetric beams," *Applied Optics*, vol. 41, pp. 5512–5518, 2002.
[63] T. Alieva and M. J. Bastiaans, "Wigner distribution and fractional Fourier transform for two-dimensional symmetric optical beams," *Journal of the Optical Society of America* {*A*}, vol. 17, no. 12, pp. 2319–2323, 2000.
[64] G. F. Boudreaux-Bartels and T. W. Parks, "Time-varying filtering and signal estimation using Wigner distribution synthesis techniques," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 34, no. 3, pp. 442–451, 1986.
[65] G. F. Boudreaux-Bartels, "Time-varying signal processing using Wigner distribution synthesis techniques," in *The Wigner Distribution—Theory and Applications in Signal Processing*, W. Mecklenbräuker and F. Hlawatsch, Eds., pp. 269–317, Elsevier, Amsterdam, The Netherlands, 1997.
[66] M. R. Teague, "Image analysis via the general theory of moments," *Journal of the Optical Society of America*, vol. 70, no. 8, pp. 920–930, 1980.
[67] M. R. Teague, "Optical calculation of irradiance moments," *Applied Optics*, vol. 19, no. 8, pp. 1353–1356, 1980.
[68] M. O. Freeman and B. E. A. Saleh, "Moment invariants in the space and frequency domains," *Journal of the Optical Society of America* {*A*}, vol. 5, no. 7, pp. 1073–1084, 1988.
[69] M. J. Bastiaans and T. Alieva, "Wigner distribution moments measured as fractional Fourier transform intensity moments," in *Proc. 19th Congress of the International Commission for Optics: Optics for the Quality of Life*, vol. 4829 of *Proceedings of SPIE*, pp. 245–246, Firenze, Italy, August 2002.
[70] G. Nemes and A. E. Siegman, "Measurement of all ten second-order moments of an astigmatic beam by the use of rotating simple astigmatic (anamorphic) optics," *Journal of the Optical Society of America* {*A*}, vol. 11, no. 8, pp. 2257–2264, 1994.
[71] B. Eppich, C. Gao, and H. Weber, "Determination of the ten second order intensity moments," *Optics & Laser Technology*, vol. 30, no. 5, pp. 337–340, 1998.
[72] N. Hodgson, T. Haase, R. Kostka, and H. Weber, "Determination of laser beam parameters with the phase space beam analyzer," *Optical and Quantum Electronics*, vol. 24, no. 9, pp. S927–S949, 1992.
[73] J. Serna, R. Martínez-Herrero, and P. M. Mejías, "Parametric characterization of general partially coherent beams propagating through ABCD optical systems," *Journal of the Optical Society of America* {*A*}, vol. 8, no. 7, pp. 1094–1098, 1991.
[74] M. A. Alonso and G. W. Forbes, "Uncertainty products for nonparaxial wave fields," *Journal of the Optical Society of America* {*A*}, vol. 17, no. 12, pp. 2391–2402, 2000.
[75] S. Lavi, R. Prochaska, and E. Keren, "Generalized beam parameters and transformation laws for partially coherent light," *Applied Optics*, vol. 27, no. 17, pp. 3696–3703, 1988.
[76] A. E. Siegman, "Defining the effective radius of curvature for a nonideal optical beam," *IEEE J. Quantum Electron.*, vol. 27, no. 5, pp. 1146–1148, 1991.
[77] P. A. Bélanger, "Beam propagation and the ABCD ray matrices," *Optics Letters*, vol. 16, no. 4, pp. 196–198, 1991.
[78] M. A. Porras, J. Alda, and E. Bernabeu, "Complex beam parameter and ABCD law for non-Gaussian and nonspherical light beams," *Applied Optics*, vol. 31, no. 30, pp. 6389–6402, 1992.
[79] M. J. Bastiaans, "Wigner distribution function applied to twisted Gaussian light propagating in first-order optical systems," *Journal of the Optical Society of America* {*A*}, vol. 17, no. 12, pp. 2475–2480, 2000.
[80] J. Serna, P. M. Mejías, and R. Martínez-Herrero, "Rotation of partially coherent beams propagating through free space," *Optical and Quantum Electronics*, vol. 24, no. 9, pp. S873–S880, 1992.
[81] R. Martínez-Herrero, P. M. Mejías, M. Sánchez, and J. L. H. Neira, "Third- and fourth-order parametric characterization of partially coherent beams propagating through ABCD optical systems," *Optical and Quantum Electronics*, vol. 24, no. 9, pp. S1021–S1026, 1992.
[82] G. Piquero, P. M. Mejías, and R. Martínez-Herrero, "Sharpness changes of Gaussian beams induced by spherically

aberrated lenses," *Optics Communications*, vol. 107, no. 3–4, pp. 179–183, 1994.

[83] R. Martínez-Herrero, G. Piquero, and P. M. Mejías, "On the propagation of the kurtosis parameter of general beams," *Optics Communications*, vol. 115, no. 3–4, pp. 225–232, 1995.

[84] D. Dragoman, "The Wigner distribution function of self-Fourier functions," *Journal of Modern Optics*, vol. 43, no. 9, pp. 1933–1938, 1996.

[85] R. Martínez-Herrero and P. M. Mejías, "Beam characterization through active media," *Optics Communications*, vol. 85, no. 2–3, pp. 162–166, 1991.

[86] J. R. Martín de Los Santos, R. Martínez-Herrero, and P. M. Mejías, "Parametric characterization of Gaussian beams propagating through active media," *Optical and Quantum Electronics*, vol. 28, no. 8, pp. 1021–1027, 1996.

[87] C. Hirlimann and J.-F. Morhange, "Wavelet analysis of short light pulses," *Applied Optics*, vol. 31, no. 17, pp. 3263–3266, 1992.

[88] L. Helczynski, D. Anderson, R. Fedele, B. Hall, and M. Lisak, "Propagation of partially incoherent light in nonlinear media via the Wigner transform method," *IEEE J. Select. Topics Quantum Electron.*, vol. 8, no. 3, pp. 408–412, 2002.

[89] C. Paré and P.-A. Bélanger, "Beam propagation in a linear or nonlinear lens-like medium using ABCD ray matrices: the method of moments," *Optical and Quantum Electronics*, vol. 24, no. 9, pp. S1051–S1070, 1992.

[90] W. Nasalski, "Scale formulation of first-order nonlinear optics," *Optics Communications*, vol. 137, no. 1–3, pp. 107–112, 1997.

[91] H. Konno and P. S. Lomdahl, "The Wigner transform of soliton solutions for the nonlinear Schrödinger equation," *Journal of the Physical Society of Japan*, vol. 63, no. 11, pp. 3967–3973, 1994.

[92] D. Dragoman and M. Dragoman, "Phase space characterization of solitons with the Wigner transform," *Optics Communications*, vol. 137, no. 4–6, pp. 437–444, 1997.

[93] D. Dragoman and M. Dragoman, "Time-frequency characterization of magnetostatic envelope soliton waves," *Applied Physics Letters*, vol. 70, no. 6, pp. 714–716, 1997.

[94] L. P. J. Kamp, "Local frequency spectra for nonlinear wave equations," *Journal of Physics A*, vol. 20, no. 11, pp. 3279–3291, 1987.

[95] L. Galleani and L. Cohen, "The Wigner distribution for classical systems," *Physics Letters A*, vol. 302, no. 4, pp. 149–155, 2002.

[96] D. Dragoman, "Wigner distribution function in nonlinear optics," *Applied Optics*, vol. 35, no. 21, pp. 4142–4146, 1996.

[97] D. Dragoman, M. Dragoman, and A. R. Trasca, "The Wigner transform of soliton solutions in $\chi^{(2)}$ media," *Optik*, vol. 111, pp. 20–24, 2000.

[98] A. Pe'er, D. Wang, A. W. Lohmann, and A. A. Friesem, "Wigner formulation of optical processing with light of arbitrary coherence," *Applied Optics*, vol. 40, no. 2, pp. 249–256, 2001.

[99] T. Iwai, N. Kawamura, and T. Takemori, "Real-time profiling of a pure phase object using an auto-Wigner distribution function," *Optics Communications*, vol. 95, no. 4–6, pp. 199–204, 1993.

[100] D. Dragoman, "Wigner distribution function applied to third-order aberrations," *Applied Optics*, vol. 35, no. 1, pp. 161–168, 1996.

[101] A. W. Lohmann, J. Ojeda-Castañeda, and N. Streibl, "The influence of wave aberrations on the Wigner distribution," *Optica Applicata*, vol. 13, pp. 465–471, 1983.

[102] R. Martínez-Herrero, G. Piquero, and P. M. Mejías, "Beam quality changes produced by quartic phase transmittances," *Optical and Quantum Electronics*, vol. 27, pp. 173–183, 1995.

[103] J. Serna, P. M. Mejías, and R. Martínez-Herrero, "Beam quality changes of Gaussian Schell-model fields propagating through Gaussian apertures," *Applied Optics*, vol. 31, no. 22, pp. 4330–4331, 1992.

[104] G. Saavedra, W. D. Furlan, E. Silvestre, and E. E. Sicre, "Analysis of the irradiance along different paths in the image space using the Wigner distribution function," *Optics Communications*, vol. 139, no. 1–3, pp. 11–16, 1997.

[105] W. D. Furlan, D. Zalvidea, and G. Saavedra, "Synthesis of filters for specified axial irradiance by use of phase-space tomography," *Optics Communications*, vol. 189, no. 1–3, pp. 15–19, 2001.

[106] J. Lancis, E. E. Sicre, A. Pons, and G. Saavedra, "Achromatic white-light self-imaging phenomenon: an approach using the Wigner distribution function," *Journal of Modern Optics*, vol. 42, no. 2, pp. 425–434, 1995.

[107] R. Castañeda, "On the relationship between the cross-spectral density and the Wigner distribution function," *Journal of Modern Optics*, vol. 45, no. 3, pp. 587–593, 1998.

[108] C. Martínez, R. Martínez-Herrero, and P. M. Mejías, "Parametric characterization of light pulses propagating through single-mode fibres," *Journal of Modern Optics*, vol. 43, no. 9, pp. 1785–1796, 1996.

[109] J. Serna, P. M. Mejías, and R. Martínez-Herrero, "Beam quality in monomode diode lasers," *Optical and Quantum Electronics*, vol. 24, no. 9, pp. S881–S887, 1992.

[110] R. Gase, "The multimode laser radiation as a Gaussian Schell model beam," *Journal of Modern Optics*, vol. 38, no. 6, pp. 1107–1115, 1991.

[111] K. B. Wolf and A. L. Rivera, "Holographic information in the Wigner function," *Optics Communications*, vol. 144, no. 1–3, pp. 36–42, 1997.

[112] L. Onural and M. T. Özgen, "Extraction of three-dimensional object-location information directly from in-line holograms using Wigner analysis," *Journal of the Optical Society of America* {*A*}, vol. 9, no. 2, pp. 252–260, 1992.

[113] M. Testorf, "Analysis of the moiré effect by use of the Wigner distribution function," *Journal of the Optical Society of America* {*A*}, vol. 17, no. 12, pp. 2536–2542, 2000.

[114] J. Paye and A. Migus, "Space-time Wigner functions and their application to the analysis of a pulsed shaper," *Journal of the Optical Society of America B*, vol. 12, no. 8, pp. 1480–1490, 1995.

[115] K. B. Wolf, D. Mendlovic, and Z. Zalevsky, "Generalized Wigner function for the analysis of superresolution systems," *Applied Optics*, vol. 37, no. 20, pp. 4374–4379, 1998.

[116] D. Mendlovic and Z. Zalevsky, "Definition, properties, and applications of the generalized temporal-spatial Wigner distribution function," *Optik*, vol. 107, no. 2, pp. 49–56, 1997.

[117] V. J. Pinto-Robledo and T. A. Hall, "Chronocyclic description of laser pulse compression," *Optics Communications*, vol. 125, no. 4–6, pp. 338–348, 1996.

[118] M. A. Muriel, J. Azaña, and A. Carballar, "Fiber grating synthesis by use of time-frequency representations," *Optics Letters*, vol. 23, no. 19, pp. 1526–1528, 1998.

[119] J. Azaña and M. A. Muriel, "Reconstruction of fiber grating period profiles by use of Wigner-Ville distributions and spectrograms," *Journal of the Optical Society of America* {*A*}, vol. 17, no. 12, pp. 2496–2505, 2000.

[120] D. Dragoman, M. Dragoman, and J.-P. Meunier, "Recovery of the refractive-index profile from the Wigner distribution of an optical waveguide," *Applied Optics*, vol. 37, no. 12, pp. 2357–2360, 1998.

[121] D. Dragoman and J. P. Meunier, "Recovery of longitudinally variant refractive index profile from the measurement of the Wigner transform," *Optics Communications*, vol. 153, no. 4–6, pp. 360–367, 1998.
[122] R. März, *Integrated Optics: Design and Modeling*, Artech House, Norwood, Mass, USA, 1995.
[123] D. Dragoman, J. P. Meunier, and M. Dragoman, "Beam-propagation method based on the Wigner transform: a new formulation," *Optics Letters*, vol. 22, no. 14, pp. 1050–1052, 1997.
[124] D. Onciul, "Efficiency of light launching into waveguides: a phase space approach," *Optik*, vol. 96, no. 1, pp. 20–24, 1994.
[125] D. Dragoman, "Wigner-distribution-function representation of the coupling coefficient," *Applied Optics*, vol. 34, no. 29, pp. 6758–6763, 1995.
[126] D. Dragoman and M. Dragoman, "Integrated optic-devices characterization with the Wigner transform," *IEEE J. Select. Topics Quantum Electron.*, vol. 2, no. 2, pp. 181–186, 1996.
[127] D. Onciul, "Waveguide launching efficiency for multimoded and partially coherent light sources," *Optik*, vol. 97, no. 2, pp. 75–77, 1994.

Daniela Dragoman was born in Raduati, Romania, in 1965. She obtained the M.S. degree from the University of Bucharest, Romania, in 1989, and the Ph.D. degree from Limerick University, Ireland, in 1993. She was a Visiting Professor at the University of Saint-Etienne, France, in 1997 and 2000 and was awarded the Alexander von Humboldt fellowship in 1998, being a Visiting Professor at the University of Mannheim, Germany, during 1998–1999 and 2001–2002. Presently, she is Professor at the Physics Faculty, University of Bucharest, where she teaches integrated optoelectronic devices and the interaction of radiation with matter. She authored about 100 papers in the area of fiber optics, optical beam characterization, quantum devices, and quantum optics. She is the coauthor of the books *Advanced Optoelecronic Devices*, Springer, 1999, *Optical Characterization of Solids*, Springer, 2002, and *Quantum-Classical Analogies*, Springer, 2004. Current research interests include beam characterization in linear and nonlinear media, optical micromechanical devices, mesoscopic devices, quantum and classical coherence, and interference. She is a reviewer at many international journals of optics and photonics and Editor of *The Frontiers Collection* published by Springer. She was awarded the Romanian Academy Prize "Gheorghe Cartianu" in 1999.

EURASIP Journal on Applied Signal Processing 2005:10, 1535–1540

Wigner Distribution Moments Measured as Intensity Moments in Separable First-Order Optical Systems

Martin J. Bastiaans
Faculteit Elektrotechniek, Technische Universiteit Eindhoven, Postbus 513, 5600 MB Eindhoven, The Netherlands
Email: m.j.bastiaans@tue.nl

Tatiana Alieva
Facultad de Ciencias Físicas, Universidad Complutense de Madrid, Ciudad Universitaria s/n, Madrid 28040, Spain
Email: talieva@fis.ucm.es

Received 31 March 2004; Revised 19 August 2004

It is shown how all global Wigner distribution moments of arbitrary order can be measured as intensity moments in the output plane of an appropriate number of separable first-order optical systems (generally anamorphic ones). The minimum number of such systems that are needed for the determination of these moments is derived.

Keywords and phrases: Wigner distribution moments, beam characterization, first-order optical systems.

1. INTRODUCTION

After the introduction of the Wigner distribution (WD) [1] for the description of coherent and partially coherent optical fields [2], it became an important tool for optical signal/image analysis and beam characterization [3, 4, 5]. The WD completely describes the complex amplitude of a coherent optical field (up to a constant phase factor) or the mutual coherence function of a partially coherent field. As the WD of a two-dimensional optical field is a function of four variables, it is difficult to analyze. Therefore, the optical field is often represented not by the WD itself, but by its global moments. Beam characterization based on the second-order moments of the WD thus became the basis of an International Organization for Standardization standard [6].

Some of the WD moments can directly be determined from measurements of the intensity distributions in the image plane or the Fourier plane, but most of the moments cannot be determined in such an easy way. In order to calculate such moments, additional information is required. Since first-order optical systems [7]—also called *ABCD* systems—produce affine transformations of the WD in phase space, the intensity distributions measured at the output of such systems can provide such additional information. The application of *ABCD* systems for the measurements of the second-order WD moments has been reported in several publications [8, 9, 10, 11, 12, 13].

It is the aim of this paper to show how all global WD moments can be measured as intensity moments only. We show that not only the second-order moments, but also all other moments of the four-dimensional WD can be obtained from measurements of only intensity distributions in an appropriate number of (generally anamorphic) separable first-order optical systems.

2. WIGNER DISTRIBUTION

Let partially coherent light be described by a temporally stationary stochastic process $f(x, y; t)$; as far as the time dependence is concerned, the ensemble average of the product $f(x_1, y_1; t_1) f^*(x_2, y_2; t_2)$, where the asterisk denotes complex conjugation, is then only a function of the time difference $t_1 - t_2$:

$$\mathrm{E}\{f(x_1, y_1; t_1) f^*(x_2, y_2; t_2)\} = \gamma(x_1, x_2; y_1, y_2; t_1 - t_2). \quad (1)$$

The function $\gamma(x_1, x_2; y_1, y_2; \tau)$ is known as the mutual coherence function [14, 15, 16, 17] of the stochastic process $f(x, y; t)$. The mutual power spectrum [16, 17] or cross-spectral density function [18] $\Gamma(x_1, x_2; y_1, y_2; \omega)$ is defined as the temporal Fourier transform of the mutual coherence function:

$$\Gamma(x_1, x_2; y_1, y_2; \omega) = \int_{-\infty}^{\infty} \gamma(x_1, x_2; y_1, y_2; \tau) \exp(j\omega\tau) d\tau. \quad (2)$$

For $x_1 = x_2 = x$, $y_1 = y_2 = y$, the cross-spectral density function reduces to the (auto) power spectrum $\Gamma(x, x; y, y; \omega)$, which represents the intensity distribution of the light for the temporal frequency ω. Since in the present discussion the explicit temporal-frequency dependence is of no importance, we will, for the sake of convenience, omit the temporal-frequency variable ω from the formulas in the remainder of the paper.

The Wigner distribution of partially coherent light is defined in terms of the cross-spectral density function by

$$W(x,u;y,v) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \Gamma\left(x+\frac{1}{2}x', x-\frac{1}{2}x'; y+\frac{1}{2}y', y-\frac{1}{2}y'\right) \times \exp\left[-j2\pi(ux'+vy')\right]dx'\,dy'. \quad (3)$$

A distribution function according to definition (3) was first introduced in optics by Walther [19, 20], who called it the generalized radiance. The WD $W(x, u; y, v)$ represents partially coherent light in a combined space/spatial-frequency domain, the so-called phase space, where u is the spatial-frequency variable associated to the space variable x, and v the spatial-frequency variable associated to the space variable y.

In this paper we consider the normalized moments of the WD, where the normalization is with respect to the total energy E of the signal:

$$E = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} W(x,u;y,v)\,dx\,du\,dy\,dv = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\Gamma(x,x;y,y)\,dx\,dy. \quad (4)$$

These normalized moments μ_{pqrs} of the WD are thus defined by

$$\mu_{pqrs}E = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} W(x,u;y,v)\,x^p u^q y^r v^s\,dx\,du\,dy\,dv \quad (p,q,r,s \geq 0) = \frac{1}{(4\pi j)^{q+s}}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x^p y^r \left(\frac{\partial}{\partial x_1}-\frac{\partial}{\partial x_2}\right)^q \left(\frac{\partial}{\partial y_1}-\frac{\partial}{\partial y_2}\right)^s \Gamma(x_1,x_2;y_1,y_2)\Big|_{x_1=x_2=x, y_1=y_2=y} dx\,dy. \quad (5)$$

Note that for $q = s = 0$ we have intensity moments, which can easily be measured:

$$\mu_{p0r0}E = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} W(x,u;y,v) \times x^p y^r\,dx\,du\,dy\,dv \quad (p,r \geq 0) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x^p y^r \Gamma(x,x;y,y)\,dx\,dy. \quad (6)$$

The WD moments μ_{pqrs} provide valuable tools for the characterization of optical beams, see, for instance [21]. First-order moments yield the position of the beam (μ_{1000} and μ_{0010}) and its direction (μ_{0100} and μ_{0001}). Second-order moments give information about the spatial width of the beam (the shape μ_{2000} and μ_{0020} of the spatial ellipse and its orientation μ_{1010}) and the angular width in which the beam is radiating (the shape μ_{0200} and μ_{0002} of the spatial-frequency ellipse and its orientation μ_{0101}); moreover, they provide information about its curvature (μ_{1100} and μ_{0011}) and its twist (μ_{1001} and μ_{0110}). Many important beam characterizers, like the overall beam quality [12]

$$(\mu_{2000}\mu_{0200} - \mu_{1100}^2) + (\mu_{0020}\mu_{0002} - \mu_{0011}^2) + 2(\mu_{1010}\mu_{0101} - \mu_{1001}\mu_{0110}), \quad (7)$$

are based on second-order moments. Higher-order moments are used, for instance, to characterize the beam's symmetry and its sharpness [21].

3. SEPARABLE FIRST-ORDER OPTICAL SYSTEMS

It is well known that the input-output relationship between the WD $W_{\text{in}}(x, u; y, v)$ at the input plane and the WD $W_{\text{out}}(x, u; y, v)$ at the output plane of a separable first-order optical system reads [3, 4, 5]

$$W_{\text{out}}(x,u;y,v) = W_{\text{in}}(d_x x - b_x u, -c_x x + a_x u; d_y y - b_y v, -c_y y + a_y v). \quad (8)$$

The coefficients a_x, b_x, c_x, d_x and a_y, b_y, c_y, d_y are the matrix entries of the symplectic ray transformation matrix [7] that relates the position x, y and direction u, v of an optical ray in the input and the output plane of the first-order optical system:

$$\begin{bmatrix} x_{\text{out}} \\ y_{\text{out}} \\ u_{\text{out}} \\ v_{\text{out}} \end{bmatrix} = \begin{bmatrix} a_x & 0 & b_x & 0 \\ 0 & a_y & 0 & b_y \\ c_x & 0 & d_x & 0 \\ 0 & c_y & 0 & d_y \end{bmatrix} \begin{bmatrix} x_{\text{in}} \\ y_{\text{in}} \\ u_{\text{in}} \\ v_{\text{in}} \end{bmatrix}. \quad (9)$$

For separable systems, symplecticity simply reads $a_x d_x - b_x c_x = 1$ and $a_y d_y - b_y c_y = 1$. Note that in a first-order optical system, with such a symplectic ray transformation matrix, the total energy E, see (4), is invariant.

As examples of first-order optical systems we mention the following in particular:

(i) a section of free space in the paraxial approximation, or "parabolic" system [22] (with $a = d = 1$, $c = 0$, and b proportional to the propagation distance z),
(ii) a fractional Fourier transform system [23], or "elliptic" system [22] (with $a = d = \cos\alpha$ and $b = -c = \sin\alpha$),
(iii) a "hyperbolic" system [22] (with $a = d = \cosh\alpha$ and $b = c = \sinh\alpha$).

These three systems are characterized by one parameter. Other one-parameter first-order optical systems are

(i) a thin lens (with $a = d = 1$, $b = 0$, and c inverse proportional to the focal distance),
(ii) an ideal magnifier (with $a = m$, $d = 1/m$, $b = c = 0$).

The latter systems however—like all systems for which the input and output planes are conjugate planes—cannot be used to determine the moments, as we will see later, because they have the property $b \equiv 0$.

The normalized moments μ^{out}_{pqrs} of the output WD $W_{\text{out}}(x,u;y,v)$ are related to the normalized moments $\mu^{\text{in}}_{pqrs} = \mu_{pqrs}$ of the input WD $W_{\text{in}}(x,u;y,v)$ as

$$\begin{aligned}
\mu^{\text{out}}_{pqrs} E &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} W_{\text{out}}(x,u;y,v) \\
&\qquad \times x^p u^q y^r v^s \, dx\, du\, dy\, dv \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} W_{\text{in}}(d_x x - b_x u, -c_x x + a_x u; \\
&\qquad\qquad d_y y - b_y v, -c_y y + a_y v) \\
&\qquad \times x^p u^q y^r v^s \, dx\, du\, dy\, dv \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} W_{\text{in}}(x,u;y,v)(a_x x + b_x u)^p \\
&\qquad \times (c_x x + d_x u)^q (a_y y + b_y v)^r \\
&\qquad \times (c_y y + d_y v)^s \, dx\, du\, dy\, dv \\
&= E \sum_{k=0}^{p}\sum_{l=0}^{q}\sum_{m=0}^{r}\sum_{n=0}^{s} \binom{p}{k}\binom{q}{l}\binom{r}{m}\binom{s}{n} a_x^{p-k} b_x^k c_x^l \\
&\qquad \times d_x^{q-l} a_y^{r-m} b_y^m c_y^n d_y^{s-n} \\
&\qquad \times \mu_{p-k+l,q-l+k,r-m+n,s-n+m},
\end{aligned} \tag{10}$$

and for the intensity moments in particular (i.e., $q = s = 0$) we have

$$\mu^{\text{out}}_{p0r0} = \sum_{k=0}^{p}\sum_{m=0}^{r}\binom{p}{k}\binom{r}{m} a_x^{p-k} b_x^k a_y^{r-m} b_y^m \, \mu_{p-k,k,r-m,m}. \tag{11}$$

The remainder of this paper is based on (11), in which the output intensity moments μ^{out}_{p0r0} are expressed in terms of the input moments μ_{pqrs} and the system parameters a_x, a_y, b_x, and b_y. Note that only the parameters a and b enter this equation; the parameters c and d can be chosen freely, as long as the symplecticity condition $a_x d_x - b_x c_x = a_y d_y - b_y c_y = 1$ is satisfied.

4. RELATIONS BETWEEN INPUT AND OUTPUT MOMENTS

4.1. First-order moments

For the first-order moments, the following two equations are relevant:

$$\mu^{\text{out}}_{1000} = a_x \mu_{1000} + b_x \mu_{0100}, \tag{12}$$

$$\mu^{\text{out}}_{0010} = a_y \mu_{0010} + b_y \mu_{0001}, \tag{13}$$

which correspond to (11) with $pqrs = 1000$ and $pqrs = 0010$, respectively, and the four input moments μ_{1000}, μ_{0100}, μ_{0010}, and μ_{0001} can be determined by measuring the intensity moments μ^{out}_{1000} and μ^{out}_{0010} in the output planes of two systems with different values of a and b, see (12) and (13), respectively.

In the case of a fractional Fourier transform system we can choose, for instance, [24, 25], the fractional angles $\alpha_x = \alpha_y = 0$ (leading to $a_x = a_y = 1$ and $b_x = b_y = 0$) and $\alpha_x = \alpha_y = \pi/2$ (leading to $a_x = a_y = 0$ and $b_x = b_y = 1$), but any other choice could be made as well, as long as it leads to four independent equations. In the case of free space propagation, we simply choose two different values of the propagation distance z, corresponding to two different values of b_x and b_y (with $a_x = a_y = 1$, of course).

Note that the two first-order optical systems can always be chosen such that they are isotropic, $a_x = a_y = a_i$, $b_x = b_y = b_i$, and so forth ($i = 1, 2$), with identical behavior in the x and the y direction.

4.2. Second-order moments

For the $3 + 4 + 3 = 10$ second-order moments, the following equations are relevant:

$$\mu^{\text{out}}_{2000} = a_x^2 \mu_{2000} + 2a_x b_x \mu_{1100} + b_x^2 \mu_{0200}, \tag{14}$$

$$\mu^{\text{out}}_{1010} = a_x a_y \mu_{1010} + a_x b_y \mu_{1001} + b_x a_y \mu_{0110} + b_x b_y \mu_{0101}, \tag{15}$$

$$\mu^{\text{out}}_{0020} = a_y^2 \mu_{0020} + 2a_y b_y \mu_{0011} + b_y^2 \mu_{0002}, \tag{16}$$

which equations correspond to (11) with $pqrs = 2000$, $pqrs = 1010$, and $pqrs = 0020$, respectively.

The three input moments μ_{2000}, μ_{1100}, and μ_{0200} can be determined by measuring the intensity moment μ^{out}_{2000} in the output planes of three systems with different values of a_x and b_x, see (14). Likewise, with the transversal coordinate x replaced by y, the three input moments μ_{0020}, μ_{0011}, and μ_{0002} can be determined by measuring the intensity moment μ^{out}_{0020}

in the output planes of three systems with different values of a_y and b_y, see (15). Note that we can choose $a_x = a_y = a_i$ and $b_x = b_y = b_i$ $(i = 1, 2, 3)$ for these three systems, in which case we are obviously using isotropic systems.

The other four input moments μ_{1010}, μ_{1001}, μ_{0110}, and μ_{0101} follow from measuring the intensity moment μ^{out}_{1010} in the output planes of four different systems, see (15). However, if we would use only isotropic systems, like we could do for (14) and (16), (15) would reduce to

$$\mu^{\text{out}}_{1010} = a^2\mu_{1010} + ab(\mu_{1001} + \mu_{0110}) + b^2\mu_{0101} \tag{17}$$

and we can only determine the combination $\mu_{1001} + \mu_{0110}$. Hence, while three systems may be isotropic again—and, for instance, be identical to the ones that we used when we were dealing with (14) and (16)—at least one system should be anamorphic.

We conclude that all ten second-order moments can be determined from the knowledge of the output intensities of four first-order optical systems, where one of them has to be anamorphic. In the case of fractional Fourier transform systems we could choose, for instance [24, 25], the fractional angles $\alpha_x = \alpha_y = 0$ (leading to $a_x = a_y = 1$ and $b_x = b_y = 0$), $\alpha_x = \alpha_y = \pi/4$ (leading to $a_x = a_y = b_x = b_y = \sqrt{2}/2$), $\alpha_x = \alpha_y = \pi/2$ (leading to $a_x = a_y = 0$ and $b_x = b_y = 1$), and the anamorphic combination $\alpha_x = \pi/2\pi$ and $\alpha_y = 0$ (leading to $a_x = b_y = 0$ and $a_y = b_x = 1$). If we decide to determine the moments using free space propagation, we should be aware of the fact that an anamorphic free space system cannot be realized by mere free space, but can only be simulated by using a proper arrangement of cylindrical lenses.

Of course, optical schemes to determine all ten second-order moments have been described before, see, for instance [8, 9, 11, 12, 13], but the way to determine these moments as presented in this paper is based on a general scheme that can also be used for the determination of arbitrary higher-order moments.

4.3. Higher-order moments

For higher-order moments we can proceed analogously. For the $4 + 6 + 6 + 4 = 20$ third-order moments, the following equations are relevant:

$$\mu^{\text{out}}_{3000} = a_x^3\mu_{3000} + 3a_x^2b_x\mu_{2100} + 3a_xb_x^2\mu_{1200} + b_x^3\mu_{0300}, \tag{18}$$

$$\begin{aligned}\mu^{\text{out}}_{2010} = {} & a_x^2a_y\mu_{2010} + a_x^2b_y\mu_{2001} + 2a_xb_xa_y\mu_{1110} \\ & + 2a_xb_xb_y\mu_{1101} + b_x^2a_y\mu_{0210} + b_x^2b_y\mu_{0201},\end{aligned} \tag{19}$$

$$\begin{aligned}\mu^{\text{out}}_{1020} = {} & a_xa_y^2\mu_{1020} + 2a_xa_yb_y\mu_{1011} + a_xb_y^2\mu_{1002} \\ & + b_xa_y^2\mu_{0120} + 2b_xa_yb_y\mu_{0111} + b_xb_y^2\mu_{0102},\end{aligned} \tag{20}$$

$$\mu^{\text{out}}_{0030} = a_y^3\mu_{0030} + 3a_y^2b_y\mu_{0021} + 3a_yb_y^2\mu_{0012} + b_y^3\mu_{0003}. \tag{21}$$

Note again that these equations correspond to (11) with $pqrs = 3000$, $pqrs = 2010$, $pqrs = 1020$, and $pqrs = 0030$, respectively. The 20 third-order moments can be determined from the knowledge of the output intensities of six first-order optical systems, where two of them have to be anamorphic. We consider in more detail how the third-order moments could be determined.

(i) The four input moments μ_{3000}, μ_{2100}, μ_{1200}, and μ_{0300} can be determined by measuring the intensity moment $\mu^{\text{out}}_{3000,i}$ $(i = 1, 2, 3, 4)$ in the output planes of four systems with different values of a_x and b_x, see (18). Likewise, with the transversal coordinate x replaced by y, the four input moments μ_{0030}, μ_{0021}, μ_{0012}, and μ_{0003} can be determined by measuring the intensity moment $\mu^{\text{out}}_{0030,i}$ $(i = 1, 2, 3, 4)$ in the output planes of four systems with different values of a_y and b_y, see (21). Note that we can choose $a_x = a_y = a_i$ and $b_x = b_y = b_i$ $(i = 1, 2, 3, 4)$ for these four different systems, in which case we are obviously using isotropic systems. This then leads to the set of four equations

$$\begin{aligned}& a_i^3\mu_{3000} + 3a_i^2b_i\mu_{2100} + 3a_ib_i^2\mu_{1200} + b_i^3\mu_{0300} \\ & \quad = \mu^{\text{out}}_{3000,i} \quad (i = 1, 2, 3, 4)\end{aligned} \tag{22}$$

based on (18) and a similar set of four equations

$$\begin{aligned}& a_i^3\mu_{0030} + 3a_i^2b_i\mu_{0021} + 3a_ib_i^2\mu_{0012} + b_i^3\mu_{0003} \\ & \quad = \mu^{\text{out}}_{0030,i} \quad (i = 1, 2, 3, 4),\end{aligned} \tag{23}$$

based on (21). Possible system choices are, for instance, four sections of free space, with $a_i = 1$ and b_i proportional to the four different propagation distances z_i $(i = 1, 2, 3, 4)$; or four isotropic fractional Fourier transform systems with $a_i = \cos\alpha_i$ and $b_i = \sin\alpha_i$, and α_i $(i = 1, 2, 3, 4)$ four different fractional angles.

(ii) Using the same four isotropic systems as above, the two input moments μ_{2010} and μ_{0201}, together with the two moment combinations $\mu_{2001} + 2\mu_{1110}$ and $2\mu_{1101} + \mu_{0210}$, follow from measuring the intensity moment $\mu^{\text{out}}_{2010,i}$ $(i = 1, 2, 3, 4)$ in the output planes of these four systems, see (19), while the two input moments μ_{1020} and μ_{0102}, together with the two moment combinations $2\mu_{1011} + \mu_{0120}$ and $\mu_{1002} + 2\mu_{0111}$, follow from measuring the intensity moment $\mu^{\text{out}}_{1020,i}$ $(i = 1, 2, 3, 4)$, see (20). This leads to the set of four equations

$$\begin{aligned}& a_i^3\mu_{2010} + a_i^2b_i(\mu_{2001} + 2\mu_{1110}) + a_ib_i^2(2\mu_{1101} + \mu_{0210}) \\ & \quad + b_i^3\mu_{0201} = \mu^{\text{out}}_{2010,i} \quad (i = 1, 2, 3, 4)\end{aligned} \tag{24}$$

based on (19) and a similar set of four equations

$$\begin{aligned}& a_i^3\mu_{1020} + a_i^2b_i(2\mu_{1011} + \mu_{0120}) + a_ib_i^2(\mu_{1002} + 2\mu_{0111}) \\ & \quad + b_i^3\mu_{0102} = \mu^{\text{out}}_{1020,i} \quad (i = 1, 2, 3, 4)\end{aligned} \tag{25}$$

based on (20).

(iii) Twelve of the 20 input moments (together with four moment combinations) can thus be determined by using four isotropic systems. To determine the remaining eight moments, we need four more equations based on (19) and (20), for which we have to use two more systems (labeled $i = 5$ and $i = 6$), which should now

Table 1: The number of nth-order moments N, and the required number of first-order optical systems to determine these N moments, given as a function of n.

n	Number of nth-order moments	N	N_t	N_a
0	1	1	1	0
1	2 + 2	4	2	0
2	3 + 4 + 3	10	4	1
3	4 + 6 + 6 + 4	20	6	2
4	5 + 8 + 9 + 8 + 5	35	9	4
5	6 + 10 + 12 + 12 + 10 + 6	56	12	6
6	7 + 12 + 15 + 16 + 15 + 12 + 7	84	16	9
⋮	⋮	⋮	⋮	⋮

be anamorphic. Among the many possibilities, an easy choice would be a system with $a_x = b_y = 0$, $b_x \neq 0$, $a_y \neq 0$, leading to

$$\begin{aligned} b_x^2 a_y \mu_{0210} &= \mu_{2010,5}^{\text{out}}, \\ b_x a_y^2 \mu_{0120} &= \mu_{1020,5}^{\text{out}}, \end{aligned} \tag{26}$$

and a system with $b_x = a_y = 0$, $a_x \neq 0$, $b_y \neq 0$, leading to

$$\begin{aligned} a_x^2 b_y \mu_{2001} &= \mu_{2010,6}^{\text{out}}, \\ a_x b_y^2 \mu_{1002} &= \mu_{1020,6}^{\text{out}}. \end{aligned} \tag{27}$$

The former system may be an anamorphic fractional Fourier transform system with fractional angles $\alpha_x = \pi/2$ and $\alpha_y = 0$ (and hence $a_x = b_y = 0$ and $b_x = a_y = 1$), while the latter may be an anamorphic fractional Fourier transform system with $\alpha_x = 0$ and $\alpha_y = \pi/2$ (and hence $b_x = a_y = 0$ and $b_y = a_x = 1$).

Altogether we have thus constructed 20 equations for the 20 third-order moments, using a total of six first-order systems: four isotropic systems where we measure the 16 output intensity moments $\mu_{3000,i}$, $\mu_{0030,i}$, $\mu_{2010,i}$, and $\mu_{1020,i}$ ($i = 1, 2, 3, 4$), and two anamorphic systems where we measure the four output intensity moments $\mu_{2010,i}^{\text{out}}$ and $\mu_{1020,i}^{\text{out}}$ ($i = 5, 6$).

For the $5 + 8 + 9 + 8 + 5 = 35$ fourth-order moments, the relevant equations follow from (11) with $pqrs = 4000$, $pqrs = 3010$, $pqrs = 2020$, $pqrs = 1030$, and $pqrs = 0040$, respectively. The 35 fourth-order moments can be determined from the knowledge of the output intensities of nine first-order optical systems spectra, where four of them have to be anamorphic. Constructing a measuring scheme along the lines described above for the second-order case and the third-order case, is rather straightforward.

To find the number of nth-order moments N, and the total number of first-order optical systems N_t (with N_a the number of anamorphic ones) that we need to determine these N moments, use can be made of the triangle presented in Table 1, which can easily be extended to higher order.

Note that N (the number of nth-order moments) is equal to the sum of the values in the nth row of the triangle, $N = (n+1)(n+2)(n+3)/6$; that N_t (the total number of first-order optical systems) is equal to the highest value that appears in the nth row of the triangle, $N_t = (n+2)^2/4$ for n = even, and $N_t = (n+3)(n+1)/4$ for n = odd; that the number of isotropic systems is $n + 1$; and that N_a (the number of anamorphic systems) follows from $N_a = N_t - (n + 1)$.

5. CONCLUSIONS

We have shown how all global WD moments of arbitrary order can be measured as intensity moments in the output planes of an appropriate number of first-order optical systems (separable, but generally anamorphic ones), and we have derived the minimum number of such systems that are needed for the determination of these moments. The results followed directly from the general relationship (11) that expresses the intensity moments in the output plane of a separable first-order optical system in terms of the moments in the input plane and the system parameters a_x, b_x, c_x, d_x and a_y, b_y, c_y, d_y.

REFERENCES

[1] E. Wigner, "On the quantum correction for thermodynamic equilibrium," *Physical Review*, vol. 40, no. 5, pp. 749–759, 1932.

[2] M. J. Bastiaans, "The Wigner distribution function applied to optical signals and systems," *Optics Communications*, vol. 25, no. 1, pp. 26–30, 1978.

[3] M. J. Bastiaans, "Wigner distribution function and its application to first-order optics," *Journal of the Optical Society of America*, vol. 69, no. 12, pp. 1710–1716, 1979.

[4] M. J. Bastiaans, "Application of the Wigner distribution function to partially coherent light," *Journal of the Optical Society of America* {*A*}, vol. 3, no. 8, pp. 1227–1238, 1986.

[5] W. Mecklenbräuker and F. Hlawatsch, Eds., *The Wigner Distribution – Theory and Applications in Signal Processing*, Elsevier, Amsterdam, The Netherlands, 1997.

[6] International Organization for Standardization, Technical Committee/Subcommittee 172/SC9, "Lasers and laser-related equipment—test methods for laser beam parameters—beam widths, divergence angle and beam propagation factor," ISO Doc. 11146: 1999, International Organization for Standardization, Geneva, Switzerland, 1999.

[7] R. K. Luneburg, *Mathematical Theory of Optics*, University of California Press, Berkeley and Los Angeles, Calif, USA, 1966.

[8] G. Nemes and A. E. Siegman, "Measurement of all ten second-order moments of an astigmatic beam by the use of rotating simple astigmatic (anamorphic) optics," *Journal of the Optical Society of America* {*A*}, vol. 11, no. 8, pp. 2257–2264, 1994.

[9] B. Eppich, C. Gao, and H. Weber, "Determination of the ten second order intensity moments," *Optics & Laser Technology*, vol. 30, no. 5, pp. 337–340, 1998.

[10] R. Simon, N. Mukunda, and E. C. G. Sudarshan, "Partially coherent beams and a generalized ABCD-law," *Optics Communications*, vol. 65, no. 5, pp. 322–328, 1988.

[11] C. Martínez, F. Encinas-Sanz, J. Serna, P. M. Mejías, and R. Martínez-Herrero, "On the parametric characterization of the transversal spatial structure of laser pulses," *Optics Communications*, vol. 139, no. 4–6, pp. 299–305, 1997.

[12] J. Serna, R. Martínez-Herrero, and P. M. Mejías, "Parametric characterization of general partially coherent beams propagating through ABCD optical systems," *Journal of the Optical Society of America* {*A*}, vol. 8, no. 7, pp. 1094–1098, 1991.

[13] J. Serna, F. Encinas-Sanz, and G. Nemes, "Complete spatial characterization of a pulsed doughnut-type beam by use of spherical optics and a cylindrical lens," *Journal of the Optical Society of America A*, vol. 18, no. 7, pp. 1726–1733, 2001.

[14] E. Wolf, "A macroscopic theory of interference and diffraction of light from finite sources. I. Fields with a narrow spectral range," *Proceedings of the Royal Society of London. Series A*, vol. 225, pp. 96–111, 1954.

[15] E. Wolf, "A macroscopic theory of interference and diffraction of light from finite sources. II. Fields with a spectral range of arbitrary width," *Proceedings of the Royal Society of London. Series A*, vol. 230, pp. 246–265, 1955.

[16] A. Papoulis, *Systems and Transforms with Applications in Optics*, McGraw-Hill, New York, NY, USA, 1968.

[17] M. J. Bastiaans, "A frequency-domain treatment of partial coherence," *Optica Acta*, vol. 24, no. 3, pp. 261–274, 1977.

[18] L. Mandel and E. Wolf, "Spectral coherence and the concept of cross-spectral purity," *Journal of the Optical Society of America*, vol. 66, no. 6, pp. 529–535, 1976.

[19] A. Walther, "Radiometry and coherence," *Journal of the Optical Society of America*, vol. 58, no. 9, pp. 1256–1259, 1968.

[20] A. Walther, "Propagation of the generalized radiance through lenses," *Journal of the Optical Society of America*, vol. 68, no. 11, pp. 1606–1610, 1978.

[21] D. Dragoman, "Applications of the Wigner distribution function in signal processing," *EURASIP Journal on Applied Signal Processing*, vol. 2005, no. 10, pp. 1520–1534, 2005.

[22] K. B. Wolf, *Integral Transforms in Science and Engineering*, chapter 9, Plenum Press, New York, NY, USA, 1979.

[23] A. W. Lohmann, "Image rotation, Wigner rotation, and the fractional Fourier transform," *Journal of the Optical Society of America* {*A*}, vol. 10, no. 10, pp. 2181–2186, 1993.

[24] M. J. Bastiaans and T. Alieva, "Wigner distribution moments measured as fractional Fourier transform intensity moments," in *Proc. 19th Congress of the International Commission for Optics, Optics for the Quality of Life (ICO-19)*, A. Consortini and G. C. Righini, Eds., vol. 4829 of *Proceedings of SPIE*, pp. 245–246, Firenze, Italy, August 2002.

[25] M. J. Bastiaans and T. Alieva, "Wigner distribution moments in fractional Fourier transform systems," *Journal of the Optical Society of America* {*A*}, vol. 19, no. 9, pp. 1763–1773, 2002.

Martin J. Bastiaans was born in Helmond, The Netherlands, in 1947. He received the M.S. degree in electrical engineering (with honours) and the Ph.D. degree in technical sciences from the Technische Universiteit Eindhoven (Eindhoven University of Technology), Eindhoven, The Netherlands, in 1969 and 1983, respectively. In 1969 he became an Assistant Professor and since 1985 he has been an Associate Professor with the Department of Electrical Engineering, Technische Universiteit Eindhoven, in the Signal Processing Systems Group, where he teaches electrical circuit analysis, signal theory, digital signal processing, and Fourier optics and holography. His main current research interest is in describing signals by means of a local frequency spectrum (for instance, the Wigner distribution function, the windowed Fourier transform, Gabor's signal expansion, etc.). He is the author and coauthor of more than 140 papers in international scientific journals and proceedings of scientific conferences. Dr. Bastiaans is a Fellow of the Optical Society of America and a Senior Member of the Institute of Electrical and Electronics Engineers.

Tatiana Alieva was born in Moscow, Russia. In 1983 and 1996 she received the M.S. and Ph.D. degrees (with honours) in physics from M. V. Lomonosov Moscow State University and University Autonoma of Madrid, respectively. From 1983 till 1999, she worked for A. L. Mintz Radiotechnical Institute, Academy of Science, Moscow. She spent several years as a Postdoctoral Fellow at the Departments of Electrical Engineering of Catholic University of Leuven, Belgium, and Eindhoven University of Technology, The Netherlands. She is currently a Researcher at the Physical Faculty of Complutense University of Madrid. She has worked in diverse fields of physics, such as theory of high energetic beams channelling in crystals, and radio waves scattering on the anisotropic formations in the ionosphere. Her current research interests are in optical information processing, fractal signal analysis, theory of fractional integral transforms, phase retrieval, and characterization of optical fields by means of phase-space distributions. She is an author of more than eighty scientific publications.

EURASIP Journal on Applied Signal Processing 2005:10, 1541–1553

Concepts for the Temporal Characterization of Short Optical Pulses

Christophe Dorrer
Bell Laboratories, Lucent Technologies, 791 Keyport-Holmdel Road, Holmdel, NJ 07733, USA
Email: dorrer@lucent.com

Ian A. Walmsley
Clarendon Laboratory, Parks Road, Oxford OX1 3PU, United Kingdom
Email: walmsley@physics.ox.ac.uk

Received 1 April 2004; Revised 15 September 2004

Methods for the characterization of the time-dependent electric field of short optical pulses are reviewed. The representation of these pulses in terms of correlation functions and time-frequency distributions is discussed, and the strategies for their characterization are explained using these representations. Examples of the experimental implementations of the concepts of spectrography, interferometry, and tomography for the characterization of pulses in the optical telecommunications environment are presented.

Keywords and phrases: optical pulse characterization, time-frequency distribution, Wigner function, spectrogram, tomography, interferometry.

1. INTRODUCTION

Ultrashort optical pulses are used in areas of science and engineering as diverse as spectroscopy, medical research, plasma physics, quantum optics, and optical telecommunications. In optical telecommunications, information is encoded in the amplitude and/or phase of an optical wave [1]. While information encoding in digital telecommunications is based on a finite number of values of a physical quantity (e.g., the presence or absence of energy in a given bit slot), the ability to measure in detail the waveform of the optical wave itself is crucial for optimizing the properties of the systems that generate the signal, and understanding the linear and nonlinear properties of the systems through which the pulses propagate. This information is critical in developing strategies to overcome the current limitations of current optical networks. For example, dispersion management compensates for the chromatic dispersion induced by linear propagation and can also be used to mitigate nonlinear effects. Similarly, the phase distortion imposed on a pulse by the modulators used in carving the pulse out of a cw- or quasi-cw source can impact the propagation of the pulse. Finally, measurements of the electric field can be used to characterize the linear or nonlinear properties of a device. There are various approaches for temporal waveform measurements. We only consider here techniques that provide self-referencing characterization of an unknown pulse or a train of unknown but identical pulses, that is, that do not use a well-characterized pulse as a reference. While test-plus-reference techniques, such as spectral interferometry [2, 3, 4], can be easier to implement in some cases, they require a well-characterized reference pulse mutually coherent with the pulse under test. Although this can be difficult to achieve over long distances, they have been used to characterize pulses in the telecommunication environment [5]. We will not deal either with sampling techniques. These techniques can provide samples of the temporal intensity of a source by fast photodetection and electronics or by nonlinear interaction with a short sampling pulse (nonlinear optical sampling) [6]. The technique of linear optical sampling [7] is also sensitive to the electric field of the source (i.e., it can measure samples of the intensity and phase of the source under test), and can therefore be used to measure constellation diagrams [8]. These techniques are particularly useful when dealing with a data-encoded optical source because of the randomness of the data stream, but they constitute a class of their own that is beyond the scope of this paper. The concepts presented here apply generally to the characterization of the temporal electric field of short optical pulses, although the details of experimental implementations are strongly dependent upon the domain of application.

Since ultrafast optical pulses are faster than the fastest electronic detection devices, they present a considerable challenge for measurement. An important additional limitation is that photodetectors respond to the intensity of the pulse,

so their output contains no information about the phase of the incident radiation. To overcome these limitations, a combination of ancillary filters can be used. The data are simply the photocurrent recorded by a time-integrating detector as a function of the parameters of the filters. These might be, for example, the passband frequency for a spectrometer (a time-stationary linear filter), the modulation index for a phase modulator (a time-nonstationary linear filter), or the relative delay between the pulse under test and the modulation induced by an electroabsorption modulator (also a time-nonstationary linear filter).

There are a number of quite general strategies for characterizing the electric field of an optical pulse using such filters. These belong to one of three categories: spectrographic, tomographic, or interferometric. The categories are distinguished by the procedure required for reconstructing the amplitude and the phase of the field from the recorded data [9, 10]. Since the analytic signal of the field is a complex function of one real variable, time, with finite support, the data must contain a finite set of complex numbers sampling the field at a finite set of time points, or equivalently a finite set of frequency points. It can sometimes be fruitful, however, to reconstruct pulses by sampling a time-frequency representation of the pulse (or its equivalent correlation function). In this case a two-dimensional set of data is obtained, from which an inversion algorithm reconstructs an estimate of the field. This is typically the case for spectrography, which makes use of a time-frequency distribution, and requires sophisticated iterative data inversion algorithms to reconstruct the field. Tomography also requires a large data set, in the form of a large number of modulated pulse spectra, but the inversion is direct (noniterative). Interferometry, in contrast, measures only a one-dimensional data set and uses direct data inversion to reconstruct the field.

In this paper, we provide examples of each of these methods that are relevant to optical telecommunications. In Section 2, we first discuss the representation of the electric field of a pulse or train of pulses, and how the various measurement techniques sample the field. Then, in Section 3, we illustrate both the data acquisition and inversion for each method.

2. REPRESENTING LIGHT PULSES

The fundamental quantity describing an isolated, individual pulse of light is the real electric field. This is a function of time and space, or equivalently frequency and wavevector. The spatial dependence of the field is often assumed uniform. This assumption is valid for optical fiber-based telecommunications if the field occupies the lowest-order mode of the fiber. Therefore in this paper, we concentrate on determining the time dependence of the electric field.

In practice, it is often difficult to characterize a single pulse, and one deals with a train of pulses instead of a single pulse. One must be careful in specifying a field for such an ensemble. If all of the pulses in the train are identical, the ensemble is deemed coherent, and the underlying electric field of an individual remains the quantity of interest. If, on the other hand, the electric field is stochastic, fluctuating from pulse to pulse, the ensemble is said to be partially coherent. When this is the case, the amplitude and phase of the electric field of an individual pulse brings little information on the train of pulses, and pulse characterization involves measurements of the statistical properties of the ensemble, for example via the two-time or two-frequency correlation function. We note that a data-modulated train of pulses is not coherent if the data modulation is random (which is the case for a deployed communication system).

2.1. Describing an optical pulse by its analytic signal

The real electric field, $\varepsilon(t)$, underlying an optical pulse is twice the real part of its analytic signal $E(t)$: $\varepsilon(t) = 2 \times \mathrm{Re}[E(t)]$. The analytic signal is the single-sided inverse Fourier transform of the Fourier transform of the field,

$$E(t) = \frac{1}{\sqrt{2\pi}} \int_0^\infty d\omega\, \tilde{\varepsilon}(\omega) \exp[-i\omega t], \tag{1}$$

where

$$\tilde{\varepsilon}(\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^\infty dt\, \varepsilon(t) \exp[i\omega t]. \tag{2}$$

The electric field is considered to have compact support in the time domain, and is further assumed to have no spectral component at $\omega = 0$ so $\tilde{\varepsilon}(0) = 0$ (since a pulse propagating in a charge-free region of space has no dc spectral component, the electric field must have zero area). The analytic signal is complex and therefore can be expressed uniquely in terms of an amplitude and phase:

$$E(t) = |E(t)| \exp[i\phi_t(t)] \exp[i\phi_0] \exp[-i\omega_0 t], \tag{3}$$

where $|E(t)|$ is the time-dependent envelope, ω_0 is the carrier frequency (usually chosen near the center of the pulse spectrum), $\phi_t(t)$ is the time-dependent phase, and ϕ_0 a constant. The square of the envelope, $|E(t)|^2$, is the time-dependent intensity of the pulse which could be measured by a detector of sufficient bandwidth. The time-dependent phase accounts for the occurrence of different frequencies at different times, and the instantaneous frequency is usually defined as $-\partial\phi_t/\partial t$. As an example, Figure 1 shows the temporal intensity and phase of a pair of chirped Gaussian pulses. The time-dependent phase structure is indicative of a variation of the instantaneous frequency across the pulse.

The frequency representation of the analytic signal is the Fourier transform of $E(t)$,

$$\begin{aligned} \tilde{E}(\omega) &= |\tilde{E}(\omega)| \cdot \exp[i\phi_\omega(\omega)] \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^\infty dt\, E(t) \exp[i\omega t] \\ &= \begin{cases} \tilde{\varepsilon}(\omega), & \omega > 0, \\ 0, & \omega \le 0. \end{cases} \end{aligned} \tag{4}$$

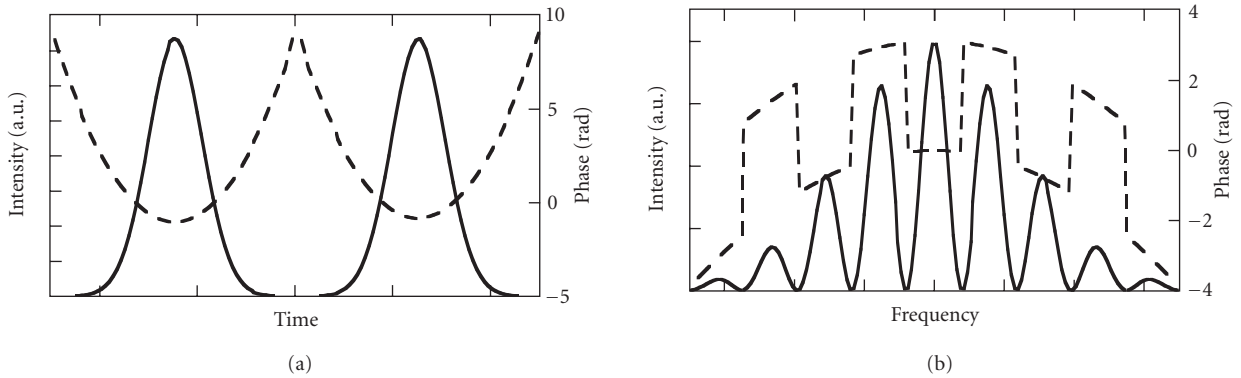


Figure 1: Intensity and phase (resp., continuous and dashed line) of a pair of chirped Gaussian pulses in (a) the temporal domain and (b) the spectral domain.

Here $|\widetilde{E}(\omega)|$ is the spectral amplitude and $\phi_\omega(\omega)$ is the spectral phase. The square of the spectral amplitude, $|\widetilde{E}(\omega)|^2$, is the spectral intensity. Strictly speaking this quantity is the spectral flux—the quantity measured in the familiar way by means of a spectrometer followed by a photodetector. The spectral phase describes the relative phases between each of the frequencies, and the group delay for frequency ω is usually defined as $\partial\phi_\omega/\partial\omega$.

The sampling requirement for the reconstruction of the electric field is given by the Whittaker-Shannon theorem [11], which asserts that if the field has compact support in the time domain over a range Δt, then a sampling of $\widetilde{E}(\omega)$ at the Nyquist frequency rate of $2\pi/\Delta t$ is sufficient for reconstructing the analytic signal $E(t)$ and consequently the electric field $\varepsilon(t)$ exactly.

Figure 1 shows the spectral intensity and spectral phase of the chirped pair of Gaussian pulses. The spectral fringes have a period of the inverse of the temporal separation of the pair of pulses. While the spectrum can reveal some properties of the waveform, both the spectral intensity and phase must be measured to fully characterize the electric field.

2.2. The two-time correlation function and its phase-space representations

If a measurement relies on averaging the detected signal over a train of pulses, then it is necessary to define the properties of the pulses in a different way. Although it is formally quite difficult to formulate rigorously even the simplest of concepts, such as the spectrum [12], for a nonstationary field, a simple-minded approach can be fruitful. If each pulse in the train is an independent realization of a stochastic ensemble, then the time average is equivalent to an ensemble average by definition. This enables the coherence of the train to be defined operationally in a reasonable way. It is important, though, to realize that the electric field amplitude and phase of an individual pulse does not bring significant information about the train of pulses and pulse characterization efforts must ultimately be directed toward measurement of the ensemble statistics.

The simplest quantity that quantifies the statistical properties of the ensemble is the nonstationary two-time field correlation function

$$C(t_1, t_2) = \langle E(t_1)E^*(t_2)\rangle, \tag{5}$$

where the angle brackets indicate an average over the ensemble of pulses, each of the electric fields being defined with respect to a local time frame. With this definition of the ensemble, we do not need to adopt procedures along the lines of those developed by Wiener and Khintchine [13] to define the correlation function.

$C(t_1, t_2)$ provides a quantitative description of fluctuations from pulse to pulse in the electric field at times t_1 relative to those at times t_2. This is a complete description of the pulse ensemble so long as the fluctuations obey normal (or Gaussian) statistics. If not, then it is the simplest of a hierarchy of multitime correlation functions defining the ensemble. Furthermore, for a train of identical pulses, $C(t_1, t_2)$ factorizes into $E(t_1)E^*(t_2)$ and the electric field amplitude and phase are readily obtained.

It is frequently useful to work with a variation of the correlation function that uses a two-dimensional space of time and frequency—the chronocyclic phase-space. The intuitive concept of chirp (i.e., time-dependent frequency in the pulse) can be most easily seen within this space. The relationships between the two-time correlation function and the chronocyclic (time/frequency) and frequency-domain representations of the ensemble may be derived by rewriting (5) in terms of a center-time coordinate, t, and a difference-time coordinate, Δt:

$$C(t, \Delta t) = C(t_1, t_2), \tag{6}$$

where $t = (t_1 + t_2)/2$ and $\Delta t = t_1 - t_2$. The two-frequency field correlation function is obtained by taking the two-dimensional Fourier transform of the two-time correlation function

$$\widetilde{\widetilde{C}}(\Delta\omega, \omega) = \frac{1}{2\pi}\iint dt\, d\Delta t\, C(t, \Delta t)\exp\left[i(t\Delta\omega + \Delta t\omega)\right]. \tag{7}$$

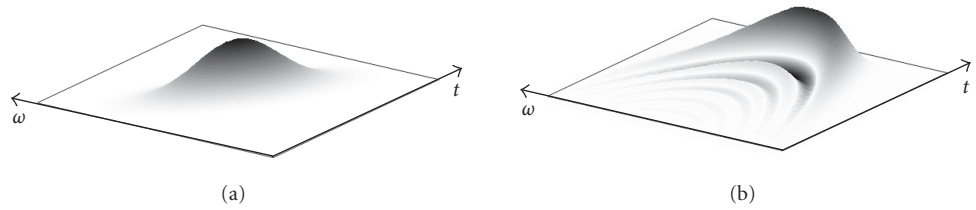


Figure 2: Wigner function of a pulse with (a) linear chirp and (b) quadratic chirp.

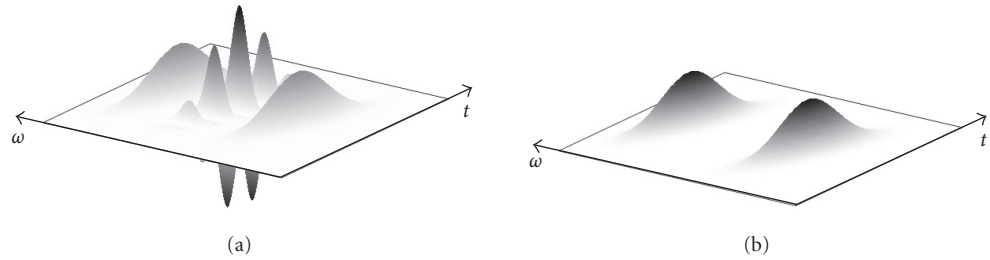


Figure 3: Wigner function of the pair of chirped Gaussian pulses when the two pulses are (a) mutually coherent and (b) mutually incoherent.

The center-frequency and difference-frequency coordinates in (7) are given by $\omega = (\omega_1+\omega_2)/2$ and $\Delta\omega = \omega_1-\omega_2$, respectively. The pulse ensemble may also be represented within the chronocyclic phase spaces defined by the complimentary variables t, ω and $\Delta\omega, \Delta t$. The chronocyclic Wigner function, $W(t,\omega)$, and ambiguity or Wigner characteristic function, $A(\Delta\omega, \Delta t)$, provide two particularly useful descriptions of the pulse train statistics in these spaces. The relationship between the various representations of the correlation function has been discussed in the context of spatially localized fields in [14], and in the context of signal analysis in [15]. Chronocyclic phase-space distributions have found increasing application in ultrafast optics. The Wigner function has been described in [16, 17, 18], and the Page distribution in [19]. Another distribution of interest is the ambiguity function that is used in radar technology [20]. These functions are also used in many areas of physics and engineering, and their relations and properties are discussed in [21]. The Wigner function is obtained by taking the one-dimensional Fourier transform of $C(t,\Delta t)$ over the time-difference coordinate

$$W(t,\omega) = \frac{1}{\sqrt{2\pi}} \int d\Delta t\, C(t,\Delta t) \exp[i\omega\Delta t], \tag{8}$$

whereas the ambiguity function is obtained from $C(t,\Delta t)$ by performing the Fourier transform over the average-time coordinate

$$A(\Delta\omega,\Delta t) = \frac{1}{\sqrt{2\pi}} \int dt\, C(t,\Delta t) \exp[i\Delta\omega t]. \tag{9}$$

These representations are uniquely related to one another by Fourier transformations. The Wigner function has many interesting properties, for example, the ability to represent a chirp, as can be seen in Figure 2a. It is a real, but not necessarily positive function, which complicates its interpretation as a density function in the time-frequency space. For example, the Wigner function of a Gaussian pulse with a quadratic chirp (i.e., a third-order spectral phase), as displayed in Figure 2b, is negative over a significant portion of the chronocyclic space. An example of the Wigner function of the pair of chirped Gaussian pulses in Figure 1 is shown in Figure 3a. The side lobes are the phase-space representations of the individual pulses, and the cross-terms between the two fields lead to a central feature with fringes that indicate that the pulses have a definite phase relation to one another—they are coherent. The Wigner function for an incoherent pair of pulses is also show in Figure 3b. In this case the ensemble has a random phase between the pair, which causes a washing out of the interference pattern when the signal is averaged over the ensemble. The coherence can be quantified using the Wigner function, as explained in the next section.

Information regarding the shapes of ultrafast optical pulses is generally inferred from the output of a square-law detector after some filtering. Therefore it is of practical interest to establish the relationship between measurable detector output and the various descriptions of the pulse ensemble. The average pulse time-dependent intensity is obtained from the two-time correlation function by setting $\Delta t = 0$. Alternatively, it is a projection of the Wigner function onto the frequency axis, or the Fourier transform of the $\Delta t = 0$ line of the ambiguity function

$$\begin{aligned} I(t) &= C(t,0) \\ &= \int d\omega\, W(t,\omega) \\ &= \int d\Delta\omega\, A(\Delta\omega,0) \exp[-i\Delta\omega t]. \end{aligned} \tag{10}$$

Furthermore, the average pulse spectral intensity is obtained from the two-frequency correlation function by setting $\Delta\omega = 0$, or by projecting the Wigner function onto the frequency axis, or by taking the Fourier transform of the $\Delta\omega = 0$ line of the ambiguity function:

$$\begin{aligned} I(\omega) &= \tilde{\tilde{C}}(0,\omega) \\ &= \int dt\, W(t,\omega) \\ &= \int d\Delta t\, A(0,\Delta t)\exp[-i\Delta t\omega]. \end{aligned} \tag{11}$$

It is important to recognize that the various time-frequency distributions and their relations are central to the characterization of pulses in the optical domain, since they are simply related to the measured data. In optics, direct measurement of the waveform is not possible. This is in contrast to the more usual application of these distributions in signal processing, where they are commonly used as mathematical tools for signal representation, for example, to track the presence of various instantaneous frequencies in a known (measured) sound waveform.

2.3. The integral degree of coherence: when is a pulse field useful?

Since a field amplitude and phase may be defined in a unique way only for identical pulses, it is important to quantify the degree to which the pulses in the ensemble are alike. A useful quantity for this purpose is the integral degree of coherence, μ. μ is readily derived from the time-domain analogue of Born and Wolf's [13] degree of coherence $\gamma(t+\Delta t/2, t-\Delta t/2)$, defined as

$$\gamma\left(t+\frac{\Delta t}{2}, t-\frac{\Delta t}{2}\right) = \frac{C(t,\Delta t)}{\left[C(t+\Delta t/2,0)C(t-\Delta t/2,0)\right]^{1/2}}. \tag{12}$$

Using the Schwarz inequality, it is straightforward to show that $0 \le |\gamma(t+\Delta t/2, t-\Delta t/2)| \le 1$. Consider the n-dimensional vectors $\vec{a}$ and $\vec{b}$, with components $\{a_i\}$ and $\{b_i\}$ where $i \in (1,n)$. Then

$$0 \le \left|\sum_{i=1}^{n} a_i^* b_i\right|^2 \le \left(\sum_{i=1}^{n} a_i^* a_i\right)\left(\sum_{i=1}^{n} b_i^* b_i\right). \tag{13}$$

Allowing $a_i = E_i(t-\Delta t/2)/\sqrt{n}$ and $b_i = E_i(t+\Delta t/2)/\sqrt{n}$, that is, the ith realization of the field amplitude at times $t-\Delta t/2$ and $t+\Delta t/2$, respectively, it is clear that

$$\begin{aligned} 0 \le & \left|\frac{1}{n}\sum_{i=1}^{n} E_i^*\left(t-\frac{\Delta t}{2}\right)E_i\left(t+\frac{\Delta t}{2}\right)\right|^2 \\ \le & \left(\frac{1}{n}\sum_{i=1}^{n} E_i^*\left(t+\frac{\Delta t}{2}\right)E_i\left(t+\frac{\Delta t}{2}\right)\right) \\ & \times\left(\frac{1}{n}\sum_{i=1}^{n} E_i^*\left(t-\frac{\Delta t}{2}\right)E_i\left(t-\frac{\Delta t}{2}\right)\right). \end{aligned} \tag{14}$$

In the limit that $n \to \infty$, the summations in (14) lead to an average over the ensemble and the inequality simplifies to

$$0 \le |C(t,\Delta t)|^2 \le C\left(t+\frac{\Delta t}{2},0\right)C\left(t-\frac{\Delta t}{2},0\right). \tag{15}$$

The upper and lower bounds on the degree of coherence follow from (15). However, it is difficult to determine $\gamma(t+\Delta t/2, t-\Delta t/2)$ experimentally since it becomes singular for times at which $C(t,\Delta t)$ is zero. A practically more useful definition is offered by integrating equation (15) over the entire t, Δt space, and dividing by the quantity on the right-hand side, leading to the integral degree of coherence, μ,

$$0 \le \mu = \frac{\iint dt\, d\Delta t\, |C(t,\Delta t)|^2}{\left[\int dt\, C(t,0)\right]^2} \le 1. \tag{16}$$

Equivalent relations follow for the frequency domain and chronocyclic representations, for example, in the case of the ambiguity function

$$\mu = \frac{\iint d\Delta t\, d\Delta\omega\, |A(\Delta\omega,\Delta t)|^2}{\left[A(0,0)\right]^2}. \tag{17}$$

An integral degree of coherence less than one corresponds to a partially coherent train in which the pulse shape and/or phase fluctuate, in which case $C(t,\Delta t)$ is the fundamental quantity of interest. When $\mu = 1$ the ensemble is said to be fully coherent (the pulses in the ensemble are identical within a constant phase factor) and $C(t,\Delta t)$ factorizes. In the latter case the electric field becomes the fundamental quantity of interest and is readily retrieved from the two-time correlation function using

$$|E(t)| = \sqrt{C(t,0)}, \tag{18}$$

and with t_2 held fixed,

$$\mathrm{Arg}\left[E(t)\right] = \tan^{-1}\left[\frac{\mathrm{Im}\,C((t+t_2)/2, t-t_2)}{\mathrm{Re}\,C((t+t_2)/2, t-t_2)}\right] + \phi_0, \tag{19}$$

where ϕ_0 is an undetermined constant. For example, the integral degree of coherence corresponding to the Wigner function in Figure 3a is equal to one, while that of the Wigner function in Figure 3b is equal to 0.5. It is important to note that (18) and (19) are valid only if the integral degree of coherence has been explicitly demonstrated to be equal to unity, which of course requires that the two-time correlation function or equivalent representation in frequency or phase space be measured. Thus in cases where an ensemble or train of pulses, rather than an individual pulse, is used for application or experimentation, pulse-shape characterization efforts must ultimately be directed toward measurement of the ensemble statistics.

For a coherent train of pulses, (18) and (19) reveal that the electric field is retrieved from a single line of the correlation function. Hence, if the ensemble is assumed a priori to be coherent, the amount of collected data can be greatly reduced. This is a luxury afforded only to measurement techniques that directly measure one of the correlation functions.

2.4. Measurement strategies

The electric field of optical pulses can be characterized using various strategies, derived from, and with implications for, other measurement problems, such as wavefront diagnosis and quantum state reconstruction. These strategies can be organized into phase-space techniques, that is, techniques that attempt to measure either the Wigner or ambiguity function by exploring the entire two-dimensional chronocyclic phase space, and direct techniques, that obtain the electric field of a coherent field from a single slice of a second-order correlation function.

The minimal requirement for the complete exploration of the chronocyclic space required in phase-space techniques is the presence of two filters. The analysis details of phase-space techniques are found in [9]. There are two subclasses of phase-space techniques; those that make simultaneous measurements of the complementary variables in an attempt to reconstruct one of the phase-space distributions, and those that record marginals of the Wigner function after rotation in the phase space, from which the Wigner function can be obtained. The former method is known as spectrographic while the latter is referred to as tomographic. Spectrography is discussed in Sections 3.1 and 3.2, and tomography is discussed in Sections 3.5 and 3.6.

In contrast, direct techniques do not require this complete exploration of the phase space occupied by the correlation function. This is a significant advantage of direct techniques compared to phase-space techniques. Moreover, if the pulse train is assumed a priori to consist of identical pulses, as is most always assumed in reconstructing pulses from spectrographic or tomographic data, only one slice of the correlation function is required to obtain the amplitude and phase of the electric field [10]. Such slices are precisely what is measured in interferometry. This is usually achieved by mixing the field under test with a modified version of itself, or more generally by mixing two modified versions of the field under test. Thus, while phase-space techniques must explore the entire chronocyclic space even when the electric field is the fundamental quantity of interest, direct techniques need only return a single slice of the correlation function in order to construct the simpler quantity. Roughly speaking, if one wishes to reconstruct the field at N time points, then at least $2N$ independent data points are required. While direct techniques are capable of reconstructing the field by recording only the necessary $2N$ points, phase-space techniques require the measurement of N^2 points. Of course, an overcomplete data set is available from direct measurement of the entire correlation function as well. Direct interferometric techniques are discussed in Sections 3.3 and 3.4.

3. SELF-REFERENCING TEMPORAL CHARACTERIZATION OF SHORT OPTICAL PULSES

From a practical point of view, one first measures an experimental trace (e.g., the current from a photodiode as a function of various parameters of the experimental setup, e.g., the central frequency of a passband spectral filter), then applies a set of mathematical operations to the measured data in order to reconstruct the electric field. The design of the experimental setup and the type of experimental trace determine the recovery algorithm, and more generally the possibility of such recovery. In this section we take a closer look at the concepts of spectrography, tomography, and interferometry, and examples of experimental implementations and results are given.

3.1. Spectrography

Spectrographic techniques make use of two sequential filters, one time-stationary (spectral filter) and one time-nonstationary (time gate) followed by a square-law detector (Figure 4). The recorded signal is either a measure of the spectrum of a series of time slices (spectrogram) or a measure of the time of arrival of a series of spectral slices (sonogram) depending upon the ordering of the filters. There is no difference in principle between the two possible filter orderings and thus this type of apparatus should be thought of as one that makes simultaneous measurements of the conjugate variables rather than sequential measurements. Since precise measurements of the conjugate variables cannot be made simultaneously, a spectrographic apparatus can measure only a smoothed out version of the Wigner function—the Wigner function convolved with an apparatus blurring or window function. In principle, if the window function is known, the Wigner function itself can be obtained via deconvolution but this is usually impractical because of the severe signal-to-noise requirements. Thus the spectrographic class of phase-space pulse characterization techniques supply only qualitative insight into pulse train statistics. However, if the pulses in the ensemble are assumed a priori to be identical, the resultant two-dimensional phase-retrieval problem can be solved iteratively. The success of this approach has been extensively demonstrated in the technique of frequency-resolved optical gating (FROG) [22].

A typical implementation of spectrography uses a temporal gate for the signal under test (e.g., the action of the pulse under test with one or several other pulses in a nonlinear optical medium [22], or a "shutter" function provided by a temporal modulator [23]) and a device capable of measuring the optical spectrum (e.g., an optical spectrum analyzer based on a diffraction grating and imaging optics, or a scanning Fabry-Perot etalon, together with a time-integrating photodiode). The spectrogram of the electric field of the test pulse is obtained by measuring the optical spectrum of the pulse after temporal gating for various relative delays between the pulse and the gate. The experimental trace is therefore

$$S(\omega,\tau) = \left| \int E(t)R(t-\tau)\exp(i\omega t)dt \right|^2, \qquad (20)$$

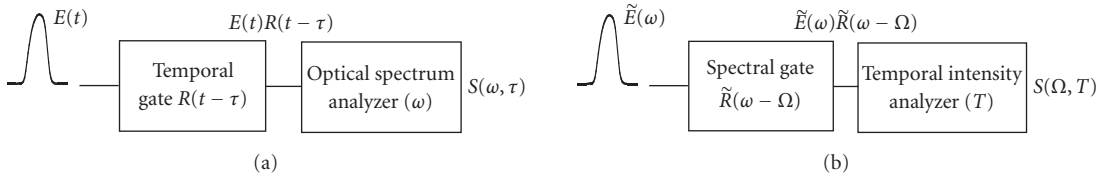


Figure 4: Conceptual implementation of (a) a spectrogram and (b) sonogram.

where ω is the optical frequency and τ the relative delay between the gate and the test pulse. It is important that the resolution of the spectral filter is very high in order to ensure that the measured trace is effectively the spectrogram of the test pulse.

A sonogram can be measured by reversing the order of the temporal and spectral gate [24, 25]. Typically, the pulse is first spectrally filtered using a spectrometer with variable central frequency Ω. The temporal intensity of the filtered pulse is then measured, and the sonogram is constructed as the set of the measured temporal intensities for various central frequencies Ω:

$$S(\Omega, T) = \left| \int \widetilde{E}(\omega)\widetilde{R}(\omega - \Omega)\exp(-i\omega T)d\omega \right|^2. \quad (21)$$

In this case the temporal resolution should be very high to ensure that the measured trace is a true sonogram. In practice, one usually implements the sonogram by means of a nonlinear cross-correlation of the spectrally gated signal with the test pulse, which has a shorter duration than the filtered pulse. Therefore, the experimental trace is given by a convolution of the sonogram of (21) with the unknown temporal intensity of the pulse under test, a fact that can be included in the inversion algorithm [25].

It can be shown that the spectrogram is the correlation of the Wigner function of the pulse with the Wigner function of the gate with a change of sign on the frequency variable [15]:

$$S(\omega, \tau) = \iint W_E(t', \omega')W_R(t' - \tau, \omega - \omega')dt'd\omega'. \quad (22)$$

The data can therefore be viewed in the chronocyclic space as a measurement of the overlap of the Wigner function of the pulse with the Wigner function of the gate (whose position in the space is related to the variables ω and τ, which must vary over the entire region of phase space occupied by the pulse), and the latter therefore appears as the apparatus window function in spectrographic measurements. As a Wigner function has a lower bound of support in the time-frequency space (i.e., it always occupies a region of phase space greater than π), the spectrogram is always a "blurred" version of the Wigner function of the pulse. In principle the test pulse can be completely characterized very simply by direct Fourier deconvolution if the gate (and therefore its Wigner function) is known [17, 21, 26, 27]. In practice such deconvolution is highly sensitive to noise, since it involves the division of the Fourier transform of the measured spectrogram with the ambiguity function of the gate.

The spectrogram or sonogram can be used to obtain the instantaneous frequency and group delay of the signal. For example, the average frequency of the pulse for a given relative delay τ between the pulse and the gate provides a measure of the chirp, and can be obtained from the spectrogram via [21]

$$\begin{aligned}\langle\omega\rangle_\tau &= \frac{\int S(\omega, \tau)\omega\, d\omega}{\int S(\omega, \tau)d\omega} \\ &= -\frac{\int |E(t)|^2 |R(t-\tau)|^2 \{\varphi_E'(t) + \varphi_R'(t-\tau)\}dt}{\int S(\omega, \tau)d\omega}.\end{aligned} \quad (23)$$

If the gate function is real (i.e., it does not have a phase) and is much shorter than any variation of the electric field of the pulse under test, then it can be replaced in the integral with a Dirac delta function, and the average frequency calculated from the spectrogram approaches $-\varphi_E'(\tau)$, that is, the instantaneous frequency of the electric field at the delay τ. Such gate can be implemented by cross-correlating the pulse under test with a much shorter optical pulse without temporal phase distortion, though this is often impractical. The difficulty with this approach is that the uncertainty in the local average frequency becomes very large, since the spectral content of the spectrogram is dominated by the broad spectrum of the gate. Such approach was initially developed for short optical pulses [28, 29], and similar approaches are still being used in optical telecommunications [30, 31].

A better way to reconstruct the pulse field from a spectrogram is to use phase retrieval. In fact, this is the only option if the gate is unknown. The spectrogram of (20) is the modulus square of the short-time Fourier transform of the pulse. The trick in phase retrieval is to estimate the phase of the transform. Once this is known, a Fourier transform directly leads to the recovery of both the pulse under test and the gating function. Phase retrieval is often ambiguous in one dimension, but is usually unique in two dimensions [32]. The excess data available in the spectrogram enables iterative reconstruction of N complex numbers specifying the field from the N^2 data points, and this can also lead to the simultaneous reconstruction of the gate [23, 33]. Furthermore, in the case of the nonlinear spectrogram, there is often a known functional relation between the pulse and the gate, since the gate is implemented as a nonlinear interaction with replicas of the pulse under test. Also, other information might be available, such as the spectrum of the pulse or the transfer function of the gate. These constraints enable the recovery of the field by means of several algorithms. A very robust inversion algorithm is based on the principal component general projections algorithm [34], which uses projections derived from the

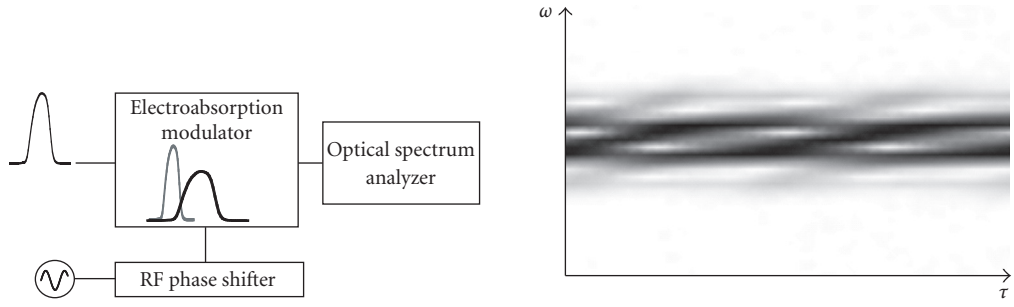


Figure 5: Example of the implementation of (a) spectrography and (b) spectrogram measured on a 40 GHz alternate-chirp return-to-zero signal with an electroabsorption modulator driven at 10 GHz .

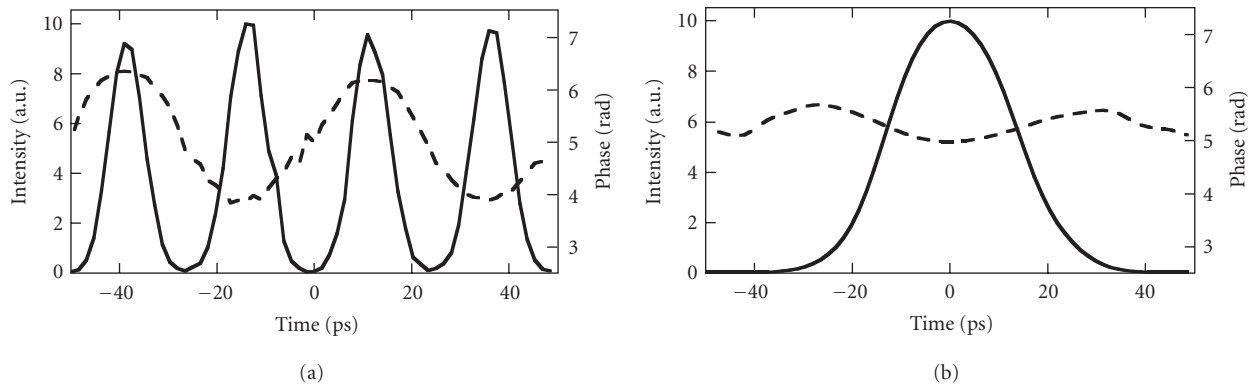


Figure 6: (a) Intensity and phase of the optical signal (resp., continuous and dashed line). (b) Transmission and phase of the gate (resp., continuous and dashed line) extracted from the previous spectrogram.

experimentally measured spectrogram and from the functional form of the spectrogram as the modulus square of the integral $\int E(t)R(t-\tau)\exp(i\omega t)dt$.

3.2. Experimental spectrography for telecommunication applications

Spectrograms and sonograms are popular tools for ultrashort optical pulse characterization. FROG generates a nonlinear spectrogram for the pulse by means of a nonlinear optical interaction of the pulse under test with one or several of its replicas [22]. This has the experimental advantage of using pump-probe geometries that are commonly used in ultrafast spectroscopy. Adaptations of FROG to pulse characterization in the telecommunication environment can be found, for example, in [35, 36, 37, 38].

For picosecond pulses, such as those used in telecommunications, a gate of smaller bandwidth than needed for femtosecond pulses, such as those typically found in ultrafast optics applications, suffices. It is possible to implement the gate using a temporal modulator, which has the important advantage of making the entire process linear. It is therefore extremely sensitive to small input pulse energies, yet insensitive to polarization and wavelength [23, 39].

The experimental implementation of spectrography with a temporal modulator is straightforward. The pulse under test is gated by a temporal modulator driven by a control signal synchronized to the pulse under test, as shown in Figure 5a. For example, an electroabsorption modulator can be driven by a sinusoidal voltage with well-defined phase relative to the test pulse. The relative delay τ between the signal under test and the gate is modified by changing the phase of the driving RF sine wave using a voltage-controlled phase shifter. The spectra after the modulator are recorded as a function of the optical frequency ω with a scanning monochromator, or with a Fabry-Perot etalon followed by a photodiode. An example of a measured spectrogram is displayed in Figure 5b. The characteristics of the pulse train and gate are extracted from the spectrogram, and examples of the retrieved signal and gate when characterizing an alternate-chirp signal [40] are plotted in Figure 6.

3.3. Interferometry

Interferometry is a well-known approach to the characterization of optical fields in the spatial domain. It is a simple method for converting phase information into amplitude information that can then be read using square-law detectors. For temporal waveform measurement, interferometry, in contrast to spectrography, measures a single line of the two-time or two-frequency correlation function of the pulse under test. This is sufficient to characterize completely a

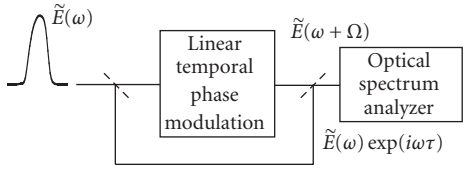


Figure 7: Conceptual approach for spectral shearing interferometry.

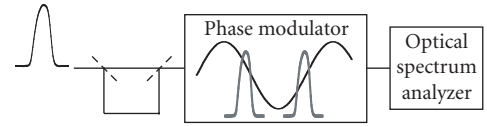


Figure 8: Experimental implementation of spectral shearing interferometry.

single pulse or a coherent pulse train, yielding N samples of the field from αN data points ($\alpha > 1$). It therefore makes optimal use of the experimental data.

It is possible to construct the spectral equivalent of spatial shearing interferometry, in which the spatial phase profile of a beam is determined by interfering it with a laterally shifted (or sheared) replica. The resulting intensity interferogram can be measured with a square-law detector and the phase simply extracted. The spectral analogue, in which two spectrally sheared pulses are interfered, also allows direct reconstruction of the electric field in the spectral domain using the measured spectral phase and a pulse spectrum. Because a single slice of the correlation function is sufficient to characterize the electric field of a pulse that has a continuous spectral support, interferometry is viable for most applications. We will focus here on techniques that use the two-time frequency correlation function (or its sampled version) $\widetilde{E}(\omega)\widetilde{E}^*(\omega-\Omega)$, whose phase $\varphi(\omega)-\varphi(\omega-\Omega)$ can be concatenated or integrated to get the spectral phase of the initial pulse [41, 42]. For a continuous spectral density, the spectral shear Ω is set by the sampling theorem, and it is typically a few percent of the total bandwidth of the pulse under test. Too large a shear would lead to undersampling of the pulse spectrum, while too small a shear could lead to increased sensitivity to noise, and thus reduced accuracy of the reconstruction. For a periodic source with high duty cycle, it suffices to measure the intensity and phase of the spectral modes, and the shear can usually be set to the value of the separation between the modes, that is, the repetition rate. The spectral intensity can be obtained either from a separate measurement using the spectrometer, or can be extracted from the correlation function directly.

A variety of interferometric techniques are known. The relative phase between adjacent spectral modes can, for example, be extracted in the time domain after spectral filtering [43, 44, 45, 46, 47], or in the spectral domain after proper temporal modulation [48]. The latter class of techniques also includes spectral shearing interferometry, where the quantity $\widetilde{E}(\omega)\widetilde{E}^*(\omega-\Omega)$ is obtained by measuring the interference of the pulse under test with its sheared replica with an optical spectrum analyzer (Figure 7). The frequency shear Ω can be implemented using a linear temporal phase modulation $\exp(i\Omega t)$. The spectral intensity of the two interfering pulses is $|\widetilde{E}(\omega)|^2+|\widetilde{E}(\omega-\Omega)|^2+\widetilde{E}(\omega)\widetilde{E}^*(\omega-\Omega)+\widetilde{E}^*(\omega)\widetilde{E}(\omega-\Omega)$. If a delay is introduced between the nonshifted and the shifted replica, this leads to spectral fringes with small spacing, by virtue of the phase $\varphi(\omega)-\varphi(\omega-\Omega)+\omega\tau$. In this case, the interferometric component can be directly extracted using Fourier processing of a single interferogram [49, 50].

3.4. Experimental interferometry for telecommunication applications

There are numerous implementations of interferometry for the self-referencing characterization of optical pulses. For spectral shearing interferometry, a spectral shear of arbitrary value can be induced by mixing the test pulse with, for example, a highly chirped pulse in a nonlinear medium. This is known as spectral phase interferometry for direct electric field reconstruction [49]. A spectral shear can also be induced, as explained above, by linear temporal phase modulation of the pulse under test. Such modulation can be obtained using an electro-optic phase modulator driven by an appropriate voltage [50, 51]. The generation of a strictly linear voltage at a high frequency is in practice difficult, and it is easier to use a sinusoidal voltage synchronized so that the pulse is at the zero-crossing of the modulation, therefore experiencing linear temporal phase modulation. One possible implementation is described in Figure 8.

The pulse under test is split into two replicas. These two replicas, separated by a delay τ are sent to a phase modulator driven by a sinusoidal voltage with period 2τ. The synchronization is performed so that the two pulses stand at different zero-crossings of the modulation. Therefore, the pulses are sheared in opposite directions along the frequency axis. If Ω is the shear imposed on one of the pulses, the extracted spectral phase difference is $\varphi(\omega+\Omega)-\varphi(\omega-\Omega)+\omega\tau$. The carrier term $\omega\tau$ can be removed either by turning the modulation off [50] or by measuring a second phase difference when the relative pulse under test and the sine wave driving the modulator has been modified by τ [51]. In this case the extracted spectral phase is $\varphi(\omega-\Omega)-\varphi(\omega+\Omega)+\omega\tau$; so that the difference between the two extracted phases is $2\varphi(\omega+\Omega)-2\varphi(\omega-\Omega)$. Figure 9 displays an experimentally measured interferogram. The rapidly varying fringes due to the delay between the two interfering pulses are evident. An example of the intensity and phase measured using spectral shearing interferometry is also shown.

3.5. Tomography

As with spectrographic methods, the so-called tomographic techniques require in-series time-stationary and time-nonstationary filters so that the entire phase space can be explored. However, unlike spectrographic techniques, the first filter in a tomographic apparatus is a phase-only filter (either a quadratic temporal phase modulator or a quadratic spectral phase modulator). The inclusion of a quadratic phase-only filter results in a distinctly different interpretation of the measurement, leading to a fundamentally different

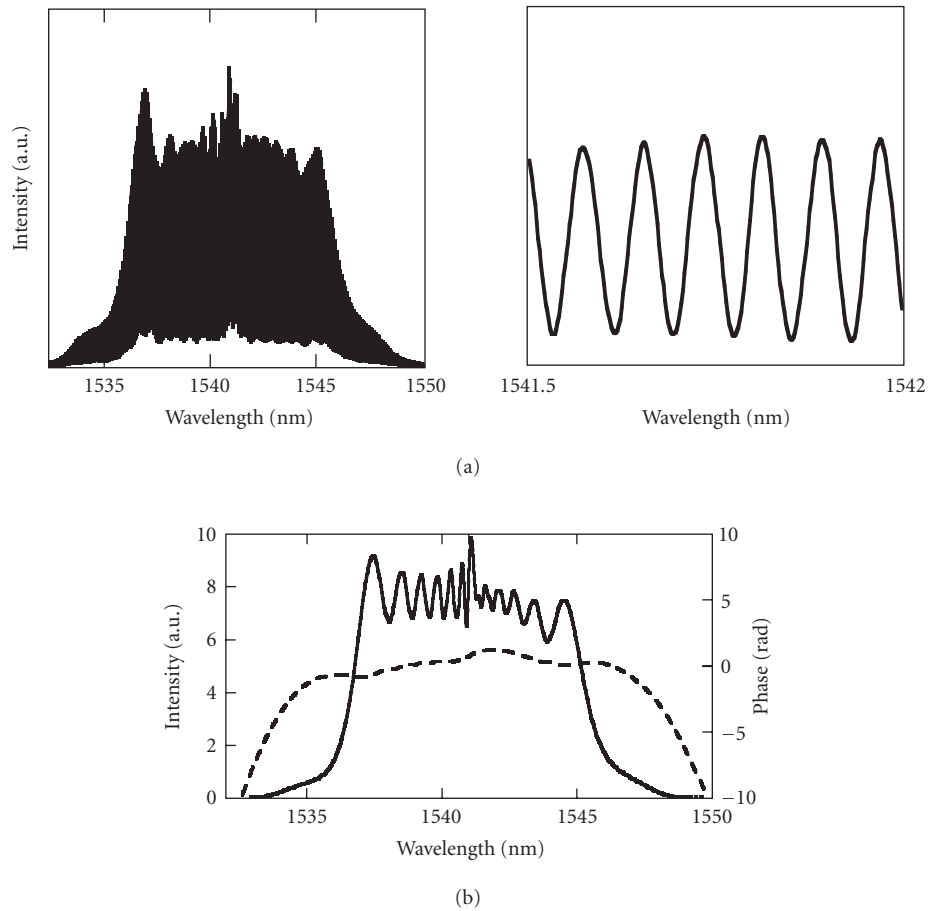


FIGURE 9: (a) Experimental interferogram and close-up on the fringes. (b) Reconstructed spectral intensity and phase (resp., continuous and dashed line).

inversion algorithm. To see this, notice that a phase-only filter does not provide any information on the frequency or the arrival time of a pulse ensemble and hence, does not constitute a measurement of either the spectral or temporal extent of the pulse. So, a tomographic apparatus does not make a simultaneous measurement of the conjugate variables time and frequency. Rather, the quadratic phase modulation acts to rotate the phase space. The square-law detector in combination with the amplitude-only filter records the resulting intensity distribution, that is, the projection of the rotated Wigner function. A sufficiently large number of phase-space rotations between $-\pi/2$ and $\pi/2$ allows, in principle, reconstruction of the Wigner function via the inverse Radon transform [52]. Numerical versions of this inversion algorithm were developed for applications of tomography in areas such as medical imaging, where one aims at reconstructing an object from a set of its projections (typically, 2D projections of a 3D object, or 1D projections of a 2D object) [53]. A typical implementation of chronocyclic tomography would use a combination of quadratic temporal and spectral phase modulations to rotate the phase space, and optical spectrum measurements to project the Wigner function. The Wigner function of the signal under test can be reconstructed from the measured projections, regardless of the degree of coherence of the pulse ensemble. This capability is unique to tomography among the techniques presented in this paper. This capability has not been realized experimentally, however, because of the relative difficulty of implementing variable temporal and spectral phase modulations.

The time-to-frequency converter [54, 55] operates by rotating the Wigner function by $\pi/2$, so that a measurement of the frequency marginal after rotation leads to the temporal marginal, that is, one obtains the temporal intensity of the signal under test via an optical spectrum measurement. However, no phase information is obtained, and this approach requires a large rotation of the Wigner function, which is difficult to obtain.

The assumption that the pulse train is coherent, as is done in other methods, reduces the requirement on the number of projections needed to reconstruct the pulse field. This leads to the concept of simplified chronocyclic tomography. This technique does not require a large rotation of the phase-space density, and reconstructs the amplitude and phase of the signal from only two projections of the Wigner function. It is based on a particular relation between the frequency marginal of the rotated Wigner function and the electric field

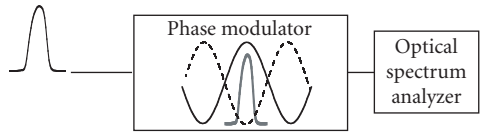
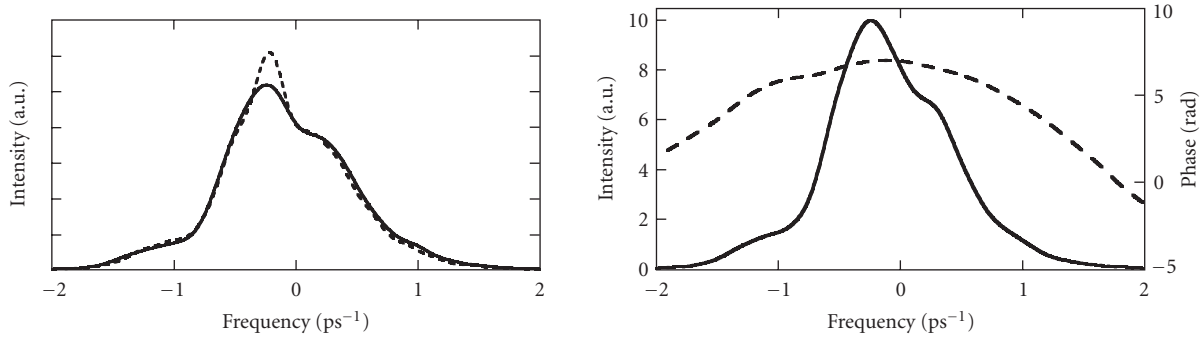


Figure 10: Experimental implementation of simplified chronocyclic tomography.



(a) (b)

Figure 11: (a) Two spectra measured for small positive and negative quadratic temporal phase modulation (continuous and dashed line). (b) Reconstructed spectral intensity and phase (continuous and dashed line).

[56, 57]. The fractional power spectrum of the pulse is obtained from the rotated Wigner function:

$$I_\alpha(\omega) = \int W[t\cos(\alpha)+\omega\sin(\alpha), \omega\cos(\alpha)-t\sin(\alpha)]dt. \quad (24)$$

The derivative of this function with respect to the angle of rotation α at $\alpha = 0$ leads to

$$\frac{\partial I_\alpha}{\partial\alpha} = \int \left[\omega\frac{\partial W}{\partial t} - t\frac{\partial W}{\partial\omega}\right]dt = -\frac{\partial}{\partial\omega}\int tW\,dt, \quad (25)$$

and therefore to

$$\frac{\partial I_\alpha}{\partial\alpha} = -\frac{\partial}{\partial\omega}\left[I\frac{\partial\varphi}{\partial\omega}\right]. \quad (26)$$

A rotation of the phase-space of the pulse requires a combination of a quadratic temporal and spectral phase modulations. However, the relation in (26) also holds for a shear of the phase-space, in which ω is transformed into $\omega + \psi t$, and the temporal coordinate is unchanged. This can be accomplished by means of a parabolic temporal phase modulation $(1/2)\psi t^2$ alone. In this case, one finds

$$\frac{\partial I_0}{\partial\psi} = \frac{\partial}{\partial\psi}\int W(\omega+\psi t, t)dt = \frac{\partial}{\partial\omega}\left[I\frac{\partial\varphi}{\partial\omega}\right]. \quad (27)$$

This is the form most amenable to experiment, since the bandwidth required to generate a small shear using a phase modulator is modest.

3.6. Experimental chronocyclic tomography

Various implementations of the time-to-frequency converter have been performed using either a phase modulator or nonlinear optics. A phase modulator driven by a sinusoidal voltage can provide quadratic temporal phase modulation to a pulse synchronized with one of its extrema. A nonlinear interaction can also provide such modulation.

Simplified chronocyclic tomography has been implemented using a temporal phase modulator [57]. The pulse under test was synchronized with a maximum of the phase modulation, and the optical spectrum after modulation measured, as displayed in Figure 10. The pulse under test was then synchronized with a minimum of the phase modulation, and the corresponding optical spectrum measured. The derivative $\partial I_0/\partial\psi$ and the spectrum $I(\omega)$ are obtained, respectively, by taking the difference and the sum of the two measured spectra. Equation (26) is then used to reconstruct the phase $\varphi(\omega)$. Figure 11 displays an experimental pair of measured spectra and the reconstructed intensity and phase for a short optical pulse. The technique can also be improved using synchronous detection of the derivative $\partial I_0/\partial\psi$, which enables an increased signal-to-noise ratio [58].

4. CONCLUSION

Correlation functions and time-frequency distributions are important concepts for both representation and characterization of ultrashort optical pulses. Aside from their utility in analyzing the properties of the electric field of the pulse and for reconstructing the field from experimental data, they are necessary for representing the coherence properties of a train of optical pulses. The concepts of spectrography, tomography, and interferometry serve as useful major classifications

for pulse-shape characterization, and each may be implemented using appropriate linear optical elements. For pulses typical in telecommunications applications, the use of linear temporal modulators as time-nonstationary filters in these classes of measurement has been detailed and experimental implementations for characterization in the telecommunication environment have been presented. As research in both signal analysis and ultrafast optics is in continual development, it can be expected that further exciting discoveries will be made in the near future.

ACKNOWLEDGMENTS

C. Dorrer would like to acknowledge his fruitful collaboration with Inuk Kang, from Bell Laboratories, Lucent Technologies, on the experiments described in [23, 39, 50, 51, 57, 58]. I. A. Walmsley acknowledges fruitful interactions with C. Iaconis and M. G. Raymer on methods of wave-field reconstruction. The work of I. A. Walmsley was supported by the US National Science Foundation and the UK Engineering and Physical Sciences Research Council.

REFERENCES

[1] I. Kaminow and T. Li, *Optical Fiber Telecommunications IV*, Academic Press, Boulder, Colo, USA, 2002.

[2] I. A. Walmsley, L. Waxer, and C. Dorrer, "The role of dispersion in ultrafast optics," *Review of Scientific Instruments*, vol. 72, no. 1, pp. 1–29, 2001.

[3] L. Lepetit, G. Cheriaux, and M. Joffre, "Linear techniques of phase measurement by femtosecond spectral interferometry for applications in spectroscopy," *Journal of the Optical Society of America B*, vol. 12, no. 12, pp. 2467–2474, 1995.

[4] D. N. Fittinghoff, J. L. Bowie, J. N. Sweetser, et al., "Measurement of the intensity and phase of ultraweak, ultrashort laser pulses," *Optics Letters*, vol. 21, no. 12, pp. 884–886, 1996.

[5] F. K. Fatemi, T. F. Carruthers, and J. W. Lou, "Characterisation of telecommunications pulse trains by fourier-transform and dual-quadrature spectral interferometry," *Electronics Letters*, vol. 39, no. 12, pp. 921–922, 2003.

[6] P. A. Andrekson, "Ultrahigh bandwidth optical sampling oscilloscopes," in *Proc. Optical Fiber Communication Conference (OFC '04)*, Los Angeles, Calif, USA, February 2004, Tuo1.

[7] C. Dorrer, D. C. Kilper, H. R. Stuart, G. Raybon, and M. G. Raymer, "Linear optical sampling," *IEEE Photon. Technol. Lett*, vol. 15, no. 12, pp. 1746–1748, 2003.

[8] C. Dorrer, J. Leuthold, and C. R. Doerr, "Direct measurement of constellation diagrams of optical sources," in *Proc. Optical Fiber Communication Conference (OFC '04)*, Los Angeles, Calif, USA, February 2004, PDP33.

[9] I. A. Walmsley and V. Wong, "Characterization of the electric field of ultrashort optical pulses," *Journal of the Optical Society of America B*, vol. 13, no. 11, pp. 2453–2463, 1996.

[10] C. Iaconis, V. Wong, and I. A. Walmsley, "Direct interferometric techniques for characterizing ultrashort optical pulses," *IEEE J. Select. Topics Quantum Electron.*, vol. 4, no. 2, pp. 285–294, 1998.

[11] J. W. Goodman, *Introduction to Fourier Optics*, McGraw-Hill, New York, NY, USA, 1996.

[12] S. A. Ponomarenko, G. P. Agrawal, and E. Wolf, "Energy spectrum of a nonstationary ensemble of pulses," *Optics Letters*, vol. 29, no. 4, pp. 394–396, 2004.

[13] M. Born and E. Wolf, *Principles of Optics*, Cambridge University Press, Cambridge, UK, 1998.

[14] K. H. Brenner and J. Ojeda-Castaneda, "Ambiguity function and Wigner distribution function applied to partially coherent imagery," *Optica Acta*, vol. 31, no. 2, pp. 213–223, 1984.

[15] T. A. C. M. Claasen and W. F. G. Mecklenbräuker, "The Wigner distribution—A tool for time-frequency signal analysis—Part III: relations with other time-frequency signal transformations," *Philips Journal of Research*, vol. 35, no. 6, pp. 372–389, 1980.

[16] K.-H. Brenner and K. Wodkiewicz, "The time-dependent physical spectrum of light and the wigner distribution function," *Optics Communications*, vol. 43, no. 2, pp. 103–106, 1982.

[17] J. Paye, "The chronocyclic representation of ultrashort light pulses," *IEEE J. Quantum Electron.*, vol. 28, no. 10, pp. 2262–2273, 1992.

[18] C. A. Hirlimann and J.-F. Morhange, "Wavelet analysis of short light pulses," *Applied Optics*, vol. 31, no. 17, pp. 3263–3266, 1992.

[19] R. Gase, "Time-dependent spectrum of linear optical systems," *Journal of the Optical Society of America A*, vol. 8, no. 6, pp. 850–859, 1991.

[20] C. E. Cook and M. Bernfeld, *Radar Signals: an Introduction to Theory and Application*, Artech House, Norwood, Mass, USA, 1993, incorporated.

[21] L. Cohen, *Time-Frequency Analysis*, Prentice Hall PTR, Englewood Cliffs, NJ, USA, 1995.

[22] R. Trebino, Ed., *Frequency-Resolved Optical Gating: the Measurement of Ultrashort Optical Pulses*, Kluwer Academic, Boston, Mass, USA , 2002.

[23] C. Dorrer and I. Kang, "Simultaneous temporal characterization of telecommunication optical pulses and modulators by use of spectrograms," *Optics Letters*, vol. 27, no. 15, pp. 1315–1317, 2002.

[24] V. Wong and I. A. Walmsley, "Ultrashort-pulse characterization from dynamic spectrograms by iterative phase retrieval," *Journal of the Optical Society of America B*, vol. 14, no. 4, pp. 944–949, 1997.

[25] D. T. Reid, "Algorithm for complete and rapid retrieval of ultrashort pulse amplitude and phase from a sonogram," *IEEE J. Quantum Electron.*, vol. 35, no. 11, pp. 1584–1589, 1999.

[26] V. Wong and I. A. Walmsley, "Phase retrieval in time-resolved spectral phase measurement", in *Generation, Amplification, and Measurement of Ultrashort Laser Pulses II*, vol. 2377 of *Proceedings of SPIE*, pp. 178–186, San Jose, Calif, USA, April 1995.

[27] K. Kikuchi, "Theory of sonogram characterization of optical pulses," *IEEE J. Quantum Electron.*, vol. 37, no. 4, pp. 533–537, 2001.

[28] E. B. Treacy, "Measurement and interpretation of dynamic spectrograms of picosecond light pulses," *Journal of Applied Physics*, vol. 42, no. 10, pp. 3848–3858, 1971.

[29] J. L. A. Chilla and O. E. Martinez, "Analysis of a method of phase measurement of ultrashort pulses in the frequency domain," *IEEE J. Quantum Electron.*, vol. 27, no. 5, pp. 1228–1235, 1991.

[30] K. Mori, T. Morioka, and M. Saruwatari, "Group velocity dispersion measurement using supercontinuum picosecond pulses generated in an optical fibre," *Electronics Letters*, vol. 29, no. 11, pp. 987–989, 1993.

[31] Y. Ozeki, Y. Takushima, and K. Kikuchi, "High dynamic range characterization of waveform and frequency-chirp of picosecond pulses based on the precise measurement of group delay in optical frequency domain," in *Proc. Optical Fiber Communication Conference (OFC '03)*, vol. 2, pp. 725–727, Atlanta, Ga, USA, March 2003, FL3.

[32] H. Stark, Ed., *Image Recovery: Theory and Application*, Academic Press, New York, NY, USA, 1987.

[33] D. J. Kane, G. Rodriguez, A. J. Taylor, and T. S. Clement, "Simultaneous measurement of two ultrashort laser pulses from a single spectrogram in a single shot," *Journal of the Optical Society of America B*, vol. 14, no. 4, pp. 935–943, 1997.

[34] D. J. Kane, "Real-time measurement of ultrashort laser pulses using principal component generalized projections," *IEEE J. Select. Topics Quantum Electron.*, vol. 4, no. 2, pp. 278–284, 1998.

[35] M. D. Thomson, J. M. Dudley, L. P. Barry, and J. D. Harvey, "Completepulse characterization at 1.5 μm by cross-phase modulation in optical fibers," *Optics Letters*, vol. 23, no. 20, pp. 1582–1584, 1998.

[36] K. Ogawa and M. D. Pelusi, "High-sensitivity pulse spectrogram measurement using two-photon absorption in a semiconductor at 1.5-μm wavelength," *Optics Express*, vol. 7, no. 3, pp. 135–140, 2000.

[37] P.-A. Lacourt, J. M. Dudley, J.-M. Merolla, H. Porte, J.-P. Goedgebuer, and W. T. Rhodes, "Milliwatt-peak-power pulse characterization at 1.55 μm by wavelength-conversion frequency-resolved optical gating," *Optics Letters*, vol. 27, no. 10, pp. 863–865, 2002.

[38] L. P. Barry, S. Del Burgo, B. Thomsen, R. T. Watts, D. A. Reid, and J. Harvey, "Optimization of optical data transmitters for 40 Gb/s lightwave systems using frequency resolved optical gating," *IEEE Photon. Technol. Lett.*, vol. 14, no. 7, pp. 971–973, 2002.

[39] C. Dorrer and I. Kang, "Real-time implementation of linear spectrograms for the characterization of high bit-rate optical pulse trains," *IEEE Photon. Technol. Lett.*, vol. 16, no. 3, pp. 858–860, 2004.

[40] P. J. Winzer, C. Dorrer, R.-J. Essiambre, and I. Kang, "Chirped return-to-zero modulation by imbalanced pulse carver driving signals," *IEEE Photon. Technol. Lett.*, vol. 16, no. 5, pp. 1379–1381, 2004.

[41] V. A. Zubov and T. I. Kuznetsova, "Solution of the phase problem for time-dependent optical signals by an interference system," *Soviet Journal of Quantum Electronics*, vol. 21, pp. 1285–1286, 1991.

[42] V. Wong and I. A. Walmsley, "Analysis of ultrashort pulse-shape measurement using linear interferometers," *Optics Letters*, vol. 19, no. 4, pp. 287–289, 1994.

[43] K. C. Chu, J. P. Heritage, R. S. Grant, et al., "Direct measurement of the spectral phase of femtosecond pulses," *Optics Letters*, vol. 20, no. 8, pp. 904–906, 1995.

[44] P. Kockaert, M. Peeters, S. Coen, Ph. Emplit, M. Haelterman, and O. Deparis, "Simple amplitude and phase measuring technique for ultrahigh-repetition-rate lasers," *IEEE Photon. Technol. Lett*, vol. 12, no. 2, pp. 187–189, 2000.

[45] M. Kwakernaak, R. Schreieck, A. Neiger, H. Jackel, E. Gini, and W. Vogt, "Spectral phase measurement of mode-locked diode laser pulses by beating sidebands generated by electrooptical mixing," *IEEE Photon. Technol. Lett.*, vol. 12, no. 12, pp. 1677–1679, 2000.

[46] P. Kockaert, J. Azana, L. R. Chen, and S. LaRochelle, "Full characterization of uniform ultrahigh-speed trains of optical pulses using fiber Bragg gratings and linear detectors," *IEEE Photon. Technol. Lett.*, vol. 16, no. 6, pp. 1540–1542, 2004.

[47] P. Kockaert, M. Haelterman, Ph. Emplit, and C. Froehly, "Complete characterization of (ultra)short optical pulses using fast linear detectors," *IEEE J. Select. Topics Quantum Electron.*, vol. 10, no. 1, pp. 206–212, 2004.

[48] J. Debeau, B. Kowalski, and R. Boittin, "Simple method for the complete characterization of an optical pulse," *Optics Letters*, vol. 23, no. 22, pp. 1784–1786, 1998.

[49] C. Iaconis and I. A. Walmsley, "Spectral phase interferometry for direct electric-field reconstruction of ultrashort optical pulses," *Optics Letters*, vol. 23, no. 10, pp. 792–794, 1998.

[50] C. Dorrer and I. Kang, "Highly sensitive direct characterization of femtosecond pulses by electro-optic spectral shearing interferometry," *Optics Letters*, vol. 28, no. 6, pp. 477–479, 2003.

[51] I. Kang, C. Dorrer, and F. Quochi, "Implementation of electro-optic spectral shearing interferometry for ultrashort pulse characterization," *Optics Letters*, vol. 28, no. 22, pp. 2264–2266, 2003.

[52] M. Beck, M. G. Raymer, I. A. Walmsley, and V. Wong, "Chronocyclic tomography for measuring the amplitude and phase structure of optical pulses," *Optics Letters*, vol. 18, no. 23, pp. 2041–2043, 1993.

[53] A. C. Kak and M. Slaney, *Principles of Computerized Tomographic Imaging*, IEEE Press, New York, NY, USA, 1988.

[54] M. T. Kauffman, W. C. Banyai, A. A. Godil, and D. M. Bloom, "Time-to-frequency converter for measuring picosecond optical pulses," *Applied Physics Letters*, vol. 64, no. 3, pp. 270–272, 1994.

[55] L. Kh. Mouradian, F. Louradour, V. Messager, A. Barthélémy, and C. Froehly, "Spectro-temporal imaging of femtosecond events," *IEEE J. Quantum Electron.*, vol. 36, no. 7, pp. 795–801, 2000.

[56] T. Alieva, M. J. Bastiaans, and L. Stankovic, "Signal reconstruction from two close fractional Fourier power spectra," *IEEE Trans. Signal Processing*, vol. 51, no. 1, pp. 112–123, 2003.

[57] C. Dorrer and I. Kang, "Complete temporal characterization of short optical pulses by simplified chronocyclic tomography," *Optics Letters*, vol. 28, no. 16, pp. 1481–1483, 2003.

[58] I. Kang and C. Dorrer, "Highly sensitive differential tomographic technique for real-time ultrashort pulse characterization," in *Proc. Conference on Lasers and Electro-Optics (CLEO '04)*, vol. 1, San Francisco, Calif, USA, May 2004, CTuZ6.

Christophe Dorrer is a Member of Technical Staff at Bell Laboratories, Lucent Technologies. He has a Ph.D. degree in optics from the École Polytechnique, France. He works in various domains of ultrafast optics related to optical telecommunications, such as optical pulse characterization, characterization of high-bit-rate data-encoded sources, characterization of devices and fibers, pulse shaping, and monitoring of optical networks.

Ian A. Walmsley is the Hook Professor of expermental physics at the University of Oxford, and the Head of the Department of Atomic and Laser Physics, Clarendon Laboratory. He has a Ph.D. degree in optics from the University of Rochester. His research concerns quantum processes on ultrafast time scales, and he has developed methods for the generation and measurement of nonclassical states of light and matter, as well as contributed to ultrafast technology through the invention of methods for ultrashort pulse characterization.

EURASIP Journal on Applied Signal Processing 2005:10, 1554–1565

Time-Frequency (Wigner) Analysis of Linear and Nonlinear Pulse Propagation in Optical Fibers

José Azaña

Institut National de la Recherche Scientifique, Énergie, Matériaux et Télécommunications, 800 de la Gauchetière Ouest, bureau 6900, Montréal, QC, Canada H5A 1K6
Email: azana@emt.inrs.ca

Received 12 April 2004; Revised 7 June 2004

Time-frequency analysis, and, in particular, Wigner analysis, is applied to the study of picosecond pulse propagation through optical fibers in both the linear and nonlinear regimes. The effects of first- and second-order group velocity dispersion (GVD) and self-phase modulation (SPM) are first analyzed separately. The phenomena resulting from the interplay between GVD and SPM in fibers (e.g., soliton formation or optical wave breaking) are also investigated in detail. Wigner analysis is demonstrated to be an extremely powerful tool for investigating pulse propagation dynamics in nonlinear dispersive systems (e.g., optical fibers), providing a clearer and deeper insight into the physical phenomena that determine the behavior of these systems.

Keywords and phrases: Wigner distributions, dispersive media, nonlinear fiber optics, optical pulse propagation and solitons.

1. INTRODUCTION

The study of optical pulse propagation in optical fibers is interesting from both fundamental and applied perspectives. Understanding the physics behind the processes that determine the evolution of optical pulses in single-mode fibers is essential for the design and performance analysis of optical fiber communication systems. As an example, it is well known that in intensity-modulated direct-detection (IM/DD) systems, the combined effects of source chirping, group velocity dispersion (GVD) and, for long-haul or high-power systems, self-phase modulation (SPM) cause distortion of the propagating signals [1]. This distortion essentially limits the maximum achievable bit rates and transmission distances. The influence of fiber GVD and fiber nonlinearities (e.g., SPM) on the performance of communication systems is becoming more critical in view of the expected evolution of fiber optics communications systems [2], in particular, (i) the channel data rates are expected to continue increasing, with 40 and 80 Gbps rate systems now under development; and (ii) the communication strategies (e.g., dense-wavelength-division-multiplexing, DWDM, strategies) tend to increase the number of channels and information (i.e., the signal power) launched into a single fiber.

For the study of the dynamics of pulse propagation in fibers, the involved signals (optical pulses) can be represented in either the temporal or the frequency domains. However, since these signals are intrinsically nonstationary (i.e., the spectrum content changes as a function of time), these conventional one-dimensional representations provide only partial information about the analyzed signals and, consequently, about the system under study. In this paper, we analyze linear and nonlinear pulse propagation in optical fibers using joint time-frequency (TF) representations [3]. Our analysis is based on the representation of the events of interest (optical pulses propagating through the fiber) in the joint TF plane, that is, the signals are represented as two-dimensional functions of the two variables time and frequency, simultaneously. For the TF representation, we will use the well-known Wigner distribution function. The Wigner distribution exhibits a lot of mathematical properties that make this approach especially attractive for the problem under consideration. For instance, as compared with other well-known methods for the TF representation of optical pulses (e.g., spectrograms [3]), the Wigner distribution provides an improved joint TF resolution. Note that this is a critical aspect for extracting detailed information about the events under analysis from the resultant images. The discussion of other attractive properties of the Wigner distribution is out of the scope of this work but the interested reader can find a good review article on the fundaments and applications of Wigner analysis by Dragoman in this same special issue or can refer to other classical papers in the subject [4].

TF representations in general, and Wigner analysis in particular, have been used in the past for the analysis of (ultra-) short light pulses and, in particular, these methods have been applied to investigating (i) simple linear optical systems (e.g., Fabry-Perot filters, fiber Bragg gratings) [5, 6], (ii) soliton waveforms [5, 7], and (iii) optical pulse-compression operations [8]. TF techniques have been

also evaluated as alternative methods for measuring optical fiber dispersion (linear regime) [9]. More recently, TF representations (spectrograms) have been applied to the analysis of specific phenomena (e.g., continuum generation) in *nonlinear* optical fiber devices [10, 11] but these recent works deal with optical pulses in the femtosecond range, a regime which is of less interest in the context of fiber optics communications (optical pulses in the picosecond range).

In this paper, the Wigner distribution is applied to the study of the *dynamics* of linear and nonlinear *picosecond* pulse propagation in optical fibers. By means of a few examples, we demonstrate that the Wigner analysis offers a simple and easy-to-interpret representation of the linear and nonlinear dynamics in fibers within the picosecond regime, providing in fact a profound insight into the physics behind the phenomena that determine the optical pulse evolution through the fibers. The information provided by the Wigner technique complements that given by other analysis methods and offers a clearer and deeper understanding of the phenomena under study. It should be also mentioned that the discussion in this present work is restricted to the case of completely coherent light distributions. The Wigner formalism has been previously applied to the analysis of propagation of partially coherent light through nonlinear media, leading in fact to the description of phenomena not discussed here [12].

The remainder of this paper is structured as follows. In Section 2, the theoretical fundaments of our analysis are established. In particular, the nonlinear Schrödinger equation (NLSE) for modeling picosecond pulse propagation in optical fibers is briefly reviewed and the Wigner distribution function used throughout the work is defined as well. In Section 3, we conduct Wigner analysis of picosecond optical pulse propagation through optical fibers operating in the linear regime. The impact of first- and second-order dispersions are analyzed in detail. Section 4 is devoted to the Wigner analysis of picosecond optical pulse propagation through nonlinear optical fibers. The interplay of GVD and SPM is analyzed in both the normal and anomalous dispersion regimes. Finally, in Section 5, we conclude and summarize.

2. THEORETICAL FUNDAMENTS

The propagation of optical pulses in the picosecond range through a lossless single-mode optical fiber can be described by the well-known NLSE [1]:

$$i\frac{\partial A(z,\tau)}{\partial z} - \frac{\beta_2}{2}\frac{\partial^2 A(z,\tau)}{\partial \tau^2} + \gamma\left|A(z,\tau)\right|^2 A(z,\tau) = 0, \quad (1)$$

where $A(z,\tau)$ is the complex envelope of the optical pulse (pulse centered at the frequency ω_0), z is the fiber length, and τ represents the time variable in the so-called retarded frame (i.e., temporal frame with respect to the pulse group delay), $\beta_2 = [\partial^2\beta(\omega)/\partial\omega^2]_{\omega=\omega_0}$ is the first-order GVD ($\beta(\omega)$ is the propagation constant of the single-mode fiber), and γ is the nonlinear coefficient of the fiber. In most cases, (1) and its modifications cannot be solved analytically and one has to use numerical approaches. Here we will use the most commonly applied numerical scheme for solving the NLSE, the so-called split-step Fourier transform (SSFT) method [13].

In order to characterize the fiber distances over which dispersive and nonlinear effects are important, two parameters are usually used, namely, the dispersion length L_D and the nonlinear length L_{NL}, [1]

$$L_D = \frac{T_0^2}{|\beta_2|}, \qquad L_{NL} = \frac{1}{\gamma P_0}, \quad (2)$$

where T_0 and P_0 are the time width and peak power of the pulse launched at the input of the fiber $A(0,\tau)$. Depending on the relative magnitudes of L_D, L_{NL}, and the fiber length z, the propagation behavior is mainly determined by dispersive effects, by nonlinear effects or by interplay between both dispersive and nonlinear effects (when both contributions are significant).

Once the NLSE in (1) is solved for the specific problem under study, our subsequent study will be based on the detailed analysis of the obtained pulse complex envelope $A(z,\tau)$. For a given fiber length ($z \equiv$ constant), this signal can be represented either in the temporal domain (as directly obtained from (1)) or in the spectral domain,

$$\tilde{A}(z,\omega) = \Im[A(z,\tau)] = \left(\frac{1}{\sqrt{2\pi}}\right)\int_{-\infty}^{+\infty} A(z,\tau)\exp(-j\omega\tau)d\tau, \quad (3)$$

where $\Im$ is the Fourier transform operator. A more profound insight into the nature of the pulse under analysis can be obtained if this pulse is represented in the joint time-frequency phase space. For this purpose, we will use the time-resolved Wigner distribution function, $W_z(\tau,\omega)$, which will be calculated as follows [3]:

$$W_z(\tau,\omega) = \int_{-\infty}^{+\infty} A\left(z,\tau+\frac{\tau'}{2}\right)A^*\left(z,\tau-\frac{\tau'}{2}\right)\exp[-i\omega\tau']d\tau'. \quad (4)$$

The Wigner distribution allows us to represent the signal propagating through the fiber $A(z,\tau)$ in the two domains, time and frequency, simultaneously, that is, the signal is mapped into a 2D image which essentially represents the signal's joint time-frequency energy distribution. This 2D image displays the link between the temporal and spectral pulse features in a very simple and direct way, thus providing broader information on the signal and system under analysis.

3. WIGNER ANALYSIS OF LINEAR PULSE PROPAGATION IN OPTICAL FIBERS

When the fiber length z is such that $z \ll L_{NL}$ and $z \ll L_D$, then neither dispersive nor nonlinear effects play a significant role during pulse propagation and as a result, the pulse

maintains its shape during propagation $A(z,\tau) = A(0,\tau)$. This case is obviously out of the interest of this work. When the fiber length is such that $z \ll L_{NL}$ but $z \approx L_D$, then the pulse evolution is governed by GVD and the nonlinear effects play a relatively minor role. More specifically, the dispersion-dominant regime is applicable when the following condition is satisfied:

$$\frac{L_D}{L_{NL}} = \frac{\gamma P_0 T_0^2}{|\beta_2|} \ll 1. \tag{5}$$

In this case, the last term in the left-hand side of the NLSE in (1) (i.e., the nonlinear term) can be neglected and the optical fiber can be modeled as a linear time-invariant system (i.e., as a filter). Specifically, the fiber operates as a phase-only filter which only affects the phase of the spectral content of the signal propagating through it. This phase-only filtering process is in fact determined by the GVD characteristic of the optical fiber and, in particular,

$$\tilde{A}(z,\omega) \propto \tilde{A}(0,\omega) \exp\left(-i\frac{\beta_2}{2} z\omega^2\right), \tag{6}$$

where the symbol $\propto$ indicates that the two terms are proportional.

The propagation regime where nonlinearities can be neglected is typical of optical communication systems when the launched signals exhibit a relatively low power. As a rough estimate, in order to ensure operation within the linear regime, the peak power of the input pulses must be $P_0 \ll 1\,\mathrm{W}$ for 1-picosecond pulses in conventional single-mode fiber operating at the typical telecommunication wavelength of $\lambda \approx 1.55\,\mu\mathrm{m}$ ($\omega_0 \approx 2\pi \times 193.4\,\mathrm{THz}$).

3.1. First-order dispersion of a transform-limited optical pulse

In the first example (results shown in Figure 1), the propagation of an optical pulse through a first-order dispersive fiber in the linear regime is analyzed. In particular, we assume a fiber with a first-order dispersion coefficient $\beta_2 = -20\,\mathrm{ps^2Km^{-1}}$ (typical value in a conventional single-mode fiber working at $\lambda \approx 1.55\,\mu\mathrm{m}$). This regime is usually referred to as anomalous dispersion regime ($\beta_2 < 0$). Figure 1 shows the Wigner representation of the optical pulse envelope $A(z,\tau)$ evaluated at different fiber propagation distances, $z = 0$ (input pulse), $z = 0.5\,L_D$, $z = 2\,L_D$, and $z = 6\,L_D$ ($L_D \approx 450\,\mathrm{m}$). For each representation, the plot at the left shows the spectral energy density of the optical pulse $|\tilde{A}(z,\omega)|^2$, the plot at the bottom shows the average optical intensity of the pulse $|A(z,\tau)|^2$, and the larger plot in the upper right of the representations shows the Wigner distribution of the pulse $W_z(\tau,\omega)$. Note that this distribution is plotted as a 2D image where the relative brightness levels of the image represent the distribution intensity and, in particular, darker regions in the image correspond to higher intensities. This 2D image provides information about the temporal location of the signal spectral components or in other words, it shows which of the spectral components of the signal occur at each instant of time.

The input pulse is assumed to be a transform-limited super-Gaussian pulse, $A(0,\tau) = \sqrt{P_0}\exp[(-1/2)(\tau/T_0)^{2m}]$, where $m = 3$, $T_0 = 3$ picoseconds and the peak power P_0 is low enough to ensure operation within the linear regime (i.e., to ensure that the fiber nonlinearities are negligible). In this paper, super-Gaussian pulses will be used as input signals because they are more suited than for instance Gaussian pulses to illustrate the effects of steep leading and trailing edges while providing similar information on the physics behind the different linear and nonlinear phenomena to be investigated. The Wigner distribution of this input pulse is typical of a transform-limited signal where all the spectral components exhibit the same mean temporal delay. Since the fiber operates as a phase-only filter, the energy spectrum of the pulse is not affected by the propagation along the optical fiber. In other words, the optical pulse propagating through the fiber retains identical spectral components to those of the incident pulse. However, due to the GVD introduced by the fiber, these spectral components are temporally realigned according to the group delay curve of the fiber. This temporal realignment of the pulse spectral components is responsible for the distortion and broadening of the temporal shape of the pulse as it propagates along the fiber and can be easily understood and visualized through the Wigner representations shown in Figure 1. The dispersion-induced pulse temporal broadening is a detrimental phenomenon for optical communication purposes. As a result of this temporal broadening, the adjacent pulses in a sequence launched at the input of the fiber (this pulse sequence can carry coded information to be transmitted through the fiber) can interfere with each other and this interference process can obviously limit the proper recovering of the information coded in the original sequence [2].

We remind the reader that the group delay in a first-order dispersive fiber is a linear function of frequency and depends linearly on the fiber distance z as well. This is in good agreement with the temporal realignment process that can be inferred from the Wigner distributions shown in Figure 1. More specifically, the pulse spectral components separate temporally from each other as they propagate through the fiber. In fact, as the Wigner representation of the pulse at $z = 6\,L_D$ shows, for a sufficiently long fiber distance, the temporal realignment process of the pulse spectral components is sufficiently strong so that only a single dominant frequency term exists at each given instant of time. This can be very clearly visualized in the corresponding Wigner representation: the signal distributes its energy along a straight line in the TF plane. In this case, there is a direct correspondence between time and frequency domains or in other words, the temporal and spectral pulse shapes are proportional, $|A(z,\tau)| \propto |\tilde{A}(z,\omega)|_{\omega=\tau/\beta_2 z}$. This frequency-to-time conversion operation induced by simple propagation of an optical pulse through a first-order dispersive medium (e.g., an optical fiber) is usually referred to as *real-time Fourier*

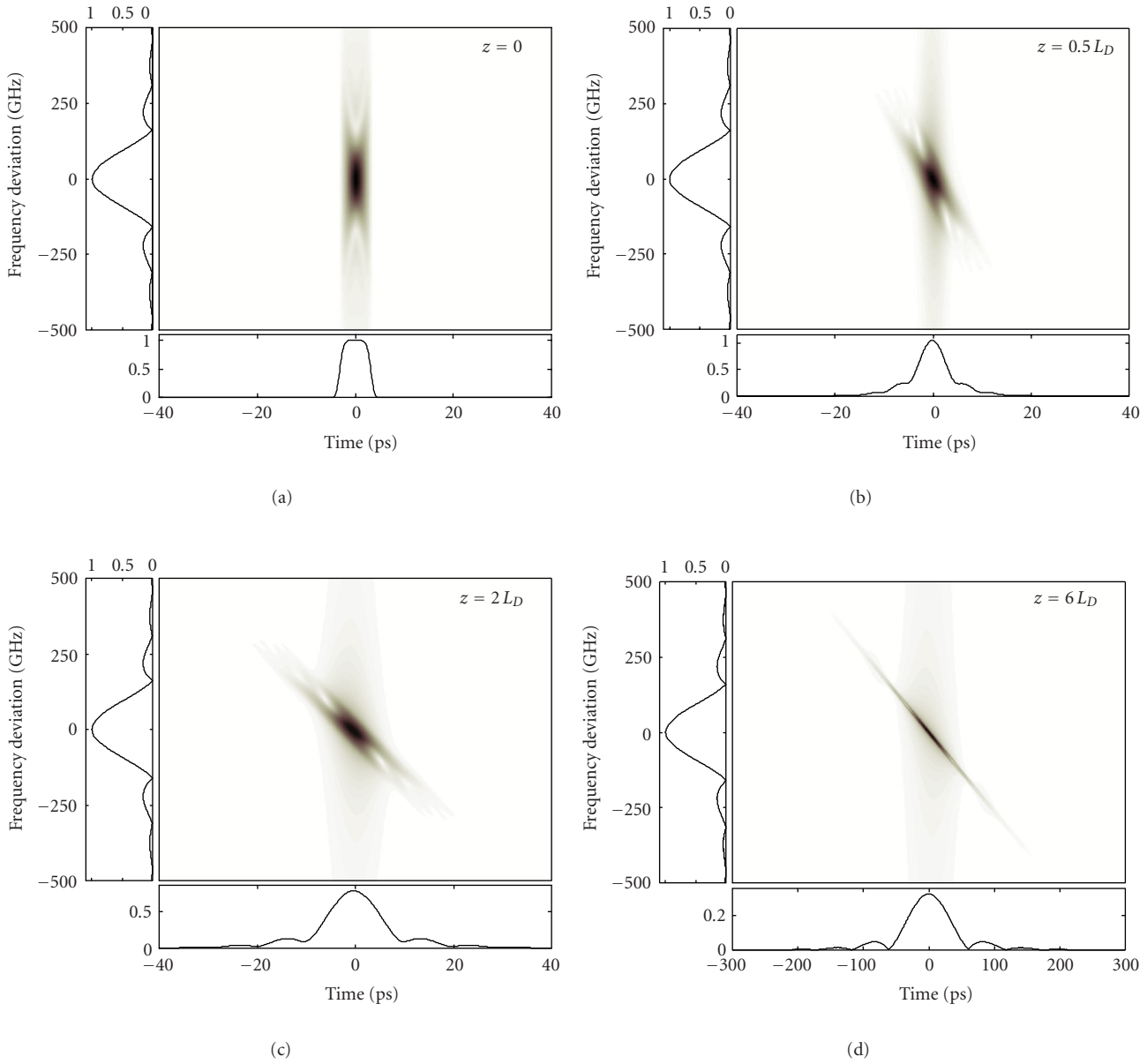


Figure 1: Wigner analysis of linear pulse propagation in an optical fiber (first-order dispersion).

transformation (RTFT) [14]. The exact condition to ensure RTFT of the input optical pulse is the following [15]:

$$z \gg \frac{T^2}{8\pi\,|\beta_2|}. \tag{7}$$

RTFT has been demonstrated for different interesting applications, including real-time optical spectrum analysis, fiber dispersion measurements [14], and temporal and spectral optical pulse shaping [15, 16]. An interesting application of the phenomenon for monitoring channel crosstalk in DWDM optical communication networks is also described in detail in the paper by Llorente et al. in this present special issue.

3.2. First-order dispersion of a chirped optical pulse

In the second example (results shown in Figure 2), the propagation of a nontransform-limited optical pulse through the same optical fiber as in the previous example is analyzed. In this case, we assume a chirped super-Gaussian input pulse $A(0,\tau) = \sqrt{P_0}\exp[(-[1+iC]/2)(\tau/T_0)^{2m}]$, where $m = 3$, $T_0 = 3$ picoseconds, and the peak power P_0 is again assumed to be low enough to ensure operation within the linear regime. The new parameter C is referred to as the chirp of the pulse and is used to model a phase variation across the temporal profile of the pulse. In our example, we fix $C = -3$. Pulses generated from semiconductor or mode-locked laser are typically chirped and that is why it is important also to evaluate the effect of pulse chirp on the dispersion process.

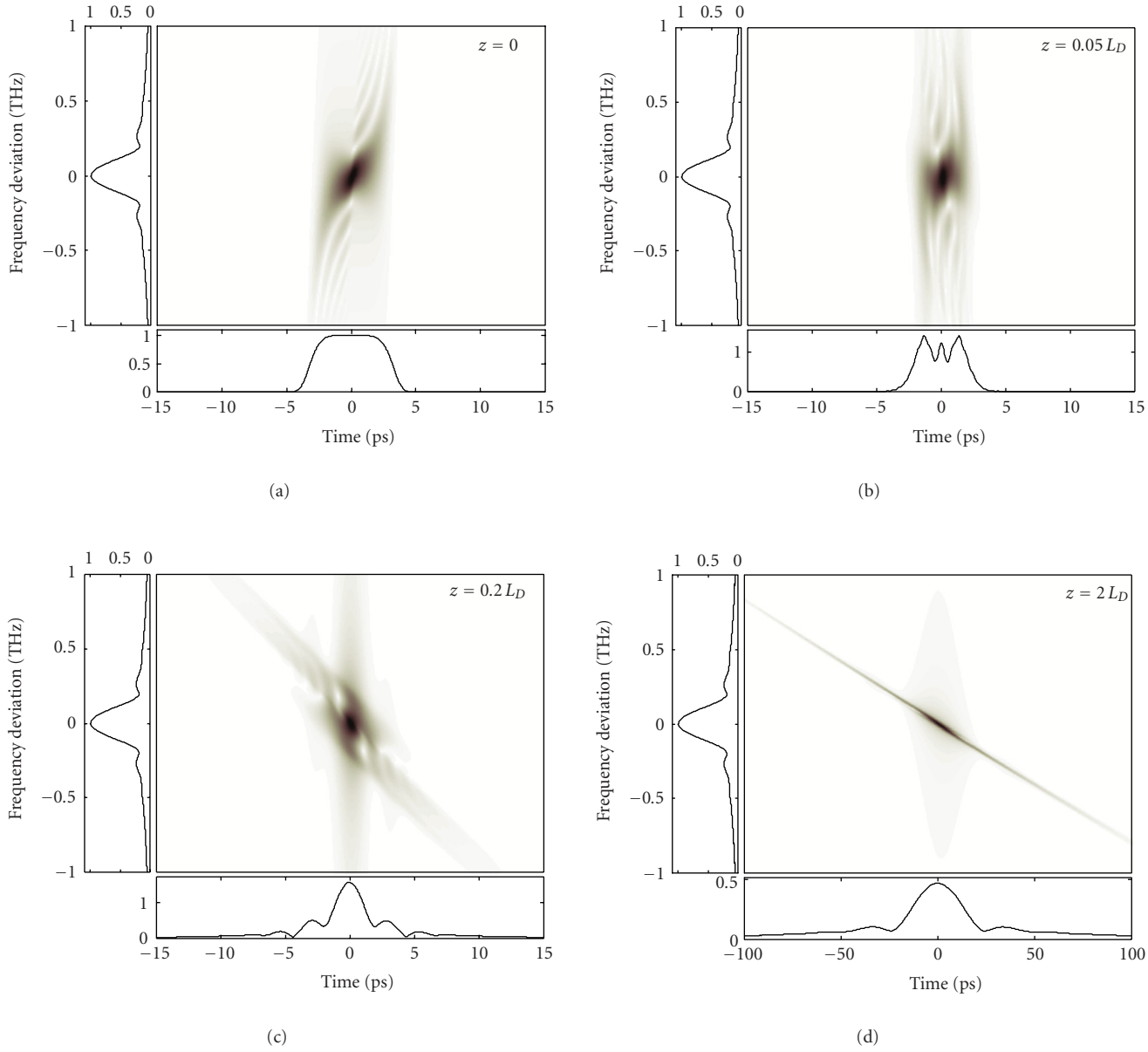


Figure 2: Wigner analysis of linear propagation of a chirped optical pulse through an optical fiber (first-order dispersion).

Figure 1 analyzes the optical pulse envelope $A(z, \tau)$ evaluated at different fiber propagation distances, $z = 0$ (input pulse), $z = 0.05\,L_D$, $z = 0.2\,L_D$, and $z = 2\,L_D$. As shown in the plot corresponding to the input pulse ($z = 0$), the temporal shape (amplitude) of this pulse is identical to that of the corresponding unchirped (transform-limited) pulse (example shown in Figure 1) but the energy spectrum differs significantly from that of the unchirped case. Similarly, the Wigner distribution clearly corresponds with a nontransform-limited pulse as the different pulse spectral components exhibit now a different mean time delay. In particular, the frequencies in the low-frequency and high-frequency sidelobes lie in the leading and trailing edges of the temporal pulse, respectively, whereas the frequencies in the main spectral lobe are associated with the central, high-energy part of the temporal pulse. The effect of propagation of the chirped pulse through the initial section of the first-order dispersive fiber is essentially different to that observed for the case of a transform-limited pulse. The effect of the fiber medium on the optical pulse can be again modeled as a phase-only filtering process as that described by (6). However, in the initial section of the fiber, the GVD introduced by the fiber will compensate partially the intrinsic positive chirp of the original pulse so that the pulse will undergo temporal compression (instead of temporal broadening as it is typical of transform-limited pulses). For a specific

fiber length, the pulse will undergo its maximum temporal compression (approximately for $z = 0.05\,L_D$, in the example shown here) when total chirp compensation is practically achieved. The Wigner distribution of the pulse confirms that in the case of maximum compression this pulse is approximately a transform-limited signal (where all the spectral components have the same mean time delay). Ideal chirp compensation with a first-order dispersive medium can be only achieved if the original pulse exhibits an ideal linear chirp (in our case, the input pulse exhibits a quadratic chirp). The described compression process of chirped optical pulses using propagation through a suitable dispersive medium has been extensively applied for pulse-compression operations aimed to the generation of (ultra-) short optical pulses [17]. In fact, optical pulse-compression operations have been analyzed in the past using Wigner representations [8]. As the plots corresponding to $z = 0.2\,L_D$ and $z = 2\,L_D$ show, further propagation in the optical fiber after the optimal compression length has a similar effect to that described for the case of transform-limited pulses. Briefly, the spectral components of the pulse are temporally separated thus causing the consequent distortion and broadening of the temporal pulse shape. For sufficiently long fiber distance, a frequency-to-time conversion process (RTFT) can be also achieved (e.g., $z = 2\,L_D$).

3.3. Second-order dispersion of a transform-limited optical pulse

The contribution of second-order dispersion on optical pulses can be introduced in the previous NLSE equation by including the corresponding term as follows:

$$i\frac{\partial A(z,\tau)}{\partial z} - \frac{\beta_2}{2}\frac{\partial^2 A(z,\tau)}{\partial \tau^2} + i\frac{\beta_3}{6}\frac{\partial^3 A(z,\tau)}{\partial \tau^3} + \gamma\left|A(z,\tau)\right|^2 A(z,\tau) = 0, \tag{8}$$

where $\beta_3 = [\partial^3\beta(\omega)/\partial\omega^3]_{\omega=\omega_0}$ is the second-order GVD. The contribution of the second-order dispersion induced by the fiber medium on optical pulses in the picosecond range can be normally neglected as compared with the contribution of the first-order dispersion factor. For optical pulses in the picosecond range, this second-order dispersion contribution becomes important only when the fibers are operated in the vicinity of the so-called zero-dispersion wavelength, where the first-order dispersion coefficient is null. Operating around the fiber zero-dispersion wavelength can be of interest for applications where fiber dispersion must be minimized, for example, to exploit some nonlinearities in the fiber [9]. Conventional single-mode fiber (such most of the fiber currently deployed for optical communication purposes) exhibits zero dispersion around 1.3 μm (the dispersion problem described above is still present at 1.55 μm in this kind of fibers) but especial fiber designs allow shifting the zero-dispersion wavelength to the desired value (e.g., 1.55 μm).

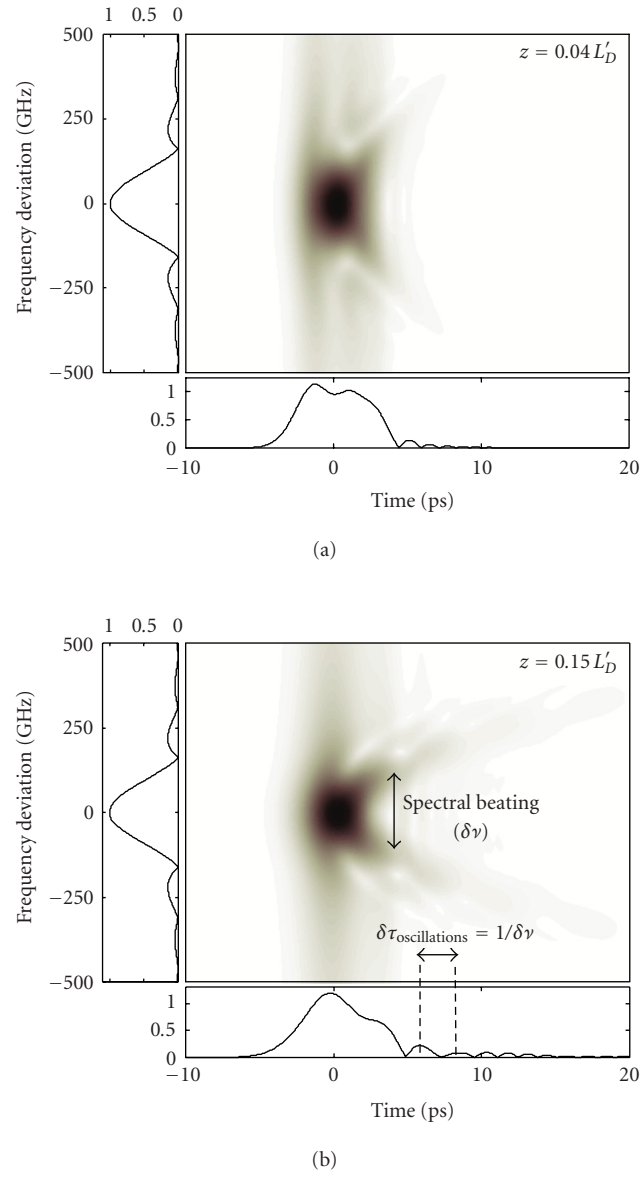


Figure 3: Wigner analysis of linear pulse propagation in an optical fiber (second-order dispersion).

If the first-order dispersion coefficient is null, then the effect of second-order dispersion must be taken into account. In order to evaluate the impact of second-order dispersion on an optical pulse, we will assume that the fiber nonlinearities are negligible as well. In this case, the second and fourth term in the left-hand side of (8) can be neglected and as a result, the optical fiber operates as a linear time-invariant system (i.e., as a filter). In particular,

$$\tilde{A}(z,\omega) \propto \tilde{A}(0,\omega)\exp\left(-j\frac{\beta_3}{2}z\omega^3\right). \tag{9}$$

In Figure 3, the propagation of a super-Gaussian optical pulse similar to that shown in Figure 1 ($z = 0$) through a

second-order dispersive fiber with $\beta_3 = -0.1\,\text{ps}^3\text{Km}^{-1}$ (typical value in a conventional single-mode fiber working at $\lambda \approx 1.3\,\mu$m) is analyzed. In particular, the pulse envelope is analyzed at the fiber distances $z = 0.04\,L_D'$ and $z = 0.15\,L_D'$ where $L_D' = T_0^3/|\beta_3| \approx 270\,\text{Km}$. As expected, the original pulse spectrum is not affected during propagation through the fiber. The Wigner distributions show that these spectral components undergo however a temporal realignment according to the GVD characteristic of the device which in turn causes the observed distortion in the temporal pulse shape. This temporal realignment of the pulse spectral components is very different from that observed in the case of first-order dispersion (compare with Figure 1) as the GVD characteristics in both fibers are different. In the case of second-order dispersion, the original pulse evolves towards a nonsymmetric temporal shape which consists of two components, a main high-energy pulse followed by a secondary component (quasiperiodic sequence of short low-intensity pulses). The oscillatory temporal structure following the main temporal component is a typical result of second-order dispersion. The Wigner distribution provides very useful information about the origin of each one of the components in the obtained temporal signal. In particular, the main temporal pulse in the resulting signal is essentially caused by the frequencies in the main spectral lobe which undergo a similar delay along the fiber (in fact, the Wigner distributions allow us to infer that this main temporal component is closely a transform-limited signal). The subsequent temporal oscillations have their origin in a spectral beating between two separated frequency bands, each one associated with each of the spectral sidelobes of the signal, which appear overlapped in time (i.e., the two beating spectral bands undergo a similar temporal delay during the fiber propagation). Note that the spectral main lobe is affected by a delay shorter than that of the spectral sidelobes (as determined by the fiber GVD). The period of the temporal oscillations is fixed by the frequency separation of the beating bands and as it can be observed, the fact that beating bands are more separated for longer delays translates into the observed oscillation period decreasing with time.

4. WIGNER ANALYSIS OF NONLINEAR PULSE PROPAGATION IN OPTICAL FIBERS

4.1. Self-phase modulation of an optical pulse

When the fiber length is such that $z \ll L_D$ but $z \approx L_{NL}$, then the pulse evolution is governed by the nonlinear effects and the GVD plays a minor role. More specifically, the nonlinear-dominant regime is applicable when the following condition is satisfied:

$$\frac{L_D}{L_{NL}} = \frac{\gamma P_0 T_0^2}{|\beta_2|} \gg 1. \tag{10}$$

In this case, the second term in the left-hand side of the NLSE in (1) (i.e., the dispersion term) can be neglected and the pulse evolution in the fiber is governed by self-phase modulation (SPM), a phenomenon that leads to spectral broadening of the optical pulse. This propagation regime will only occur for relatively high peak power when the dispersion effects can be neglected either because the fiber is operated around the zero-dispersion wavelength or because the input pulses are sufficiently wide (in a conventional single-mode fiber working at $\lambda \approx 1.55\,\mu$m, typical values for entering the SPM regime are $T_0 > 100$ picoseconds and $P_0 \approx 1\,\text{W}$).

SPM has its origin in the dependence of the nonlinear refractive-index with the optical pulse intensity (Kerr effect), which induces an intensity-dependent phase shift along the temporal pulse profile according to the following expression:

$$\begin{aligned} A(z,\tau) &= A(0,\tau)\exp\left(i\phi_{NL}(z,\tau)\right), \\ \phi_{NL}(z,\tau) &= \gamma\left|A(0,\tau)\right|^2 z. \end{aligned} \tag{11}$$

Equation (11) shows that during SPM the pulse shape remains unaffected as the SPM only induces a temporally-varying phase shift. This phase shift implies that an additional frequency chirp is induced in the optical pulse so that new frequency components are generated along the pulse profile. In particular, the SPM-induced instantaneous frequency along the pulse duration is

$$\delta\omega(z,\tau) = -\frac{\partial\phi_{NL}(z,\tau)}{\partial\tau} = -\gamma z\frac{\partial\left|A(0,\tau)\right|^2}{\partial\tau}. \tag{12}$$

Note that according to (11), the maximum SPM-induced phase-shift across the pulse is $\phi_{\text{MAX}} = \gamma P_0 z$. Figure 4 analyzes SPM of a long super-Gaussian pulse ($m = 3$ and $T_0 = 90$ picoseconds) for different values of ϕ_{MAX} (i.e., evaluated at different fiber lengths or for different pulse peak powers). The input pulse is also shown ($\phi_{\text{MAX}} = 0$). The expected spectral pulse distortion and broadening is observed in the plots. The Wigner distribution is an ideal tool to visualize the process of generation of new spectral components as it associates these new spectral components with the temporal features of the optical pulse. The Wigner distribution confirms the generation of new spectral content according to (12). In general, this spectral content generation process is more significant as ϕ_{MAX} increases. Specifically, the steeping edge of the pulse is responsible for the generation of new frequency components in the low-frequency sidelobe (negative side) whereas the trailing edge is responsible for the generation of new frequency components in the high-frequency sidelobe (positive side). The central part of the pulse, where the intensity keeps approximately constant, is only responsible for the generation of new spectral content in the narrow, central frequency band (spectral main lobe). The Wigner distribution reveals that this spectral main lobe is not a transform-limited signal but rather it exhibits a pronounced chirp which becomes more significant as ϕ_{MAX} increases. It is important to note that such important feature of the generated optical pulses cannot be inferred from the basic SPM theory presented above or through the representation of the instantaneous frequency of the signals (i.e., by calculating the derivative of the pulse phase profile).

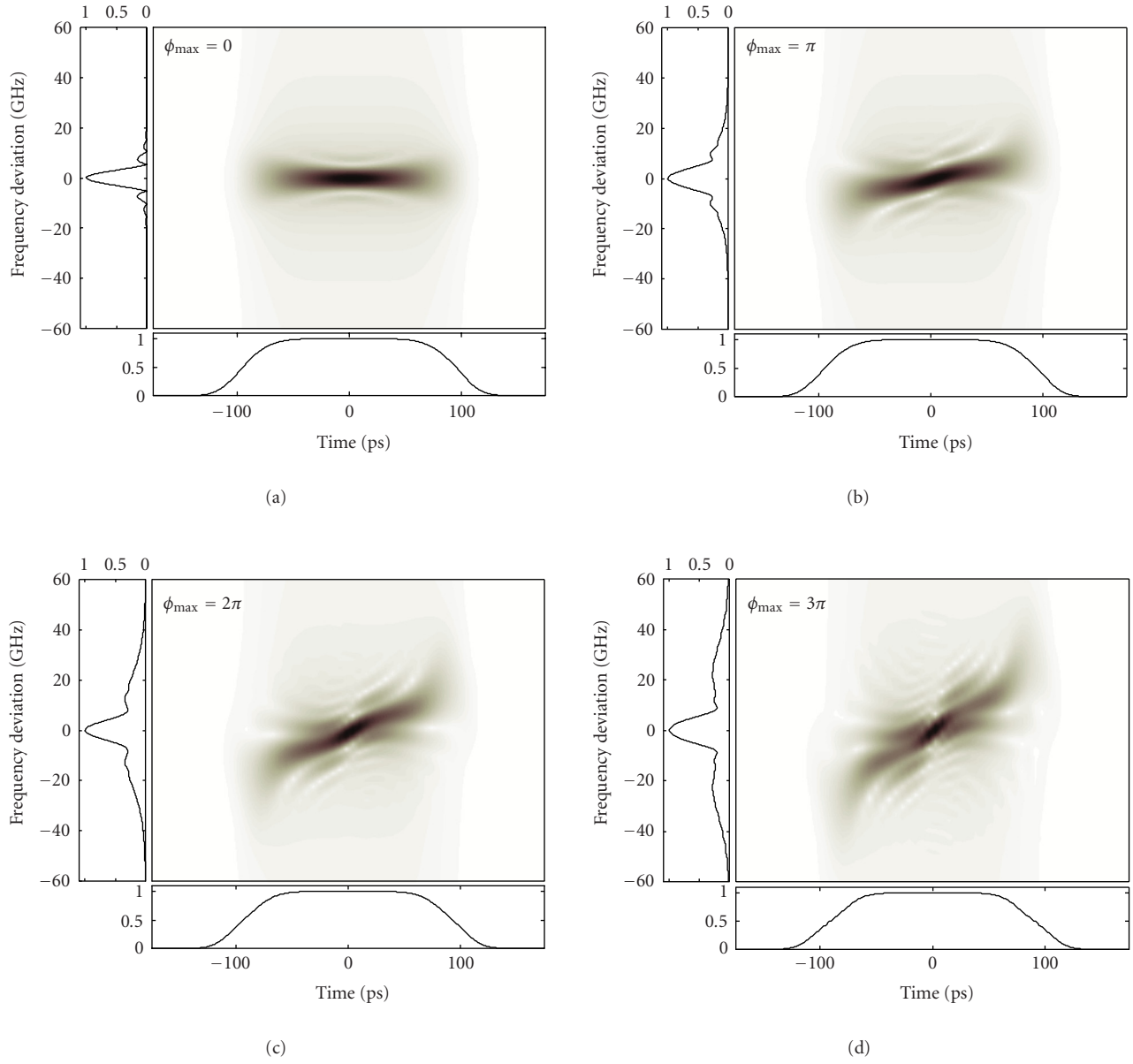


Figure 4: Wigner analysis of pulse self-phase modulation in an optical fiber.

4.2. *Dynamics of temporal soliton formation in the anomalous dispersion regime*

When the fiber length z is longer or comparable to both L_D and L_{NL}, then dispersion and nonlinearities act together as the pulse propagates along the fiber. The interplay of the GVD and SPM effects can lead to a qualitatively different behavior compared with that expected from GVD or SPM alone. In particular, in the anomalous dispersion regime ($\beta_2 < 0$) the fiber can support temporal solitons (bright solitons). Basically, if an optical pulse of temporal shape $A(0,\tau) = \sqrt{P_0}\,\mathrm{sec}\,h(\tau/T_0)$ is launched at the input of the fiber and the pulse peak power is such that it satisfies exactly the following condition $L_D = L_{NL}$, then the pulse will propagate undistorted without change in shape for arbitrarily long distances (assuming a lossless fiber). It is this feature of the fundamental solitons that makes them attractive for optical communication applications. As a generalization, if an optical pulse of arbitrary shape and a sufficiently high peak power (peak power higher than that required to satisfy the fundamental soliton condition) is launched at the input of an optical fiber in the anomalous dispersion regime, then a temporal soliton (*sech* temporal shape) will form after propagation through a sufficiently long section of fiber. The analysis of the dynamics of formation of a temporal soliton is a topic of paramount importance in understanding the nonlinear

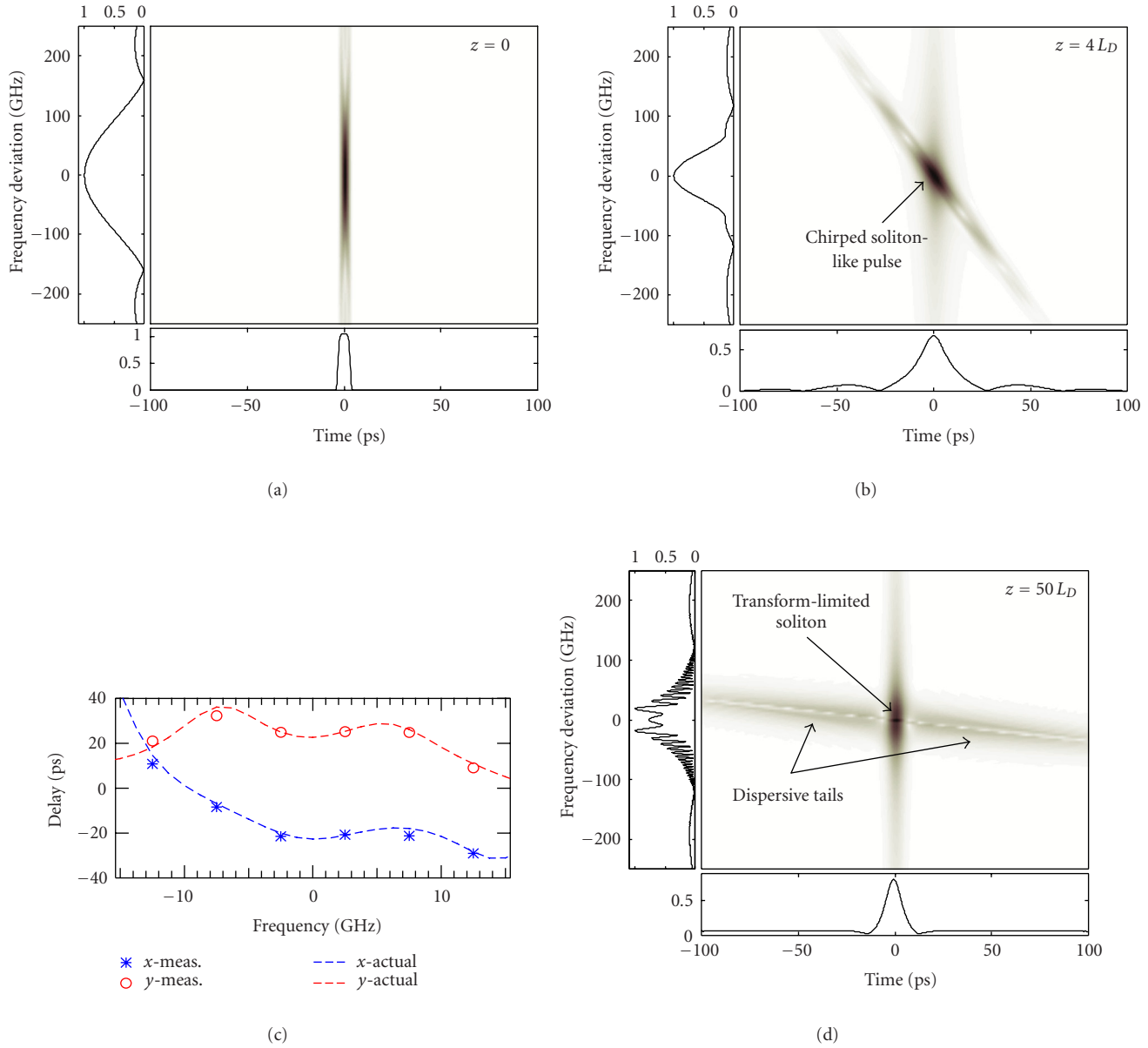


FIGURE 5: Wigner analysis of fundamental soliton formation in an optical fiber (anomalous dispersion regime).

dynamics in optical fibers and has attracted considerable attention [1, 18, 19]. Wigner analysis has been proposed as a simple and powerful method for characterizing optical soliton waveforms (e.g., to evaluate the quality of an optical soliton) [7].

Here, we analyze soliton formation dynamics when the pulse launched at the input of a fiber is not an exact soliton solution in that fiber (deviation in temporal shape). In the example of Figure 5, we assume a fiber with parameters $\beta_2 = -20\,\text{ps}^2\text{Km}^{-1}$ and $\gamma = 2\,\text{W}^{-1}\text{Km}^{-1}$ (typical values for a conventional single-mode fiber working at $\lambda \approx 1.55\,\mu$m). The input pulse is assumed to be a super-Gaussian pulse with $m = 3$, $T_0 = 3$ picoseconds and the peak power P_0 is fixed to satisfy exactly the basic first-order soliton condition, that is, $P_0 = 1.11\,\text{W}$. The Wigner representation of the optical pulse envelope $A(z, \tau)$ is evaluated at different fiber propagation distances, $z = 0$ (input pulse), $z = 4\,L_D$, $z = 20\,L_D$, and $z = 50\,L_D$. The representations in Figure 5 show that for sufficiently long distance ($z > 20\,L_D$) the original super-Gaussian pulse evolves into a signal consisting of (a) a transform-limited first-order temporal soliton and (b) two long dispersive tails. These two well-known features (soliton solution and radiation solution of the NLSE, respectively) can be visualized very clearly in the corresponding two-dimensional Wigner representations. For distances shorter than that required for soliton formation (e.g., $z = 4\,L_D$),

the characteristics patterns of the two mentioned signal components (soliton + dispersive tails) can be already distinguished in the Wigner representation but this representation shows that the main component is still a nontransform-limited (chirped) soliton-like pulse. This component is the one which finally evolves into a transform-limited soliton by virtue of the interplay between GVD and SPM. Note that when the transform-limited soliton is formed, the pulse spectrum exhibits significant oscillations. These oscillations are typical of soliton formation when the pulse launched at the input of the fiber does not satisfy the exact fundamental soliton conditions (i.e., when the input pulse is slightly different in shape, power, or chirp to the ideal soliton) [18, 19] and can be detrimental for practical applications. The fact that the input pulse must satisfy exactly the soliton conditions in order to avoid the presence of these and other detrimental effects have in part precluded the use of soliton-based techniques for communication applications. The spectral oscillations observed in our plots have been observed experimentally and a physical explanation based on complicated analytical studies has been also given [19]. The Wigner representation provides a simple and direct physical understanding of such spectral oscillations and their more significant features. In particular, these oscillations can be interpreted as Fabry-Perot-like resonance effects associated with interference between the frequencies lying in the transform-limited soliton pulse and those in the dispersive tails (i.e., same frequencies with different delays). The period of these oscillations is then fixed by the temporal delay between the interfering frequency bands. Note that the delay between interfering bands (horizontal distance in the Wigner plane) increases for a higher frequency deviation and this translates into the observed oscillation period decreasing as the frequency deviation increases. A similar explanation can be found for the observed variations in the period of the spectral oscillations as a function of the fiber length. Since the dispersive tails are affected by the fiber GVD whereas the soliton pulse is unaffected, the temporal distance between interfering bands increases as the fiber distance increases and this results into the observed oscillation period decreasing with fiber length.

4.3. Optical wave breaking phenomena in the normal dispersion regime

Soliton phenomena can also occur when the optical fiber exhibits normal dispersion ($\beta_2 > 0$) at the working wavelength. In this case, a different class of temporal solitons is possible, that is, the so-called dark soliton, which consists of an energy notch in a continuous, constant light background [1]. Although the dark soliton is of similar physical and practical interest than the bright soliton, in this section, we have preferred to focus on other similarly interesting phenomena that are typical of nonlinear light propagation in the normal dispersion regime (e.g., optical wave breaking [20, 21]) and have no counterpart in the anomalous dispersion regime.

In Figure 6, we analyze the combined action of dispersion and nonlinearities on a Gaussian pulse ($m = 1$, $T_0 = 3$ picoseconds) along a fiber with normal dispersion ($\beta_2 = +0.1\,\mathrm{ps^2Km^{-1}}$). The peak power of the pulse is fixed to ensure that the nonlinear effects (self-phase modulation, SPM) are much more significant than the dispersive effects and, in particular, $L_D = 900\,L_{NL} \approx 90\,\mathrm{Km}$, so that $P_0 \approx 5\,\mathrm{W}$. Figure 2 shows the Wigner representation of the optical pulse envelope $A(z,\tau)$ evaluated at different propagation distances, $z = 0$ (input pulse), $z = 0.02\,L_D$, and $z = 0.06\,L_D$ (note that the figure at the bottom right is a detailed analysis or "zoom" of the temporal response at $z = 0.06\,L_D$). At short distances (e.g., $z = 0.02\,L_D$), the pulse is mainly affected by SPM and as expected, the temporal variation of the spectral content (i.e., instantaneous frequency) is determined by the temporal function $\partial|A(z,\tau)|^2/\partial\tau$, see (12). The oscillations in the pulse spectrum can be interpreted again as a Fabry-Perot-like resonance effects (i.e., these oscillations have their origin in interference between the same spectral components located at different instants of time). For a distance $z = 0.06\,L_D$, the pulse energy is temporally and spectrally redistributed as a result of the interplay between dispersion and SPM [20]. The Wigner distribution provides again a simple understanding of the observed temporal and spectral pulse features. In particular, the Fabry-Perot resonance effects described above appear again and are responsible for the observed oscillations in the main spectral band. The temporal pulse evolves nearly into a square shape slightly broader than the input Gaussian pulse. This square pulse exhibits a linear frequency chirp practically along its total duration. This fact has been used extensively for pulse compression applications [17]. It is also important to note that the pulse spectrum exhibits significant sidelobes. From the Wigner representation, it can be easily inferred that these sidelobes are responsible for the observed oscillations in the leading and trailing edges of the temporal pulse. A more detailed analysis of the temporal oscillations in the trailing (leading) edge of the pulse shows that these oscillations have their origin in a spectral beating between two separated spectral bands located in the high-frequency (low-frequency) sidelobe of the pulse spectrum. The whole process by virtue of which the temporal pulse develops the described temporal oscillations in its edges associated with sidelobes in the spectral domain is usually referred to as optical wave breaking [20, 21]. Our results show that the Wigner analysis constitutes a unique approach for visualizing and understanding the physics behind this well-known phenomenon.

5. CONCLUSIONS

In summary, the Wigner analysis has been demonstrated to be a powerful tool for investigating picosecond pulse propagation dynamics in optical fibers in both the linear and nonlinear propagation regimes. This analysis provides a simple, clear, and profound insight into the nature of the physical phenomena that determine the pulse evolution in an optical fiber, in some cases revealing details about these physical phenomena which otherwise cannot be inferred.

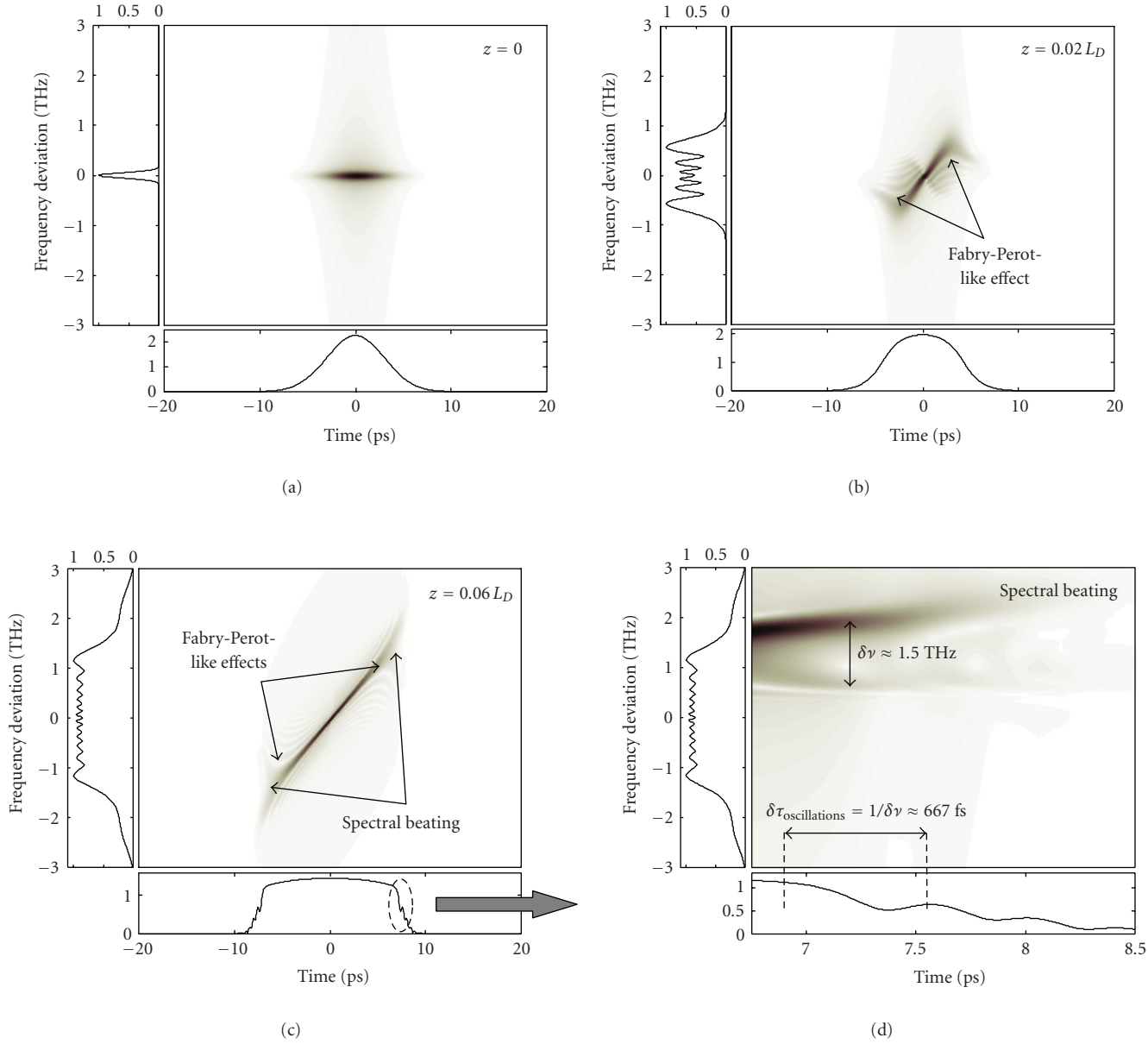


Figure 6: Wigner analysis of nonlinear pulse propagation through an optical fiber in the normal dispersion regime.

The examples in this paper demonstrate the efficiency of the TF (Wigner) techniques for the analysis of linear and nonlinear optical systems and should encourage the application of these techniques to a variety of related problems, such as the systematic study of femtosecond pulse propagation in optical fibers (i.e., influence of high-order dispersion and nonlinear effects) or spatiotemporal dynamics.

In general, the results presented here clearly illustrate how advanced signal processing tools (e.g., TF analysis) can be applied to investigating physical systems of fundamental or practical interest and how the unique information provided by these advanced analysis tools can broaden our understanding of the systems under study.

REFERENCES

[1] G. P. Agrawal, *Nonlinear Fiber Optics*, Academic Press, San Diego, Calif, USA, 3rd edition, 2001.

[2] G. P. Agrawal, *Fiber-Optic Communication Systems*, John Wiley & Sons, New York, NY, USA, 3rd edition, 2002.

[3] L. Cohen, "Time-frequency distributions—a review," *Proc. IEEE*, vol. 77, no. 7, pp. 941–981, 1989.

[4] D. Dragoman, "The Wigner distribution function in optics and optoelectronics," *Progress in Optics*, vol. 37, pp. 1–56, 1997.

[5] J. Paye, "The chronocyclic representation of ultrashort light pulses," *IEEE J. Quantum Electron.*, vol. 28, no. 10, pp. 2262–2273, 1992.

[6] J. Azaña and M. A. Muriel, "Study of optical pulses—fiber gratings interaction by means of joint time-frequency signal representations," *IEEE/OSA J. Lightwave Technol.*, vol. 21, no. 11, pp. 2931–2941, 2003.
[7] D. Dragoman and M. Dragoman, "Phase space characterization of solitons with the Wigner transform," *Optics Communications*, vol. 137, no. 4–6, pp. 437–444, 1997.
[8] V. J. Pinto-Robledo and T. A. Hall, "Chronocyclic description of laser pulse compression," *Optics Communications*, vol. 125, no. 4–6, pp. 338–348, 1996.
[9] K. Mochizuki, M. Fujise, M. Kuwazuru, M. Nunokawa, and Y. Iwamoto, "Optical fiber dispersion measurement technique using a streak camera," *IEEE/OSA J. Lightwave Technol.*, vol. 5, no. 1, pp. 119–124, 1987.
[10] N. Nishizawa and T. Goto, "Experimental analysis of ultrashort pulse propagation in optical fibers around zero-dispersion region using cross-correlation frequency resolved optical gating," *Optics Express*, vol. 8, no. 6, pp. 328–334, 2001.
[11] J. M. Dudley, X. Gu, L. Xu, et al., et al., "Cross-correlation frequency resolved optical gating analysis of broadband continuum generation in photonic crystal fiber: simulations and experiments," *Optics Express*, vol. 10, no. 21, pp. 1215–1221, 2002.
[12] L. Helczynski, D. Anderson, R. Fedele, B. Hall, and M. Lisak, "Propagation of partially incoherent light in nonlinear media via the Wigner transform method," *IEEE J. Select. Topics Quantum Electron.*, vol. 8, no. 3, pp. 408–412, 2002.
[13] D. Yevick and B. Hermansson, "Soliton analysis with the propagation beam method," *Optics Communications*, vol. 47, no. 2, pp. 101–106, 1983.
[14] Y. C. Tong, L. Y. Chan, and H. K. Tsang, "Fibre dispersion or pulse spectrum measurement using a sampling oscilloscope," *Electronics Letters*, vol. 33, no. 11, pp. 983–985, 1997.
[15] J. Azaña and M. A. Muriel, "Real-time optical spectrum analysis based on the time-space duality in chirped fiber gratings," *IEEE J. Quantum Electron.*, vol. 36, no. 5, pp. 517–526, 2000.
[16] P. C. Chou, H. A. Haus, and J. F. Brennan III, "Reconfigurable time-domain spectral shaping of an optical pulse stretched by a fiber Bragg grating," *Optics Letters*, vol. 25, no. 8, pp. 524–526, 2000.
[17] W. J. Tomlinson, R. H. Stolen, and C. V. Shank, "Compression of optical pulses chirped by self-phase modulation in fibers," *Journal of the Optical Society of America B*, vol. 1, no. 2, pp. 139–149, 1984.
[18] C. Desem and P. L. Chu, "Effect of chirping on solution propagation in single-mode optical fibers," *Optics Letters*, vol. 11, no. 4, pp. 248–250, 1986.
[19] M. W. Chbat, P. R. Prucnal, M. N. Islam, C. E. Soccolich, and J. P. Gordon, "Long-range interference effects of soliton reshaping in optical fibers," *Journal of the Optical Society of America B*, vol. 10, no. 8, pp. 1386–1395, 1993.
[20] W. J. Tomlinson, R. H. Stolen, and A. M. Johnson, "Optical wave breaking of pulses in nonlinear optical fibers," *Optics Letters*, vol. 10, no. 9, pp. 457–459, 1985.
[21] J.-P. Hamaide and Ph. Emplit, "Direct observation of optical wave breaking of picosecond pulses in nonlinear single-mode optical fibres," *Electronics Letters*, vol. 24, no. 13, pp. 818–819, 1988.

José Azaña was born on December 8, 1972, in Toledo, Spain. He received the Ingeniero de Telecomunicación degree (a six-year engineering program) and the Ph.D. degree (in the areas of optical signal processing and fiber Bragg gratings) from the Universidad Politécnica de Madrid (UPM) in 1997 and 2001, respectively. From September 2001 to mid 2003, he worked as a Postdoctoral Research Associate in the Department of Electrical and Computer Engineering, McGill University, and he was recently appointed by the Institut National de la Recherche Scientifique, Montreal, as an Assistant Research Professor. His current research interests focus on fiber and integrated technologies for ultrafast optical signal processing and for optical pulse shaping.

EURASIP Journal on Applied Signal Processing 2005:10, 1566–1573

A Novel Optical Vector Spectral Analysis Technique Employing a Limited-Bandwidth Detector

C. K. Madsen

Department of Electrical Engineering, Texas A&M University, College Station, TX 77843, USA
Email: cmadsen@ee.tamu.edu

Received 23 April 2004; Revised 6 December 2004

A new technique for obtaining frequency-dependent magnitude and phase information across an optical channel is presented using tunable allpass optical filters and a detector with a small bandwidth relative to the full channel width. This technique has application to optical monitoring of intersymbol interference distortions, including chromatic and polarization-mode dispersion effects, and thus can provide vector information for input to the control of adaptive optical filters. A method for generating a test signal spanning the spectrum of a modulated data signal without introducing intersymbol interference is discussed. This technique can also be used to characterize an optical pulse source and does not scale in complexity or cost as the bandwidth of the source increases.

Keywords and phrases: intersymbol interference, chromatic and polarization-mode dispersion, optical channel monitoring, optical channel estimation.

1. INTRODUCTION

A novel optical vector spectral analysis (OVSA) technique is proposed that measures the phase and amplitude of a signal using a detector having a limited bandwidth with respect to the full signal bandwidth. When a known signal with spectral components spanning the channel width is launched into a wavelength-division-multiplexed (WDM) optical communication system, it allows the frequency response of the system across the channel bandwidth to be determined. Given the importance of chromatic dispersion and polarization-mode dispersion (PMD) in high-bitrate optical communication systems, the OVSA can be implemented to characterize both the frequency-dependent relative phase and the signal's polarization-dependence. The goal of characterizing the channel, or system response across a limited spectral width, is to provide accurate information on signal distortion. This information may then be used as a signal quality monitor or to update an adaptive filter that compensates for distortions [1, 2] and subsequently allows a higher bitrate times distance product to be achieved for the system. The proposed technique also allows a periodically modulated optical source to be fully characterized.

A previous OVSA technique suggested the use of narrowband optical filters to extract sidebands equally spaced around a return-to-zero (RZ) modulated carrier [3], from which relative delay and dispersion information could be obtained without introducing any special test signals at the transmitter. One narrowband filter (NBF) was required per tone, so a minimum of three filters were required, one at the carrier and one for each sidetone. The detector bandwidth was necessarily on the order of the signal spectral width. It is desired to keep the detector bandwidth as small as possible for low cost and to minimize the complexity of the optical circuit as well. It is also desirable to arbitrarily increase the number of discrete frequencies at which measurements are made to increase the allowable PMD compensation range or baseband data rate or both. The new technique is proposed with these considerations in mind. This approach does not necessarily change the input polarization at the transmitter, unlike [4].

In this paper, we begin with a description of the optical source properties that are critical to our new OVSA method and some applicable source generation techniques. The analyzer is then described theoretically, followed by a presentation of system simulations to verify the expected performance. While the new OVSA can be used to characterize a periodic input pulse stream with an arbitrary waveform shape using a much simpler implementation that assumes a fixed input state of polarization, the explanation is given for an implementation that allows a channel with polarization-dependent loss and polarization-mode dispersion to be characterized. In the latter case, vector analysis implies characterization with respect to the phase response as in a traditional vector network analyzer as well as with respect to the vector nature of the incoming polarization.

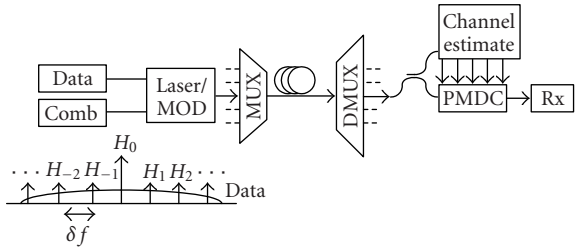


FIGURE 1: Modified transmitter consisting of a laser/modulator driven with data and test frequency comb embedded in a WDM system with a tap on the receiver side for channel estimation. The output spectrum at the transmitter is shown in the inset.

2. OPTICAL SOURCE GENERATION

While phase and magnitude information can be obtained from the modulated signal without modification to the transmitter [3], methods to enhance the signal-to-noise ratio (SNR) and decrease the complexity and cost of the measurement are needed. Test, or training, signals are used in wireless transmission on dedicated frequency and/or time slots to characterize the channel response. In optical communications, burst-mode transmission is unlikely to be introduced for inserting time slot monitors and frequency allocation within the spectrum of a high-bitrate signal, where we wish to measure the system response, since it could produce more distortion than we're trying to measure. Consequently, we focus on adding information to the optical signal that will minimally impact its transmission in terms of intersymbol interference (ISI). Sensitivity to nonlinearities is also important, but it is beyond the scope of this paper. It is proposed to add a test signal to the data signal in the form of a comb with known relative magnitude and phase relationships at the transmitter as shown in Figure 1, where δf may be on the order of 2.5 to 5 GHz for practical purposes. This provides information with a higher SNR than trying to extract the desired information from the equivalent of a pseudorandom bit sequence (PRBS). With the test comb and proposed analyzer circuit, the fast photodetectors need only have enough bandwidth to accommodate δf and not the whole signal bandwidth!

The tones are assumed to be equally spaced in this paper. Equally spaced tones are easy to generate using a step-recovery diode. For completeness, it is noted that chirped tones, whereby each pair of adjacent tones has a slightly different spacing, would produce a beat signal between adjacent pairs at a slightly different frequency so the magnitude and phase between each pair could be resolved by mixing with a swept local oscillator. Mixing two combs with slightly different δf, for example at 2.4 and 2.5 GHz, will produce a series of chirped tones that could be useful as well. Note that the data encoder could be modified to produce tones with a desired frequency spacing in the output spectrum; however, data encoding is typically undertaken over a relatively large block of bits which would produce much lower δf than desired for our purposes. Optical phase modulation, discussed in the last section, appears promising for generating a test signal that introduces minimal ISI on the data signal.

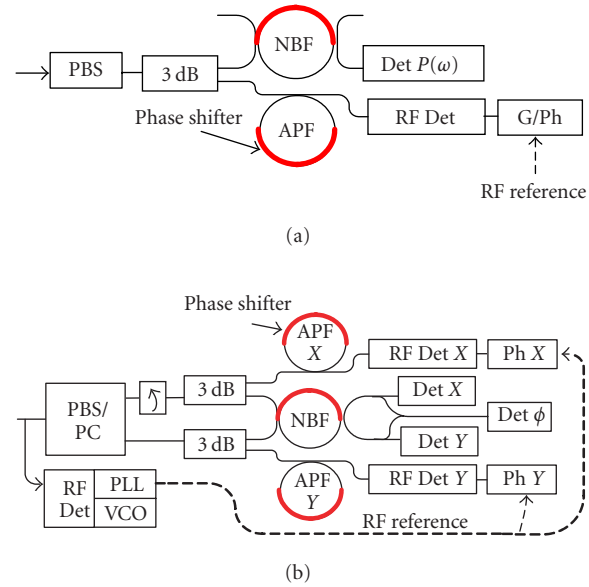


FIGURE 2: Optical vector spectrum analyzer based on a planar lightwave circuit with ring resonators for (a) a single-input polarization and (b) for measuring PMD.

3. A NOVEL OPTICAL VECTOR SPECTRUM ANALYZER

The goal of our OVSA is to determine the amplitude of the tones and relative phase between each pair of tones when they are all present at the detector. While a tunable narrowband filter solves the amplitude discrimination problem, determining the relative phase is more challenging. An obvious solution is to have a bandpass filter with width δf or two narrowband optical filters, ring resonators for example, separated by δf, which scan across the signal spectrum. The relative phase between each pair of tones is then determined via quadrature detection; however, one challenge is having the filters precisely separated as they are tuned, since any undesired offset will contribute phase error arising from the ring dispersion. As the filter is tuned, the tone will be substantially attenuated intermediate to the desired sampling points, which could make the tone difficult to track with a phase-locked loop. Our proposed technique relies on the narrowband phase response of allpass filters instead.

The proposed analyzer circuit for a single-input polarization, which uses a single detector with minimal bandwidth and can easily be realized with planar waveguide ring resonators, is shown in Figure 2a. In this architecture, it is assumed that the input polarization is aligned to a principal state of the device, referred to as the x- or y-polarization for a planar waveguide, either by using polarization maintaining components or a polarization controller. For an input signal where information may be carried by either polarization or any combination thereof, simultaneous measurement of orthogonal polarizations is required as shown in Figure 2b. The transfer function of a polarization-dependent system is described by its Jones matrix. In the simplified case of a lossless

system, the transfer matrix is unitary and described by

$$\mathbf{M}(\omega) = \begin{bmatrix} U(\omega) & -V^*(\omega) \\ V(\omega) & U^*(\omega) \end{bmatrix}, \tag{1}$$

where $U(\omega)$ and $V(\omega)$ are the complex transfer functions for each orthogonal polarization. For a system which is polarization independent or where only a single polarization is important, the transfer function is denoted by $H(\omega)$.

In Figure 2b, each polarization is split into a separate path by a polarization beam splitter (PBS) and polarization controller (PC). The PC allows the power in the x- and y-outputs to be controlled so that all of the power is not in one output or the other. One polarization on the output of the PBS/PC is rotated by 90 degrees so that the outputs going into the 3 dB couplers have the same polarization, that is, either x- or y-polarization. An integrated polarization beam splitter with a polyimide half-waveplate have been demonstrated [5]. After the 3 dB coupler, one portion of each polarization is analyzed by a tunable narrowband filter (NBF) to obtain the magnitude across the channel, while the other portion is transmitted through a tunable allpass filter (APF) before being detected. Note that both previously orthogonally-polarized signals experience the same NBF response but in counterpropagating directions, so the polarization dependence of the NBF is not a limiting factor. An additional measurement that is critical for PMD-impaired channels is the relative phase between polarizations [6], which is indicated by the Det ϕ in Figure 2b. The relative phase is measured by tapping off a portion of the x- and y-outputs signal and interfering them in a directional coupler. Varying the phase of one of the signals before the coupler, by introducing a phase shifter, allows the relative phase to be determined without ambiguity. Alternatively, a full 90-degree hybrid coupler arrangement as used in coherent detection, where it is used to mix the signal and local oscillator in phase and in quadrature, may be used. The measurement of relative phase and magnitude provides the Jones vector up to a common phase term at each tone.

The filters' resonant frequencies are tuned, for example, by thermo-optic phase shifters. The APFs are identical in principle, but any variations can be compensated via calibration. Each APF is designed to provide a very sharp transition in its phase response from 0 to 2π near resonance [7] as shown in Figure 3. On-resonance, the phase is π. Off-resonance, the phase quickly approaches 0 or 2π. As the resonant frequency is shifted via a phase shifter in the feedback path, the phase response is translated across the channel spectrum and the RF detector records different linear combinations of beats between adjacent tones.

In linear systems terms, the transmission channel for a single polarization has a complex frequency response $H(f)$ which we wish to measure at discrete points defined by the tones. For simplicity, let the tone amplitudes and phases be equal at the transmitter. The response of the system at the nth tone is designated by $H(n\delta f) = h_n e^{j\varphi_n}$, which consists of an amplitude h_n and phase ϕ_n. The allpass filter adds an additional phase term, call it θ_n. The detected photocurrent $p(t)$ is proportional to the square magnitude of the field:

$$p(t) \propto \left| \sum_{-N}^{N} h_n e^{j(\varphi_n+\theta_n)} e^{jn2\pi\delta f t} \right|^2. \tag{2}$$

When the detected photocurrent is mixed with a reference RF signal at the tone difference frequency and lowpass filtered to retain just the δf mixing terms, the in-phase and quadrature photocurrent components are given by

$$\begin{aligned} i &= \langle p(t)\cos(2\pi\delta f t)\rangle = \sum_{n=-N}^{N-1} h_n h_{n+1} \cos\left(\Delta\varphi_{n+1} + \Delta\theta_{n+1}\right), \\ q &= \langle p(t)\sin(2\pi\delta f t)\rangle = \sum_{n=-N}^{N-1} h_n h_{n+1} \sin\left(\Delta\varphi_{n+1} + \Delta\theta_{n+1}\right), \end{aligned} \tag{3}$$

where $\Delta\varphi_{n+1} = \varphi_{n+1} - \varphi_n$ and similarly for $\Delta\theta_n$. The brackets indicate lowpass filtering by averaging over time. In the ideal case of a lossless APF whose pole magnitude approaches unity, $\Delta\theta_n$ approaches 0 or π (modulo 2π) and the amplitude response is unity. The phase response becomes more gradual around the resonant frequency as the pole magnitude decreases as shown in Figure 3 for a single-stage APF at several values of the pole magnitude r and for zero frequency offset. Independent of the pole magnitude, the phase at resonance is shifted by π relative to the antiresonant frequency ($\omega = +/-\pi$). The idealized case is assumed in the following discussion for illustrative purposes. In practice, the tone spacing relative to the APF's free spectral range (FSR) determines the pole magnitude required for the phase to change fast enough to make the phase difference approximation by 0 or π valid. After the initial explanation, we show how to remove this restriction entirely and also accommodate filters with loss.

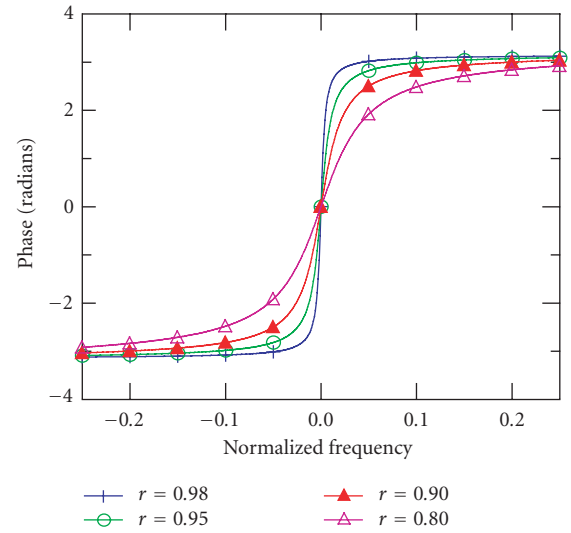


Figure 3: A single-stage APF phase response for various values of the pole magnitude.

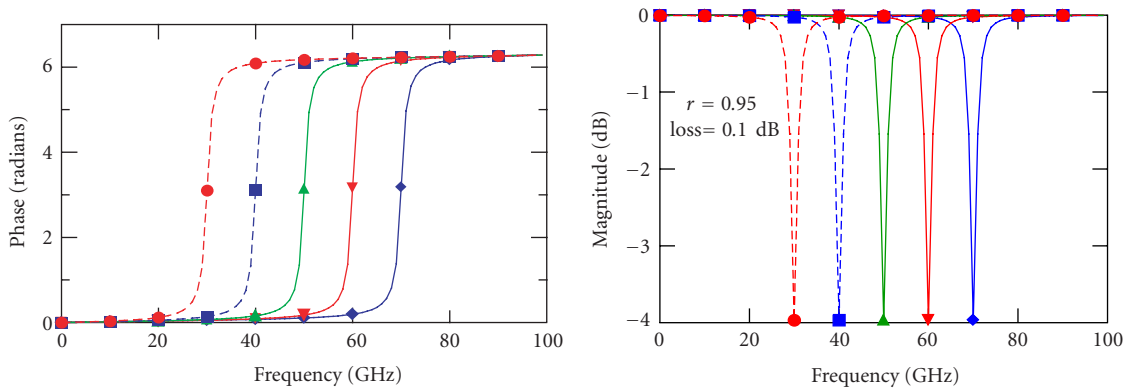


FIGURE 4: Allpass filter phase (left) and magnitude (right) response at several different frequency offsets with the markers indicating the tone locations.

As the APF response is tuned across the tones, the detected i and q change. For the case of a discrete input spectrum, we can represent the phase difference resulting from discrete frequency shifts of an idealized APF response, where zero shift corresponds to the center tone's frequency, using a Kronecker delta function. Let the APF's resonant frequency offset be $m\delta f$ for a particular i_m and q_m measurement; then the phase difference introduced by an idealized APF (lossless and pole magnitude approaches unity) between the nth and $n+1$st tones is given by $\Delta\theta_{m,n} \approx \pi[\delta(n-m) - \delta(n-1-m)]$ modulo 2π. Consider a case with three tones. For the first measurement, the APF resonant frequency is set out of band, let $m = -2$, so $\Delta\theta_{0,0} = \Delta\theta_{0,1} = 0$ and the quadrature components are

$$\begin{aligned} i_1 &= h_{-1}h_0 \cos(\Delta\varphi_0) + h_0h_1 \cos(\Delta\varphi_1), \\ q_1 &= h_{-1}h_0 \sin(\Delta\varphi_0) + h_0h_1 \sin(\Delta\varphi_1). \end{aligned} \tag{4}$$

For the second measurement, the resonant frequency is set at $f_0 - \delta f$, or $m = -1$ in this example, so that $\Delta\theta_{1,0} = \pi$ and $\Delta\theta_{1,1} = 0$. Then, the quadrature components are

$$\begin{aligned} i_2 &= -h_{-1}h_0 \cos(\Delta\varphi_0) + h_0h_1 \cos(\Delta\varphi_1), \\ q_2 &= -h_{-1}h_0 \sin(\Delta\varphi_0) + h_0h_1 \sin(\Delta\varphi_1). \end{aligned} \tag{5}$$

Thus,

$$h_0h_1 \cos(\Delta\varphi_1) = \frac{i_1 + i_2}{2}, \qquad h_{-1}h_0 \cos(\Delta\varphi_0) = \frac{i_1 - i_2}{2}, \tag{6}$$

and similarly for the q components so that

$$\Delta\varphi_1 = \tan^{-1}\left[\frac{q_1 + q_2}{i_1 + i_2}\right], \qquad \Delta\varphi_0 = \tan^{-1}\left[\frac{q_1 - q_2}{i_1 - i_2}\right]. \tag{7}$$

By detecting i and q at $2N$ frequency offsets of the APF for an input signal with $2N+1$ tones, two matrix equations result: $\mathbf{i} = \mathbf{Ax}$ and $\mathbf{q} = \mathbf{Ay}$, where $\mathbf{x} = \{h_nh_{n-1}\cos(\Delta\varphi_n)\}$, $\mathbf{y} = \{h_nh_{n-1}\sin(\Delta\varphi_n)\}$, and $\mathbf{A} = \{\cos(\Delta\theta_{mn})\}$ for $-N+1 \le n, m \le N$. Note that the elements of $\mathbf{A}$ for this ideal case are either 1 or -1. Magnitude and phase information are given by $h_nh_{n-1} = \sqrt{x_n^2 + y_n^2}$ and $\Delta\varphi_n = a\tan(y_n/x_n)$, respectively. In the ideal three-tone case, $\mathbf{A} = \left[\begin{smallmatrix} 1 & 1 \\ -1 & 1 \end{smallmatrix}\right]$. Once the magnitude response is determined independently at one frequency, then the $\mathbf{x}$ (or $\mathbf{y}$) vector is used to calculate the remaining magnitude values.

The APFs provide a functionality that would otherwise have to be obtained by separating the tones into separate paths (i.e., demultiplexing closely spaced tones) and using a phase shifter in each path to obtain the desired series of linear combinations. Note that only the resonant frequency is tuned for both the NBF and APFs, so the coupling ratios are fixed. This allows simple waveguide layouts for the rings, which can achieve very low roundtrip losses. An underlying assumption is that the APF can be translated over $2N$ points in a time short compared to changes in the system response. An advantage of working with δf's in the range of wireless carrier frequencies, such as 2.5 GHz, is that low-cost RF signal processing chips are available. The AD8302 gain and phase measurement circuit by analog devices that operates up to 2.7 GHz is one example. By synchronizing the sweep of the NBF and APFs, the location of the tones in the magnitude response can be used to trigger the sampling of i and q. For quadrature detection, a reference signal at the tone frequency is required. This may be obtained by locking a voltage-controlled oscillator (VCO) to the detected RF tone at δf using a phase-locked loop (PLL) as shown in Figure 2 in a manner similar to clock recovery on a data signal. Since we are tuning the APF, the tones will never completely drop out during the sweep.

With practical APFs, the pole has a value less than unity, causing the phase to vary more gradually across the resonant frequency region, and the feedback path has some loss, which produces a frequency-dependent amplitude response as shown in Figure 4. The elements of $\mathbf{A}$ are no longer 1 or -1. The nonidealities of the APF are now included in matrix equations given by

$$\begin{aligned} \mathbf{i} &= \mathbf{Ax} - \mathbf{By}, \\ \mathbf{q} &= \mathbf{Bx} + \mathbf{Ay}, \end{aligned} \tag{8}$$

where the elements of $\mathbf{A}$ and $\mathbf{B}$ are given by $a_na_{n-1}\cos(\Delta\theta_{mn})$

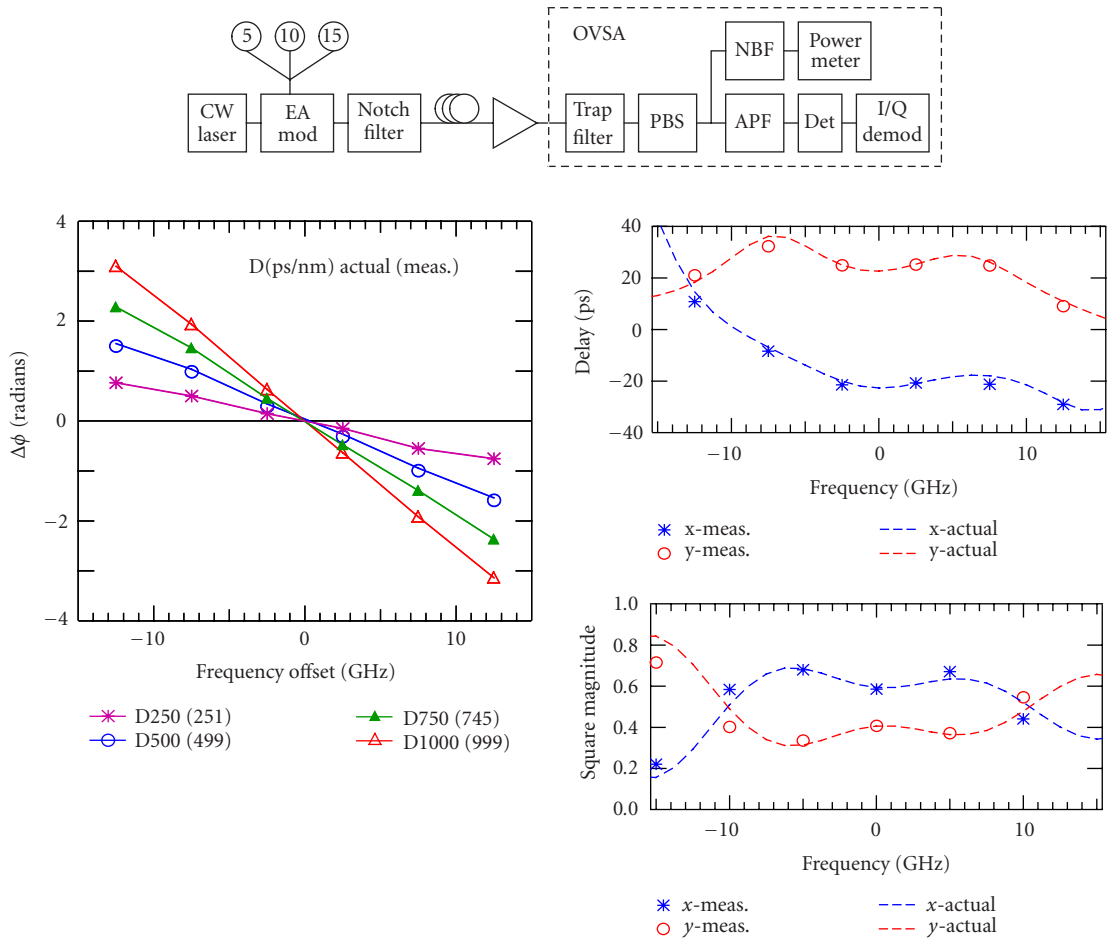


FIGURE 5: (Top) Simulated system. (Left) Dispersion measurement using 7 tones spaced at 5 GHz and (right) measurement of the Jones matrix frequency responses for a system with PMD.

and $a_n a_{n-1} \sin(\Delta\theta_{mn})$, respectively, and the magnitude response of the APF at the tone frequencies is represented by the a_n terms. Note that **A** and **B** depend only on the APF response and not the system, so they can be determined by a one-time calibration procedure. Formula (8) can be rearranged to solve for **x** and **y** as follows:

$$\begin{aligned} \mathbf{x} &= \mathbf{C}^{-1}(\mathbf{B}^{-1}\mathbf{i} + \mathbf{A}^{-1}\mathbf{q}), \\ \mathbf{y} &= \mathbf{C}^{-1}(-\mathbf{A}^{-1}\mathbf{i} + \mathbf{B}^{-1}\mathbf{q}), \end{aligned} \tag{9}$$

where $\mathbf{C} = (\mathbf{A}^{-1}\mathbf{B} + \mathbf{B}^{-1}\mathbf{A})$. The requirement for the tunable APF is that it sample the tones in a manner so that **A** and **B** are nonsingular and therefore invertible. This condition is certainly met with nonideal, but practical, parameters for the APF as shown in Figure 4 of 0.1 dB/roundtrip loss and the pole at 0.95. When the input tones occur mainly on one side or the other of the allpass filter's resonant frequency, the APF's response is dominantly quadratic in phase and it acts like chromatic dispersion from an optical fiber. It is well known that chromatic dispersion, which is characterized by a quadratic phase response, causes fading of RF tones [8].

4. SYSTEM SIMULATIONS

System simulations using commercial software[1] were first run using a comb produced by an electro-absorption modulator driven by 3 frequencies at 5, 10, and 15 GHz (Figure 5). A notch filter was used to attenuate the carrier by 20 dB so that all tones had approximately the same magnitude. The output power after the notch filter was −24 dBm. The relative phase between each pair of adjacent tones was calculated and is plotted in Figure 5 for various cumulative system dispersions. The linear fit for the measured dispersion is within 1% of the actual. Then, the frequency response of a system with simulated PMD, consisting of a cascade of 1000 birefringent plates with random orientations, was tested. The results are shown for both the delay and magnitude in Figure 5. The group delay difference between orthogonal polarizations varies across the channel and has a value of 45 picoseconds at the center frequency. The dominant noise source was out-of-band tones generated from nonlinearities in the modulator. A trapezoidal filter was used to attenuate these unwanted

[1] VPItransmissionMaker by Vitual Photonics, Inc.

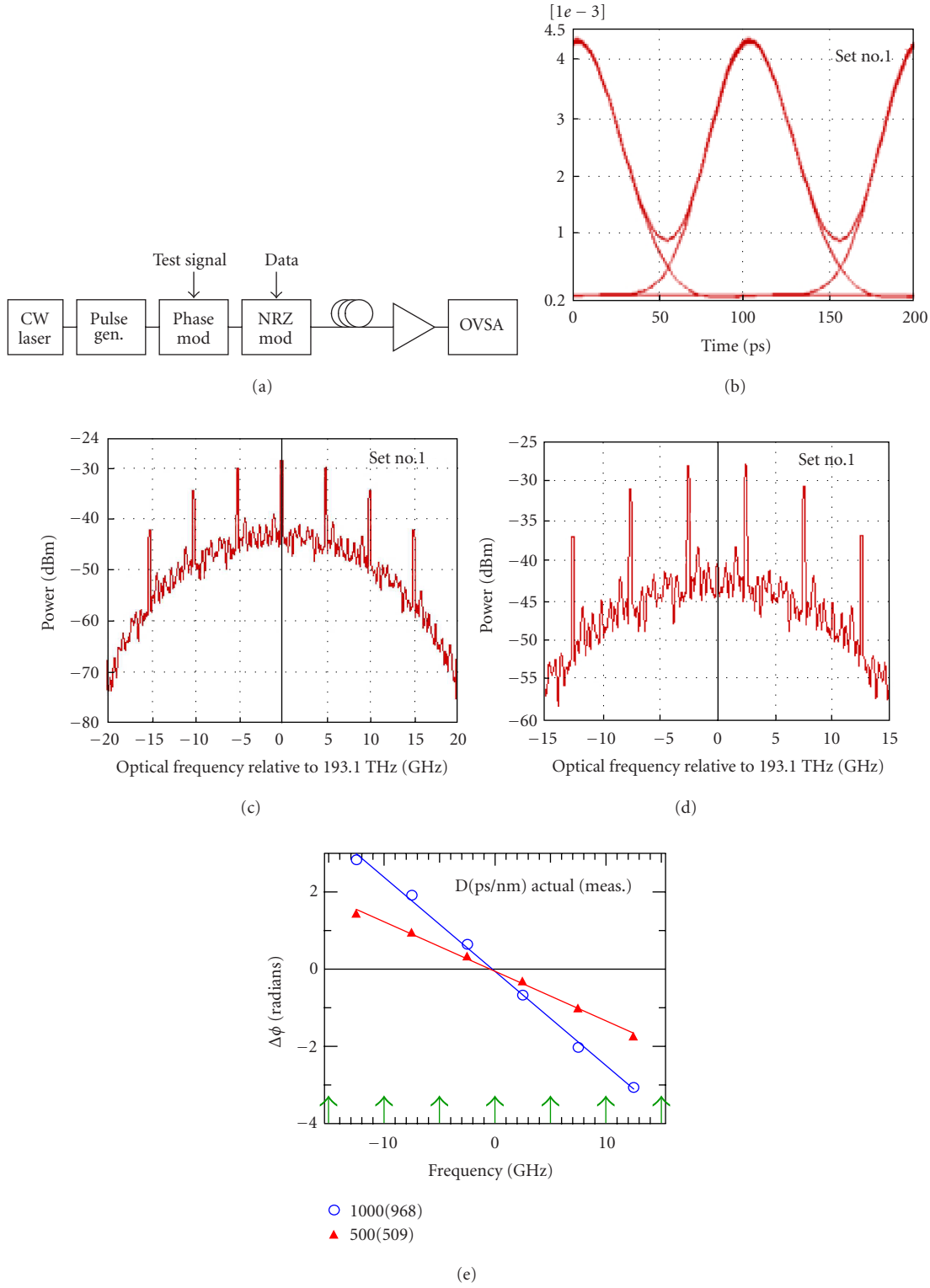


FIGURE 6: (c), (d) An RZ 33% duty cycle signal phase modulated to produce tones at half the bitrate, and (e) measured dispersion. (a) The system simulation schematic and (b) the transmitted eye diagram are shown.

tones by at least 20 dB. A 7th-order bandpass Bessel filter centered at the tone frequency was placed after the detector and before the quadrature demodulation circuit. The input power to the optical amplifier was reduced to increase the

amplified spontaneous emission noise, but it did not impact the measurements, which illustrates an advantage of using narrowband quadrature detection.

In practice, we wish to have the test and data signals coexist. For amplitude modulation and direct detection, phase modulation represents an orthogonal space for introducing the test signal, as shown in Figure 6. An RZ pulse generator was followed by a phase modulator with the repeating pattern $[1010]\pi/2$ and then by an NRZ amplitude modulator for the signal (PRBS in this case) to produce the spectrum in Figure 6c. The resulting test signal is orthogonal to the data using direct detection so it does not introduce ISI. By varying the pattern, which is effectively at half the bitrate in this example, and modulation depth, the tone locations can be translated in frequency. A phase modulation of $[1100]\pi$ produced the spectrum in Figure 6d. Simulated dispersion measurements with 3% or less error from the actual dispersion are shown in Figure 6e for a single polarization, 10 Gbp/s RZ (33% duty cycle) signal with a 24 dB OSNR. Similar results to Figure 5 were also obtained for a system with PMD.

If there is no dispersion in the system, the RF spectrum has tones only at the clock rate. To obtain a reference signal for quadrature detection at half the bitrate, either the clock can be recovered and halved, for example by a divide-by-2 circuit, or APFs can be introduced as in the analyzer circuit itself to break the deconstructive interference condition between adjacent tones under direct detection and produce a detectable RF tone. Another practical consideration is the phase calibration. The relative phases between adjacent tones at the transmitter must be known. A back-to-back measurement was used as the calibration for these simulations.

For polarization multiplexed signals, a slightly different comb frequency difference δf can be used for each polarization and retrieved by mixing with the appropriate local oscillator after detection. Alternatively, the tones can be staggered as indicated in Figures 6c and 6d. Let the Jones vector at the input for each polarization be denoted by $\mathbf{S}_X = [S_X(\omega)\ \ 0]^T$ and $\mathbf{S}_Y = [0\ \ S_Y(\omega)]^T$, where $S(\omega)$ is the input spectrum. The output, measured by a circuit such as shown in Figure 2, is then given by $\mathbf{H} = \mathbf{M}(\mathbf{S}_X + \mathbf{S}_Y)$, where $\mathbf{H} = [H_X(\omega)\ \ H_Y(\omega)]^T$, and

$$\begin{aligned} H_X(\omega) &= U(\omega)S_X(\omega) - V^*(\omega)S_Y(\omega), \\ H_Y(\omega) &= V(\omega)S_X(\omega) + U^*(\omega)S_Y(\omega). \end{aligned} \tag{10}$$

Thus, it is possible to measure the system Jones matrix with this technique without varying the input polarization to the system. For PMD compensation, this provides the information needed to estimate the ideal compensating function, $\mathbf{M}^{-1}(\omega)$.

In summary, a new technique was presented for obtaining phase and magnitude information from a signal using tunable, optical allpass filters. While a particular application was cited, the underlying approach of applying allpass filtering to obtain spectrally dependent phase information without optical heterodyning or without bandlimiting the signal in the optical domain is general. Low-coherence techniques allow one to determine the phase of a device embedded in an interferometer; however, the proposed technique allows one to obtain phase information on the source and, subsequently, of the device or system. Optical modulation is proposed to introduce a test signal that is continuous in time and does not introduce ISI. Channel estimation at discrete frequency points across a noisy signal spectrum was demonstrated via simulation; thus, the available information for controlling PMD compensators and polarization demultiplexers is substantially enhanced. The technique is also applicable to characterizing a periodic pulse stream with an arbitrary waveform shape. The first experimental results for characterization of a periodic pulse stream were reported in [9].

REFERENCES

[1] C. K. Madsen, "Optical all-pass filters for polarization mode dispersion compensation," *Optics Letters*, vol. 25, no. 12, pp. 878–880, 2000.

[2] A. Eyal and A. Yariv, "Design of broad-band PMD compensation filters," *IEEE Photon. Technol. Lett.*, vol. 14, no. 8, pp. 1088–1090, 2002.

[3] C. K. Madsen, "Chromatic and polarization mode dispersion measurement technique using phase-sensitive sideband detection," in *Proc. Optical Fiber Communication Conference (OFC '01)*, vol. 1, pp. MO6-1–MO6-3, Anaheim, Calif, USA, March 2001.

[4] L. Moller and L. Buhl, "Method for PMD vector monitoring in picosecond pulse transmission systems," *J. Lightwave Technol.*, vol. 19, no. 8, pp. 1125–1129, 2001.

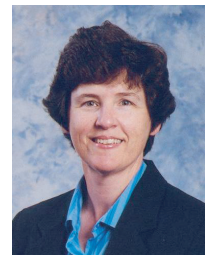
[5] T. Saida, Y. Orihara, H. Yamada, K. Takiguchi, T. Goh, and K. Okamoto, "Integrated optical polarisation analyser on planar lightwave circuit," *Electronics Letters*, vol. 35, no. 22, pp. 1948–1949, 1999.

[6] C. K. Madsen, E. Laskowski, J. Bailey, et al., et al., "An integrated wavelength-sensitive polarimeter," in *Proc. European Conference on Optical Communications (ECOC '02)*, Copenhagen, Denmark, September 2002.

[7] C. K. Madsen and J. Zhao, *Optical Filter Design and Analysis: A Signal Processing Approach*, John Wiley, New York, NY, USA, 1999.

[8] B. Christensen, J. Mark, G. Jacobsen, and E. Bodtker, "Simple dispersion measurement technique with high resolution," *Electronics Letters*, vol. 29, no. 1, pp. 132–134, 1993.

[9] C. K. Madsen, M. Cappuzzo, E. Chen, et al., "A novel optical vector spectrum analysis technique," in *Conference on Lasers and Electro-Optics (CLEO '04)*, San Francisco, Calif, USA, May 2004, Postdeadline CPDD5.



C. K. Madsen, a Professor of electrical engineering at Texas A&M University, College Station, IEEE Senior Member, and an OSA Fellow, received the B.S. degree from the University of Texas at Austin in 1986, the M.S. degree from Stanford University, Standford, Calif, in 1987, and the Ph.D. degree from Rutgers University, Piscataway, NJ, in 1996, all in electrical engineering. She joined AT&T Bell Laboratories in 1987 and worked for the submarine systems business unit. After completing her Ph.D., she transferred to the Integrated Photonics Research Department at Bell Laboratories. Since then, her research has focused on the application of digital filter and signal processing techniques

to optical filters for high-speed, high-capacity optical communication systems. In 1998, Madsen invented a class of tunable, multistage optical allpass filters that allow any phase response to be approximated and have application in chromatic dispersion compensation and polarization-mode dispersion compensation. She has given a short course on "Optical Filters for WDM Systems: Theory, Technologies, and Applications" at OFC and was the 2004 General Chair for the Integrated Photonics Research (IPR) Conference. She was promoted to Distinguished Member of the Technical Staff at Bell Laboratories in 2002 and achieved a Fellow ranking in the Optical Society of America in 2003. She holds 16 US patents and has given over 70 technical talks and papers. In 2004, she joined Texas A&M University as a Full Professor.

EURASIP Journal on Applied Signal Processing 2005:10, 1574–1583

Optical Wavelet Signals Processing and Multiplexing

Gabriella Cincotti
Department of Applied Electronics, University Rome Tre, Via della Vasca Navale 84, 00146 Roma, Italy
Email: g.cincotti@uniroma3.it

Michela Svaluto Moreolo
Department of Applied Electronics, University Rome Tre, Via della Vasca Navale 84, 00146 Roma, Italy
Email: svaluto@uniroma3.it

Alessandro Neri
Department of Applied Electronics, University Rome Tre, Via della Vasca Navale 84, 00146 Roma, Italy
Email: neri@uniroma3.it

Received 31 March 2004; Revised 1 February 2005

We present compact integrable architectures to perform the discrete wavelet transform (DWT) and the wavelet packet (WP) decomposition of an optical digital signal, and we show that the combined use of planar lightwave circuits (PLC) technology and multiresolution analysis (MRA) can add flexibility to current multiple access optical networks. We furnish the design guidelines to synthesize wavelet filters as two-port lattice-form planar devices, and we give some examples of optical signal denoising and compression/decompression techniques in the wavelet domain. Finally, we present a fully optical wavelet packet division multiplexing (WPDM) scheme where data signals are waveform-coded onto wavelet atom functions for transmission, and numerically evaluate its performances.

Keywords and phrases: optical communication system, wavelet transforms, wavelet packet division multiplexing, planar lightwave circuits, optical signal processing.

1. INTRODUCTION

Global communication networks infrastructure has been expanded thanks to recent advances in optical technology, such as transparent photonic switches and ultra-long-haul transmission systems, and the increase of multimedia data traffic is strengthening the demand for ultra-high-capacity photonic networks. The key feature of ultrafast optical networks is that the electrical conversion is avoided until data signals reach the most external edge node. Optical signal processing is not only faster than the electrical one, but it can also support a larger throughput, and current optical networks are evolving toward solutions where photonic routers are replacing electronic routers in the intermediate nodes [1].

In the present paper, we show that the use of wavelet functions in optical communications can provide an overall capacity improvement thanks to the two-dimensional processing capability of multiresolution analysis (MRA), and we present different architectures in standard planar lightwave circuits (PLC) technology that perform the wavelet analysis and multiplexing of data signals directly in the optical domain.

The wavelet transform involves joint time-frequency representation of nonstationary signals using compactly supported basis functions, and MRA has been extensively used to solve a large variety of problems in different research areas. The enormous flexibility in the choice of the wavelet allows the use of optimal wavelets for specifical applications, such as image compression, signal denoising, human vision, radar, earthquake prediction, and computer vision problems, such as range detection or motion estimation [2, 3, 4]. In optics communications, wavelets have been used for time-frequency multiplexing [5, 6] and ultrafast image transmission [7]. In general, wavelet signal analysis is associated with an effective computational algorithm, even faster and simpler than the fast Fourier transform (FFT) algorithm [8].

The aim of the present paper is to present a complete overview of the capabilities of wavelet signal processing and multiplexing in optical communications. In Section 2, we give a brief description of the MRA, illustrating the basic properties of the subband filtering process of both discrete wavelet transform (DWT) and the wavelet packet (WP) decomposition. In Section 3, we present the guidelines to synthesize optical wavelet filters using PLC technology, and

describe some optical implementations for the DWT and WP decomposition. Some numerical examples for optical signal denoising and compression are given in Section 4; in particular, we show that the SNR of an optical signal can be enhanced by performing the DWT of the optical signal and thresholding the detail coefficients. In addition, we evidence the correspondence of the wavelet analysis and the optical packets compression/decompression method using delay line lattice structures [9]. Finally, in Section 5 we present an innovative full optical wavelet packets division multiplexing (WPDM) scheme and numerically test its performance.

The huge bandwidth provided by optical fibers and the capability of signal processing directly in the optical layer make an attractive combination for future multiple access networks. The feasibility of optical code division multiple access (O-CDMA) systems has been demonstrated where the spectral [10] or the temporal [11] encoding results in spreading the optical pulse in the frequency or the time domain. However, a more efficient use of the time-frequency plane has the potential of proving flexible access to a larger number of multiple end users. WPDM is an emerging technique, used in wireless communications, that transmits message signals overlapped in time and frequency domains, and recovers them with very low multiple access interference (MAI) noise, thanks to the orthogonal properties of the wavelet packet functions [12, 13, 14, 15]. Data signals from different users are waveform-coded onto wavelet atom functions and transmitted in a multiple access network. Wavelet atom functions are self-orthogonal against integer translations and mutual-orthogonal due to different subbands occupancy; these properties ensure a better use of time-frequency plane, with respect to standard time division multiplexing (TDM), wavelength division multiplexing (WDM), and CDMA systems [12]. In its standard form, WPDM requires N waveform shapers to code each user bit sequence with a different waveform, and its optical implementation is impracticable [5]. We present a compact, full optical WPDM scheme, where the transmitter is a single encoder that multiplies N optical data signals from N simultaneous users. The encoder gives the equivalent sequence at the root of the WP decomposition tree, and it is followed by a single modulator that completes the waveform coding, shaping the optical pulses of the composite signal with the scaling function profile. The receiver is a reversal version of the transmitter, and the same optical devices are used for both encoding and decoding the binary sequences. In addition, we show that both the WP encoder/decoder and the waveform modulator can be fabricated using PLC technology and integrated on a single device: the WP encoder/decoder can be realized as a tree of lattice-form optical delay line filters [16], whereas the modulator as a weight/phase-programmable tapped delay line filter [17].

2. MULTIRESOLUTION ANALYSIS

The analysis of nonstationary signals involves a compromise between how well transitions or discontinuities are located, and how finely long-term behavior can be identified. For instance, standard Fourier transform decomposes a continuous time-signal $s(t)$ into individual frequency components, using complex sinusoidal basis functions $\exp(i\omega t)$, that are infinite in extent. As these basis functions are periodic, any short duration signal spreads over the whole basis. In contrast, the MRA decomposes a signal at different scales or resolutions, using a basis whose elements are localized in both time and frequency domains, and the representation of short duration and nonstationary signals focuses on a few components, that immediately enlighten the predominant frequencies and the time occurrence of abrupt changes.

Specifically, the continuous wavelet transform (CWT) performs correlations between the signal $s(t)$ and scaled versions of the *mother wavelet* $\psi(t)$:

$$\mathrm{CWT}_s(a,\tau) = \frac{1}{|a|}\int s(t)\psi^*\left(\frac{t-\tau}{a}\right)dt, \tag{1}$$

where a is a real nonzero scale parameter, and τ the translation factor. If $\psi(t)$ satisfies the *admissibility condition*

$$c_\psi = \int \frac{|\tilde{\psi}(f)|^2}{|f|}df < \infty, \tag{2}$$

where $\tilde{\psi}(f)$ is the Fourier transform of the mother wavelet $\psi(t)$, the signal $s(t)$ can be reconstructed by means of the inverse wavelet transform

$$s(t) = \frac{1}{c_\psi}\iint \mathrm{CWT}_s(a,\tau)\frac{1}{\sqrt{|a|}}\psi\left(\frac{t-\tau}{a}\right)\frac{da\,d\tau}{a^2}. \tag{3}$$

Usually $\psi(t)$ is localized both in time and frequency domains, and the CWT displays the time evolution of the frequency components of a signal. In fact, by applying the Fourier Parseval formula [18], (1) can be rewritten as

$$\mathrm{CWT}_s(a,\tau) = \sqrt{|a|}\int \tilde{s}(f)\tilde{\psi}^*(-af)e^{-j2\pi f\tau}df, \tag{4}$$

where $\tilde{s}(f)$ is the Fourier transform of the signal. Therefore, the CWT can be seen as the output from a bank of filters which are constructed by dilatations/compressions of the mother wavelet. Filters obtained by dilatations of $\psi(t)$ process the low-frequency information of the signal $s(t)$, whereas the filters related to the compressed version of $\psi(t)$ analyze the high-frequency content.

More useful in digital signal processing and multiplexing are the orthogonal wavelet series expansions, derived from the CWT when scale and translation factors are constrained to discrete values. Let $\phi(t)$ be a (smooth) scaling function such that the discrete set of functions $\{2^{-\ell/2}\phi(2^{-\ell}t - k\Delta\tau)\ (\ell, k \in Z)\}$ forms an orthonormal basis for a subspace $\mathbf{V}_\ell \in \mathbf{L}^2(\mathbf{R})$; here $\Delta\tau$ is a time interval that will coincide with the inverse of the *free spectral range* (FSR). Let $\psi(t)$ be an admissible mother wavelet such that (a) the discrete set

of functions $\{2^{-\ell/2}\psi(2^{-\ell}t - k\Delta\tau)\ (\ell, k \in Z)\}$ forms an orthonormal basis for a subspace $\mathbf{W}_\ell \in \mathbf{L}^2(\mathbf{R})$; (b) the subspaces $\mathbf{V}_\ell$ and $\mathbf{W}_\ell$ are mutually orthogonal, that is, $\mathbf{W}_\ell \perp \mathbf{V}_\ell$; (c) the subspace $\mathbf{V}_{\ell-1}$ can be expressed as direct sum of $\mathbf{V}_\ell$ and $\mathbf{W}_\ell$, namely

$$\mathbf{V}_{\ell-1} = \mathbf{V}_\ell \oplus \mathbf{W}_\ell. \tag{5}$$

Then, a signal $s(t) \in \mathbf{V}_0$ is represented by a smoothed approximation at resolution 2^M, obtained by combining translated versions of the basic scaling function $\phi(t)$, and M details at the dyadic scales $a = 2^\ell$ $(\ell = 1, 2, \ldots, M-1)$ obtained by combining shifted and dilated versions of the mother wavelet $\psi(t)$:

$$\begin{aligned} s(t) = &\sum_k 2^{-M/2} c_M[k]\phi(2^{-M}t - k\Delta\tau) \\ &+ \sum_{\ell=1}^{M}\sum_k 2^{-\ell/2} d_\ell[k]\psi(2^{-\ell}t - k\Delta\tau). \end{aligned} \tag{6}$$

The wavelet and the scaling function satisfy the *dilation equations*

$$\begin{aligned} \phi(t) &= \sqrt{2}\sum_k h[k]\phi(2t - k\Delta\tau), \\ \psi(t) &= \sqrt{2}\sum_k g[k]\phi(2t - k\Delta\tau), \end{aligned} \tag{7}$$

where $g[k]$ and $h[k]$ are the coefficients of two quadrature mirror filters (QMFs)

$$\begin{aligned} H(\omega) &= \frac{1}{\sqrt{2}}\sum_k h[k]e^{-j\omega k\Delta\tau}, \\ G(\omega) &= \frac{1}{\sqrt{2}}\sum_k g[k]e^{-j\omega k\Delta\tau}, \end{aligned} \tag{8}$$

and $\Delta\tau$ is the inverse of their FSR [19, 20].

The DWT is computed by using Mallat's pyramidal algorithm [8]. Let $c_0[n]$ be the expansion coefficients of the signal $s(t) \in \mathbf{V}_0$, with respect to the orthonormal basis $\phi(t - \Delta n\tau)$:

$$c_0[n] = \langle s(t), \phi(t - n\Delta\tau)\rangle = \int s(t)\phi^*(t - n\Delta\tau)dt, \tag{9}$$

where angular brackets indicate inner product. Then the coefficients $c_0[n]$ can be decomposed into the scaling coefficients $c_1[n]$ and detail coefficients $d_1[n]$, via recursive discrete convolutions with the lowpass $h[n]$ and highpass $g[n]$ filters, respectively, followed by subsampling of factor 2:

$$\begin{aligned} c_1[n] &= \sum_k c_0[k]h[2n - k], \\ d_1[n] &= \sum_k c_0[k]g[2n - k]. \end{aligned} \tag{10}$$

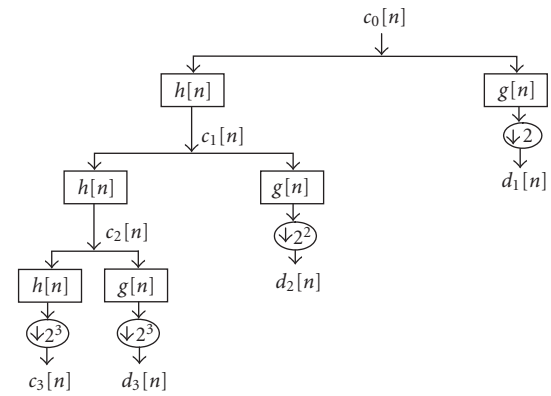


Figure 1: Pyramidal decomposition scheme for the DWT: an optical signal is driven into the device input and the optical wavelet detail and scaling coefficients are obtained at the device outputs. The output signals have to be subsampled according to their decomposition level.

The DWT decomposition halves the time resolution and doubles the frequency resolution, because the frequency band of the output signals spans only half of the frequency band of the original signal; therefore, half of the samples can be discarded by subsampling, without any information loss. On the other hand, the reconstruction process yields the input sequence by the inverse filtering and upsampling:

$$c_0[n] = \sum_k c_1[k]h[2k - n] + d_1[k]g[2k - n]. \tag{11}$$

The decomposition procedure can be further repeated, and, at each decomposition level ℓ, the filtering and subsampling halves both the number of the samples and the frequency band:

$$\begin{aligned} c_\ell[n] &= \sum_k c_{\ell-1}[k]h[2n - k], \\ d_\ell[n] &= \sum_k c_{\ell-1}[k]g[2n - k]. \end{aligned} \tag{12}$$

Here, $c_\ell[k]$ and $d_\ell[k]$ are the scaling and the details coefficients, respectively, at resolution 2^ℓ. This approach reduces the computational load, with respect to the standard FFT, since at each decomposition level a reduced number of samples are processed, and, at the same time, improves both the time and frequency resolution. In fact, high frequencies are resolved better in the time domain and low frequencies better in the frequency domain. The pyramidal scheme for DWT is depicted in Figure 1 and we observe that only the scaling coefficients $c_\ell[n]$ are recursively filtered, whereas the detail coefficients $d_\ell[n]$ are never reanalyzed. In addition, the decimation of the wavelet coefficients is performed by subsampling according to the decomposition level, so that the number of bits in the input signal coincides with the total number of bits in all the DWT coefficients.

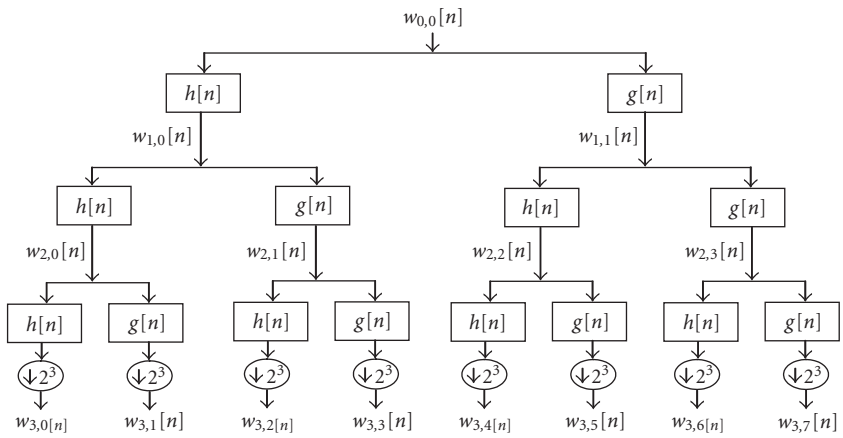


Figure 2: Pyramidal scheme for the WP decomposition: an optical signal is driven into the device input and the optical wavelet packet coefficients are obtained at the device input. The decomposition level is $\ell = 3$, and all the output signals have to be subsampled at $2^\ell = 8$.

The WP decomposition offers a richer signal analysis, with respect to DWT, as both the scaling $c_\ell[n]$ and the detail coefficients $d_\ell[n]$ are recursively decomposed, following the same filtering and subsampling scheme [3]. The decomposition tree corresponding to a full WP analysis is illustrated in Figure 2: in this case, all the outputs have the same number of samples and span over the same frequency bandwidth, because they correspond to the same decomposition level. The WP decomposition is performed by computing the convolution of the input signal $c_0[n]$ with the wavelet atom functions that are a set of functions defined in the following recursive manner:

$$w_{\ell+1,2m}(t) = \sum_k h[k] w_{\ell,m}(t - 2^\ell k\Delta\tau),$$
$$w_{\ell+1,2m+1}(t) = \sum_k g[k] w_{\ell,m}(t - 2^\ell k\Delta\tau). \tag{13}$$

Here ℓ is the decomposition level, m $(0 \le m \le 2^\ell - 1)$ the wavelet atom position in the tree [4]; in addition, the function $w_{0,0}(t)$ coincides with the scaling function $\phi(t)$. By exploiting the recursive structure of (13), a wavelet atom function can be expressed as

$$w_{\ell,m}(t) = \sum_k f_{\ell,m}[k]\phi(t - k\Delta\tau), \tag{14}$$

with $f_{\ell,m}[n]$ being the equivalent filter from the root to the (ℓ, m)th terminal, that can be recursively evaluated using (13). The WP atoms are self- and mutual-orthogonal functions at integer multiples of dyadic intervals [21]:

$$\langle w_{\ell,m}(t - 2^\ell n\tau) w_{\lambda,\mu}(t - 2^\lambda k\tau)\rangle = \delta[\ell - \lambda]\delta[m - \mu]\delta[n - k],$$
$$\ell, \lambda \in Z, \quad 0 \le m \le 2^\ell - 1, \quad 0 \le \mu \le 2^\lambda - 1, \quad n, k \in Z. \tag{15}$$

This property, that is, the waveform orthogonality is used in the WPDM to transmit multiple message signals overlapped in time and frequency domains.

3. SYNTHESIS OF OPTICAL WAVELET FILTERS

To implement the DWT or the WP decomposition of an optical signal, we consider the tree structures of Figures 1 and 2, where each pair $H(\omega)$ and $G(\omega)$ of QMFs is synthesized as a two-port lattice-form planar filter. If an optical digital signal, with bit rate B is driven into the device input, at the device outputs we obtain the DWT or the WP decomposition, up to the ℓth level. In the standard MRA, the outputs have to be subsampled at a factor 2^ℓ, so that the total number of the bits in the input signal coincides with the total number of bits in all the wavelet coefficient signals.

In [22], we presented the design guidelines to synthesize optical wavelet filters, following the algorithm proposed by Jinguji and Oguma [23] and using the power *half-band* (HB) property of wavelet filters:

$$|H(\omega)|^2 + |G(\omega)|^2 = 1,$$
$$G(\omega) = e^{-j\omega\Delta\tau} H^*\left(\omega + \frac{\pi}{\Delta\tau}\right). \tag{16}$$

Jinguji and Oguma showed that an optical HB filter of length M can be fabricated using a Mach-Zehnder interferometer (MZI) with a path delay difference $\Delta\tau$, and $M/2 - 1$ MZIs with a path delay difference of $2\Delta\tau$ [23]. The first MZI has an input 3 dB coupler and no phase shifter inserted in its arms; the remaining circuit parameters can be calculated by applying a recursive algorithm. The synthesis procedure of QMFs is quite immediate, as the circuit parameters are directly related to the filter coefficients [22]. For instance, the Daubechies wavelet filters of length $M = 2$, also known as the Haar wavelet, are

$$\begin{pmatrix} H_{dB1} \\ G_{dB1} \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}. \tag{17}$$

In this case, the optical architecture for the DWT is depicted in Figure 3: each wavelet filter is synthesized as a single MZI

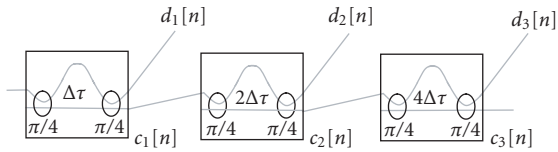


Figure 3: Optical architecture for the DWT with Daubechies wavelet of length $M = 2$. Each optical wavelet filter is an MZI with input/output 3 dB couplers, and FSR that decreases from the root to the leaves in a logarithmic way.

with input/output 3 dB couplers; furthermore, the FSRs in the decomposition tree decrease from the root to the leaves in a logarithmic way, since at each decomposition level ℓ, the subsampling halves the signal frequency band.

The QMFs of Daubechies wavelet of length $M = 4$ are

$$\begin{pmatrix} H_{dB2} \\ G_{dB2} \end{pmatrix} = \frac{1}{4\sqrt{2}} \begin{pmatrix} 1-\sqrt{3} & 3-\sqrt{3} & 3+\sqrt{3} & 1+\sqrt{3} \\ 1+\sqrt{3} & -(3+\sqrt{3}) & 3-\sqrt{3} & -(1-\sqrt{3}) \end{pmatrix}, \tag{18}$$

and the WP decomposition corresponds to the circuital scheme of Figure 4.

4. OPTICAL WAVELET SIGNAL PROCESSING

Wavelets are a powerful tool to denoise signal corrupted by white Gaussian noise [24, 25, 26]. The denoising scheme is based on the principle of selective wavelet reconstruction: in fact an inhomogeneous signal compacts into just a few wavelet coefficients, whereas white noise is distributed over a large number of coefficients, and, therefore, it can be reduced by thresholding the detail wavelet coefficients. The denoised signal is then obtained by inverse wavelet transforming the thresholded coefficients.

The time and frequency localization properties of the wavelet transform can also take the chromatic dispersion into account, since dispersion affects only the detail coefficients that represent highpass-filtered versions of the original signal. In this case it is necessary to perform a selective reconstruction of the wavelet coefficients that will be the subject of a next paper.

To give a numerical example of the proposed denoising method, we consider a 128-bit-long pseudorandom bit sequence (PSRS) at $B = 10$ Gbps modulated by an external Mach-Zehnder modulator with 30 dB extinction ratio; the signal SNR ratio is 24 dB. The device of Figure 3 performs the DWT of the optical signal at $\lambda = 1550$ nm, composed of Gaussian pulses of $\delta t = 20$ ps width; the FSR is chosen equal to $16/\delta t$. At the device outputs we obtain the scaling coefficients $c_3[n]$ at the third level of decomposition ($\ell = 3$), and all the detail coefficients $d_\ell[n](\ell = 1, 2, 3)$, that are plotted in Figure 5. It is evident that all the detail coefficients are the highpass-filtered version of the signal, and they can be suppressed to eliminate the noise; therefore, the denoised signal is reconstructed using only the scaling coefficients $c_3[n]$. The eye diagram of the original and the denoised signals are reported in Figure 6 and we observe a significative improvement in the eye opening. We also remark that the efficiency of the denoising method depends on the wavelet choice, and the threshold selection rules. In the previous example, we consider a very simple case, where all the detail coefficients have been completely eliminated, but better performances can be achieved with optimal thresholding methods.

To quantify the SNR improvement, in Figure 7 we plot both the SNR and the BER of the denoised signal, as functions of the SNR of the input signal, and we observe that the SNR is increased by more than 10 dB.

The device of Figure 3 can be also used to compress/decompress optical data in ultrafast packet-switched networks. In fact the delay lattice scheme proposed by Toliver et al. in [9] is an incomplete DWT scheme that evaluates only the scaling coefficients $c_\ell[n]$. In this case, the unit delay $\Delta\tau$ equates the difference between the pulse periods in the uncompressed and compressed signals. It is evident that a standard DWT or a full WP decomposition allows a larger variety of compression/decompression methods.

5. OPTICAL WAVELET PACKET DIVISION MULTIPLEXING

The success and widespread use of code division multiple access (CDMA) in the wireless domain has renewed interest in exploring its use in the optical domain, which, however, presents a different set of challenges [27]. In an optical CDMA (O-CDMA) network, the information sequences are codified in time, using temporal codes, or in frequency domain with standard diffraction-based spread-spectrum techniques. Recently, hybrid techniques have been proposed to manage burst changes of data traffic or increasing requests of new data services [28]. O-CDMA can support high-capacity services, broadband signals processing, and multiplexing of a large number of users; in addition, many approaches have been proposed to avoid the optical-electrical-optical conversion, that severely limits the transmission system performances.

We present an all-optical WPDM system that makes a very efficient use of time-frequency plane. The digital sequences from each user are encoded by a set of orthogonal waveforms [12]: the orthogonal properties of the wavelet atoms and their overlapping nature in time and frequency yield an overall system capacity improvement [5].

The TDM sequence $\sigma_{\ell,m}[n]$ from the mth user at bit rate $B = 1/T$ is waveform-coded by the wavelet atom function $w_{\ell,m}(t)$:

$$s_{\ell,m}(t) = \sum_n \sigma_{\ell,m}[n] w_{\ell,m}(t - 2^\ell n\Delta\tau), \tag{19}$$

and summed together with the waveform-coded signals from the other $2^\ell - 1$ users:

$$s_c(t) = \sum_{m=0}^{2^\ell - 1} \sum_n \sigma_{\ell,m}[n] w_{\ell,m}(t - 2^\ell n\Delta\tau). \tag{20}$$

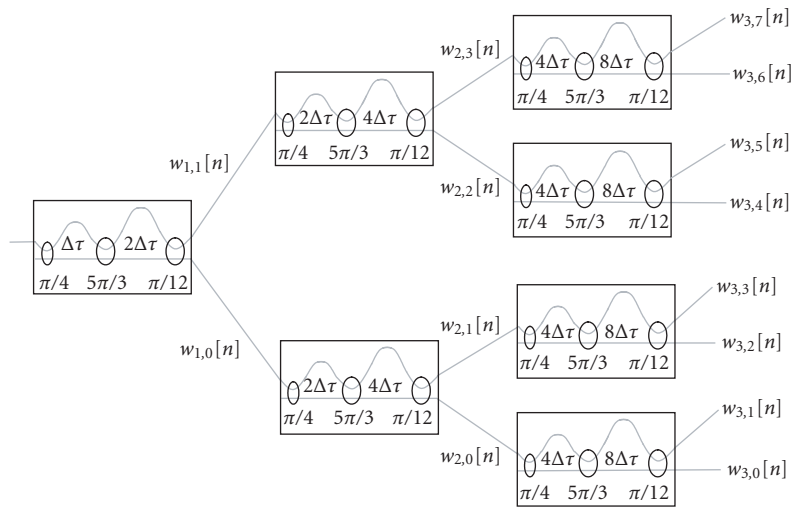


Figure 4: Optical architecture for the WP decomposition with Daubechies wavelet filter of length $M = 4$. Each optical wavelet filter is a chain of two MZIs, with FSR that decreases from the root to the leaves in a logarithmic way.

The *processing gain* of the WPDM system is equivalent to the O-CDMA parameter and is expressed as the ratio of the time duration of the waveform and the input signal width δt [29]:

$$F = \frac{2^{\ell}\Delta\tau}{\delta t}. \tag{21}$$

For a correct waveform shaping, it is necessary that the input bit duration δt equates the inverse of the FSR, that is, $\delta t = \Delta\tau$: in this way, each optical pulse is transformed into the corresponding wavelet atom function at the device output. Therefore, the processing gain $F = 2^{\ell}$ equates the number of simultaneous users.

A standard WPDM transmission system consists of a bank of waveform modulators, one for each user, and its optical implementation is really complex [5]. A more compact WPDM architecture, with a single WP encoder and only one waveform modulator, can be designed by "reversing" the expression of (20) for the composite signal $s_c(t)$. In fact, if we substitute (14) into (20), we obtain

$$\begin{aligned} s_c(t) &= \sum_{m=0}^{2^{\ell}-1} \sum_{n} \sigma_{\ell,m}[n] \sum_{k} f_{\ell,m}[k]\phi[t - (2^{\ell}n + k)\Delta\tau] \\ &= \sum_{i} \sigma[i]\phi(t - i\Delta\tau), \end{aligned} \tag{22}$$

where

$$\sigma[i] = \sum_{m=0}^{2^{\ell}-1} \sum_{n} f_{\ell,m}[i - 2^{\ell}n]\sigma_{\ell,m}[n] \tag{23}$$

is the equivalent sequence from all the users obtained at the root of the WP tree [12]. From an inspection of this equation, it is evident that the composite signal $s_c(t)$ can be generated by first encoding all the input sequence $\sigma_{\ell,m}[n]$ from each user with a WP encoder, thus obtaining $\sigma[i]$, and then waveform shaping all the composite bits by the scaling function $\phi(t)$ [6]. This WPDM architecture is schematically illustrated in Figure 8: the WPDM transmission system is implemented by the WP encoder of Figure 2, followed by a single modulator that shapes the multiplexed optical pulses $\sigma[i]$ with the scaling function profile $\phi(t)$. The receiver is a time reversal version of the transmitter, and the WP decoder follows the pulse shaper; therefore, the same device can be used for both encoding and decoding the binary sequences.

The time gating opens a time window to extract the autocorrelation peak and it is somewhat equivalent to narrowbandpass filtering in the wireless CDMA. The use of a time gating before detection significantly relaxes the requirements for the detector bandwidth to the bit rate of $2^{\ell}\Delta t$. The synchronization from the time gating can be derived from the recovered clock.

The optical architecture for a WP encoder/decoder is identical to an optical WP decomposition scheme, so that the scheme of Figure 4 can be thought of as optical E/D for the Daubechies wavelets of length $M = 4$.

In a WPDM system, each bit from each user is waveform-coded by wavelet atom functions, and all the data streams from the users are transmitted simultaneously. On the other hand, in a code-based O-CDMA system, a different code is assigned to each user; therefore, if we remove the pulse shapers in the transmission scheme of Figure 8, we obtain a standard O-CDMA system, where the input data signals are coded by optical orthogonal codes (OOCs) sequences [30]. In fact, the device of Figure 4 is a full optical E/D that codes/decodes binary sequences from 2^{ℓ} users simultaneously [31, 32]; of course, for an O-CDMA system, it is $\delta t < \Delta\tau$.

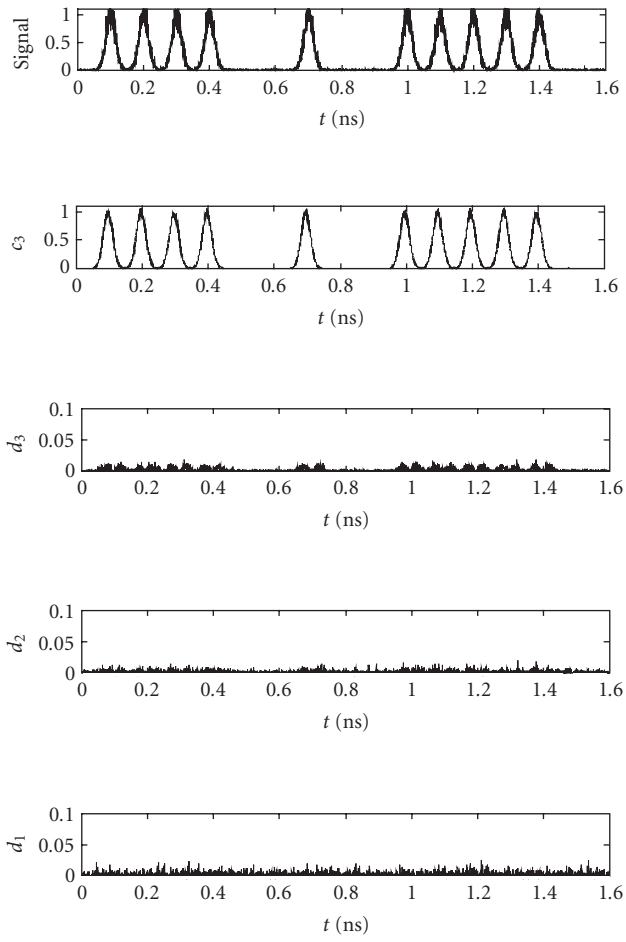


FIGURE 5: Input optical signal at $B = 10$ Gbps, $\lambda = 1550$ nm, and SNR = 24 dB; the Gaussian optical pulses have 20 ps width. Scaling coefficients c_3 that coincide with the denoised signal and thresholded detail coefficients at levels 1, 2, 3 are obtained. The DWT is performed with the device of Figure 3, with no subsampling.

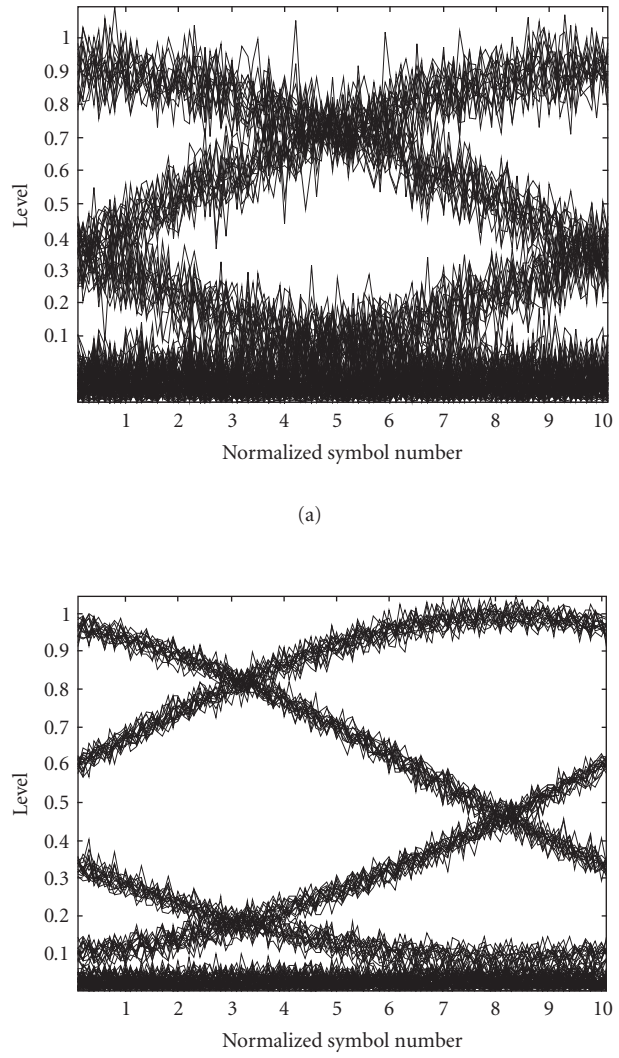


FIGURE 6: Eye pattern diagrams of the input signal and the denoised signal of Figure 5.

To complete the waveform coding in a WPDM system, an optical pulse shaper is necessary. We consider two different devices; the first one is a diffractive pulse shaper, identical to an O-CDMA spread-spectrum encoder, composed of two diffraction gratings and an amplitude/phase filter [33]. The incoming pulses are spatially decomposed in their spectral components by the first grating, and an amplitude/phase mask gives the scaling function spectral profile to the radiation; the second grating reassembles the waveform-coded pulses. Otherwise, an optical pulse shaper can be fabricated as a weight/phase-programmable optical-tapped delay line filter [17]. In this case, both the WP E/D and the waveform shaper can be integrated on a single silica substrate.

To evaluate the detection capabilities of a full optical WPDM system, we generate N independent data signals at $B = 10$ Gbps modulated by $2^7 - 1$ word length PRBSs, using external Mach-Zehnder modulators with 30 dB extinction ratio. Each user bit sequence is forwarded to the N inputs of the encoder of Figure 4 and the encoder output is waveform-modulated by the pulse shaper corresponding to the scaling wavelet of the Daubechies filters of length $M = 4$, as illustrated in Figure 8. Figure 9 shows the signal-to-MAI ratio as a function of the number of the simultaneous users N, that is evaluated as the ratio between the signal corresponding to a "1" from the ith user and the signal detected at the same output when the ith user is transmitting a "0" and all the other $N - 1$ users are transmitting a "1."

We finally observe that the system performances can be enhanced by a suitable choice of the wavelet decomposition [12].

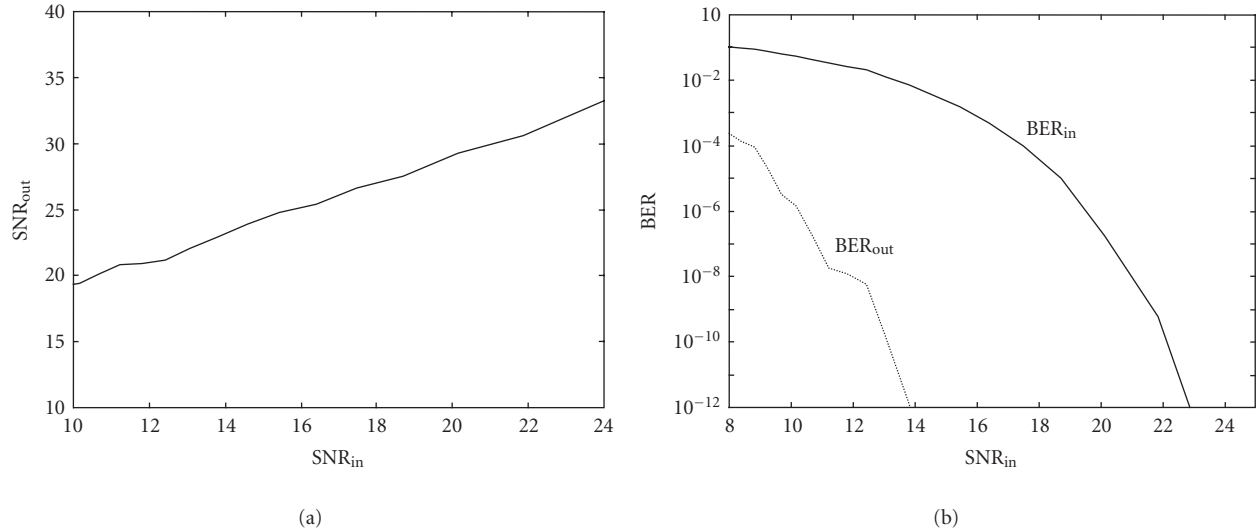


FIGURE 7: (a) SNR of the denoised signal (that coincides with the scaling coefficients c_3 from the device of Figure 3) versus the SNR of the input optical signal. (b) Bit error rate of the input signal (solid line) and the denoised signal (dotted line) versus the SNR of the input signal.

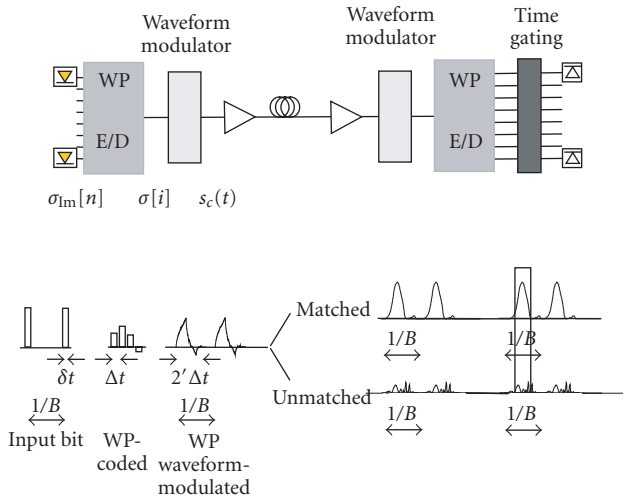


FIGURE 8: Schematic of an optical WPDM system: the data sequences from N users are WP-encoded by a single optical device; the composite signal is then waveform-modulated by a pulse shaper. The receiver is a reverse version of the transmitter.

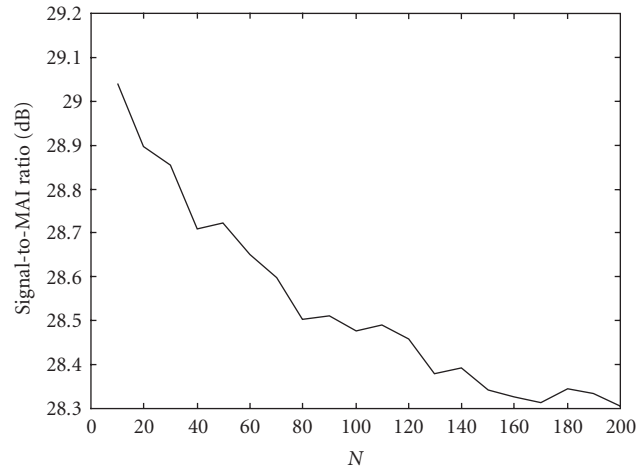


FIGURE 9: Signal-to-multiple-access-interference ratio (dB) versus the number N of simultaneous users.

6. CONCLUSION

We present a complete overview of the wavelet signal processing and multiplexing in the optical domain, using passive PLC devices.

We describe the PLC architectures to perform the DWT and the WP decomposition of an optical digital signal, and furnish the design guidelines to synthesize a wavelet filter as a two-port lattice-form planar device. We demonstrate that within the MRA it is possible to both denoise and compress/decompress data streams directly in the optical domain; in addition we evidence the large flexibility of the proposed approaches, thanks to the possibility to choose the optimal wavelet.

Furthermore, we present an innovative full optical WPDM scheme that transmits multiple signals by waveform coding, yielding an overall capacity improvement and a more efficient use of the common shared resources, with respect to standard TDM and WDM [12]. We show that the encoded sequences are recovered with very low MAI noise, since multiple signals transmitted, overlapped in both time and frequency domains.

The WP encoder/decoder and the pulse shaper are low-loss compact devices that can be fabricated using standard PLC technology and integrated together on a single common mechanically rigid substrate, avoiding undesirable variations of optical path lengths over time, temperature, and exposure to mechanical vibrations. The possibility of growing or pruning the wavelet decomposition tree allows to adapt the multiple access system to different traffic requirements, adding or dropping data streams.

The optical signal processing fully exploits the fibre bandwidth, and it is suitable for broadband multiple access networks. The proposed devices can be employed in ultrafast and highly-robust multiple access networks and their low cost, compactness, and fabrication simplicity, make them attractive for an ever-increasing number of different applications.

REFERENCES

[1] K.-I. Kitayama and N. Wada, "Photonic IP routing," *IEEE Photon. Technol. Lett.*, vol. 11, no. 12, pp. 1689–1691, 1999.

[2] E. A. Rosenfeld, *Multiresolution Techniques in Computer Vision*, Springer-Verlag, New York, NY, USA, 1984.

[3] J. J. Benedetto and M. W. Frazier, *Wavelets: Mathematics and Applications*, CRC Press, Boca Raton, Fla, USA, 1994.

[4] A. K. Louis and P. Maaß, *Wavelets: Theory and Applications*, John Wiley & Sons, New York, NY, USA, 1997.

[5] T. Olson, D. Healy, and U. Österberg, "Wavelets in optical communications," *IEEE Computing in Science & Engineering*, vol. 1, no. 1, pp. 51–57, 1999.

[6] G. Cincotti, M. S. Moreolo, and A. Neri, "All optical multiplexing scheme for multiple access networks based on wavelet packets filter banks," in *Proc. SPIE International Symposium Integrated Photonics Europe*, Strasbourg, France, 2004.

[7] T. Konishi and Y. Ichioka, "Ultrafast image transmission by optical time-to-two-dimensional-space-to-time-to-two-dimensional-space conversion," *Journal of the Optical Society of America {A}*, vol. 16, no. 5, pp. 1076–1088, 1999.

[8] S. G. Mallat, "A theory for multiresolution signal decomposition: the wavelet representation," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 11, no. 7, pp. 674–693, 1989.

[9] P. Toliver, K.-L. Deng, I. Glesk, and P. R. Prucnal, "Simultaneous optical compression and decompression of 100-Gb/s OTDM packets using a single bidirectional optical delay line lattice," *IEEE Photon. Technol. Lett.*, vol. 11, no. 9, pp. 1183–1185, 1999.

[10] R. P. Scott, et al., "Demonstration of an error-free 4 X 10-Gb/s multi-user SPEC O-CDMA network test bed," *IEEE Photon. Technol. Lett.*, vol. 16, pp. 2186–2188, 2004.

[11] K.-I. Kitayama, "Code division multiplexing lightwave networks based upon optical code conversion," *IEEE J. Select. Areas Commun.*, vol. 16, no. 7, pp. 1309–1319, 1998.

[12] K. M. Wong, J. Wu, T. N. Davidson, and Q. Jin, "Wavelet packet division multiplexing and wavelet packet design under timing error effects," *IEEE Trans. Signal Processing*, vol. 45, no. 12, pp. 2877–2890, 1997.

[13] A. R. Lindsey, "Wavelet packet modulation for orthogonally multiplexed communication," *IEEE Trans. Signal Processing*, vol. 45, no. 5, pp. 1336–1339, 1997.

[14] K. M. Wong, J. Wu, T. N. Davidson, Q. Jin, and P.-C. Ching, "Performance of wavelet packet-division multiplexing in impulsive and Gaussian noise," *IEEE Trans. Commun.*, vol. 48, no. 7, pp. 1083–1086, 2000.

[15] R. E. Learned, H. Krim, B. Claus, A. S. Willsky, and W. C. Karl, "Wavelet-packet-based multiple access communication," in *Wavelet Applications in Signal and Image Processing II*, vol. 2303 of *Proceedings of SPIE*, pp. 246–259, San Diego, Calif, USA, October 1994.

[16] G. Cincotti, "Fiber wavelet filters," *IEEE J. Quantum Electron.*, vol. 38, no. 10, pp. 1420–1427, 2002.

[17] K.-I. Kitayama, S. Osawa, N. Wada, and W. Chujo, "Optical pulse train synthesis of arbitrary waveform using weight/phase-programmable 32-tapped delay line waveguide filter," in *Optical Fiber Communication Conference and Exhibit (OFC '01)*, vol. 3, pp. WY3/1–WY3/3, Anaheim, Calif, USA, March 2001.

[18] R. N. Bracewell, *The Fourier Transform and Its Applications*, McGraw-Hill, New York, NY, USA, 1986.

[19] M. Vetterli and C. Herley, "Wavelets and filter banks: theory and design," *IEEE Trans. Signal Processing*, vol. 40, no. 9, pp. 2207–2232, 1992.

[20] C. Herley and M. Vetterli, "Orthogonal time-varying filter banks and wavelet packets," *IEEE Trans. Signal Processing*, vol. 42, no. 10, pp. 2650–2663, 1994.

[21] B. S. Krongold, K. Ramchandran, and D. L. Jones, "Frequency-shift-invariant orthonormal wavelet packet representations," *IEEE Trans. Signal Processing*, vol. 47, no. 9, pp. 2579–2582, 1999.

[22] G. Cincotti, M. S. Moreolo, and A. Neri, "Synthesis of optical wavelet filters," *IEEE Photon. Technol. Lett.*, vol. 16, no. 7, pp. 1679–1681, 2004.

[23] K. Jinguji and M. Oguma, "Optical half-band filters," *J. Lightwave Technol.*, vol. 18, no. 2, pp. 252–259, 2000.

[24] D. L. Donoho and I. M. Johnstone, "Ideal spatial adaptation by wavelet shrinkage," *Biometrika*, vol. 81, no. 3, pp. 425–455, 1994.

[25] L. M. Johnstone and B. W. Silverman, "Wavelet threshold estimators for data with correlated noise," *Journal of the Royal Statistical Society: Series B*, vol. 59, pp. 319–351, 1997.

[26] M. Lang, H. Guo, J. E. Odegard, C. S. Burrus, and R. O. Wells Jr., "Noise reduction using undecimated discrete wavelet transform," *IEEE Signal Processing Lett.*, vol. 3, no. 1, pp. 10–12, 1996.

[27] J. A. Salehi, "Code division multiple-access techniques in optical fiber networks. I. Fundamental principles," *IEEE Trans. Commun.*, vol. 37, no. 8, pp. 824–833, 1989.

[28] T. Pfeiffer, J. Kissing, J.-P. Elbers, et al., "Coarse WDM/CDM/TDM concept for optical packet transmission in metropolitan and access networks supporting 400 channels at 2.5 Gb/s peak rate," *J. Lightwave Technol.*, vol. 18, no. 12, pp. 1928–1938, 2000.

[29] K.-I. Kitayama, N. Wada, and H. Sotobayashi, "Architectural considerations for photonic IP router based upon optical code correlation," *J. Lightwave Technol.*, vol. 18, no. 12, pp. 1834–1844, 2000.

[30] T. Dennis and J. F. Young, "Optical implementation of bipolar codes," *IEEE J. Quantum Electron.*, vol. 35, no. 3, pp. 287–291, 1999.

[31] G. Cincotti, "Full optical encoders/decoders for photonic IP routers," *J. Lightwave Technol.*, vol. 22, no. 2, pp. 337–342, 2004.

[32] PCT patent Application No. PCT/IT03/000879 filed on December 30th 2003 by University Roma TRE, inventor Gabriella Cincotti.

[33] J. A. Salehi, A. M. Weiner, and J. P. Heritage, "Coherent ultrashort light pulse code-division multiple access communication systems," *J. Lightwave Technol.*, vol. 8, no. 3, pp. 478–491, 1990.

Gabriella Cincotti was born in Naples, Italy, in 1966. She received the Laurea (M.S.) degree (cum laude) in electronic engineering from University of Rome "La Sapienza" in April 1992. She was a project engineer at the Microwave Laboratory, Alenia, Aeritalia & Selenia S.p.A., Rome, from 1992 to 1994, and joined the Department of Applied Electronics, University Rome Tre, as an Assistant Professor, in October 1994. Her ongoing research areas include passive optical devices, as optical filters, wavelength demultiplexing, encoders/decoders, and polarizing devices. Her research results have been recorded in over 50 refereed papers and summarized in nearly 20 conference and symposium presentations. She is a Member of the IEEE Lasers and Electro-Optics Society (LEOS), the Consorzio Nazionale Interuniversitario per le Telecomunicazioni (CNIT), and the National Institute for the Physics of the Matter (INFM), University Rome Tre, Rome, Italy.

Michela Svaluto Moreolo was born in Rome, Italy, in 1977. She received the Laurea (M.S.) degree (cum laude) in electronic engineering from University Rome Tre in May 2003. She is currently a Ph.D. student in telecommunication engineering at the University Rome Tre, Rome, Italy, since November 2003. Her research interests are in the field of passive optical devices, such as planar lightwave circuits and photonic bandgap structures. Her studies are directed towards all-optical signal processing in WDM systems and all-optical beamforming in radio-over-fiber systems. Her research in nanophotonics involves also quantum dot lasers/LEDs. She is a Student Member of the IEEE Lasers and Electro-Optics Society (LEOS).

Alessandro Neri was born in Viterbo in 1954. In 1977 he received the Doctoral degree in electronic engineering from the University of Rome "La Sapienza." In 1978 he joined the Research and Development Department, Contraves Italiana S.p.A., where he gained a specific expertise in the field of radar signal processing and in applied detection and estimation theory, thus becoming the Chief of the Advanced Systems Group. In 1987 he joined the INFOCOM Department, University of Rome "La Sapienza," as an Associate Professor in signal and information theory. In 1992 he joined the Department of Applied Electronics, University Rome Tre, as an Associate Professor. He became a Full Professor in telecommunications in September 2001. Since 1992, he has been responsible for the coordination and management of research and teaching activities in the telecommunication fields at the University Rome Tre. His research activity has mainly been focused on information theory, signal theory, and signal and image processing, and their applications to both telecommunications systems and remote sensing. His current research is focused on 3rd- and 4th-generation cellular systems and multimedia communications.

EURASIP Journal on Applied Signal Processing 2005:10, 1584–1592

Adaptive Electronic Dispersion Compensator for Chromatic and Polarization-Mode Dispersions in Optical Communication Systems

Ut-Va Koc

Bell Labs, Lucent Technologies, 600 Mountain Avenue, Murray Hill, NJ 07974, USA
Email: koc@bell-labs.com

Received 1 April 2004; Revised 30 November 2004

The widely-used LMS algorithm for coefficient updates in adaptive (feedforward/decision-feedback) equalizers is found to be suboptimal for ASE-dominant systems but various coefficient-dithering approaches suffer from slow adaptation rate without guarantee of convergence. In view of the non-Gaussian nature of optical noise after the square-law optoelectronic conversion, we propose to apply the higher-order least-mean $2N$th-order (LMN) algorithms resulting in OSNR penalty which is 1.5–2 dB less than that of LMS. Furthermore, combined with adjustable slicer threshold control, the proposed equalizer structures are demonstrated through extensive Monte Carlo simulations to achieve better performance.

Keywords and phrases: electronic PMD compensation, adaptive equalization, signal processing, optical communication, least-mean fourth-order algorithm, least-mean-square algorithm.

1. INTRODUCTION

Optical communication forms the backbone of modern telecom and Internet networks around the globe. Due to its enormous inherent channel capacity [1], it is anticipated that this trend will continue or even accelerate. In this ongoing evolution, adaptive electronic equalization for combating impairments in fiber-optic communication may play an important role in pushing from the core of networks all the way to the edge by providing cost-effective solution. Two major impairments commonly encountered in modern fiber-optic systems are chromatic dispersion (group velocity dispersion or GVD) and polarization-mode dispersion (PMD). Chromatic dispersion can be compensated effectively by an optical dispersion compensation module (DCM) due to its static nature. However, at substantially high data rates (10 or even 40 Gbps), especially in long-haul networks, residual chromatic dispersion amount remains problematic and thus electronic equalization against residual chromatic dispersion is still important [2]. In cost-sensitive metro networks, electronic solution is considered a viable option to replace the expensive optical solution. On the other hand, PMD is dynamic in nature and substantial unpredictable PMD is accumulated over a long distance of old fibers, enough to cause network outage [3]. Currently it is extremely expensive to be compensated optically by bulky optical PMD compensators (OPMDCs) and thus electronic solution is vigorously sought in recent years.

Adaptive electronic equalizers for impairment compensation in fiber-optic networks have been studied for decades. In early work [4], the dominant noise was quantum, shot, or electronic thermal noise, which can be modeled effectively as additive Gaussian noise. After the advent of efficient and low-noise fiber amplifiers in 1987 [5], optical amplifiers (EDFA or Raman) were used extensively to increase the transmission distance without O-E-O conversion. Since then, a number of studies were undertaken to explore a variety of equalizer structures for adaptive optical-channel impairment compensation ranging from feedforward-type equalizers to maximum-likelihood estimators [6, 7, 8, 9, 10, 11]. Among many studies in electronic PMD compensation in recent years, [12, 13, 14, 15, 16, 17, 18, 19], the first decision-feedback loop and tapped-delay-line equalizer at 10 Gbps were implemented in [20] and [21] respectively. The maximum-likelihood estimation for PMD compensation was also investigated in [22, 23].

In a fiber-optic link, a number of optical amplifiers, either erbium-doped fiber amplifiers (EDFAs) or Raman amplifiers (RAs), are employed to strengthen the optical signal, but at the same time add in the incoherent amplified spontaneous emission (ASE) noise (commonly called optical noise) [5]. In the quantum regime, the ASE noise follows the Bose-Einstein distribution [24] and the optically amplified coherent light exhibits a noncentral-negative-binomial distribution [25]. However, in the classical limit when the signal and noise involve many photons, this optical noise can be

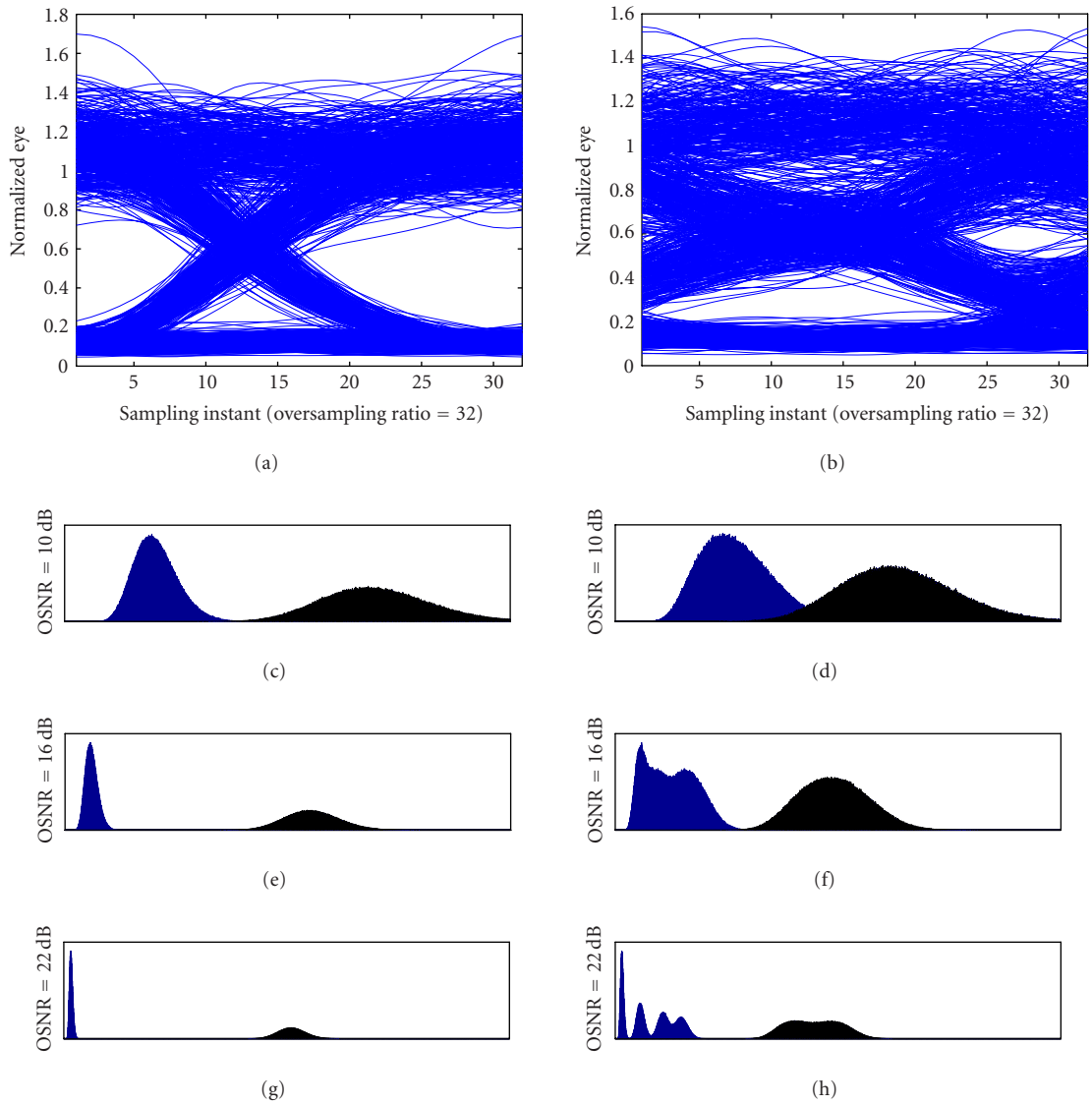


FIGURE 1: (a) and (b) Eye diagrams when OSNR = 16 dB and (c), (d), (e), (f), (g), and (h) histograms of noise distribution of bits 1 and 0 at the decision point for NRZ when OSNR is equal to (c), (d) 10 dB; (e), (f) 16 dB; and (g), (h) 22 dB. (Left column: DGB=0 picoseconds; Right column: DGB = 75 picoseconds; both are at 10 Gbps NRZ.)

modeled as complex additive white Gaussian noise (AWGN) in the optical field of each orthogonal polarization mode of the optically amplified signal based on the central limit theorem [26, 27]. An optical filter $H_o(f)$ is usually placed before a photodetector to limit the optical noise reaching the receiver and also to filter out other unselected channels in a wavelength-division multiplexing (WDM) system.

At the receiver side, after square-law detection using a photodiode, the noise becomes non-Gaussian. As a result, the noisy input signal $u(t)$ contains a *signal-dependent* nonsymmetric Gaussian term having a variance determined by the signal levels of bits (0 and 1), and a nonzero-mean noncentral χ^2 term. It should be noted that the mean of $u(t)$ is nonzero and varies at different optical signal-to-noise ratio levels (OSNR), as depicted in Figures 1a and 1b without PMD (i.e. differential group delay or DGD = 0 picosecond) and with substantial amount of PMD (DGD = 75 picoseconds), respectively, where the eye diagram and noise distributions of bits 0 and 1 at the decision point for NRZ at different OSNR levels are plotted to demonstrate the varying nonzero mean values. This noise property is very different from electronic thermal noise or even optical noise after the combination of an interferometer and a balanced receiver [28].

Because of this non-Gaussian noise property, the well-studied *least-mean-square* (LMS) algorithm becomes *suboptimal* for optical noise in the sense that OSNR penalty is observed even without the presence of PMD-induced intersymbol interference (ISI), though it is optimal for additive Gaussian noise [29]. Various coefficient-dithering approaches

were devised to replace LMS but generally suffer from slow adaptation rate and no guarantee of convergence. Typically, tap coefficients are adjusted in a trial-and-error manner with feedback from various types of error monitors such as minimization of eye-monitoring pseudoerror count [30], maximization of eye-opening [31], Q-factor optimization [32], bit-pattern-dependent threshold control [9], or FEC-error-count minimization [33].

The higher-order statistic techniques, the *least-mean* $2N$*th-order* (LMN) family, were first proposed in [34] as an extension to the LMS and were found better suited than the LMS for non-Gaussian noise [34, 35]. In this paper, we propose to apply these higher-order statistic algorithms to coefficient update for combating the non-Gaussian optical noise after photodetection. Through Monte Carlo simulation, we demonstrate that 1.5–2 dB better OSNR penalty can be attained for the LMN than the LMS [36]. Furthermore, due to the nonsymmetric nature of optical noise, we also propose to combine these higher-order algorithms with automatic slicer threshold control (ATC-8LICER) and show that the threshold-optimized LMN still achieves better results than the conventional LMS with optimized thresholds. In Section 2, we briefly introduce the higher-order algorithms (LMN) in comparison to the LMS algorithm for feedforward (FFE) and decision-feedback (DFE) equalizers. In Section 3, the results through extensive Monte Carlo simulation are presented to evaluate the performance of the LMN algorithms combined with automatic threshold control compared against the LMS. Finally, the conclusion is drawn in Section 4.

2. ARCHITECTURES AND ALGORITHMS OF ELECTRONIC PMD COMPENSATORS

In a digital receiver architecture, an analog-to-digital converter (ADC) is required to digitize the incoming bandlimited analog signal for further processing in the digital adaptive equalizer. Though more complex functionality can be achieved in low-power CMOS digital circuitry coupled with pipelined/parallel data flows after digitization, the ADC is considered one of the bottlenecks for digital implementation of advanced equalizers. The current state-of-the-art ADC can provide about 6-bit accuracy at 10 Gbps but it becomes challenging at 40 Gbps. Furthermore, the ADC power consumption at such a high speed may offset the low-power benefit of digital CMOS circuits. On the other hand, the analog equalizer, particularly the simple analog tapped-delay-line equalizer, is an attractive alternative in the absence of ADC for providing a practical electronic dispersion compensation solution. No matter what implementation (digital or analog) is adopted, the operations of an equalizer can be described in the discrete-time or z-domain. In this paper, we express equalizer algorithms in the discrete-time domain without explicitly referring to a digital or analog implementation.

A number of electronic equalizer structures, such as feedforward equalizer (FFE), decision-feedback equalizer (DFE), and maximum-likelihood sequence estimator (MLSE), have been considered over the years. Though MLSE achieves the best performance, substantial complexity and compromises are often associated with practical implementation. Specifically, MLSE has exponential complexity $O(e^N)$ but FFE/DFE requires only linear complexity $O(N)$. However, the best reason for using relatively simple architectures (FFE/DFE) in most wireless/wireline electronic systems instead of MLSE is that they usually offer entirely adequate performance [37]. In this paper, we examine two basic equalizer architectures: feedforward equalizers (FFEs) and decision-feedback equalizers (DFEs) with FFE, as shown in Figures 2a and 2b, respectively.

2.1. *Least-mean* $2N$*th-order algorithms*

It has been shown that the LMS algorithm is H^∞-optimal under the assumption of temporal whiteness and Gaussian disturbances [38], providing theoretical justification for the excellent *robustness* of LMS. However, after square-law detection by a photodiode, optical noise becomes nonzero-mean signal-dependent χ^2. As a result, the conventional LMS algorithm becomes *suboptimal* for optical noise in the sense that equalization introduces OSNR penalty even without any ISI [29], though it is *optimal* for additive Gaussian noise (e.g. thermal noise). The higher-order statistic techniques, the least-mean $2N$th-order (LMN) family, were proposed in [34] as a higher-order extension of LMS and found to be better suited for non-Gaussian noise than LMS [35]. Due to the non-Gaussian nature of optical noise, we propose to employ the higher-order counterparts of LMS which are found to yield better results than LMS.

After equalization, a slicer makes a *hard decision* to determine the estimated symbol $\hat{d}(k)$ (or bit in the OOK case) from the compensated slicer input $s(k)$ in reference to a slicer threshold. For the case of only two signal levels, the slicer is simply a high-speed comparator as used in a conventional CDR. The LMS minimizes the cost function of the square of the slicer error $e(k)$: $J_{\mathrm{LMS}}(k) = E\{e^2(k)\}$, where $e(k) = a(k) - s(k)$ is the difference between the target signal $a(k)$ and the equalized signal $s(k)$ before decision. In the normal mode, $a(k) = \hat{d}(k)$ but, in the training mode, $a(k) = d(k)$, the transmitter input bit. Usually a training sequence is required at the startup stage of equalization and then the equalizer is switched to the normal mode. However, it is possible to start the equalizer without any training sequence through blind startup which is out of the scope of this paper. In fiber-optic communication, fixed bytes such as SONET A1/A2 bytes are sent in every frame and can be utilized as the training sequence. As an extension of the LMS, the LMN is a class of adaptive algorithms to minimize the cost function of the $2N$th order of $e(k)$ $J(k) = E\{e^{2N}(k)\}$. For $N = 1$, the LMN algorithm becomes LMS. For $N = 2$, it is called the least-mean fourth-order (LMF) algorithm. When $N = 3$, it is the least-mean sixth-order (LM6) algorithm. Since the LMS belongs to the LMN family, the extensive knowledge of the LMS in both theory and implementation can also be applied to the LMN family.

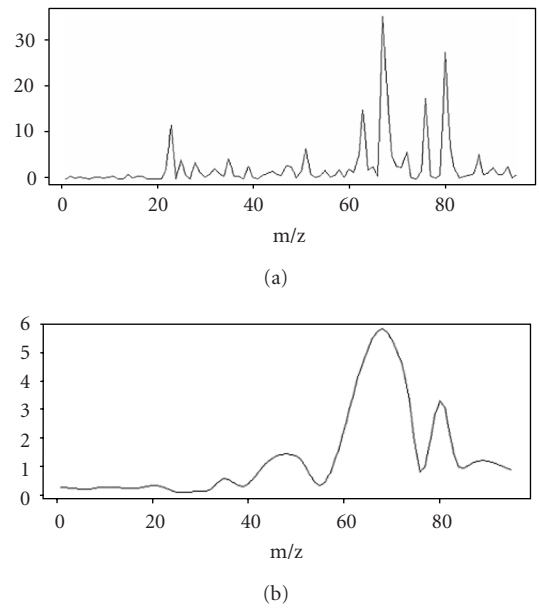


Figure 2: Different equalizer architectures are considered. The filter response of FFE is $C(z)$ while the response of DFE is $B(z)$. The decision level is fixed to the midpoint of the bit values $\{0, 1\}$. (a) Feedforward equalizer (FFE). (b)Decision feedback equalizer (DFE) coupled with an FFE.

Because the first and second moments (equivalently mean and variance) are sufficient to describe Gaussian distributions, it is not necessary to account for higher moments in the cost function and the second-order cost function in the LMS is enough for Gaussian noise. For non-Gaussian noise, higher moments are required. Intuitively, the higher-order cost function wages a large penalty for a large slicer error to account for higher moments and causes the noise distribution to be more compact for non-Gaussian noise, as will be seen in Section 3.

2.2. Feedforward equalizers

Consider the case of FFE depicted in Figure 2a where the $(2L+1)$-tap FFE coefficients are $\{c_{-L}(k), \ldots, c_L(k)\}$ at $t = kT_s$ in the discrete-time domain. Here $T_s = 1/f_s$ is the sample period. If the oversampling ratio $R = T/T_s > 1$, then the FFE is called a fractionally-spaced equalizer. In vector form, these weights can be denoted as $\vec{\mathbf{c}}^T(k) = [c_{-L}(k), \ldots, c_L(k)]$. The receiver input data sequence is $\vec{\mathbf{u}}^T(k) = [u(k+L), \ldots, u(k-L)]$. Therefore, the slicer input is $s(k) = x(k) = \vec{\mathbf{c}}^T(k)\vec{\mathbf{u}}(k)$. The weight updating unit of $C(z)$ (WUD$[C]$) adapts $\vec{\mathbf{c}}(k)$ based on an adaptive algorithm, among which the LMS is the most widely used. As in Figure 2a, the weight updating unit $C(z)$ adopts the LMN of Nth order. It can be shown that

$$\begin{aligned}\vec{\mathbf{c}}(k+1) &= \vec{\mathbf{c}}(k) - \frac{\beta}{2}\nabla_C\left\{[e(k)]^{2N}\right\} \\ &= \vec{\mathbf{c}}(k) + \beta N[e(k)]^{2N-1}\vec{\mathbf{u}}(k),\end{aligned} \tag{1}$$

where β is the preset step size.

2.3. Decision-feedback equalizers with FFE

For the case of a decision-feedback equalizer (DFE) coupled with an FFE, as in Figure 2b, where $s(k) = x(k) - y(k) = \vec{\mathbf{c}}^T(k)\vec{\mathbf{u}}(k) - \vec{\mathbf{b}}^T(k)\vec{\mathbf{a}}(k)$ with the DFE coefficients defined as $\vec{\mathbf{b}}^T(k) = [b_1(k), \ldots, b_M(k)]$. Here $\vec{\mathbf{a}}^T(k) = [a(k-1), \ldots, a(k-M)]$. The weight updating units for FFE and DFE are denoted as WUD(C) and WUD(B), respectively and can employ a variety of adaptive algorithms including the LMN. In Figure 2b, $s(k) = x(k) - y(k) = \vec{\mathbf{c}}^T(k)\vec{\mathbf{u}}(k) - \vec{\mathbf{b}}^T(k)\vec{\mathbf{a}}(k)$. Define $\vec{\mathbf{w}}^T(k) = [\vec{\mathbf{c}}(k), -\vec{\mathbf{b}}(k)]$ and $\vec{\mathbf{r}}^T(k) = [\vec{\mathbf{u}}(k), -\vec{\mathbf{a}}(k)]$, where $\vec{\mathbf{c}}^T(k) = [c_{-L}(k), \ldots, c_L(k)]$, $\vec{\mathbf{u}}^T(k) = [u(k+L), \ldots, u(k-L)]$, $\vec{\mathbf{b}}^T(k) = [b_1(k), \ldots, b_M(k)]$, $\vec{\mathbf{a}}^T(k) = [a(k-1), \ldots, a(k-M)]$. We can rewrite $e(k) = a(k) - \vec{\mathbf{w}}^T(k)\vec{\mathbf{r}}(k)$. Therefore, the coefficients of the FFE and DFE can be updated as follows to minimize the cost function, $J(k) = E\{[e(k)]^{2N}\}$:

$$\begin{aligned}\vec{\mathbf{w}}(k+1) &= \vec{\mathbf{w}}(k) - \frac{\beta}{2}\nabla_w\left\{[e(k)]^{2N}\right\} \\ &= \vec{\mathbf{w}}(k) + \beta N[e(k)]^{2N-1}\vec{\mathbf{r}}(k).\end{aligned} \tag{2}$$

For the first-order PMD channel impulse response $h(t) = \gamma\delta(t) + (1-\gamma)\delta(t-\tau_{\mathrm{DGD}})$, it is easy to show that there exist dips in the frequency response at locations inversely proportional to τ_{DGD}, especially for $\gamma = 0.5$ [19]. It has been known that an FFE amplifies the noise in the spectral dips whereas a DFE does not [37]. Thus, the performance of FFE for PMD compensation in the high-DGD region is anticipated to be less than that of DFE.

2.4. Slicer with automatic threshold control

In the uncompensated OOK case, the two noiseless signal levels for bits 0 and 1 are normalized to 0 and 1 with the midpoint defined at 0.5 (or equivalently to $\{-1, 1\}$ with the midpoint at 0). In view of the nonsymmetric distributions for bits 0 and 1 (having different variances) reflected in the signal-dependent term after square-law detection, the optimal threshold point is usually not the midpoint. Therefore, an automatic threshold control algorithm (ATC) is designed to track the incoming signal profile and automatically adjust the threshold in the CDR for the minimum bit error rate [39]. In this paper, the *ideal* ATC algorithm is employed by searching the entire simulation sequence for the best threshold value to obtain the least number of errored bits. In a practical system, the ATC algorithm usually accumulates signal distribution information within a window of finite duration for optimal threshold calculation to allow tracking of slowly varying nonstationary channels [40].

In the conventional equalizer structures illustrated in Figures 2a and 2b, the decision point (slicer threshold θ) is fixed at the midpoint of two noiseless signal levels. Similar to the uncompensated case, the slicer input distributions for bits 0 and 1 are also nonsymmetric and thus result in the optimal threshold point being away from the midpoint. Therefore, we propose alternative architectures for FFE/DFE with adjustable threshold control in the slicer (ATC-SLICER) as shown in Figures 3a and 3b. The ATC-SLICER's function is to

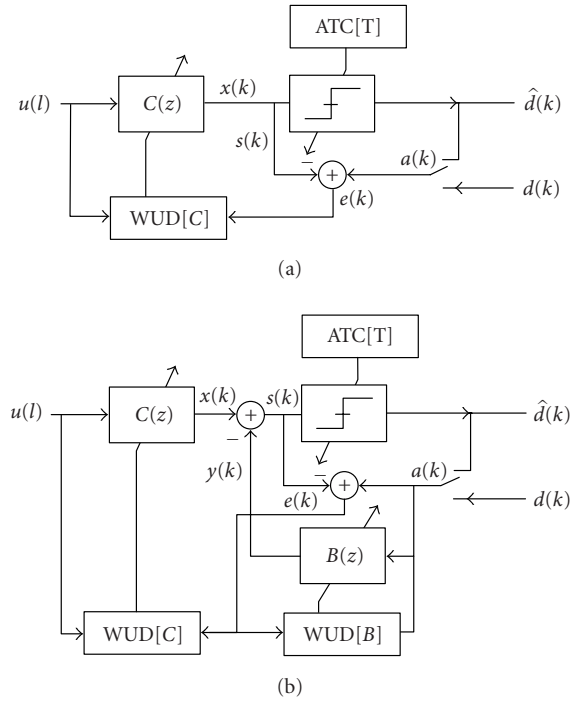


Figure 3: Equalizers with automatic threshold adjustment (ATC-SLICER). The decision threshold is controlled adaptively by an automatic threshold control (ATC) algorithm instead of being fixed at the midpoint of two noiseless signal levels. (a) FFE with ATC-SLICER. (b) DFE with FFE and ATC-SLICER.

track the slicer input histogram and adjust the slicer threshold accordingly. In the steady state of equalization, the ATC-SLICER can start to build up a slicer input profile to determine the best threshold for the slicer, and then dynamically track and adjust the optimal slicer threshold.

There are many ways readily available in the literature to implement the automatic threshold control technique (ATC). In our case of adaptive equalization, the input to the ATC-SLICER control block is the slicer input $s(k)$. In one of many possible implementations, a histogram is established in two arrays of memory, $\text{bin}_1(i)$ and $\text{bin}_0(i)$, for $i = 0, \ldots, B+1$ corresponding to the intervals $\{(-\infty, v_L), [v_L, v_L + d), [v_L + d, v_L+2d), \ldots, [v_L+(i-1)d, v_L+id), \ldots, [v_H-d, v_H], (v_H, \infty)\}$, where $d = (v_H - v_L)/B$ and B is the number of bins that form the range from v_L to v_H. The value of B dictates the accuracy of the final threshold determination but a large B value requires more memory space. In general, B could be 128, 256, or even 1024. If $s(k)$ is in the interval of $[v_L+(i-1)d, v_L+id)$, then the count in the memory $\text{bin}_1(i)$ is incremented by one for the slicer output being 1 and $\text{bin}_0(i)$ is incremented for the slicer output being 0. The bit error can be estimated as follows: $\text{error}[0] = \text{bin}_1[0] - \text{bin}_0[0]$, and $\text{error}[i] = \text{error}[i-1] + \text{bin}_1[i] - \text{bin}_0[i]$ for $i = 1, \ldots, B+1$. The optimal threshold is determined by finding the minimum of $\text{error}[i]$. To avoid the detrimental memory effect for a changing optical channel, a reset signal is sent periodically to clear the memory arrays and a histogram is rebuilt fresh in accordance with the above rule. Since PMD drifts very slowly, it is possible to refresh this profile once one or few minutes without encountering significant change in channel response while the number of bits collected in that time frame is large enough for meaningful profiling. For an analog implementation of the equalizer, the compensated signal can be sampled at a low speed to build the profile of $s(k)$ over time. The optimal threshold can be determined externally through the method described above.

3. RESULTS AND DISCUSSION

An extensive Monte Carlo simulation is conducted in NRZ and RZ formats at 10 Gbps to compare the performance of the LMN algorithms with/without automatic threshold control against the common LMS algorithm. The full-width half-maximum (FWHM) of the RZ pulse is 50%. The simulation length is chosen dynamically at each OSNR level with a given amount of DGD to provide sufficient samples for reliable BER estimation. It usually requires a longer simulation for an expected low BER value in a given OSNR-DGD condition, but a shorter one for an expected high BER. The ATC algorithm finds the optimum threshold value on the uncompensated electronic signal over the *entire* simulation sequence. Its performance curves provide the benchmark, against which the performance charts of all the equalizer structures are compared. In order to isolate the effect of compensation on other impairments such as poor extinction ratio (ER), we deliberately set ER = 30 dB and use a wideband optical bandpass filter for the simulation. The electronic lowpass filter is set to be a third-order Bessel filter of bandwidth 7 GHz for 10 Gbps, whereas the optical filter is modeled as a tenth-order Gaussian filter with a bandwidth of 150 GHz to weakly filter any receiving signal without causing any distortion. The peak optical power of the transmitter laser is controlled at 19.95 mW to avoid any nonlinear distortion during transmission, though no fiber nonlinearity is modeled in the interest of fast simulation. Only the first-order PMD and GVD are considered in the simulation. The *first-order* PMD is commonly characterized by the amount of DGD in picoseconds and the *first-order* GVD is characterized by the amount of dispersion in picoseconds per a nanometer of wavelength (ps/nm) without explicitly specifying the actual fiber type and length.

For the sake of brevity, we pick a fractionally spaced FFE of 9 taps at an oversampling ratio 2 coupled with a DFE of 2 feedback taps as the typical equalizer architecture (designated as FFE9 + DFE2) to illustrate the advantage of the higher-order LMN algorithms over the LMS, even though we have simulated a large variety of equalizer structures. The OSNR penalty curves of this equalizer are plotted in Figure 4 with and without ATC-SLICER (for optimized and fixed thresholds). The dash lines represents the use of the LMS, and the solid lines for the LMF. These OSNR penalty curves are aggregated from the OSNR-BER plots, of which two examples are shown in Figures 5a and 5b for the uncompensated and the FFE9 + DFE2 with ATC-SLICER and LMF respectively in 10 Gbps NRZ. In Figure 4a with GVD only, the LMF is 1.5 dB better than the LMS for fixed thresholds, whereas with ATC-SLICER, it is 0.5 dB better than the LMS.

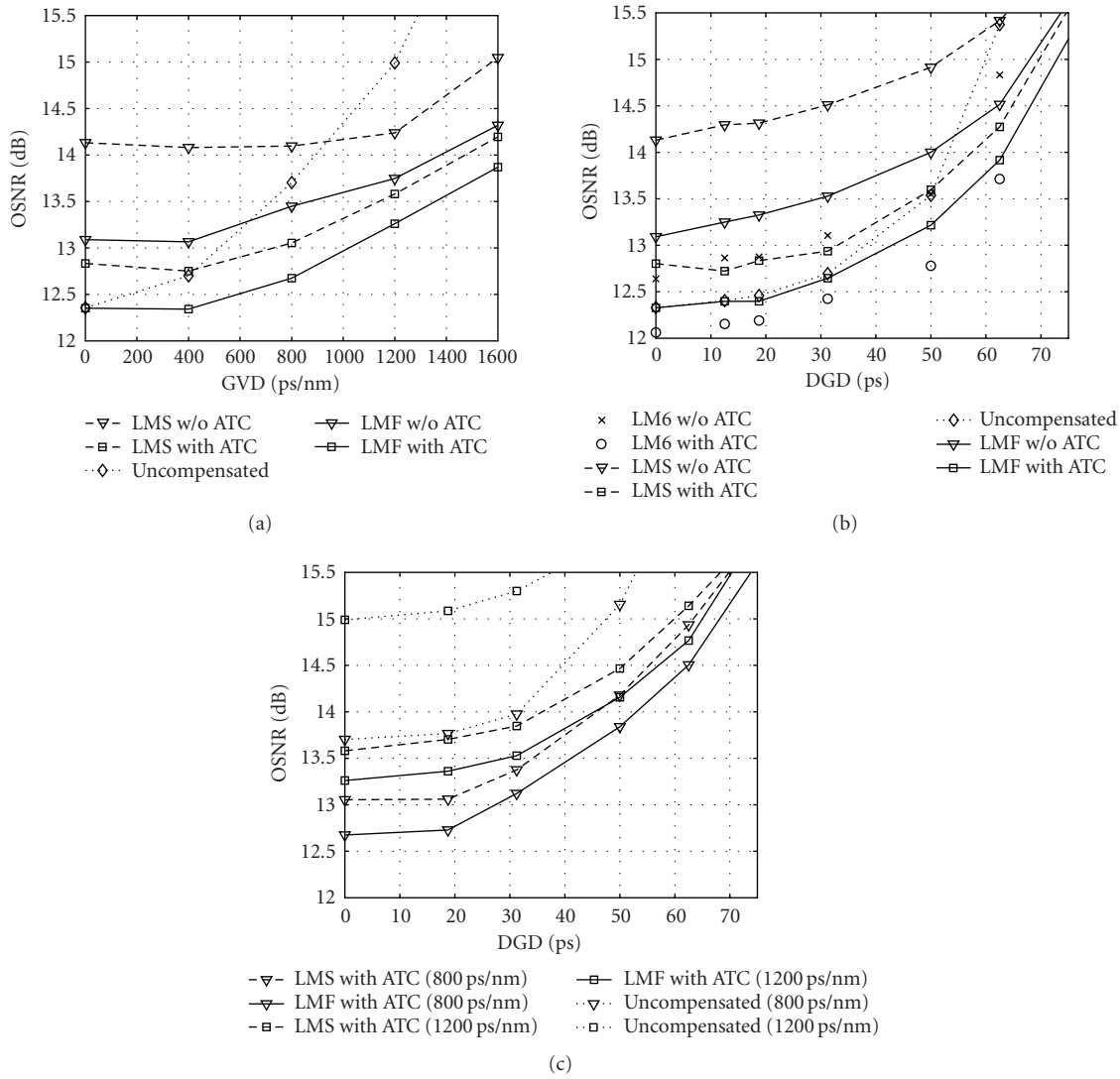


Figure 4: OSNR penalty when BER is set at 6×10^{-5} (FEC-correctable error-rate level) for 10 Gbps NRZ under single and combined impairment conditions. The equalizer has a 9-tap fractionally spaced FFE and a 2-tap DFE with/without ATC-SLICER. (a) GVD only. (b) PMD only. (c) PMD combined with GVD (800 ps/nm).

It should be noted that a 3 dB gain means the possibility of doubling the transmission distance. In Figure 4b with PMD only, the LMF and LM6 are 1.5–2 dB better than the LMS for fixed thresholds, whereas the LM6 is almost 1 dB better than the LMS with ATC-SLICER. In addition, the LM6 with fixed thresholds has comparable performance to the LMS with optimized thresholds. In both cases, the LMF and LM6 with ATC-SLICER show less penalty than the uncompensated under no or mild dispersion condition while the LMS has higher penalty than the uncompensated. This implies that the LMS is *potentially not penalty-free* in a distortion-free environment, no matter how we optimize the threshold points. For the combined effect of PMD and GVD (800 or 1200 ps/nm) depicted in Figure 4c, the LMF/LM6 are almost 0.5 dB better than the LMS with ATC. The trend indicates that the advantage of the LMN algorithms is more obvious in an environment of less distortion. Another implication is that the implementation of an ATC-SLICER can be simplified or avoided by employing the LMF or LM6 with minor increase in complexity of one or two multiplications as indicated in (1) and (2).

The advantage of the LMN algorithms over LMS is further illustrated in Figure 6, of which (a) and (b) demonstrate that the LMN can achieve much lower BER than the LMS. Note that the symbols such as "FFE9:0-3" in Figures 6a and 6b mean the FFE equalizer type with 9 feedforward taps, no feedback taps, and the LM6 ($N = 3$). Figure 6c depicts the slicer input histograms for a 9-tap fractionally spaced FFE and shows how the LMN (LMF or LM8) can compact the distribution more effectively than the LMS. At DGD = 0 picoseconds, LMS actually spreads the noise, causing lower BER and higher OSNR penalty, but the LMF maintains the noise variance.

Similar plots can be obtained for the 10 Gbps RZ case in Figure 7 where the performance of an FFE9 + DFE2 is shown to further demonstrate the advantage of the higher-order

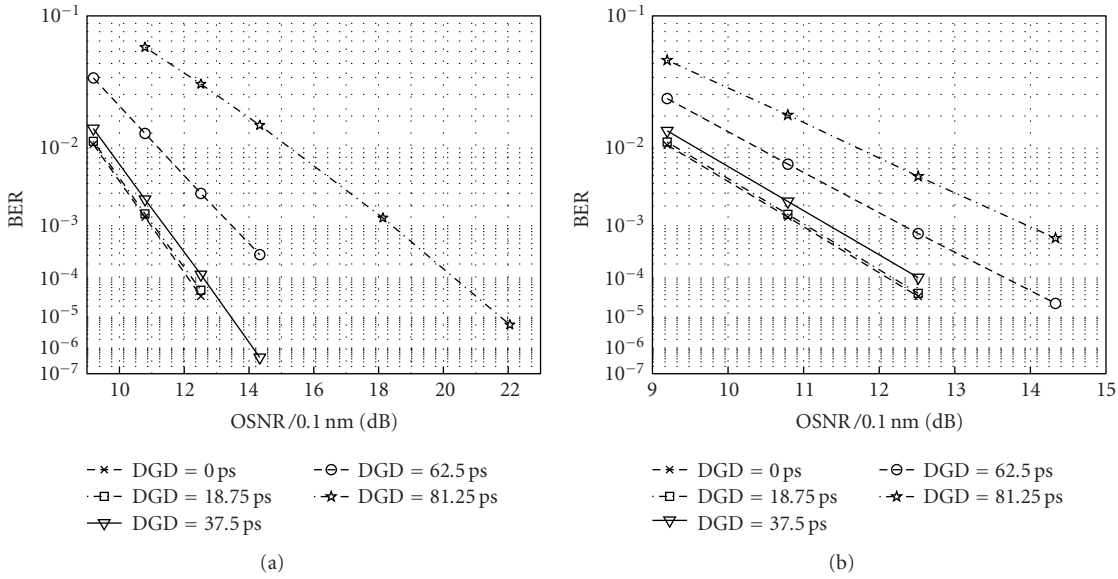


Figure 5: OSNR-BER plots with 1st-order PMD for NRZ format at 10 Gbps in Monte Carlo simulation. The equalizer has a 9-tap fractionally spaced FFE and a 2-tap DFE with ATC-SLICER optimized via the LMF. (a) Uncompensated with ATC. (b) FFE with DFE, ATC-SLICER, and LMF.

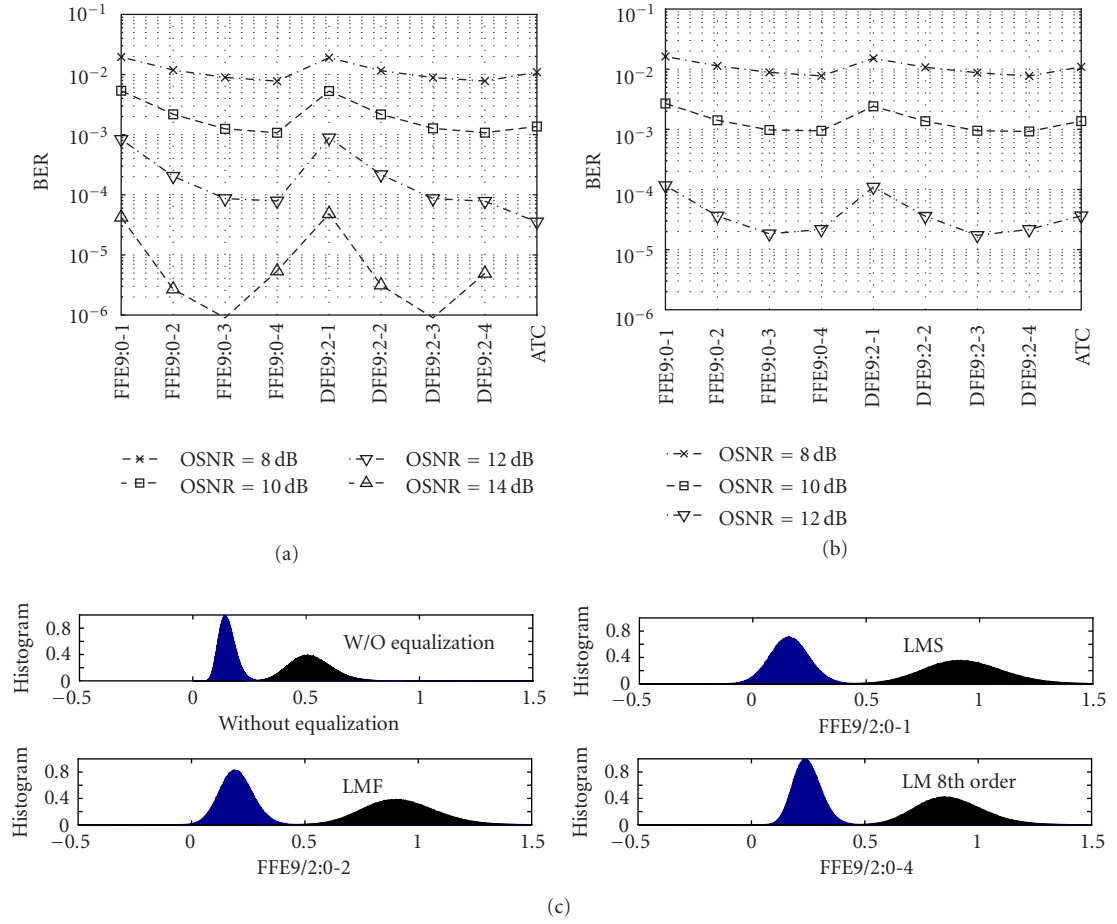


Figure 6: BER and slicer input histograms at DGD = 0 picosecond of the first-order PMD for NRZ at 10 Gbps. The numbers in the equalizer type represent the following: number of FFE taps (*N*FFE), oversampling rate ($R > 1$: fractionally-spaced), number of DFE taps (*N*DFE), LMN order ($N = 1$: LMS; $N = 2$: LMF; $N = 3$: LM6). (a) BER without ATC-SLICER. (b) BER with ATC-SLICER. (c) Slicer histogram for T/2-spaced 9-tap FFE.

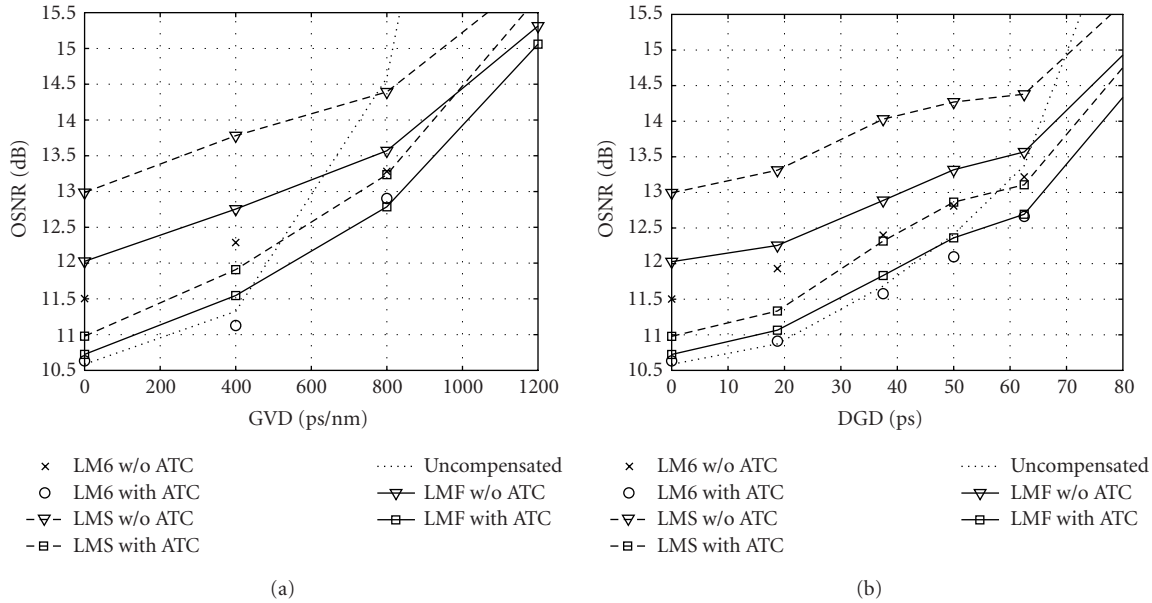


Figure 7: OSNR penalty when BER is set at 6×10^{-5} (FEC-correctable error-rate level) for 10 Gbps RZ under single and combined impairment conditions. (a) GVD only. (b) PMD only.

LMN algorithms over the conventional LMS for two different cases: only GVD and only PMD. With GVD only in Figure 7a, the LMN is 1.5 dB better without ATC-SLICER and 0.5 dB better with ATC-SLICER than the LMS. A similar trend can be found in Figure 7b affected by PMD alone.

4. CONCLUSION

LMS is well studied over the past several decades and considered optimal for a *linear* system with additive Gaussian noise. In a fiber-optic communication system, the dominant noise is optical noise (ASE). Due to the non-Gaussian nature of optical noise after nonlinear quadratic detection, the LMS is no longer optimal in ASE-dominant systems. The usual adaptive equalization techniques, so successful in electronic systems, cannot be directly applied to ASE-dominant fiber-optic systems in view of nonlinear square-law detection. In this paper, we demonstrate that its higher-order cousins in the LMN family are 1.5–2 dB better without ATC-SLICER and 0.5–1 dB better with ATC-SLICER than the LMS. In certain cases, the LMN with fixed thresholds even has comparable performance to the LMS with optimized thresholds. It suggests that the ATC-SLICER control unit may be eliminated in use of the LMN for the sake of less complexity. On the other hand, with a demanding performance criterion, the LMN with ATC-SLICER can be used to provide *no back-to-back penalty* (i.e. in absence of ISI-induced impairments). In comparison, the LMS is potentially penalized in a distortion-free environment, no matter how we optimize the threshold points, which is usually not acceptable in most receiver design goals.

REFERENCES

[1] P. P. Mitra and J. B. Stark, "The nonlinear limits to the information capacity of optical fibre communications," *Nature*, vol. 411, no. 28, pp. 1027–1030, 2001.

[2] S. Otte and W. Rosenkranz, "Electrical dispersion compensation for 10 Gb/s transmission systems: simulation results," in *Proc. 2nd Working Conference on Optical Network Design and Modelling (ONDEM '98)*, pp. 71–74, Rome, Italy, February 1998.

[3] H. Bulow, "Polarisation mode dispersion (PMD) sensitivity of a 10 Gbit/s transmission system," in *Proc. European Conference on Optical Communication (ECOC '96)*, vol. 2, pp. 211–214, Oslo, Norway, September 1996.

[4] D. G. Messerschmitt, "Minimum MSE equalization of digital fiber optic systems," *IEEE Trans. Commun.*, vol. 26, no. 7, pp. 1110–1118, 1978.

[5] E. Desurvire, *Erbium-Doped Fiber Amplifiers: Principles and Applications*, Wiley-Interscience, New York, NY, USA, 1994.

[6] R. Hopper, "Optimum equalisers for binary digital transmission systems and their application to optical receivers," *IEE Proceedings J-Optoelectronics*, vol. 136, no. 2, pp. 137–140, 1989.

[7] J. H. Winters and R. D. Gitlin, "Electrical signal processing techniques in long-haul fiber-optic systems," *IEEE Trans. Commun.*, vol. 38, no. 9, pp. 1439–1453, 1990.

[8] J. H. Winters, "Equalization in coherent lightwave systems using a fractionally spaced equalizer," *J. Lightwave Technol.*, vol. 8, no. 10, pp. 1487–1491, 1990.

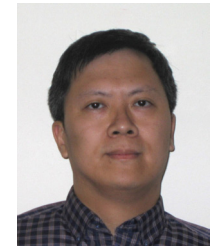
[9] S. Kasturia and J. H. Winters, "Techniques for high-speed implementation of nonlinear cancellation," *IEEE J. Select. Areas Commun.*, vol. 9, no. 5, pp. 711–717, 1991.

[10] N. Swenson, B. Shoop, and J. Cioffi, "Precluding nonlinear ISI in direct detection long-haul fiber optic systems," *IEEE Photon. Technol. Lett.*, vol. 3, no. 2, pp. 182–184, 1991.

[11] J. Cartledge, R. McKay, and M. Nowell, "Performance of smart lightwave receivers with linear equalization," *J. Lightwave Technol.*, vol. 10, no. 8, pp. 1105–1109, 1992.

[12] H. Bulow, D. Schlump, J. Weber, B. Wedding, and R. Heidemann, "Electronic equalization of fiber PMD-induced distortion at 10 Gbit/s," in *Optical Fiber Communication Conference and Exhibit (OFC '98)*, pp. 151–152, San Jose, Calif, USA, February 1998.

[13] F. Cariali, F. Martini, P. Chiappa, and R. Ballentin, "Electronic compensation of PMD and chromatic dispersion with an IC in 10 Gbit/s transmission system," *Electronics Letters*, vol. 36, no. 10, pp. 889–891, 2000.
[14] H. Bulow, F. Buchali, W. Baumert, R. Ballentin, and T. Wehren, "PMD mitigation at 10 Gbit/s using linear and nonlinear integrated electronic equaliser circuits," *Electronics Letters*, vol. 36, no. 2, pp. 163–164, 2000.
[15] F. Buchali, H. Bulow, and W. Kuebart, "Adaptive decision feedback equalizer for 10 Gbit/s dispersion mitigation," in *Proc. European Conference on Optical Communication (ECOC '00)*, Munich, Germany, September 2000, p. 5.2.5.
[16] H. Bulow and G. Thielecke, "Electronic PMD mitigation—from linear equalization to maximum-likelihood detection," in *Optical Fiber Communication Conference and Exhibit (OFC '01)*, vol. 3, pp. WAA3-1–WAA3-3, Anaheim, Calif, USA, March 2001.
[17] H. Bulow, "PMD mitigation by optic and electronic signal processing," in *Proc. IEEE Lasers and Electro-Optics Society (LEOS '01)*, vol. 2, pp. 602–603, San Diego, Calif, USA, November 2001.
[18] M. Win, J. H. Winters, and G. Vitetta, *Equalization Techniques for Mitigating Transmission Impairments*, chapter 18, Academic Press, New York, NY, USA, 2002.
[19] T. Adali, "Applications of signal processing to optical fiber communications," in *Proc. IEEE Lasers and Electro-Optics Society (LEOS '02)*, vol. 2, pp. 623–624, Glasgow, UK, November 2002.
[20] L. Moller, A. Thiede, S. Chandrasekhar, et al., "ISI mitigation using decision feedback loop demonstrated with PMD distorted 10 Gbit/s signals," *Electronics Letters*, vol. 35, no. 24, pp. 2092–2093, 1999.
[21] H. Bulow, R. Ballentin, W. Baumert, F. Buchali, and W. Kuebart, "Adaptive PMD mitigation at 10 Gbit/s using an electronic SiGe equaliser IC," in *Proc. European Conference on Optical Communication (ECOC '99)*, pp. 138–139, Nice, France, September 1999.
[22] H. F. Haunstein, K. Sticht, A. Dittrich, M. Lorang, W. Sauer-Greff, and R. Urbansky, "Implementation of near optimum electrical equalization at 10 Gbit/s," in *Proc. European Conference on Optical Communication (ECOC '00)*, vol. 3, pp. 223–224, Munich, Germany, September 2000.
[23] H. F. Haunstein, K. Sticht, A. Dittrich, W. Sauer-Greff, and R. Urbansky, "Design of near optimum electrical equalizers for optical transmission in the presence of PMD," in *Optical Fiber Communication Conference and Exhibit (OFC '01)*, vol. 3, pp. WAA4-1–WAA4-3, Anaheim, Calif, USA, March 2001.
[24] K. Shimoda, G. Takahashi, and C. H. Townes, "Fluctuations in amplification of quanta with application to maser amplifiers," *Journal of the Physical Society of Japan*, vol. 12, no. 6, pp. 686–700, 1957.
[25] J. Perina, "Superposition of coherent and incoherent fields," *Physics Letters A*, vol. 24, no. 6, pp. 333–334, 1967.
[26] H. A. Haus, *Electromagnetic Noise and Quantum Optical Measurements*, Springer-Verlag, New York, NY, USA, 2000.
[27] L. Kazovsky, S. Benedetto, and A. Willner, *Optical Fiber Communication Systems*, Artech House, Norwood, Mass, USA, 1996.
[28] U.-V. Koc, "Noise property and optimum electronic PMD compensation for strongly/weakly optically filtered differential phase shift keying (DPSK) modulation," in *Optical Fiber Communication Conference and Exhibit (OFC '03)*, vol. 2, pp. 596–599, Atlanta, Ga, USA, March 2003.
[29] H. Bulow, "Electronic equalization of transmission impairments," in *Optical Fiber Communication Conference and Exhibit (OFC '02)*, pp. 24–25, Anaheim, Calif, USA, March 2002.
[30] G. L. Frazer, M. W. Goodwin, K. E. Leonard, J. P. Moffatt, and F. Zhang, "Static and dynamic performance of an adaptive receiver for 10 Gbps optical transmission," in *Proc. European Conference on Optical Communication (ECOC '00)*, vol. 2, pp. 113–114, Munich, Germany, September 2000.
[31] H. Bulow, W. Baumert, F. Buchali, and W. Kuebart, "Adaptation of an electronic PMD mitigator by maximization of the eye opening," in *Proc. European Conference on Optical Communication (ECOC '00)*, Munich, Germany, September 2000, paper P3.10.
[32] M. Fregolent, S. Herbst, H. Soehnle, and B. Wedding, "Adaptive optical receiver for performance monitoring and electronic mitigation of transmission impairments," in *Proc. European Conference on Optical Communication (ECOC '00)*, Munich, Germany, September 2000, paper 2.1.2.
[33] K. Sticht, H. F. Haunstein, M. Lorang, W. Sauer-Greff, and R. Urbansky, "Adaptation of electronic PMD equalizer based on BER estimation derived from FEC decoder," in *Proc. 27th European Conference on Optical Communication (ECOC '01)*, pp. 454–455, 2001.
[34] E. Walach and B. Widrow, "The least mean fourth (LMF) adaptive algorithm and its family," *IEEE Trans. Inform. Theory*, vol. 30, no. 2, pp. 275–283, 1984.
[35] C. L. Nikias and A. P. Petropulu, *Higher-Order Spectra Analysis—A Nonlinear Signal Processing Framework*, PTR Prentice Hall, Englewood Cliffs, NJ, USA, 1993.
[36] U.-V. Koc, K. Tu, and N. Kaneda, "Adaptive electronic equalization using higher-order statistics for PMD compensation in long-haul fiber-optic systems," in *European Conference on Optical Communication (ECOC '02)*, Copenhagen, Denmark, September 2002, paper 7.1.5.
[37] R. D. Gitlin, J. F. Hayes, and S. B. Weinstein, *Data Communications Principles*, Plenum Press, New York, NY, USA, 1992.
[38] B. Hassibi, A. H. Sayed, and T. Kailath, "H^∞ optimality of the LMS algorithm," *IEEE Trans. Signal Processing*, vol. 44, no. 2, pp. 267–280, 1996.
[39] N. Kaneda, E. C. Burrows III, B. Le, K. Sticht, M. Hartlaub, and R. M. Jopson, "Polarization mode dispersion tolerance of an adaptive threshold receiver", in *LEOS Topical Meeting*, pp. TuN3-39–TuN3-40, 2002.
[40] M. Kawai, H. Watanabe, T. Ohtsuka, and K. Yamaguchi, "Smart optical receiver with automatic decision threshold setting and retiming phase alignment," *J. Lightwave Technol.*, vol. 7, no. 11, pp. 1634–1640, 1989.



Ut-Va Koc received the Ph.D. degree from the University of Maryland, College Park, and the B.S. degree from National Chiao Tung University. He is currently a Distinguished Member of Technical Staff (DMTS) at High Speed Electronics Research, Bell Labs, Lucent Technologies, Murray Hill, New Jersey. He has published one book, a number of peer-reviewed papers, and book chapters on signal processing in communications and multimedia, in addition to 5 patents granted or in submission. He has been active in serving as a reviewer for various journals and conferences, Editor for EURASIP Journal on Applied Signal Processing (JASP), Guest Editor for a special issue in JASP, and Guest Cochair in various international conferences. His current research interests include electronic and optical signal processing for optical/electronic wireline/wireless communication, analog/mixed signal processing for high-speed data conversion, and multimedia signal processing.

EURASIP Journal on Applied Signal Processing 2005:10, 1593–1602

Linear and Nonlinear Crosstalk Evaluation in DWDM Networks Using Optical Fourier Transformers

R. Llorente
Fibre-Radio Group, Nanophotonics Technology Centre, Polytechnic University of Valencia, 46022 Valencia, Spain
Email: rllorent@dcom.upv.es

R. Clavero
Fibre-Radio Group, Nanophotonics Technology Centre, Polytechnic University of Valencia, 46022 Valencia, Spain
Email: raclaga@ntc.upv.es

F. Ramos
Fibre-Radio Group, Nanophotonics Technology Centre, Polytechnic University of Valencia, 46022 Valencia, Spain
Email: framos@dcom.upv.es

J. Marti
Fibre-Radio Group, Nanophotonics Technology Centre, Polytechnic University of Valencia, 46022 Valencia, Spain
Email: jmarti@ntc.upv.es

Received 1 April 2004; Revised 11 October 2004

A novel DWDM channel monitoring technique based on the conversion from wavelength domain to time domain by performing a real-time optical Fourier transform over the whole DWDM system bandwidth is proposed and experimentally demonstrated. The use of chromatic dispersion-based optical Fourier transformers has been validated in the case of a spectrum comprising light from different uncorrelated sources. Linear and nonlinear crosstalks between the DWDM channels appear as amplitude noise at specific time positions. The correspondence of this amplitude noise with the crosstalk spectral distribution is evaluated theoretically and experimentally.

Keywords and phrases: crosstalk monitoring, optical Fourier transform, optical signal processing, DWDM transmissions, ultrahigh-speed optical transmissions, OTDM.

1. INTRODUCTION

The increasing demand for higher transport capacity in DWDM core networks can be fulfilled with different complementary approaches: by increasing the number of channels, increasing the transported bit rate per channel or decreasing the channel spacing. The later two approaches lead to an augment of the spectral efficiency of the DWDM network. As the spectral efficiency increases, the crosstalk between the DWDM channels arises as important transmission and also node functionality impairment. Next generation DWDM networks operating at ultrahigh bit rates, that is, 160 Gbps per channel, require precise channel transmission quality assessment systems inside the network infrastructure. For this reason fibre link condition and optical node functionalities have to be carefully monitored in order to enable ultracapacity (greater than 10 Tbps) all-optical DWDM networks [1].

From the network node point of view, second-generation all-optical nodes are especially susceptible to crosstalk [2, 3] which is accumulated at each node along the optical path [4]. Crosstalk in the optical node can be classified as either heterodyne (crosstalk between signals at different wavelengths) or homodyne (crosstalk between signals at the same nominal wavelength, also known as in-band crosstalk). Homodyne crosstalk can be further subdivided into homodyne coherent crosstalk, if it is produced between phase-correlated signals, and homodyne incoherent crosstalk, if produced between signals which are not phase-correlated [3]. The most important noise contribution in the optical node is the homodyne noncoherent crosstalk [3, 4]. This noise is originated inside the network node due to nonperfect blocking of channels at the same nominal wavelength during channel extraction, multiplex, demultiplex, or channel wavelength conversion operations. This crosstalk is incoherent, since it is originated by a different channel and

cannot be eliminated once generated. This incoherent homodyne crosstalk exhibits the same spectral power distribution present in any DWDM channel in the system and has to be properly monitored in order to guarantee node operation. Regarding the possible presence of additional homodyne intrachannel coherent crosstalk (originated by channel power leaking and propagation by different paths inside the optical node), it does not possess a real problem since the difference between the light paths inside the network node is usually longer than the coherence length of the laser source [4].

In the DWDM network the optical nodes are connected by optical transmission links. In order to reach maximum transport capacity, it is necessary to combine tight channel spacing and compact-spectrum modulations [5]. The link capacity limiting factor is the optical bandwidth of the transmission amplifiers. Commercially available erbium doped fibre amplifiers (EDFA) allow a usable bandwidth of 30–40 nm. To achieve bandwidths larger than 80 nm, novel amplifier architectures based on hybrid Raman amplification/erbium, tellurite doping, or hybrid Raman amplification/tellurite have been proposed [6] but these technologies are not readily available. Ultra-high-capacity DWDM networks require to fit the maximum number of channels in the amplifier bandwidth, which in turn require the use of spectrally efficient modulations like single sideband (SSB), vestigial sideband (VSB), nonreturn to zero (NRZ) or optical duobinary, or compact-spectrum modulation like carrier-suppressed return to zero (CS-RZ), alternate polarization RZ (AP-RZ) or alternate chirp RZ (AC-RZ), and/or polarization interleaving between the different WDM channels [5, 7]. Under these circumstances (close packaging of the WDM channels and compact spectrum of the DWDM channels) crosstalk monitoring is a key element for the optical link reliability as transmitter, or optical path degradation will impact the link availability very quickly [1].

Crosstalk in the optical link arises from different sources: linear crosstalk between adjacent DWDM channels due to insufficient channel spacing, nonlinear crosstalk arising from the wavelength interaction due to the fibre nonlinearities, and also intrachannel time crosstalk produced by the pulse broadening when pulses propagate along the dispersion map of the transmission link. Linear crosstalk is important as the degradation of network equipment, channel central frequency shift due to temperature drift in DFB-LD lasers, or degradation in optoelectronic transmitters will strongly impact the performance if tight channel spacing is used [7]. Nonlinear crosstalk is produced in the DWDM link from nonlinear effects mainly four-wave mixing (FWM) between different channels [8]. This effect although mitigated by the dispersion map is enhanced when the channel separation is very tight and nonuniform channel spacing can not be used [8, 9].

In the case of intrachannel crosstalk, this noise is due to cross-phase modulation (XPM) and four-wave mixing in the time domain. Time-domain FWM is reflected in amplitude fluctuation as reported in [10]. In dispersion-managed transmission links with small average dispersion this noise increases with distance and its mitigation requires the use of loss compensation by means of Raman amplification. In the case of intrachannel XPM this is reflected in a mean frequency shift of the pulses which is translated to timing jitter through fibre dispersion [11]. If negligible average dispersion is achieved along the transmission link, this effect is suppressed [11, 12]. This is the particular case of ultracapacity DWDM networks. Transmission at 160 Gbps channel bit rate requires that the optical link between nodes be carefully planned in advance as pointed out in the small number of trials performed over actual SSMF [13]. The dispersion map used requires extremely low residual dispersion (1.2 ps/nm max. deviation for 160 km SSMF transmission) and also dispersion slope compensation [14]. It is also worth noting that the complete chromatic dispersion compensation along the system makes the accumulated chirp in the pulses negligible [12].

Common channel monitoring techniques rely on the use of arrayed-wavelength gratings (AWG) [15], Fabry-Perot filters [16], acousto-optic tunable filters [17], or tunable active filters [18] in order to separate the different WDM channels and to evaluate the channel noise power in absence of transmitted power, that is, stopping the channel operation. Other nondisruptive techniques propose the use of pilot sideband tones besides the digital data spectrum in order to evaluate the degradation of the transmitted information for each wavelength [19]. This method lacks transparency through the optical network nodes, like optical add-drop multiplexers (OADM) and optical cross-connects (OXC), as the pilot tones are eliminated after regeneration inside the network element. Another technique proposes the use of low-speed modulated signals overimposed on the main bit stream [20]. This technique requires the use of specific modulation and demodulation systems which can increase system complexity.

Our target is to monitor the crosstalk level in the DWDM network node either if it is originated inside the node or in the transmission link. The herein proposed system is based on optically time gating the DWDM signal during a specific time duration (one-bit time-slot) and to perform a continuous optical Fourier transform (OFT) comprising the whole set of transmitted wavelengths. This approach does not require to stop the channel operation. Once the spectral information has been brought to time domain, the basic parameters such as amplitude (channel power) or central wavelength can be evaluated. The key advantage of the proposed technique is that, additionally, it is feasible to evaluate the crosstalk spectral distribution (the interfering optical power distribution with frequency) that any DWDM channel is suffering. This is accomplished by the real-time OFT operation. Once the spectral data is in the time domain, it is photodetected and sampled to be postprocessed. In presence of crosstalk, the samples captured exhibit a uniformly distributed noise at specific time positions. By measuring the noise power and the noise time position we can evaluate the crosstalk level and the crosstalk spectral location, respectively.

The crosstalk spectral distribution gives us more information than the total crosstalk figure in dB usually given in crosstalk measurements. The knowledge of the spectral distribution opens up the opportunity to identify the crosstalk source, that is, if it comes (in the case of linear crosstalk) from the channel at the right or at the left. It is clear that if the crosstalk source is an adjacent channel, its spectral distribution will be at one side of the channel spectrum. If it is originated by nonlinear crosstalk due to FWM (evenly spaced channels), then the noise spectral maximum will be at the central frequency of the crosstalked channel. In this case, once the crosstalk type is identified, we can correlate the presence of crosstalk in a specific channel with the presence of power in a given set of channels. This correlation allows us to identify the crosstalk origin channels.

As discussed before, the OFT is the key operation of the proposed system and it is performed over the whole optical system bandwidth. The OFT operation is implemented here by means of a dispersive element. The use of chromatic dispersion effect in standard single-mode fibre (SSMF) to convert the spectral information to time domain has been previously reported in [21, 22]. Linearly chirped fibre Bragg gratings (LCFBG) have been also proposed as dispersive elements capable of performing this operation [23]. The Fourier transformers using SSMFs give the advantage of nearly no bandwidth limitation, which makes this approach especially interesting for large-channel-count DWDM systems. The use of dispersive devices in real-time spectrum analysis systems was proposed for DWDM systems in [21] and experimentally demonstrated for a phase coherent spectrum [24] (coherency means here that all the lights present at the spectrum have the same phase reference, i.e., the whole spectrum comes from only one optical source). Actual DWDM systems, however, exhibit lack of phase coherency between the different WDM channels as they come from different optical sources. This lack of spectral coherency leads to the presence of an amplitude noise after the OFT that depends on the crosstalk level. The amplitude noise power level gives us information about the crosstalk level that the channel under evaluation is suffering from the rest of the WDM channels. This correspondence will be evaluated in the next section.

Another interesting feature of the proposed OFT approach is the time expansion effect introduced by the dispersive element. In the proposed system, the output from the OFT cell is a time-stretched version of the Fourier transform of the input signal envelope. This expansion in time is a convenient feature as it is possible to use commercially available electrical analog-to-digital conversion (ADC) circuitry in order to capture (convert the electrical photodetected signal to digital data) the spectrum and to postprocess the results.

2. PRINCIPLE OF OPERATION

We consider a generic crosstalk situation in a DWDM network employing Gaussian RZ modulation. The network uses K channels (wavelengths) with channel spacing Δ (rad/s). In conventional DWDM networks the channel separation Δ is wide enough so that the spectral overlap between the channels is negligible. This is not the case of high-spectral efficiency networks where the adjacent channel power can leak due to not perfect channel filtering at multiplexing stage, during channel extraction operation or because of transmitter degradation.

In the general case, for the sake of simplicity we will consider that the DWDM channels are evenly spaced around a central angular frequency ω_0 so the central angular frequency for the kth channel is $\omega_0 + \omega_k = \omega_0 + \Delta \cdot k$, where $k = -(K-1)/2 \cdots (K-1)/2$. This DWDM spectral space can be described by equation (1), where $s_A(t)$ is the bit-synchronised received optical complex field envelope around the central angular frequency of the DWDM transmission system (ω_0). The channels are considered synchronised in time as this is the worst-case crosstalk scenario. In equation (1) A_k stands for the optical field amplitude and $b_{k,n}$ is the transmitted data (mark or space) by the channel k in the nth-bit time-slot. The parameter T_k stands for the Gaussian pulse half-width at $1/e$ fall from peak. $R = 1/T_b$ is the system bit rate. Figure 1 shows the DWDM system in a bidimensional space (angular frequency, time) described by

$$s_A(t) = \sum_k \sum_n A_k b_{n,k} \exp\left(j\omega_k t\right) \exp\left[-\frac{1}{2}\left(\frac{t - nT_b}{T_k}\right)^2\right]. \quad (1)$$

Our analysis targets crosstalk evaluation in a generic network node. Such node, optical add-drop multiplexer (OADM) or optical cross-connect (OXC), is depicted in Figure 2 (top diagram). This generic node can perform DWDM insertion/extraction or wavelength conversion whether it is an OADM (optical add-drop multiplexer) or an OXC (optical cross-connect). The channel monitoring block is shown before the node operation, controlling a channel equalizer and monitoring the crosstalk introduced in the link. This block can be placed inside the node or directly in the transmission link sending information through a supervisory channel.

The proposed DWDM crosstalk monitoring system architecture is shown in Figure 2 (bottom schematic). The system consists of a gating switch which selects only one bit period (T_b time duration) of every N bits. This gating switch can be implemented in the optical domain by using saturable absorbers [25, 26], interferometric devices like the nonlinear optical loop mirror (NOLM) [27] and the ultrafast nonlinear interferometer (UNI) [28], or a Mach-Zender architecture employing semiconductor optical amplifiers (SOA) [28]. These devices can work at 160 Gbps core network line rates. The purpose of the gating switch is to select only one bit-slot of every N bits (one bit-slot contains only one bit from all the wavelengths of the DWDM system as the optical gate does not impose any bandwidth limitation) and to feed it into the OFT cell which performs an optical Fourier transform. The exact value of N depends on the chromatic dispersion of the OFT cell as will be discussed later. This gating can be performed in a continuous basis or every time that a channel monitoring is desired. Equation (2) describes

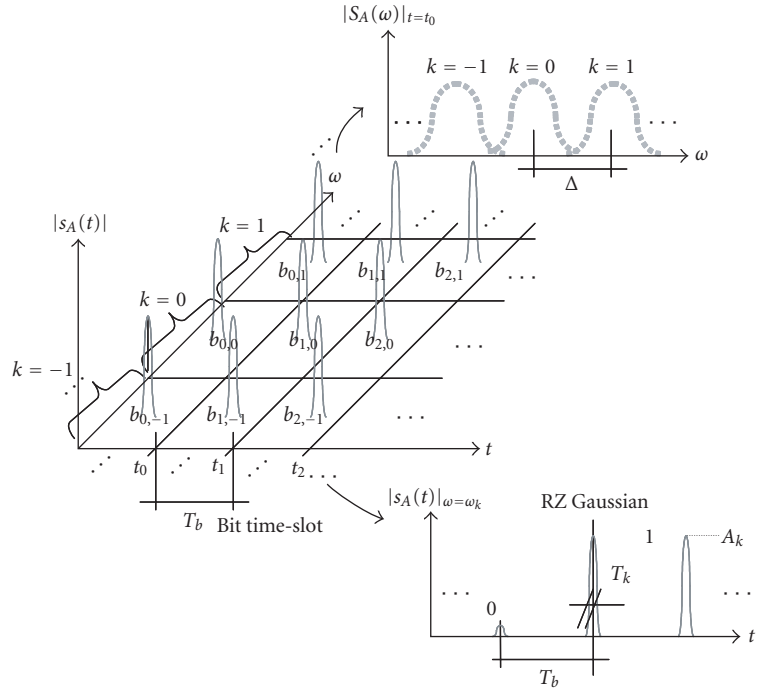


FIGURE 1: Bidimensional (angular frequency, time) DWDM space in modulus, $|s_A(t,\omega)|$. Gaussian RZ signaling is used. A spectrum cut (in modulus) $|s_A(\omega)|_{t=t_0}$ is shown on top for a given time t_0. This spectrum comprises K wavelengths with Gaussian shape. The time signal (in modulus) is also shown (bottom) for a given channel k. The optical Fourier transform operation is performed over the T_b bit time-slot shown.

the complex field envelope after theoretical gating of an arbitrary nth bit and its Fourier transform. The time reference has been placed in the centre of the pulse to simplify calculations. After the optical gating, one bit of the DWDM channels is fed to the Fourier transformer to bring the spectral information into the time domain. Once in time domain this spectrum is sampled at R_s rate:

$$s_{B,n}(t) = \sum_k A_k b_{n,k} \exp(j\omega_k t) \exp\left[-\frac{1}{2}\left(\frac{t}{T_k}\right)^2\right],$$
$$S_{B,n}(\omega) = \sum_k A_k b_{n,k} T_k \sqrt{2\pi} \exp\left[-\frac{1}{2}(\omega-\omega_k)^2 T_k^2\right]. \quad (2)$$

In Figure 3 the gating and optical Fourier transform process is shown in detail. As mentioned before, at the output of the OFT cell, the spectral information is obtained as a time signal. Therefore, we have performed the spectrum-to-time conversion of one time-slot comprising the whole set of K wavelengths of the DWDM system. The time signal at the output of the OFT cell [23] is described by (3). The output pulses width depends on the first-order dispersion coefficient $\ddot{\Phi}$. The pulses amplitudes involve the factor L standing for the OFT device optical losses. This signal is then photodetected and converted from electrical continuous signal to digital data in order to evaluate the spectrum. This conversion can be performed by a commercial ADC:

$$\begin{aligned} s_{C,n}(t) &= L \cdot \exp\left(-\frac{jt^2}{2\ddot{\Phi}}\right)\{\mathrm{FT}[s_B(t)]\}_{\omega=t/\ddot{\Phi}} \\ &= L \cdot \exp\left(-\frac{jt^2}{2\ddot{\Phi}}\right) S_B\left(\frac{t}{\ddot{\Phi}}\right) \\ &= L \cdot \exp\left(-\frac{jt^2}{2\ddot{\Phi}}\right) \\ &\quad \times \sum_k A_k b_{n,k} T_k \sqrt{2\pi} \exp\left\{-\frac{1}{2}\left[\frac{(t-k\cdot\Delta\cdot\ddot{\Phi})^2 T_k^2}{\ddot{\Phi}^2}\right]\right\}. \end{aligned} \quad (3)$$

From (3), if the dispersive element presents a first-order dispersion coefficient of $\ddot{\Phi} = -\partial^2\Phi/\partial\omega^2$, the resulting time signal after OFT comprises Gaussian shape pulses (as corresponds to the Fourier transform of a Gaussian pulse) of width $\ddot{\Phi}/T_k$ separated in time by the factor $\Delta \cdot \ddot{\Phi}$. Each channel spectrum outputs from the OFT cell with $k \cdot \Delta \cdot \ddot{\Phi}$ separation. The time necessary to output all the wavelengths of the DWDM system is $K \cdot \Delta \cdot \ddot{\Phi}$. The gating ratio N must be high enough to produce a complete operation, so N has to be larger than $K \cdot \Delta \cdot \ddot{\Phi}/T_b$, where T_b is the system bit period. By photodetecting and sampling $s_{C,n}(t)$, we can evaluate the spectrum from the K system

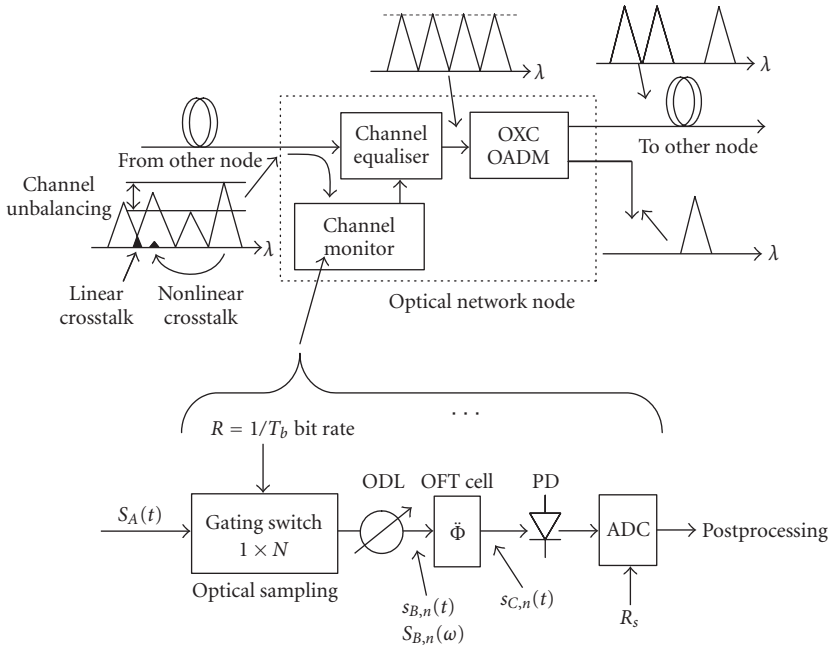


Figure 2: Top diagram shows generic network node structure. Proposed channel monitoring system architecture based in OFT operation using SSMF is shown in the schematic (bottom).

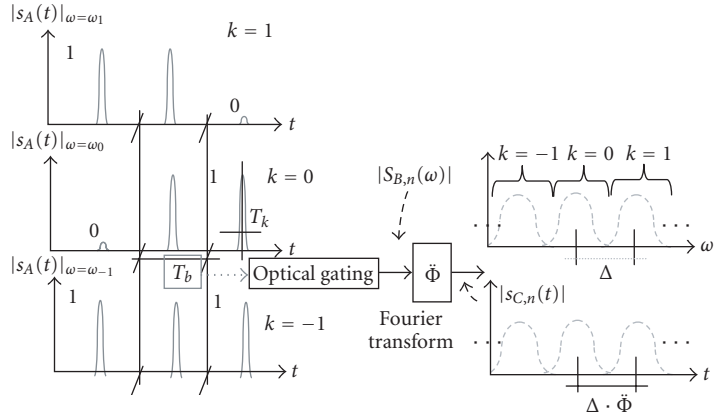


Figure 3: Channel monitoring principle of operation: T_b time-slot id-gated and fed to the optical Fourier transformer. After Fourier transform we obtain the spectrum components in time domain.

channels. The minimum sampling rate should allow one sample per channel, that is, $R_{s\_\min} = 1/(\Delta \cdot \Phi)$. After OFT, photodetection, and sampling, given the spectrum information, we can evaluate the channel amplitude (power) level, the channel separation (time difference between power peaks), and the crosstalk the channel is suffering, as will be discussed later. By example, the use of a commercial ADC with a sampling rate of 800 Msps (readily available from different vendors) leads to 1.25 nanosecond temporal resolution. The accumulated dispersion from 150 km SSMF (−19 296 ps/THz) OFT gives a spectral resolution (frequency separation between samples) as low as 64 GHz. To achieve a fine sampling of the spectrum envelope, the proposed system includes an optical delay line (ODL) shown in Figure 2 which delays the input to the ADC in a computer-controlled way. Sweeping the optical delay, the different points of the DWDM spectrum will be sampled and we can reconstruct its shape completely. The sampled waveform corresponds to a spectral-domain to time-domain conversion taking place, and the power overlapping in the spectrum will appear as an amplitude noise in the Gaussian pulses after Fourier transform.

In order to characterize the noise in the time domain, we will evaluate now the output from the OFT when two generic wavelengths with Δ channel separation overlap to some extent in spectrum: we consider two adjacent DWDM channels, k and $k+1$, bearing Gaussian RZ pulses in the DWDM system of amplitude A_k and A_{k+1} and pulse width T_k and T_{k+1}, respectively. This is a general case from equation (2) where the interaction of any two channels is evaluated. This generic situation can be particularized for different kinds of crosstalk described in the previous section. For linear crosstalk, the noise origin is the spectrum overlap from the adjacent channel, the spectral separation between lights of channel k and $k+1$ is the channel spacing $\Delta \gg 0$, and the amplitudes A_k and A_{k+1} will be in the same order of magnitude, $A_k \approx A_{k+1}$. In the case of nonlinear crosstalk, the noise origin is another channel which, due to nonlinear effect like FWM, appears spectrally overimposed (the channels are equally spaced) exactly at the centre of the channel under study k. In this case the spectral separation between channel k and the interference $k+1$ is negligible $\Delta \approx 0$, but due to limited FWM efficiency, the interference amplitude will be much lower $A_{k+1} \ll A_k$.

Following the general analysis, the channels k and $k+1$ come from different optical sources through different optical paths. As these channels are mutually noncoherent, the optical phase difference is considered by the term ϕ which is a uniformly distributed random variable. Then, the complex optical envelope $s(t)$ and spectrum envelope $S(\omega)$ for channels k and $k+1$ can be written as

$$\begin{aligned} s_k(t) &= A_k \exp(j\omega_k t) \exp\left[-\frac{1}{2}\left(\frac{t}{T_k}\right)^2\right] \\ &\xrightarrow{\text{FT}} A_k T_k \sqrt{2\pi} \exp\left(-\frac{1}{2}((\omega-\omega_k)T_k)^2\right) \\ &\triangleq S_k(\omega), \\ s_{k+1}(t) &= A_{k+1} \exp\left[j(\omega_k+\Delta)t + j\phi\right] \exp\left[-\frac{1}{2}\left(\frac{t}{T_{k+1}}\right)^2\right] \\ &\xrightarrow{\text{FT}} A_{k+1} T_{k+1} \sqrt{2\pi} \\ &\quad \times \exp\left(-\frac{1}{2}((\omega-\omega_k-\Delta)T_{k+1})^2\right) \exp(j\phi) \\ &\triangleq S_{k+1}(\omega)\exp(j\phi), \end{aligned} \tag{4}$$

where A_{k+1}, ω_{k+1}, and T_{k+1} stand for the optical field amplitude, the central angular frequency, and the pulse width at $1/e$, respectively, for wavelength $k+1$. From (3), after the Fourier transform element we have the combination of both spectra:

$$s_c'(t) = L \cdot e^{-jt^2/2\ddot{\Phi}} \cdot \left\{S_k(\omega) + S_{k+1}(\omega)\exp(j\phi)\right\}_{\omega=t/\dot{\Phi}}. \tag{5}$$

The term $L \cdot e^{-jt^2/2\ddot{\Phi}}$ is a complex constant whose phase will be lost after photodetection. The photodetected signal at the OFT output is proportional to the optical intensity $I_{\text{out}}(t)$ as described by (6). This intensity corresponds to the combination of the wavelengths k and $k+1$ power spectra with noise

$$\begin{aligned} I_{\text{out}} &\sim \left| S_k(\omega) + S_{k+1}(\omega)\exp(j\phi)\right|^2_{\omega=t/\dot{\Phi}} \\ &= S_k^2\left(\frac{t}{\ddot{\Phi}}\right) + S_{k+1}^2\left(\frac{t}{\ddot{\Phi}}\right) + n(t). \end{aligned} \tag{6}$$

The noise $n(t)$ involves the random variable ϕ which is the phase difference between both wavelengths at any time position. Except for if both lights are provided by the same optical source and are guided through the same optical path, we can only say that ϕ adopts a random value between $-\pi$ and π with uniform distribution. This noise is a random process and is described by equation (7). Every time a channel estimation is done, we obtain one realization of the process. For simplicity, the realization number is not shown in

$$n(t) = 2S_k\left(\frac{t}{\ddot{\Phi}}\right) S_{k+1}\left(\frac{t}{\ddot{\Phi}}\right)\cos(\phi). \tag{7}$$

From (7) the OFT output at any specific time position t_i, that is, $n(t_i)$, is a random variable. This random variable will fluctuate according to the random nature of the phase difference ϕ. The noise variance (fluctuation) for time position is given by (8) and shows that the noise power in time domain follows (except by a proportionality factor) the optical spectrum overlap $S_k(t_i/\ddot{\Phi})S_{k+1}(t_i/\ddot{\Phi})$:

$$\sigma^2_{n,t_i} = 4\pi S_k^2\left(\frac{t_i}{\ddot{\Phi}}\right) S_{k+1}^2\left(\frac{t_i}{\ddot{\Phi}}\right). \tag{8}$$

Equation (8) reflects that the noise power at the specific time position t_i depends only on the crosstalk at the specific spectrum angular frequency $\omega_0 + t_i/\ddot{\Phi}$. Sweeping the time position t_i, we sweep over the complete spectral range. In this way, accumulating the samples for any time position t_i, we can evaluate its mean and variance for the corresponding frequency. The variance represents the crosstalk noise as this is produced by the crosstalk (if no crosstalk is present, no variance can be observed) as is discussed in the next section. In this way we can calculate the crosstalk spectral distribution. This result is difficult to achieve with current spectrum analysis techniques and is enabled by the feasible real-time Fourier Transform operation proposed in [21].

The crosstalk spectral distribution gives us total crosstalking power (noise integrated over the channel under study bandwidth) and, from the envelope of the noise spectral distribution, we can identify the crosstalk nature: if it is located at one side of the channel bandwidth, then the origin is linear crosstalk from the adjacent channel as is demonstrated in the experimental work in the next section. If the noise distribution is centered in the channel bandwidth under study and follows the Gaussian power profile, then the

source is nonlinear FWM (provided the channels are equally spaced). In this case, performing a statistical correlation of this noise with the presence of power (marks) transmitted in any other DWDM channel, we can identify the crosstalk source.

This result can be extended to more than two interfering channels: we evaluate now the influence of multiple channels interfering a given channel k. Noncoherent crosstalk, which means that the interfering channels are uncorrelated, will be reflected again as the power spectrum overlap. In the case of M interferers over the kth channel, the noise from any channel of the system will also be additive in power. Equation (9) is an generalization of equation (8) for the interaction of channel m over channel k:

$$\sigma^2_{k,m,t_i} = 4\pi S_k^2\left(\frac{t_i}{\ddot{\Phi}}\right) S_m^2\left(\frac{t_i}{\ddot{\Phi}}\right). \tag{9}$$

The total noise power, if M channels affect, will be given by (10), as the different $n^2_{k,m}(t)$ for each m interfering channel are uncorrelated noise processes for each m channel as the channels are not mutually coherent:

$$\sigma^2_{k,M,t_i} = \sum_M \sigma^2_{k,m,t_i} = S_k^2\left(\frac{t_i}{\ddot{\Phi}}\right) \sum_M S_m^2\left(\frac{t_i}{\ddot{\Phi}}\right). \tag{10}$$

Equation (10) reflects that the total crosstalk noise spectral profile is the sum of the spectrum envelope overlaps of all channels $m \neq k$. In the case of channels with large frequency separation, its influence will be quickly dismissed, as the Gaussian spectral shape decays quickly. In the case of nonlinear crosstalk, as mentioned at the beginning of the section, the channel separation is almost zero, but due to the low nonlinear process efficiency the interference amplitude A_m will be much lower than the channel under study A_k. The particular FWM efficiency value depends on the particular dispersion and amplification map [9] of the system, and will be reflected in the A_m amplitude implicit in the term $S_m^2(t_i/\ddot{\Phi})$ in (10).

3. EXPERIMENTAL RESULTS AND DISCUSSION

As a proof of concept, a set-up has been arranged to evaluate the proposed monitoring technique operation on two DWDM channels transporting Gaussian RZ pulses with strong linear crosstalk following the calculations in the previous section. This linear crosstalk is induced arranging the different channels with enough frequency overlap. If we would consider in-band crosstalk, we would simply overlap the wavelengths completely. The set-up is shown in Figure 4 and consists of two tunable mode-locked laser sources with a wavelength range of +/−12 nm around 1552 nm. The generated pulses are mutually noncoherent and exhibit $T_{\mathrm{FWHM}} =$ 1.6 picoseconds. This pulse width corresponds to 25% RZ signaling at 160 Gbps bit rate per channel. The repetition rate is set to $R = 1.04150$ GHz using an external RF reference signal. A RF phase shifter was employed for fine tuning in order to vary the relative pulse positions. An optical delay line for coarse tuning may be also used, as shown in Figure 4. After combining both pulsed lights, the signal is passed through a dispersive device (OFT cell). In the experiment a coil of 2.1 km of SSMF was used, providing a total dispersion of $\ddot{\Phi} = -42.855\ \mathrm{ps}^2/\mathrm{rad}$, which is large enough to perform a successful Fourier transform, as it meets the condition $|T_b^2/\ddot{\Phi}| \ll 1$ reported in [23]. The large optical bandwidth of the fibre-based OFT cell presents the advantage of being wide enough to allocate the whole system bandwidth. This would be difficult to accomplish if a linearly chirped fibre Bragg grating was used for FT operation in a large WDM system, as the system bandwidth might exceed the grating bandwidth. Furthermore, the signal is photo-detected and monitored in the sampling scope.

Figure 5a shows the spectrum of two DWDM channels from the experimental set-up described above with channel spacing of 3.72 nm, whereas Figure 5b shows the output of the OFT cell after photodetection (electrical output). We can observe the good agreement between the optical spectrum at the input and the time electrical waveforms shape at the output, hence demonstrating the proper FT operation.

In order to evaluate the fibre-based Fourier transformer behavior in the presence of linear crosstalk, a strong wavelength overlap is introduced by getting closer (spectral spacing of 2.3976 nm) the DWDM wavelengths. Figure 5c shows the spectrum in this situation, whereas Figure 5d shows the output from the OFT cell. The presence of the predicted amplitude noise dependent on the linear crosstalk at the input is clearly shown as was expected from (7). If no crosstalk is present (e.g., $A_k = 0$) all the amplitude noise vanishes as can be seen in the insets (e) and (f) in Figure 5.

This amplitude noise (marked with the rectangle in Figure 5) is present only at the time positions where there was spectrum overlap in the frequency domain. In order to assess this correspondence, we have evaluated the accumulated noise after several successive OFT (channel evaluation) operations. Figure 6 shows the optical power spectrum overlap $4\pi S_k^2(\omega)S_{k+1}^2(\omega)$ over the electrical amplitude noise at OFT output $\sigma^2_{n,t}$. The correspondence of the spectrum overlap with the amplitude noise rms (noise power distribution over time) value shows good agreement, thus validating (8).

4. CONCLUSIONS

A novel channel monitoring technique for high-speed DWDM networks based on performing the Fourier transform with a simple dispersive element has been proposed and validated. Gaussian 1.6-picosecond pulses, typical in 25% RZ 160 Gbps transmission, have been used for demonstration purposes. Proper OFT cell operation has been demonstrated using 2.1 km of SSMF. Amplitude noise correspondence, after OFT operation and photodetection, with the optical spectral overlap (linear crosstalk) profile has been evaluated and experimentally validated.

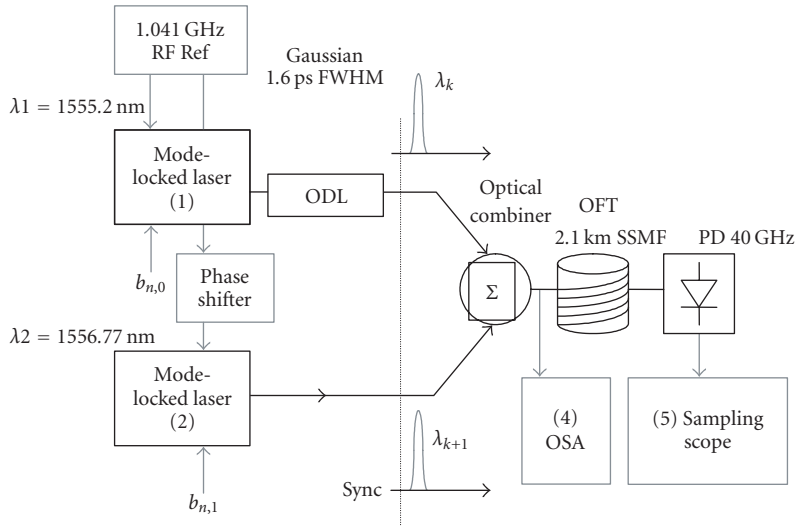


FIGURE 4: Experimental set-up. Combination of two nonmutually phase coherent 1.6-picosecond Gaussian RZ DWDM channels. OFT cell used comprises 2.1 km of SSMF.

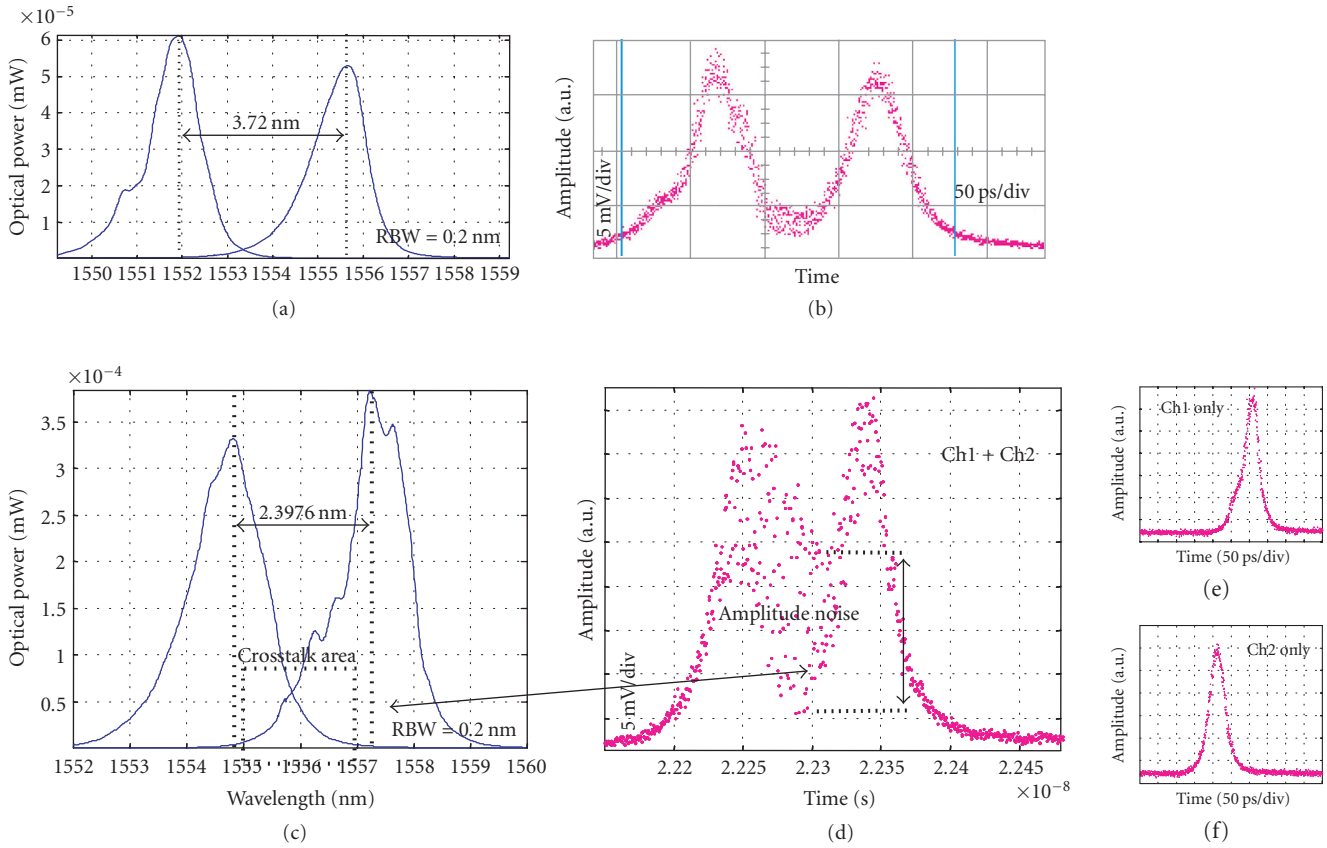


FIGURE 5: (a) Optical spectra at the input of the OFT cell. (b) Electrical signal after OFT and photodetection for 3.72 nm channel spacing. (c) Optical spectra with 2.3976 nm channel separation (11.62 dB crosstalk). (d) Corresponding electrical traces at the output of the OFT cell. Insets (e) and (f) show the two channels when no crosstalk is present.

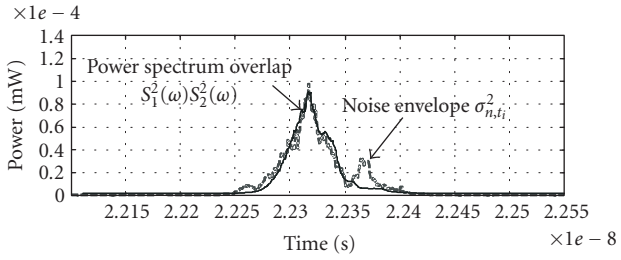


Figure 6: Optical power overlap profile at 2.3976 nm wavelength separation (solid line). Electrical amplitude noise power distribution envelope in time (dashed line).

ACKNOWLEDGMENTS

This work was carried out in the framework of the IST-2000-28657 TOPRATE project partially funded by the European Commission. The regional Valencian Government is also acknowledged for partly funding this project.

REFERENCES

[1] D. C. Kilper, R. Bach, D. J. Blumenthal, et al., "Optical Performance Monitoring," *J. Lightwave Technol.*, vol. 22, no. 1, pp. 294–304, 2004.

[2] R. Ramaswami and K. Sivarajan, *Optical Networks: A Practical Perspective*, Morgan Kaufmann, San Francisco, Calif, USA, 2001.

[3] E. Iannone, R. Sabella, M. Avattaneo, and G. De Paolis, "Modeling of in-band crosstalk in WDM optical networks," *J. Lightwave Technol.*, vol. 17, no. 7, pp. 1135–1141, 1999.

[4] E. L. Goldstein, L. Eskildsen, and A. F. Elrefaie, "Performance implications of component crosstalk in transparent lightwave networks," *IEEE Photon. Technol. Lett.*, vol. 6, no. 5, pp. 657–660, 1994.

[5] K. Fukuchi, "Wideband and ultra-dense WDM transmission technologies toward over 10-Tb/s capacity," in *Optical Fiber Communication Conference and Exhibit (OFC '02)*, pp. 558–559, March 2002.

[6] A. Mori, H. Masuda, K. Shikano, and M. Shimizu, "Ultra-wide-band tellurite-based fiber Raman amplifier," *J. Lightwave Technol.*, vol. 21, no. 5, pp. 1300–1306, 2003.

[7] T. Ono and Y. Yano, "Key technologies for terabit/second WDM systems with high spectral efficiency of over 1 bit/s/Hz," *IEEE J. Quantum Electron.*, vol. 34, no. 11, pp. 2080–2088, 1998.

[8] R. W. Tkach, A. R. Chraplyvy, F. Forghieri, A. H. Gnauck, and R. M. Derosier, "Four-photon mixing and high-speed WDM systems," *J. Lightwave Technol.*, vol. 13, no. 5, pp. 841–849, 1995.

[9] E. A. Golovchenko and N. S. Bergano, "Four-wave mixing in multispan dispersion-managed transmission links," *IEEE Photon. Technol. Lett.*, vol. 10, no. 10, pp. 1481–1483, 1998.

[10] M. J. Ablowitz and T. Hirooka, "Intrachannel pulse interactions in dispersion-managed transmission systems: energy transfer," *Optics Letters*, vol. 27, no. 3, pp. 203–205, 2002.

[11] M. J. Ablowitz and T. Hirooka, "Intrachannel pulse interactions in dispersion-managed transmission systems: timing shifts," *Optics Letters*, vol. 26, no. 23, pp. 1846–1848, 2001.

[12] M. J. Ablowitz and T. Hirooka, "Resonant intrachannel pulse interactions in dispersion-managed transmission systems," *IEEE J. Select. Topics Quantum Electron.*, vol. 8, no. 3, pp. 603–615, 2002.

[13] E. Lach, K. Schuh, M. Schmidt, et al., "7 × 170 Gbit/s (160 Gbit/s + FEC overhead) DWDM transmission with 0.53 Bit/s/Hz spectral efficiency over long haul distance of standard SMF," in *Proc. European Conference on Optical Communication (ECOC '03)*, pp. Th4.3.5-1–Th4.3.5-2, Rimini, Italy, September 2003.

[14] R. Ludwig, U. Feiste, C. Schmidt, et al., "Enabling transmission at 160 Gbit/s," in *Optical Fiber Communication and Exhibit (OFC '02)*, pp. 1–2, Anaheim, Calif, USA, March 2002.

[15] M. Teshinma, M. Koga, and K. I. Sato, "Performance of multiwavelength simultaneous monitoring circuit employing arrayed-waveguide grating," *J. Lightwave Technol.*, vol. 14, no. 10, pp. 2277–2285, 1996.

[16] G. Cocorullo, F. Della-Corte, M. Iodice, and I. Rendina, "Simple and low-cost silicon Fabry-Perot filter for WDM channel monitoring," in *Digest of the LEOS Summer Topical Meetings*, pp. IV45–IV46, Aventura, Fla, USA, July 2000.

[17] Q. Li, A. A. Au, C.-H. Lin, I. V. Tomov, and H. P. Lee, "Performance characteristics of a WDM channel monitor based on an all-fiber AOTF with an on-fiber photodetector," *IEEE Photon. Technol. Lett.*, vol. 15, no. 5, pp. 718–720, 2003.

[18] J. H. Chen, Y. Chani, J. Y. Fan, et al., "WDM channel monitoring and signal power control/equalization using integrated tuneable active filters," in *Digest of the IEEE/LEOS Summer Topical Meetings*, pp. 48–49, Montreal, Quebec, Canada, August 1997.

[19] R. Gaudino and D. J. Blumenthal, "WDM channel equalization based on subcarrier signal monitoring," in *Optical Fiber Communication Conference and Exhibit (OFC '98)*, pp. 167–168, San Jose, Calif, USA, February 1998.

[20] R. A. Jensen and D. G. Duff, "Monitoring undersea WDM systems with optical add/drop multiplexers," in *Optical Fiber Communication Conference and Exhibit (OFC '98)*, pp. 274–275, San Jose, Calif, USA, February 1998.

[21] J. Azaña and M. A. Muriel, "Real-time Fourier transformations performed simultaneously over multiwavelength signals," *IEEE Photon. Technol. Lett.*, vol. 13, no. 1, pp. 55–57, 2001.

[22] Y. C. Tong, L. Y. Chan, and H. K. Tsang, "Fibre dispersion or pulse spectrum measurement using a sampling scope," *IEE Electronics Letters*, vol. 33, no. 11, pp. 983–985, 1997.

[23] M. A. Muriel, J. A. Azaña, and A. Carballar, "Real-time Fourier transformer based on fiber gratings," *Optics Letters*, vol. 24, no. 1, pp. 1–3, 1999.

[24] N. K. Berger, B. Levit, A. Beckker, and B. Fisher, "Real-time optical spectrum analyzer based on chirped Bragg gratings," *IEE Electronics Letters*, vol. 36, no. 14, pp. 1189–1191, 2000.

[25] J. Mangeney, N. Stelmakh, A. Shen, et al., "Sub-picosecond wideband efficient saturable absorber created by high energy (200 MeV) irradiation of Au^+ ions into bulk GaAs," *IEE Electronics Letters*, vol. 34, no. 8, pp. 818–820, 1998.

[26] H. Kobayashi, R. Takahashim, Y. Matsuoka, and H. Iwamura, "1 Tbit/s demultiplexing using low temperature grown InGaAs/InAlA smultiple quantum wells," *IEE Electronics Letters*, vol. 34, no. 9, pp. 908–910, 1998.

[27] N. Chi, L. Oxenlowe, A. Siahlo, and P. Jeppesen, "All-optical fiber signal processing based on balanced NOLM and imbalanced NOLM," in *Proc. European Conference on Optical Communication (ECOC '02)*, Copenhagen, Denmark, September 2002, paper 6.3.5.

[28] C. Schubert, S. Diez, J. Berger, et al., "160-Gb/s all-optical demultiplexing using a gain-transparent ultrafast-nonlinear interferometer (GT-UNI)," *IEEE Photon. Technol. Lett.*, vol. 13, no. 5, pp. 475–477, 2001.

[29] M. Heid, S. L. Jansen, S. Spalter, E. Meissner, W. Vogt, and H. Melchior, "160-Gbit/s demultiplexing to base rates of 10 and 40 Gbit/s with a monolithically integrated SOA-Mach-Zehnder interferometer," in *Proc. European Conference on Optical Communication (ECOC '02)*, Copenhagen, Denmark, September 2002, paper 8.4.3.

R. Llorente was born in Valencia, Spain. He received the M.S. degree in telecommunication engineering from the Polytechnic University of Valencia in 1998. Since then, he has been with the Fibre-Radio Group at the same university. Currently he is a Lecturer at this university in the Communications Department. He has participated in several national and European research projects on areas such as biophotonics, optical signal processing, and OTDM/DWDM transmission systems. He has authored or coauthored more than 20 papers in international journals and conferences and has authored two patents. He has acted as a reviewer for the IEE Institute. His research interests include hybrid electro-optical signal processing and high-capacity optical links and networks.

R. Clavero received the Ingeniero de Telecomunicación degree from the Polytechnic University of Valencia, Spain, in 2002 where she is currently working as a researcher at the Valencia Nanophotonics Technology Centre. She is also working towards the Ph.D. degree. She has been actively involved in European-level projects, such as IST TOPRATE and IST LASAGNE. Her research interests include OTDM and OWDM transmission systems, PMD monitoring, all-optical signal processing, and optical networking. She has coauthored over 5 papers in international journals and conferences.

F. Ramos was born in Valencia, Spain, on April 2, 1974. He received the M.S. and Ph.D. degrees in telecommunication engineering from the Polytechnic University of Valencia in 1997 and 2000, respectively. Since 1998, he has been with the Department of Communications at the same university, where he is now an Associate Professor. He has participated in several national and European research projects on areas such as optical access networks, broadband wireless systems, and optical networking. Professor Ramos has coauthored more than 60 papers in international journals and conferences and he has acted as a reviewer for the IEEE and Taylor and Francis publishers. He is also the recipient of the Prize of the Telecommunication Engineering Association of Spain for his dissertation on the application of optical nonlinear effects in microwave photonics. His research interests include nonlinear fiber optics, optical-phase conjugation, microwave and millimeter-wave optical systems, broadband access networks, and high-speed optical networks. During the last years, as a member of the Valencia Nanophotonics Technology Centre, his research is focused on the application of active Mach-Zehnder interferometers to all-optical signal processing in OTDM/DWDM networks and all-optical label swapping networks.

J. Marti received the Ingeniero de Telecomunicación degree from the Polytechnic University of Catalunya (UPC), Catalunya, Spain, in 1991, and the Doctor Ingeniero de Telecomunicación degree (Ph.D.) from the Polytechnic University of Valencia, Valencia, Spain, in 1994. During 1989 and 1990, he was an Assistant Lecturer at the UPC. Since 1991 to 2000, he obtained the positions of Lecturer and Associate Professor at the Telecommunication Engineering Faculty, Polytechnic University of Valencia, where he is currently a Full Professor and leads the Fibre-Radio Group. Recently he has been appointed as a Director of the Nanophotonics Technology Centre. He has authored 7 patents and over 150 papers in refereed international technical journals and leading international conferences in the fields of fibre-radio systems, access networks, ultra-high bit-rate WDM networks, advanced optical processing techniques, fibre gratings, and planar photonic crystals. Nowadays, he is a Project Coordinator of FP6 IST projects GANDALF and LASAGNE. He has led many other national and international research projects, as a Coordinator of the FP5 IST-OBANET project. He is currently participating in IST-TOPRATE (Terabit/s Optical Transmission Systems). Professor Marti is or has been a member of the Technical Program Committee of several conferences. He is also the recipient of several academic and industrial awards in Spain.

EURASIP Journal on Applied Signal Processing 2005:10, 1603–1616

Analysis of Optical CDMA Signal Transmission: Capacity Limits and Simulation Results

Aminata A. Garba
Department of Electrical and Computer Engineering, McGill University, 3480 University Street, Montreal, QC, Canada H3A 2A7
Email: agarba@tsp.ece.mcgill.ca

Raymond M. H. Yim
Division of Engineering and Applied Sciences, Harvard University, Cambridge, MA 02138, USA
Email: ryim@fas.harvard.edu

Jan Bajcsy
Department of Electrical and Computer Engineering, McGill University, 3480 University Street, Montreal, QC, Canada H3A 2A7
Email: jbajcsy@tsp.ece.mcgill.ca

Lawrence R. Chen
Department of Electrical and Computer Engineering, McGill University, 3480 University Street, Montreal, QC, Canada H3A 2A7
Email: chen@photonics.ece.mcgill.ca

Received 5 April 2004; Revised 18 January 2005

We present performance limits of the optical code-division multiple-access (OCDMA) networks. In particular, we evaluate the information-theoretical capacity of the OCDMA transmission when single-user detection (SUD) is used by the receiver. First, we model the OCDMA transmission as a discrete memoryless channel, evaluate its capacity when binary modulation is used in the interference-limited (noiseless) case, and extend this analysis to the case when additive white Gaussian noise (AWGN) is corrupting the received signals. Next, we analyze the benefits of using nonbinary signaling for increasing the throughput of optical CDMA transmission. It turns out that up to a fourfold increase in the network throughput can be achieved with practical numbers of modulation levels in comparison to the traditionally considered binary case. Finally, we present BER simulation results for channel coded binary and M-ary OCDMA transmission systems. In particular, we apply turbo codes concatenated with Reed-Solomon codes so that up to several hundred concurrent optical CDMA users can be supported at low target bit error rates. We observe that unlike conventional OCDMA systems, turbo-empowered OCDMA can allow overloading (more active users than is the length of the spreading sequences) with good bit error rate system performance.

Keywords and phrases: optical CDMA communication, multiple-access channels, M-ary modulation, channel capacity, turbo codes.

1. INTRODUCTION

Optical CDMA can provide multiple attractive features for data access networks with bursty traffic, that is, flexibility of asynchronous and decentralized network operation, potential of high data throughputs, inherent data security, and potential of total bandwidth utilization by all network users (when utilizing two-dimensional spreading sequences). Ultrahigh transmission rates envisioned for all-optical networks as well as the limited availability of optical domain digital logic and signal processing have made OCDMA a challenging research area. Multiple advances and novel techniques have been developed in the past two decades to enable OCDMA technology for access networks [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18]. Significant research efforts have been made in the design and performance analysis of optical CDMA spreading sequences. The main objective has been to construct new families of optical spreading codes with low cross-correlation values in order to maximize the signal-to-interference ratio and obtain good bit error rate (BER) performances. However, despite the introduction to many families of new OCDMA spreading sequences, severe multiuser interference has usually limited the maximum number of active OCDMA users to between 10 and 20 (at

target BER of 10^{-9}) and this fact has placed a limit on achieving high aggregate data throughputs in OCDMA networks.

Recently, channel (error-correcting) codes have been considered as a means to reduce the effects of the multiuser interference and improve performance in OCDMA networks [19, 20, 21, 22]. It is thus of interest to know the theoretical (Shannon) capacity of an OCDMA network so that the overall system performance can be properly optimized. Channel capacity of optical communication systems has been explored for single-user transmission, for example, in [23, 24], taking into account physical limitations of the optical transmission link. Nonetheless, different techniques and approaches have been needed and utilized so far to analyze OCDMA capacity limitations in the presence of multiple users. Authors in references [25, 26] explored fundamental limitations under specific OCDMA architectures, for example, on-off keying, specific optical receivers, and code correlation constraints in [25]. Hence, the motivation of this paper is first to determine capacity limits on binary OCDMA transmission assuming that the best possible spreading sequences, single-user receiver, and channel codes are utilized. Consequently, we study the effects of using more than two modulation levels on increasing the capacity of OCDMA transmission with single-user detection at the receiver. Finally, we use powerful channel codes, such as turbo codes concatenated with Reed-Solomon codes, to achieve up to several hundred users in OCDMA systems with a target BER of 10^{-9}.

This paper is organized as follows. Section 2 presents channel models for binary and M-ary OCDMA transmission when single-user detection is applied at the receiver and considers the noiseless, interference-limited transmission. Section 3 calculates the channel capacity limits on OCDMA network throughput under these assumptions and analyzes the benefit of using M-ary OCDMA modulation to improve upon the capacity of the binary systems. Furthermore, for high modulation levels, we propose a specific input distribution that satisfactorily increases the throughput of the M-ary OCDMA compared to the equiprobable signaling. In Section 4, we evaluate the impact of Gaussian (photodetector) noise on the throughput limits derived previously. Section 5 proposes and explores coded OCDMA architectures utilizing turbo codes and these are shown to significantly improve the overall system performance in OCDMA networks. Finally, concluding remarks and directions of further work are given in Section 6.

2. CHANNEL MODELS

2.1. General assumptions

We consider the general case of M-ary OCDMA transmission ($M \geq 2$) with bit-asynchronous, chip-synchronous transmission when K users are sending information simultaneously through the network using a laser source that can be intensity modulated to one of M modulation levels $(0, 1, \ldots, M-1)$. Each user's data are spread and encoded prior to being sent over the shared channel and the users behave in an independent (i.e., no collaboration is allowed) and symmetric manner (i.e., they all have the same probability of using a given modulated symbol in their spreading sequence). We will also assume that the receiver performs single-user detection (SUD), where we can assume without any loss of generality that the first user is the decoded user. Our capacity analysis is carried out on the chip level in a general sense, assuming that the best possible single-user detector, spreading and error-control coding are used. In Section 5, we will deal with specific architectures and explore their performances through simulations results.

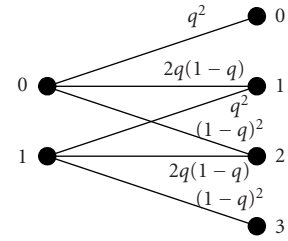


Figure 1: Chip-level discrete memoryless channel model of a 3-user asynchronous OCDMA system with single-user detection. Please note that the channel input symbols (chips) are used with probabilities $p_0 = q$ and $p_1 = 1 - q$.

We will start our analysis with the interference-limited (noiseless) transmission case, that is, the effects of receiver and channel noise are assumed negligible when compared to the interference of other users who are also transmitting their data over the shared multi-access channel. Under these assumptions, the OCDMA transmission can be modeled using an appropriate discrete memoryless channel.

2.2. Examples of channel models for OCDMA transmission with SUD

Consider the case of 3 users transmitting binary OCDMA chip symbols 1 and 0 over the optical channel with probabilities $p_0 = q$ and $p_1 = 1 - q$, respectively. We can model such transmission using a discrete memoryless channel (DMC) shown in Figure 1. At the output of the channel, the input signal of user 1 is corrupted by interference due to other simultaneous users of the OCDMA network. For this 3-user binary network, if user 1 transmits chip symbol 0, the receiver will receive 0 if and only if the other 2 users also transmit zero. Hence the transition probability is q^2. Similarly, the receiver takes the value 1 if one interfering user transmits the chip symbol 1 and the other transmits 0, which occurs with probability $2q(1-q)$.

As a second example, we consider the case of 2 OCDMA users with ternary modulation, with corresponding DMC representation shown in Figure 2. At the chip level, each user can send symbols 0, 1, or 2 with probabilities p_0, p_1, and $p_2 = (1 - p_0 - p_1)$, and the channel transition matrix is hence given by

$$P_{Y|X_1} = \begin{pmatrix} p_0 & p_1 & p_2 & 0 & 0 \\ 0 & p_0 & p_1 & p_2 & 0 \\ 0 & 0 & p_0 & p_1 & p_2 \end{pmatrix}. \quad (1)$$

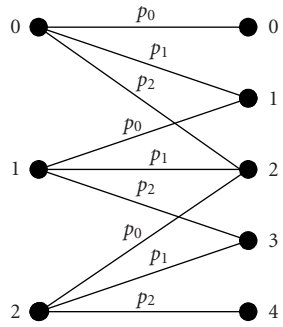


Figure 2: Chip-level DMC model for a 2-user ternary ($M = 3$) OCDMA transmission with single-user detection. Channel input symbols are used with probabilities p_0, p_1, and $p_2 = (1 - p_0 - p_1)$.

The channel output Y is the real sum of the symbol intensities sent by all simultaneous users and can hence take values 0, 1, 2, 3, or 4 with probabilities p_0^2, $2p_0p_1$, $p_1^2 + 2p_0p_1$, $2p_1p_2$, and p_2^2 respectively. If we evaluate the information-theoretic capacity of this specific channel, we find out that such a system is limited to throughput of 0.74 bits per OCDMA chip for the decoded user and thus the cumulative throughput for the 2 independent users is limited to 1.48 bits per OCDMA chip.

It is also interesting to note from these examples that the channel matrix depends on the input probability distribution. Consequently, the DMC representing the OCDMA system is not a constant (fixed) channel like the majority of channels considered in information theory [27]. As we will see in Section 4, this variability of the channel will lead to a difficult optimization problem of enumerating the numerical value of the capacity.

2.3. M-ary OCDMA transmission model

More generally, at the chip level, the DMC representation of a K-user M-ary modulated OCDMA transmission has as input the random variable X_1 corresponding to the information sent by the desired user and taking value on the alphabet $\mathbf{X} = \{0, 1, \ldots, M-1\}$; and as an output the real sum

$$Y = \sum_{i=1}^{K} X_i, \tag{2}$$

where X_i for $i = 1, 2, \ldots, K$ is a sequence of independent identically distributed random variables corresponding to chip symbols transmitted by users $1, 2, \ldots, K$ after OCDMA spreading and error-control coding. Therefore, the output alphabet is given by $\mathbf{y} = \{0, 1, 2, \ldots, K \times (M-1)\}$.

The entries of the channel matrix $P_{Y|X_1}$ are the conditional probabilities of the output symbol, given the input symbol and can be described for all $y \in \mathbf{y}$ and $x_1 \in \mathbf{X}$ as follows:

$$P_{Y|X_1}(Y = y|X_1 = x_1) = \sum_{\substack{(x_2, x_3, \ldots, x_K) \in \mathbf{X}^{K-1} \\ \text{s.t.} \\ x_2 + x_3 + \cdots + x_{K-1} = (y - x_1)}} p_{x_2} p_{x_3} \cdots p_{x_K}. \tag{3}$$

Consequently, it can be again noticed that the channel depends on the input probability mass function for a general K-user, M-ary OCDMA transmission with SUD. Furthermore, the output probability distribution is given by the following expression for all $y \in \mathbf{y}$:

$$P_Y(Y = y) = \sum_{\substack{(x_1, x_2, x_3, \ldots, x_K) \in \chi^K \\ \text{s.t.} \\ x_1 + x_2 + x_3 + \cdots + x_K = y}} p_{x_1} p_{x_2} \cdots p_{x_K}. \tag{4}$$

The values of the entries of the channel transition matrix $P_{Y|X_1}(Y = y|X_1 = x_1)$ as well as the output probabilities $P_Y(Y = y)$ can be found using generating functions. For K users sending data over an optical M-ary modulated channel, the conditional probabilities assuming that the input symbol is $x_1 \in \mathbf{X}$ are generated by the following polynomial in the symbolic variable z:

$$R(z) = z^{x_1} \left(p_0 z^0 + p_1 z^1 + \cdots + p_{M-1} z^{M-1}\right)^{K-1}. \tag{5}$$

To obtain the resulting probabilities, we expand this polynomial in the form

$$R(z) = \alpha_{x_1} z^{x_1} + \alpha_{x_1+1} z^{x_1+1} + \cdots + \alpha_{x_1+(K-1)(M-1)} z^{x_1+(K-1)(M-1)} \tag{6}$$

and the conditional probabilities on the channel are given by $P_{Y|X_1}(Y = y|X_1 = x_1) = \alpha_{y+x_1}$ for all $y \in \mathbf{y} = \{0, 1, \ldots, K \times (M-1)\}$ and $x_1 \in \mathbf{X}$. Similarly, the generating polynomial in the symbolic variable z for the output distribution is given by

$$Q(z) = \left(p_0 z^0 + p_1 z^1 + \cdots + p_{M-1} z^{M-1}\right)^K. \tag{7}$$

To obtain the output probabilities, we expand the previous polynomial in the form

$$Q(z) = \alpha_0 z^0 + \alpha_1 z^1 + \cdots + \alpha_{K(M-1)} z^{K(M-1)} \tag{8}$$

and the output probabilities are given by $P_Y(Y = y) = \alpha_y$ for $y \in \mathbf{y} = \{0, 1, \ldots, K \times (M-1)\}$.

If we consider, for instance, the case of binary modulation ($M = 2$) with input alphabet $\mathbf{X} = \{0, 1\}$ and the output alphabet $\mathbf{y} = \{0, 1, 2, \ldots, K\}$, the channel conditional probabilities are given by

$$P_{Y|X_1}(Y = y|X_1 = x_1) = \binom{K-1}{y - x_1} (1-q)^{y-x_1} q^{K-1-y+x_1}, \tag{9}$$

where $p_0 = q$, $p_1 = 1 - q$ and, for notational convenience, we define $\binom{K-1}{-1} = \binom{K-1}{K} = 0$. Due to the symmetry of the problem, the channel matrix of conditional probabilities can be written as

$$P_{Y|X_1} = \begin{pmatrix} \beta_0 & \beta_1 & \cdots & \beta_{K-1} & 0 \\ 0 & \beta_0 & \beta_0 & \cdots & \beta_{K-1} \end{pmatrix}, \tag{10}$$

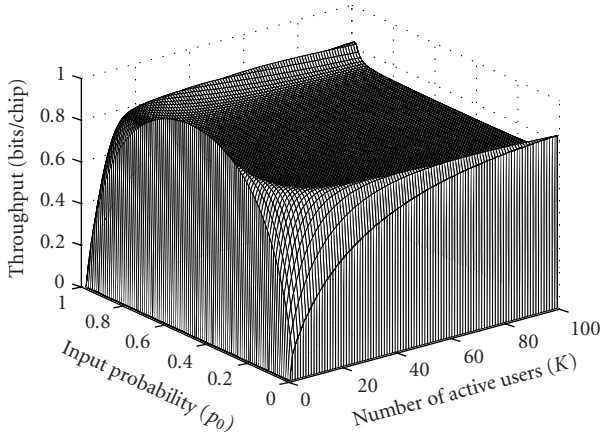


Figure 3: Fundamental limits on asynchronous binary OCDMA transmission.

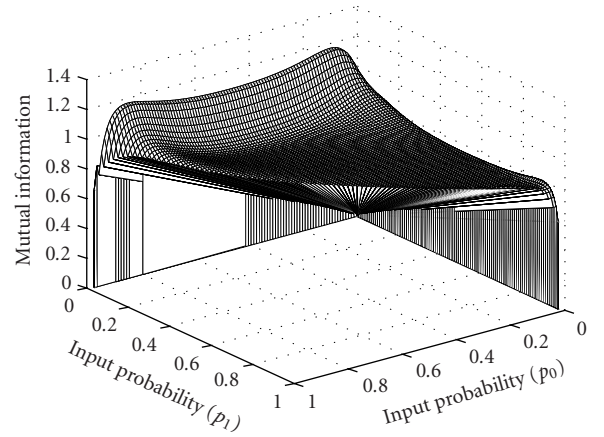


Figure 4: Aggregate mutual information as a function of input probability for 10 users, ternary OCDMA transmission.

where $\beta_i = \binom{K-1}{i}(1-q)^i q^{K-1-i}$. The channel output probability is consequently given by

$$P_Y(Y=y) = \binom{K}{y}(1-q)^y q^{K-y}. \tag{11}$$

3. CAPACITY EVALUATION

3.1. Information capacity of the binary OCDMA transmission with SUD

The information-theoretical (Shannon) capacity of the binary optical CDMA transmission with single-user detection at the receiver can be calculated using the channel model described in (9). Due to the symmetrical and independent operation of the K users on the OCDMA channel, the cumulative throughput by all the K simultaneous users on the channel is limited in (information) bits per transmitted OCDMA chip by

$$C_{2,K} = K \sup_{q\in[0,1]} I_{2,K}(X_1, Y), \tag{12}$$

where $I_{2,K}(X_1, Y)$ is given by (13) below and represents overall mutual information of a K-user binary OCDMA transmission with single-user detection.

The aggregate mutual information of such a bit-asynchronous binary OCDMA channel is plotted as a function of the input probability $q = P(X_1 = 0)$ in Figure 3 for various numbers of users. When only 1 user is active in the system, without the presence of noise, the channel reduces to a perfect noiseless binary channel and the capacity is 1 bit per chip. As the number of users increases, the interference reduces the theoretical capacity limit until it reaches a steady-state value. For example, for a 50-user transmission, the aggregate capacity is 0.8374 bits per chip achieved when the input probability is $q = 0.03$. Note that the uniform distribution does not achieve the capacity and for the same system, it yields to a throughput limit of around 0.7288 bits per OCDMA chip.

$$\begin{aligned} I_{2,K}(X_1;Y) &= -\sum_{i=0}^{K}\binom{K}{i}(1-q)^i q^{K-i}\log_2\left(\binom{K}{i}(1-q)^i q^{K-i}\right) \\ &\quad + \sum_{i=0}^{K-1}\binom{K-1}{i}(1-q)^i q^{K-1-i} \\ &\quad \times \log_2\left(\binom{K-1}{i}(1-q)^i q^{K-1-i}\right). \end{aligned} \tag{13}$$

3.2. Information capacity with nonbinary modulation

Due to the symmetrical and independent operation of the K users on the channel, the cumulative throughput for all the K users is limited by

$$C_{M,K} = K \sup_{P_{X_1}} I_{M,K}(X_1; Y). \tag{14}$$

Moreover, due to the symmetry of the channel matrix $P_{Y|X_1}$ in (3) (rows are cyclical shift of each other), it follows that $H(Y = y|X_1 = x_1) = H(Y = y|X_1 = 0)$ for all $x_1 = 1, 2, \ldots, M-1$. Consequently, the mutual information $I_{M,K}(X;Y)$ for a K-user M-ary OCDMA transmission with single-user detection can be expressed as in (15).

Generally, the variability of the discrete memoryless channel modeling the M-ary OCDMA transmission on the input probability mass function causes a non-concavity of the mutual information $I_{M,K}(X_1;Y)$. For example, Figure 4 shows the overall mutual information as a function of the input probability distribution and its non-concave nature in the case of ternary ($M = 3$) modulation and transmission by $K = 10$ active users. This distinguishes the M-ary multiaccess channel with single-user detection from the standard class of fixed channels where the mutual information is concave in the channel input distribution [27], and hence finding this capacity requires global optimization tools.

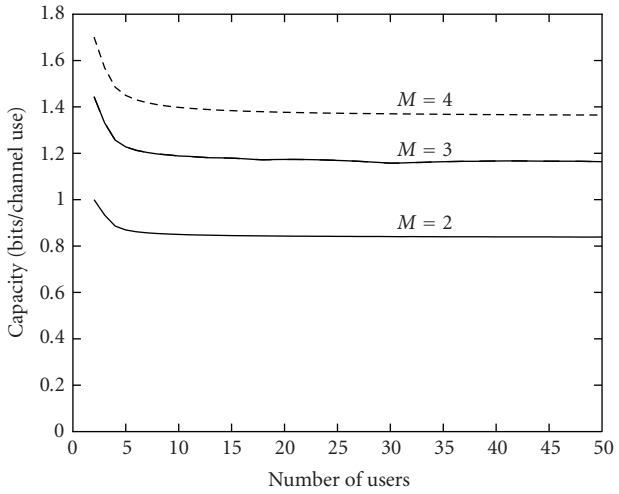


Figure 5: Aggregate capacity of OCDMA transmission as a function of the number of users K shown for $M = 2, 3$, and 4 modulation levels.

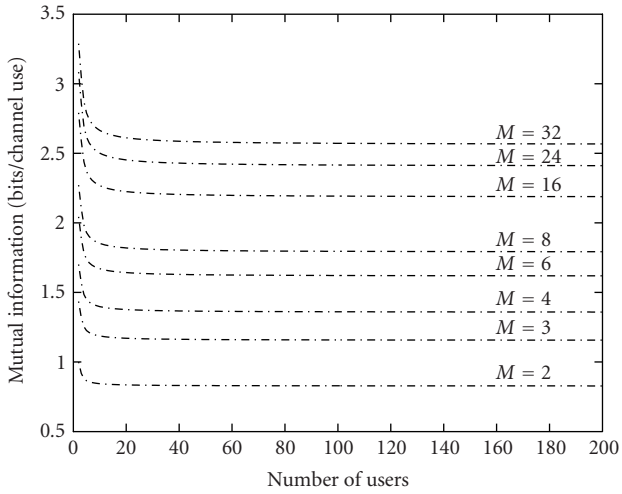


Figure 6: Mutual information using the proposed input probability distribution.

Indeed, the commonly used numerical algorithms for evaluating the channel capacity (e.g., the Arimoto-Blahut algorithm [28]) do not apply in this case, since they generally assume and utilize concavity of the mutual information in the channel input distribution.

$$
\begin{aligned}
&I_{M,K}(X_1;Y)\\
&= \sum_{y=0}^{(K-1)(M-1)} \left(\sum_{\substack{(x_2,x_3,\ldots,x_K)\in\chi^{K-1}\\ \text{s.t.}\\ x_2+x_3+\cdots+x_K=y}} p_{x_2}p_{x_3}\cdots p_{x_K} \right)\\
&\quad\times \log_2 \left(\sum_{\substack{(x_2,x_3,\ldots,x_K)\in\chi^{K-1}\\ \text{s.t.}\\ x_2+x_3+\cdots+x_K=y}} p_{x_2}p_{x_3}\cdots p_{x_K} \right)\\
&\quad- \sum_{y=0}^{K(M-1)} \left(\sum_{\substack{(x_1,x_2,x_3,\ldots,x_K)\in\mathbf{X}^K\\ \text{s.t.}\\ x_1+x_2+x_3+\cdots+x_K=y}} p_{x_1}p_{x_2}\cdots p_{x_K} \right)\\
&\quad\times \log_2 \left(\sum_{\substack{(x_1,x_2,x_3,\ldots,x_K)\in\mathbf{X}^K\\ \text{s.t.}\\ x_1+x_2+x_3+\cdots+x_K=y}} p_{x_1}p_{x_2}\cdots p_{x_K} \right).
\end{aligned}
\tag{15}
$$

Due to the difficulty of having an exact analytical expression for the capacity of the OCDMA transmission with single-user detection, we numerically compute the capacity of the channel and these results are shown in Figure 5 for selected numbers of modulation levels. It can then be seen that the throughput of this system is maximum for a 2-user system and reaches a limiting value for increasing number of users. Consequently, when the number of users is sufficiently large ($K \geq 10$), the aggregate throughput of the OCDMA transmission is stable, quite independent of the total number of users. This is a highly desirable feature in a network with bursty traffic, since congestion does not occur as the number of users increases.

It is interesting to compare the capacity with the throughput achieved for uniformly distributed inputs. In this case, the aggregate mutual information for large enough number of users is much lower than the capacity of the channel shown in Figure 5 and is approximately

$$KI_{M,K}(X_1;Y) \approx K\log_2\left(\sqrt{\frac{K}{K-1}}\right) \longrightarrow 0.72 \text{ bits/chip}. \tag{16}$$

3.3. Near-capacity-achieving input distributions

Although the numerical evaluation gives a way of determining the capacity of an M-ary OCDMA network, this process can still be difficult and computational intensive, especially for many modulation levels. This is explained by the high dimensionality of the problem added to the difficulty of global optimization needed to maximize the mutual information. Hence, to allow less complex, approximate evaluation, we heuristically propose the following class of input probability distributions that satisfactorily increases the throughput of the system:

$$
\begin{aligned}
p_0 = p_{M-1} &= \frac{K-1}{2K},\\
p_1 = p_2 = \cdots = p_{M-2} &= \frac{1}{K(M-2)}.
\end{aligned}
\tag{17}
$$

The system throughput can then be viewed as the binary channel information due to the two most used symbols,

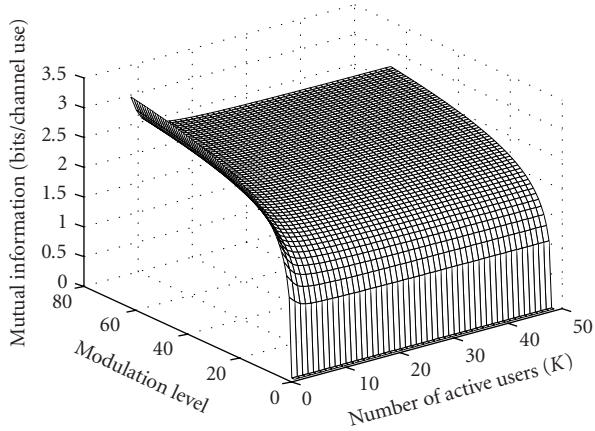


FIGURE 7: Mutual information as a function of the modulation levels and the number of users.

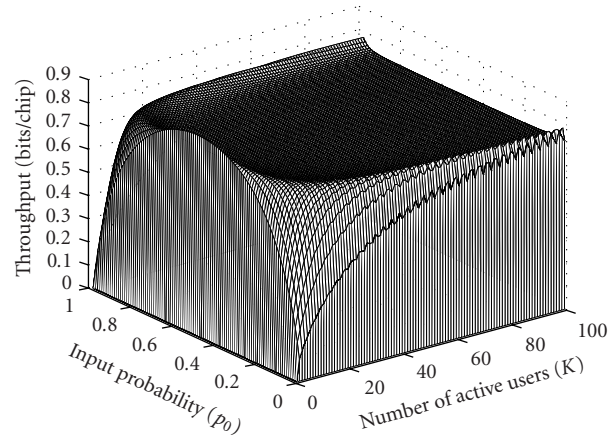
0 and $M-1$, plus additional information corresponding to the contribution of the $M-2$ unlikely symbols $1, 2, \ldots, M-2$.

The mutual information, obtained using this proposed input probability distribution, is plotted as a function of the number of active users in Figure 6 and as a joint function of the number of users and the M-ary modulation levels in Figure 7. For $M = 2, 3$, and 4 modulation levels, this mutual information is approximately equal to the actual capacity of the system (shown originally in Figure 5). As an example, the mutual information using this distribution for a 25-user ternary communication system is about 1.159, identical within the first 4 digits to the capacity of the system found by global optimization.

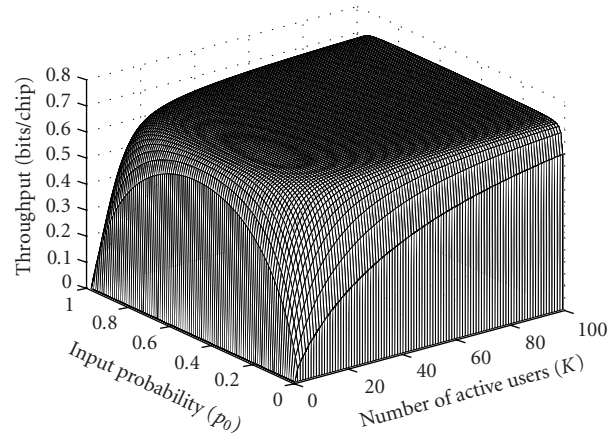
It is interesting to note that the throughput increases approximately logarithmically with the increasing number of modulation levels M as $KI_{M,K}(X,Y) \approx 0.36\log_2(M-1) + 0.83$ bits per chip for $M > 2$ and $K > 10$. It may be possible to increase the throughput of the OCDMA by increasing the M-ary modulation level as desired. However, as the number of the modulation levels M increases, the information gained becomes less important. Therefore, an OCDMA system may be implemented with a suitable choice of M, creating a compromise between complexity and throughput.

4. IMPACT OF NOISE ON THE THROUGHPUT OF OCDMA TRANSMISSION WITH SUD

So far, we have omitted the presence of channel noise. Although this gives an upper bound on the capacity of OCDMA transmission under perfect conditions, noise should be included in the analysis to assess the robustness of such results. We will assume that the channel is corrupted by an additive white Gaussian noise (AWGN) which models the electronic (photodetector) noise which may degrade the sent information [11]. The input of the channel is still the random variable X_1 taking values on the alphabet $\mathbf{X} = \{0, 1, \ldots, M-1\}$ with the probabilities derived in the previous section. However, the channel output is now the continuous random variable $Z = Y + N$, where Y is the output of the noiseless channel from (2) and N is a Gaussian random variable with zero mean and variance of σ^2.



(a)



(b)

FIGURE 8: Throughput limit on binary asynchronous OCDMA transmission when AWGN is present for noise variance: (a) $\sigma^2 = 0.09$ and (b) $\sigma^2 = 0.25$.

4.1. Impact of noise on the binary OCDMA transmission

The aggregate Shannon capacity of the binary OCDMA channel with SUD has the form of (12), but the mutual information term is modified to $I_{2,K}^{\mathrm{AWGN}}(X_1; Z)$ due to the additive white Gaussian noise (AWGN) corrupting the received data and is described by (18) below. This new mutual information is plotted as a function of the input distribution and the number of active users in Figure 8 for two different values of noise variance. In Figure 9, the overall throughput limit for a 20-user binary OCDMA transmission is shown as a function of the input probability for selected noise variances. Note that for high enough signal-to-noise ratios, the result approaches the interference-limited scenario, however for extremely low SNR (for noise variance of 0.25, the SNR is

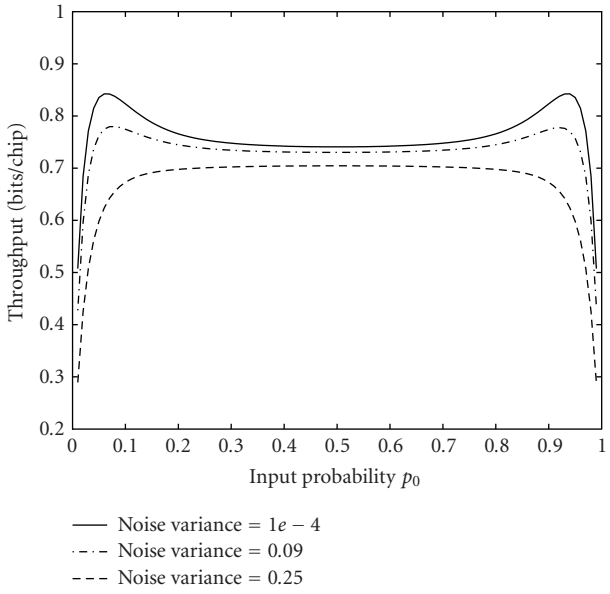


Figure 9: Limit on aggregate throughput of a 20-user binary OCDMA system at different noise variances with different input distribution.

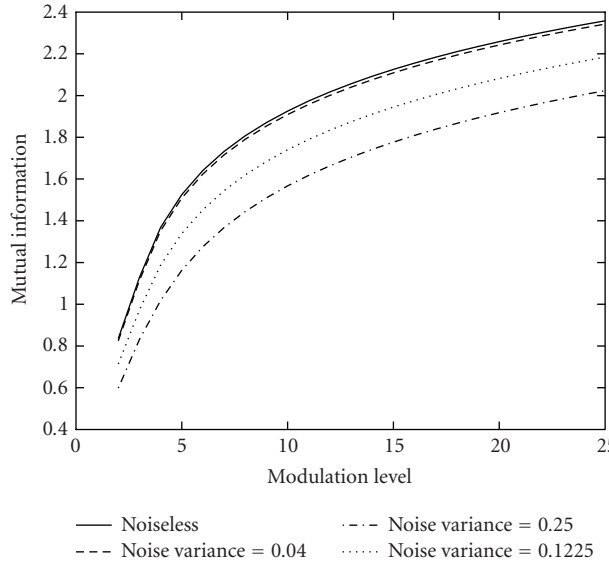


Figure 10: Mutual information as a function of the number of modulation levels M for a 20-user system.

approaching zero dB), the throughput starts decreasing:

$$
\begin{aligned}
&i_{2,K}^{\mathrm{AWGN}}(X_1;Z)\\
&= -\int_{-\infty}^{\infty}\left(\sum_{j=0}^{K}\binom{K}{j}(1-q)^j q^{K-j}\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{(z-j)^2}{2\sigma^2}\right)\right)\\
&\quad\times\log_2\left(\sum_{j=0}^{K}\binom{K}{j}(1-q)^j q^{K-j}\frac{1}{\sqrt{2\pi\sigma^2}}\right.\\
&\qquad\left.\times\exp\left(-\frac{(z-j)^2}{2\sigma^2}\right)\right)dz\\
&\quad+\int_{-\infty}^{\infty}\left(\sum_{i=0}^{K-1}\binom{K-1}{i}(1-q)^i q^{K-1-i}\frac{1}{\sqrt{2\pi\sigma^2}}\right.\\
&\qquad\left.\times\exp\left(-\frac{(z-i)^2}{2\sigma^2}\right)\right)\\
&\quad\times\log_2\left(\sum_{i=0}^{K-1}\binom{K-1}{i}(1-q)^i q^{K-1-i}\frac{1}{\sqrt{2\pi\sigma^2}}\right.\\
&\qquad\left.\times\exp\left(-\frac{(z-i)^2}{2\sigma^2}\right)\right)dz.
\end{aligned}
\tag{18}
$$

4.2. Impact of noise on the M-ary OCDMA transmission

To explore the noise resistance/sensitivity of the results computed so far, we compute the throughput limits on the OCDMA transmission in the presence of AWGN. We will use the noiseless capacity-achieving distribution in case of $M = 2, 3, 4$ modulation levels and the proposed near-capacity-achieving input distributions from Section 3.3 for

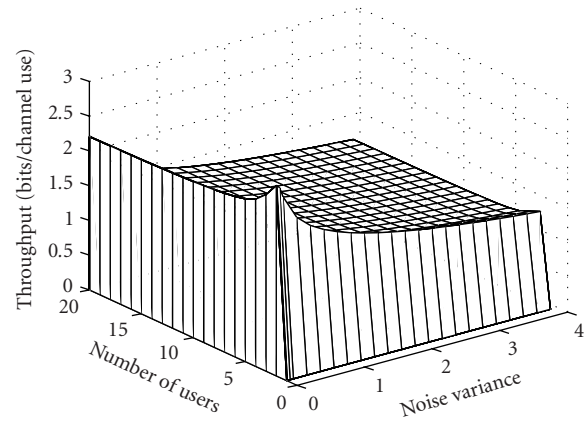


Figure 11: Mutual information as a function of the number of users and the noise variance for a 16-ary OCDMA transmission.

$M > 4$ modulation levels. The aggregate throughput of the OCDMA transmission under these assumptions is limited by $KI_{M,K}^{\mathrm{AWGN}}(X_1;Z)$, where the mutual information $I_{M,K}^{\mathrm{AWGN}}(X_1;Z)$ is given by (18).

Figure 10 shows the aggregate throughput limits on a 20-user OCDMA network for selected noise variances, plotted as a function of the modulation levels number. For small noise variances, the noisy throughput limits almost overlap with the noiseless system throughput limits.

The throughput limits are reduced for increasing noise power, but even in this case, the throughput increase with nonbinary modulation is satisfactory. In Figure 11, we show the aggregate throughput limit against the noise variance and the number of users for 16-ary OCDMA modulation. It can be seen that even for high noise variance, the transmission is still reasonably robust to noise, since the throughput limit does not experience a sharp drop due to the noise.

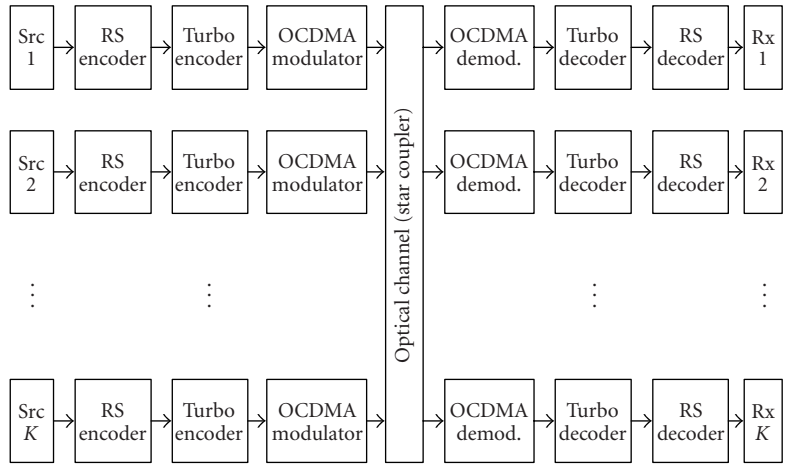


Figure 12: Schematic block diagram of the proposed OCDMA transmission scheme for binary and nonbinary modulation.

5. CODED OCDMA TRANSMISSION WITH TURBO AND REED-SOLOMON CODES

Multi-user interference usually severely limits performance of OCDMA systems without error-control coding. Systems based on optical spreading sequences only can usually support between 10 to 20 simultaneously transmitting users with bit error rates below 10^{-9} (standard performance benchmark for OCDMA transmission). Alternatively, one can compare OCDMA systems using normalized spectral efficiency, that is, aggregate network throughput measured in information bits per OCDMA chip. Traditional (uncoded) OCDMA systems achieve aggregate system throughputs of about 0.03 bits per OCDMA chip, much below the predicted capacity limits. Consequently, we propose several coded OCDMA transmission schemes to increase the OCDMA throughput.

5.1. Coded binary OCDMA systems

In the proposed transmission scheme in Figure 12 data of each user are first encoded by a high-rate Reed-Solomon code and interleaved (permuted), then encoded by a turbo code and finally modulated by a user-specific binary optical spreading sequence, such as two-dimensional wavelength-time balanced codes for differential detection (BCDD) [17]. The role of the optical spreading sequence is to match the data to the channel and to provide low cross-correlation between data sent by different users. The role of the turbo code is to achieve an intermediate bit error rate of 10^{-3} to 10^{-5}, while the Reed-Solomon code provides final error rate at the receiver below the desired error rate of 10^{-9}. At the receiver end, optical matched filtering and consequent sampling are performed using the optical spreading sequence of the desired user. Turbo decoding followed by Reed-Solomon decoding are used to recover data from the effects of multi-user interference and channel noise.

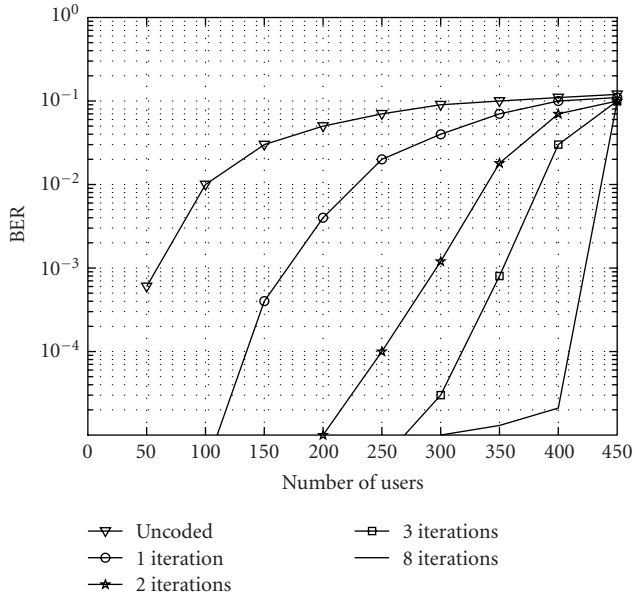
We first present BER performance of the proposed binary OCDMA system when each user employs a $(255, 239)$ Reed-Solomon code over $GF(256)$ [29], a rate 1/2 (or 1/3) turbo code and a BCDD$(32, 16, 0.156)$ wavelength-time optical spreading sequence (please see [17] for notation). The turbo code is based on the original rate 1/3 turbo coding scheme [30] which uses a 16-state recursive systematic code with generator polynomial $(37, 21)$. The rate of 1/2 is obtained through puncturing; a packet of length of 10 000 bits is used with an s-rand interleaver of spread 40. Consequent application of the Reed-Solomon decoder decreases the overall BER below 10^{-9}, as shown in Figure 13b. In terms of channel efficiency, this corresponds to an aggregate system throughput of 0.37 information bits per OCDMA chip.

Figure 14 shows the BER performance when the rate 1/3 turbo code is utilized. Simulation results show that up to 700 users can be supported with a BER below 10^{-9} after 8 turbo decoding iterations and RS decoding are performed. This corresponds to an aggregate throughput of 0.42 bits per OCDMA chip. In terms of transmission rate over the optical channel, if each user transmits using 32 wavelengths at OC-12 chip rate and 16-time chips per bit, the proposed scheme would achieve an effective aggregate throughput of 8.4 Gbps$= (700\times(1/3)\times(239/255)\times(1/16)\times 620$ Mbps).

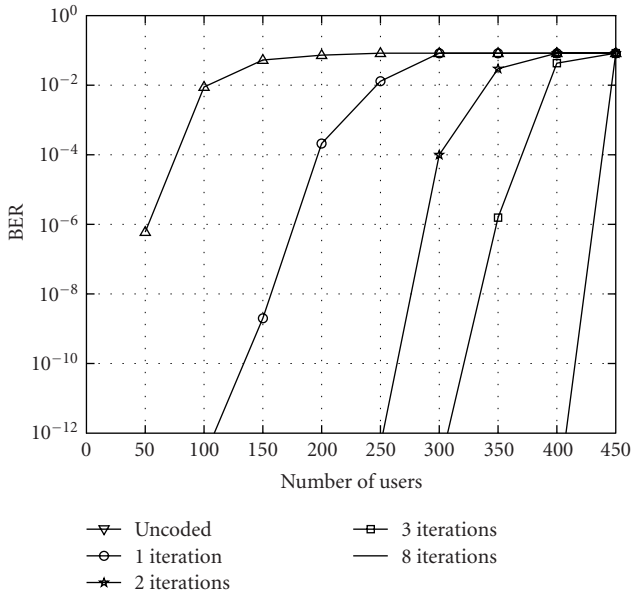
Finally, Figure 15 shows the BER performance for the system when noise is present with $\sigma^2 = 0.09$. With rate 1/3 turbo code, spreading length of 512, more than 650 users can be supported, which corresponds to a throughput of 0.40 bits per OCDMA chip.

5.2. Coded M-ary OCDMA transmission

We considered the coded M-ary OCDMA transmission scheme shown in Figure 12. Each user's data are first encoded by a $(255, 239)$ Reed-Solomon code over $GF(256)$ followed by a rate 1/3 turbo code, then modulated using a user-specific M-ary optical spreading sequence corresponding to bits 0 and 1. The M-ary OCDMA spreading sequences are one-dimensional codes composed of symbols $m = 0, 1, \ldots, M-1$ pseudorandomly generated with probabilities chosen according to the capacity-achieving input probability distribution of (17). For example, for ternary OCDMA transmission with 16 aggregate users, the spreading codes are generated

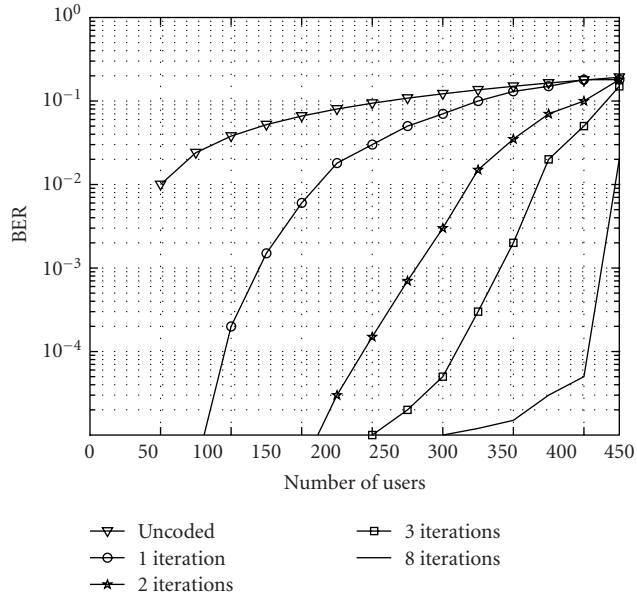


(a)

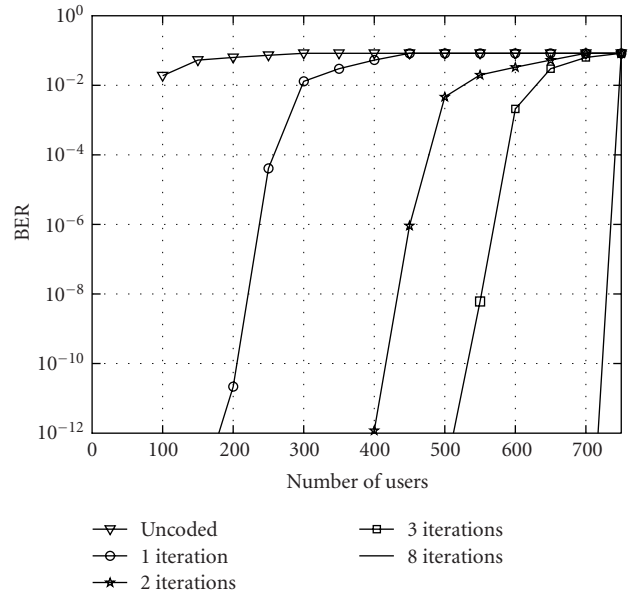


(b)

FIGURE 13: BER performance of the proposed coded OCDMA system as a function of the number of active users: (a) after rate 1/2 turbo decoding; (b) after turbo and RS decoding.



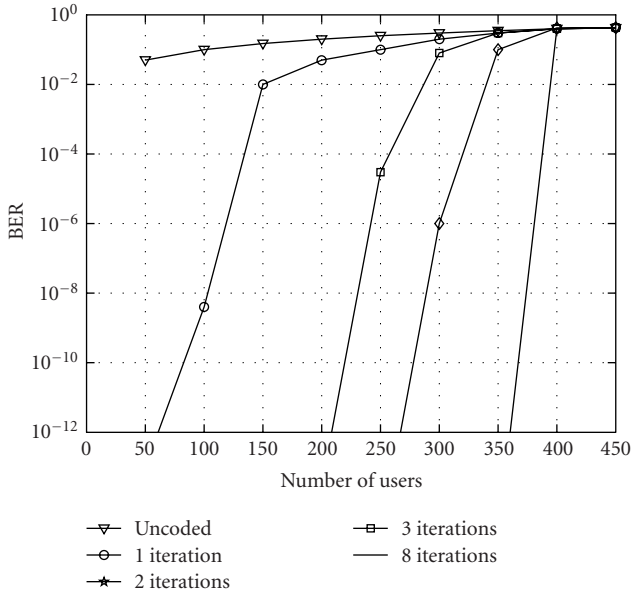
(a)



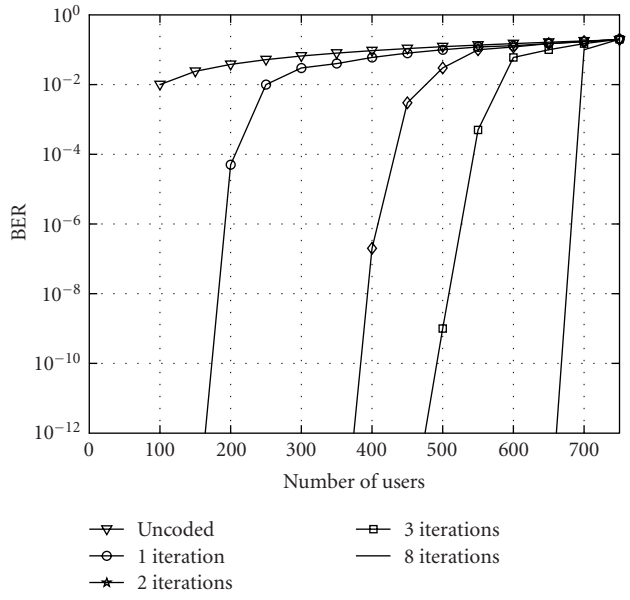
(b)

FIGURE 14: BER performance of the proposed coded OCDMA system as a function of the number of active users: (a) after rate 1/3 turbo decoding; (b) after turbo and RS decoding.

using symbols 0,1, and 2 with probability distribution corresponding to $p_1 \approx 1/16$ and $p_0 \approx p_2 \approx 15/32$. The OCDMA modulator serves as an interface between the user's data and the optical channel and the purpose of using such spreading sequences is to aim at achieving the capacity from the theoretical model of the previous sections. The optical channel is modeled by a K-user adder channel affected by AWGN channel noise. Given the channel observations on the chip level, the M-ary OCDMA demodulator uses an appropriate channel model described in Section 2 to estimate the a posteriori probabilities of chip-symbol from set $\{0, 1, 2, M\text{-}1\}$ being transmitted by the desired user. Consequently, these estimates are combined using the M-ary optical spreading sequence of this user via multiplying the appropriate

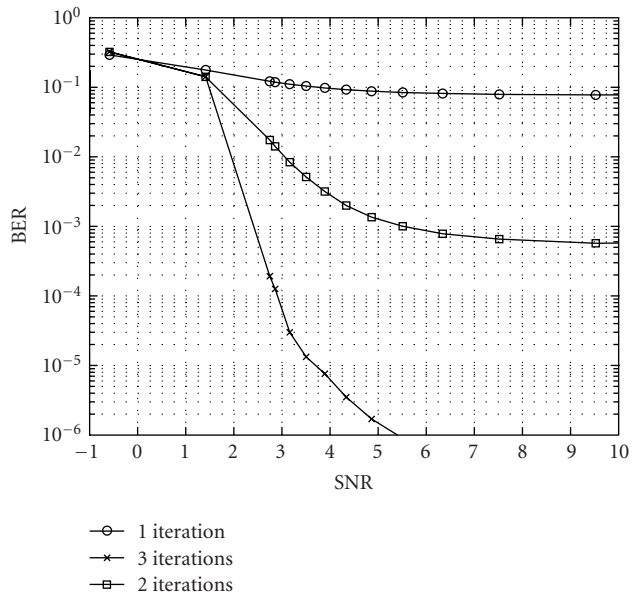


(a)

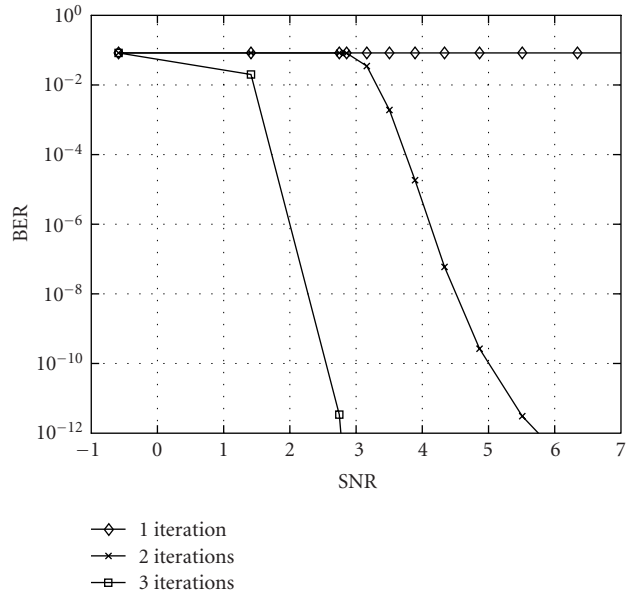


(b)

Figure 15: BER performance of the proposed coded OCDMA system in the presence of noise ($\sigma^2 = 0.09$) as a function of the number of active users: (a) after rate 1/2 turbo and RS decoding; (b) after rate 1/3 turbo and RS decoding.



(a)

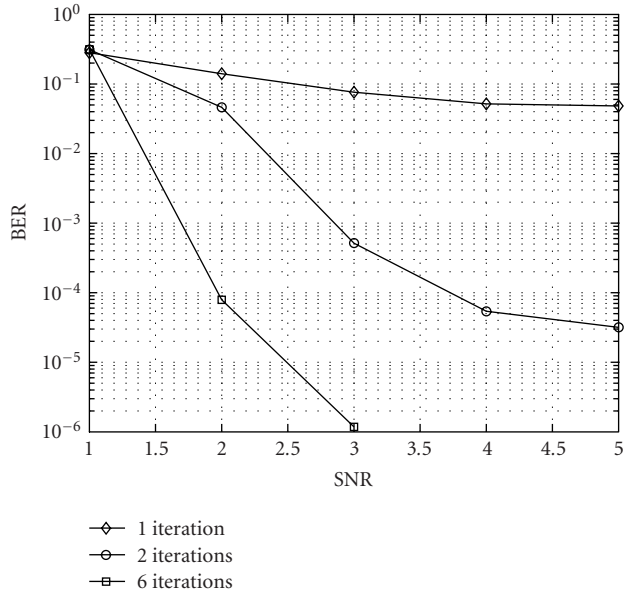


(b)

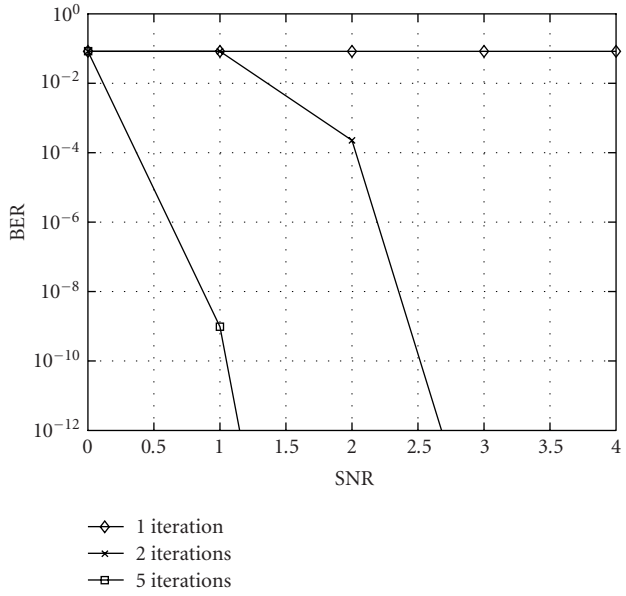
Figure 16: BER for a 67-user coded OCDMA transmission with a ternary spreading sequence of length 48: (a) after turbo decoding; (b) after turbo as well as RS decoding.

probabilities, thus estimating the probability of transmitted coded bits. The performance of the system is measured in terms of the bit error rate of a desired user.

We simulated the bit error rates of selected turbo- and RS-coded M-ary OCDMA systems for the transmission model described in Figure 12 and for optical signal-to-noise ratio (SNR) normalized to E_b/N_0. To increase the performance of the systems, we have constrained all the users' spreading sequences to have the same energy and to be separated from one another by at least a minimum Euclidean distance corresponding to 0.85 times their constant energy. In Figures 16 and 17, the BER is simulated using ternary and quaternary spreading sequences of length 48, respectively.

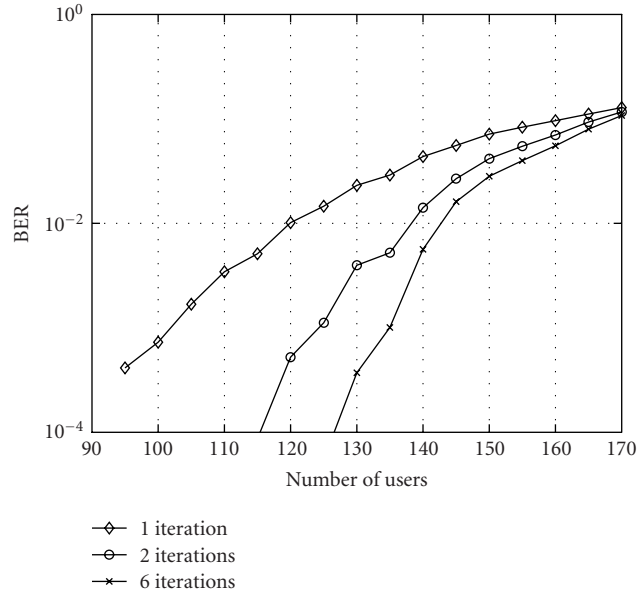


(a)

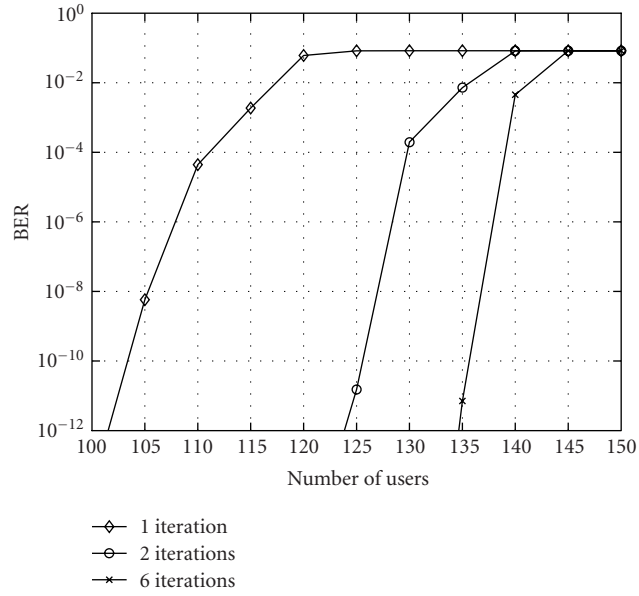


(b)

Figure 17: BER for a 57-user OCDMA transmission with a quaternary spreading sequence of length 48: (a) after turbo decoding; (b) after turbo as well as RS decoding.



(a)

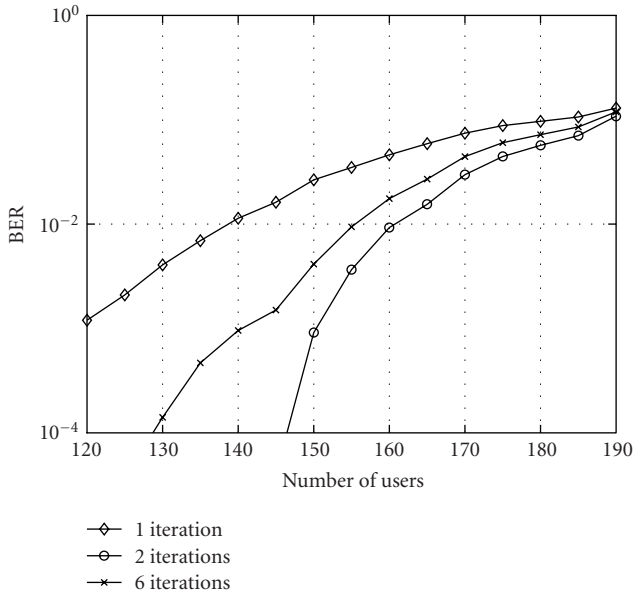


(b)

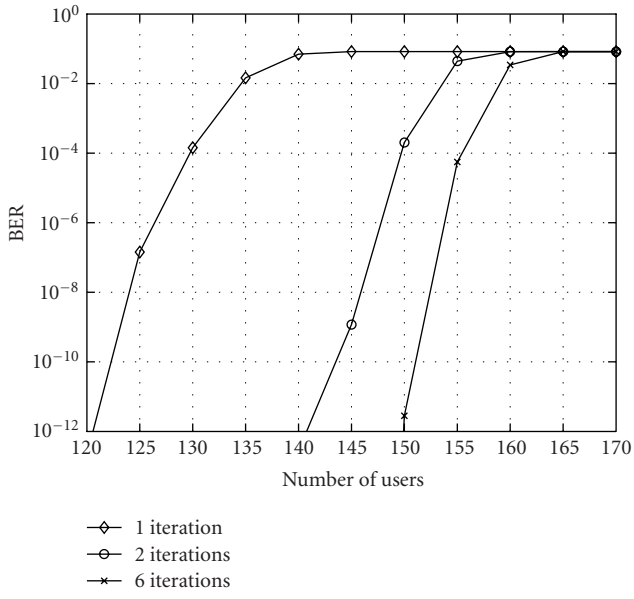
Figure 18: BER against the number of users for ternary OCDMA network with spreading sequences of length 128 and 10 dB SNR: (a) after turbo decoding; (b) after turbo as well as RS decoding.

It can be noticed that the transmission is overloaded in both cases with 67 and 57 active users and that the system still performs well at low SNRs. It is also interesting to note that these transmissions are particularly interference limited, since, at a certain point, an increasing SNR does not improve the performance. This occurs as the multi-user interference becomes much higher than the noise variance and causes the performance degradation.

In Figures 18 and 19, the simulated BER is plotted against the number of active OCDMA users for ternary and quaternary spreading sequences of length 128 and SNR = 10 dB. In the case of quaternary modulation, for example, up to 150 total users (after 6 turbo iterations) can be supported with almost error-free transmission. It can be noted that the system is then overloaded since we have used spreading sequences with length 128 but still performs well for 150 users.



(a)



(b)

Figure 19: BER against the number of users for quaternary OCDMA network with spreading sequences of length 128 and 10 dB SNR: (a) after turbo decoding; (b) after turbo as well as RS decoding.

In addition, if the spreading sequences are constrained to have better cross-energy (higher minimum distance constraint), more users can be accommodated in the system with probability of error tending to zero.

It is important to note that in this part that the spreading sequences have been randomly generated and are hence not necessarily optimal. Therefore, the aggregate number of users can then be increased by using more elaborate spreading sequences. In our future work we will focus on the design of M-ary spreading sequences with better correlation properties for data recovery.

6. CONCLUSION

In this work, we have evaluated the channel capacity limits on optical CDMA transmission, when single-user detection is used at the receiver. We have shown that the theoretical network throughput of binary optical CDMA is limited by about 0.84 bits per OCDMA chip and that it can be improved by up to 4 times using M-level OCDMA spreading sequences. Hence, the M-ary modulation for OCDMA systems can be seen as one practical way to increase the capacity limitations of the binary optical CDMA.

In addition to the capacity calculations, we have also proposed and explored specific architectures for bit-asynchronous OCDMA transmission. Using a concatenation of Berrou's turbo codes and Reed-Solomon codes, we have proposed a coded binary OCDMA system that can support several hundred active users on a multi-access network. Furthermore, we have also explored a turbo coded M-ary OCDMA transmission with single-user detection. Using proposed pseudorandomly generated spreading sequences based on near-capacity-achieving distributions, the M-ary OCDMA transmission schemes achieve good BER performance and support overloading at sufficiently low SNRs. The attractiveness of the proposed scheme lies in its reasonable complexity (e.g., due to the use of single-user detection), the utmost flexibility of the OCDMA network access scheme and ability to overload the system.

Our future work will focus on improving the proposed coding scheme to close the gap to the Shannon limit. We will also further explore a coding scheme for higher M-ary modulation levels and evaluate the optical CDMA system performance limits under different modes of channel noise.

ACKNOWLEDGMENTS

The authors would like to thank the anonymous reviewers for their constructive comments that helped to improve this manuscript. This research was supported by the Natural Sciences and Engineering Research Council (Canada), the Québec Fonds pour la Formation de Chercheurs et l'Aide à la Recherche, and the Canadian Foundation for Innovation. This work was presented in part at the Biennial Symposium on Communications, Kingston, Ontario, June 2002, the International Symposium on Turbo Codes, Brest, France, September 2003, and the IASTED International Multi-Conference on Wireless and Optical Communications, Banff, Alberta, July 2004.

REFERENCES

[1] A. A. Shaar and P. A. Davies, "Prime sequences: quasi-optimal sequences for optical channel code division multiplexing," *Electronics Letters*, vol. 19, no. 21, pp. 888–889, 1983.

[2] P. Prucnal, M. Santoro, and T. Fan, "Spread spectrum fiber-optic local area network using optical processing," *J. Lightwave Technol.*, vol. 4, no. 5, pp. 547–554, 1986.

[3] J. A. Salehi, "Code division multiple-access techniques in optical fiber networks. I. Fundamental principles," *IEEE Trans. Commun.*, vol. 37, no. 8, pp. 824–833, 1989.

[4] R. M. Gagliardi and A. J. Mendez, "Pulse combining and time-space coding for multiple-accessing with fiber arrays," in *Proc. IEEE/LEOS Summer Topical Meeting on Optical Multiple Access Networks*, pp. 45–46, Monterey, Calif, USA, July 1990.

[5] L. Tančevski and I. Andonovic, "Wavelength hopping/time spreading code division multiple access systems," *Electronics Letters*, vol. 30, no. 17, pp. 1388–1390, 1994.

[6] A. J. Mendez and R. M. Gagliardi, "Code division multiple access (CDMA) enhancement of wavelength division multiplexing (WDM) systems," in *Proc. IEEE International Conference on Communications (ICC '95)*, vol. 1, pp. 271–276, Seattle, Wash, USA, June 1995.

[7] M. Brandt-Pearce and B. Aazhang, "Performance analysis of single-user and multiuser detectors for optical code division multiple access communication systems," *IEEE Trans. Commun.*, vol. 43, no. 234, pp. 435–444, 1995.

[8] K. Yu and N. Park, "Design of new family of two-dimensional wavelength-time spreading codes for optical code division multiple access networks," *Electronics Letters*, vol. 35, no. 10, pp. 830–831, 1999.

[9] H. M. H. Shalaby, "Chip-level detection in optical code division multiple access," *J. Lightwave Technol.*, vol. 16, no. 6, pp. 1077–1087, 1998.

[10] H. Fathallah, L. A. Rusch, and S. LaRochelle, "Passive optical fast frequency-hop CDMA communications system," *J. Lightwave Technol.*, vol. 17, no. 3, pp. 397–405, 1999.

[11] R. F. Ormondroyd and M. M. Mustapha, "Optically orthogonal CDMA system performance with optical amplifier and photodetector noise," *IEEE Photon. Technol. Lett.*, vol. 11, no. 5, pp. 617–619, 1999.

[12] A. Stok and E. H. Sargent, "Lighting the local area: optical code-division multiple access and quality of service provisioning," *IEEE Network*, vol. 14, no. 6, pp. 42–46, 2000.

[13] S. Zahedi and J. A. Salehi, "Analytical comparison of various fiber-optic CDMA receiver structures," *J. Lightwave Technol.*, vol. 18, no. 12, pp. 1718–1727, 2000.

[14] E. Inaty, H. M. H. Shalaby, P. Fortier, and L. A. Rusch, "Multirate optical fast frequency hopping CDMA System using power control," *J. Lightwave Technol.*, vol. 20, no. 2, pp. 166–176, 2002.

[15] R. M. H. Yim, L. R. Chen, and J. Bajcsy, "Design and performance of 2-D codes for wavelength-time optical CDMA," *IEEE Photon. Technol. Lett.*, vol. 14, no. 5, pp. 714–716, 2002.

[16] A. J. Mendez and R. M. Gagliardi, "Combined effectiveness of optical hard-limiting and guard-time in optical CDMA systems," in *Proc. IEEE Lasers and Electro-Optics Society (LEOS '02)*, vol. 2, pp. 857–858, Glasgow, UK, November 2002.

[17] R. M. H. Yim, J. Bajcsy, and L. R. Chen, "A new family of 2-D wavelength-time codes for optical CDMA with differential detection," *IEEE Photon. Technol. Lett.*, vol. 15, no. 1, pp. 165–167, 2003.

[18] A. J. Mendez, R. M. Gagliardi, V. J. Hernandez, C. V. Bennett, and W. J. Lennon, "Design and performance analysis of wavelength/time (W/T) matrix codes for optical CDMA," *J. Lightwave Technol.*, vol. 21, no. 11, pp. 2524–2533, 2003.

[19] M. R. Dale and R. M. Gagliardi, "Channel coding for asynchronous fiberoptic CDMA communications," *IEEE Trans. Commun.*, vol. 43, no. 9, pp. 2485–2492, 1995.

[20] J. Y. Kim and H. V. Poor, "Turbo-coded packet transmission for an optical CDMA network," *J. Lightwave Technol.*, vol. 18, no. 12, pp. 1905–1915, 2000.

[21] P. Azmi, M. Nasiri-Kenari, and J. A. Salehi, "Soft-input decoder for decoding of internally channel coded fiber-optic CDMA communication systems," *IEEE Trans. Commun.*, vol. 50, no. 12, pp. 1994–2002, 2002.

[22] R. M. H. Yim, J. Bajcsy, and L. R. Chen, "Optical CDMA transmission enabled by turbo codes," in *Proc. International Symposium on Turbo Codes*, pp. 571–574, Brest, France, September 2003.

[23] A. Mecozzi and M. Shtaif, "On the capacity of intensity modulated systems using optical amplifiers," *IEEE Photon. Technol. Lett.*, vol. 13, no. 9, pp. 1029–1031, 2001.

[24] E. E. Narimanov and P. Mitra, "The channel capacity of a fiber optics communications system: perturbation theory," *J. Lightwave Technol.*, vol. 20, no. 3, pp. 530–537, 2002.

[25] H. M. H. Shalaby, "Complexities, error probabilities, and capacities of optical OOK-CDMA communication systems," *IEEE Trans. Commun.*, vol. 50, no. 12, pp. 2009–2017, 2002.

[26] R. M. H. Yim, J. Bajcsy, L. R. Chen, and C. Beainy, "On the capacity limit of asynchronous OCDMA with single user detection," in *Proc. Biennial Symposium on Communications*, pp. 485–489, Kingston, Ontario, Canada, June 2002.

[27] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, John Wiley & Sons, New York, NY, USA, 1991.

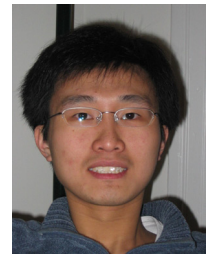
[28] S. Arimoto, "An algorithm for computing the capacity of arbitrary discrete memoryless channels," *IEEE Trans. Inform. Theory*, vol. 18, no. 1, pp. 14–20, 1972.

[29] S. B. Wicker and V. K. Bhargava, Eds., *Reed-Solomon Codes and Their Applications*, IEEE Press, Piscataway, NJ, USA, 1994.

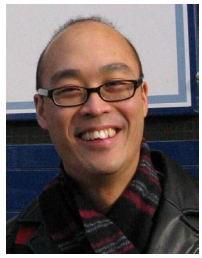
[30] C. Berrou, A. Glavieux, and P. Thitimajshima, "Near shannon limit error-correcting coding and decoding: turbo-codes (1)," in *Proc. IEEE International Conference on Communications (ICC '93)*, pp. 1064–1070, Geneva, Switzerland, May 1993.

Aminata A. Garba received a B.Eng. degree from Laval University in 2002 and an M.Eng. degree from McGill University in 2004, both in electrical engineering. She is presently a Ph.D. student at McGill University and her research interests include optical communications, information theory, and coding.

Raymond M. H. Yim received his B.Eng. and M.Eng. degrees in electrical engineering from McGill University in 2001 and 2002, respectively. He is currently pursuing his Ph.D. studies at Harvard University and his current research interests include switching and scheduling, high-speed access networks, and wireless communications.



Jan Bajcsy received a B.S. degree in engineering sciences from Harvard University in 1994 and M.A. and Ph.D. degrees in electrical engineering from Princeton University in 1997 and 1999, respectively. He is presently an Assistant Professor of electrical engineering at McGill University and his research interests include optical and wireless communication systems, coding, and information theory.



Lawrence R. Chen received the B.Eng. degree in electrical engineering and mathematics from McGill University in 1995 and the M.A.S. and Ph.D. degrees in electrical and computer engineering from the University of Toronto in 1997 and 2000, respectively. He is presently an Assistant Professor in the Department of Electrical and Computer Engineering, McGill University, and his research interests focus on ultrafast photonics, fiber optics, and optical communication systems.

EURASIP Journal on Applied Signal Processing 2005:10, 1617–1623

Design of Extended Depth-of-Focus Laser Beams Using Orthogonal Beam Expansions

David P. Goren
Symbol Technologies Inc., One Symbol Plaza, Holtsville, NY 11742-1300, USA
Email: gorens@optonline.net

Joseph Katz
Symbol Technologies Inc., One Symbol Plaza, Holtsville, NY 11742-1300, USA
Mitsubishi Electric Research Laboratories, 201 Broadway, Cambridge, MA 02139, USA
Email: katz@merl.com

Leonard Bergstein
Symbol Technologies Inc., One Symbol Plaza, Holtsville, NY 11742-1300, USA
Email: drssb2000@aol.com

Received 16 April 2004; Revised 3 November 2004

Laser beams with extended depth of focus have many practical applications, such as scanning printed bar codes. Previous work has concentrated on synthesizing such beams by approximating the nondiffracting Bessel beam solution to the wave equation. In this paper, we introduce an alternate novel synthesis method that is based on maintaining a minimum MTF value (contrast) over the largest possible distance. To achieve this, the coefficients of an orthogonal beam expansion are sequentially optimized to this criterion. One of the main advantages of this method is that it can be easily generalized to noncircularly symmetrical beams by the appropriate choice of the beam expansion basis functions. This approach is found to be very useful for applications that involve scanning of the laser beam.

Keywords and phrases: lasers, beam shaping, optical transfer functions, propagation, extended depth of focus, orthogonal beam expansion.

1. INTRODUCTION

Laser beams are commonly used to read digital information that has been encoded as a sequence of alternating light and dark regions on a reflective media. One such application is the reading of printed bar codes [1]. In this application, the distance from the bar code reader to the bar code label is usually variable and unknown, and in many cases it is desirable to read it over the largest possible distance, a feature that greatly improves the ergonomics of the reader and reduces operator training.

For such a system, the laser beam should have a large depth of focus, loosely defined as the region where the beam is "narrow enough" to resolve the fine structure of the bar code. A proper definition of the depth of focus is crucial to optimize such a system, and it must take into account that the spatially encoded information can typically withstand moderate distortions and still be properly decoded. In addition, when scanning bar codes whose aspect ratio (i.e., the ratio between the height and width of a bar code element) is high, it is the line-spread function (LSF) of the beam, rather than its point-spread function (PSF), that determines the overall system performance. In such cases, a highly elliptical beam is desirable in order to take advantage of the vertical redundancy in the code, and to reduce printing or laser speckle noise. (In the case of two-dimensional bar codes [1], a nearly circular beam is required in order to avoid interrow interference.)

Most previous work on synthesizing extended depth-of-focus laser beams have been based on either approximating the nondiffracting Bessel beam solution to the wave equation [2, 3, 4], or applying more general 3D synthesis techniques [5, 6]. Compared to the more general 3D synthesis techniques, the technique developed in this paper uses a simpler optimization criterion, and does not require samples of the desired beam profile at various planes. In addition, the use of a modal beam expansion guarantees the resulting beam satisfies the paraxial wave equation.

The outline of this paper is as follows. First an appropriate definition of depth of focus based on the modulation

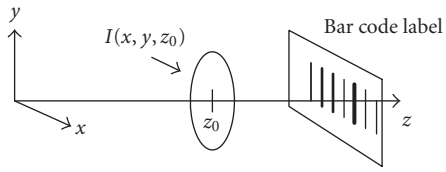


Figure 1: Schematic layout of the optical system.

transfer function (MTF) is presented. Using this definition, a Gaussian beam is optimized to achieve a maximum depth of focus, followed by the optimization of more general beams through the use of orthogonal beam expansions. Both circular and noncircular beams are considered by using the appropriate basis functions.

2. MTF-BASED OPTIMIZATION CRITERION

Consider a two-dimensional laser beam intensity profile at a fixed distance z_0 given by $I(x, y, z_0)$ with Fourier transform $U(u, v, z_0)$, where v is spatial frequency and x, y, and z are distances in the x-, y-, and z-directions, respectively, of the beam profile. The overall setup of the system is shown schematically in Figure 1. The line-spread function (LSF) of the beam is defined as [7]

$$s(x, z_0) = \int_{-\infty}^{\infty} I(x, y, z_0)\, dy. \tag{1}$$

The modulation transfer function (MTF) of the beam is defined as the Fourier transform of the LSF and can be expressed as

$$F(u, z_0) = U(u, 0, z_0) = \int_{-\infty}^{\infty} s(x, z_0) e^{-jux} dx. \tag{2}$$

The MTF describes the spatial filtering effect of scanning the laser beam over a one-dimensional spatial pattern extended infinitely in the y-direction.

An important class of (approximately) such spatial patterns is the ubiquitous printed bar code. These signals encode digital information through the use of alternating white and dark regions (bars and spaces) of varying widths. Laser scanning systems designed to read such patterns typically require a minimum contrast level for all spatial frequencies up to the highest fundamental spatial frequency of the narrowest bar/space pair. Higher signal-to-noise ratios allow lower contrast levels to be used.

For such applications, it is convenient to define the depth of focus of a scanning laser beam to be the region on the z-axis that maintains a minimum contrast level, C, for all frequencies up to and including the highest fundamental frequency. In what follows, all MTF curves will be normalized so that $F(u = 0, z) = 1$ for all z.

Let u_0 denote the fundamental spatial frequency of the narrowest elements of a bar code (i.e., narrow bar/space pair). Such a spatial pattern will be considered within focus at a distance z_0 if

$$F(u, z_0) \geq C, \quad 0 \leq u \leq u_0. \tag{3}$$

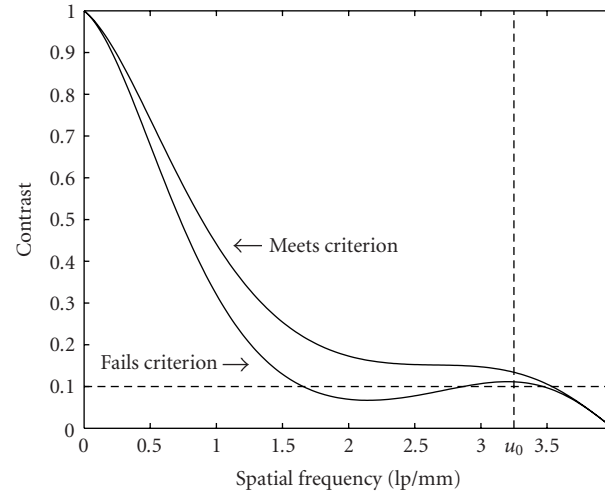


Figure 2: Example of MTF curves.

This criterion guarantees that there will be an adequate level of signal modulation present at the highest fundamental bar/space frequency u_0 and that no contrast reversals or loss of modulation occurs at lower frequencies. An illustration of MTF curves that both meet and fail the criterion is shown in Figure 2.

This definition of depth of field will be used to design laser beams that maximize the depth of focus given a highest spatial frequency u_0 and the minimum contrast level C.

3. GAUSSIAN BEAM OPTIMIZATION

The optimization of a simple Gaussian beam [8] will serve both as a simple demonstration of the method, as well as the first step required for the optimization of more general beams. Note that the LSF of a two-dimensional Gaussian beam is also Gaussian, thus a one-dimensional beam optimization is sufficient.

The LSF of a Gaussian beam at a fixed distance z_0 is given by

$$s(x, z_0) = \frac{\sqrt{2}}{\sqrt{\pi}\omega(z_0)} e^{-2x^2/\omega^2(z_0)}, \tag{4}$$

where $\omega(z_0)$ is the beam radius at z_0 given by

$$\omega(z_0) = \omega_0 \sqrt{1 + \left(\frac{\lambda z_0}{\pi \omega_0{}^2}\right)^2}. \tag{5}$$

For this case, it is assumed that the only free parameter to optimize is the minimum beam radius ω_0 occurring at $z = 0$.

The optimum beam radius ω_0 that maximizes the depth of focus for a given contrast C and maximum spatial frequency u_0 is calculated to be

$$\omega_0 = \frac{1}{u_0\pi}\sqrt{-\ln C}. \tag{6}$$

The resulting depth of focus is

$$d_{\text{opt}} = \frac{-2\ln C}{{u_0}^2\pi\lambda} \tag{7}$$

with the distances $|z| \le d_{\text{opt}}/2$ meeting the criterion for focus as defined by (3). Note that the depth of focus is inversely proportional to the square of the maximum spatial frequency u_0. This inverse relationship also applies to higher-order modes.

In the following section, the optimization criterion will be applied to the more general set of Hermite-Gaussian beams.

4. OPTIMIZATION OF THE HERMITE-GAUSSIAN EXPANSION

The Gaussian beam is the lowest-order member of the family of Hermite-Gaussian (H-G) beams (also known as H-G modes [8]). The H-G beams form a complete orthogonal set of functions satisfying the paraxial wave equation in rectangular coordinates, and are thereby capable of representing an arbitrary propagating beam. These beams also have the desirable property that they remain H-G as they propagate, thus providing a simple model for the propagation and optimization of complex beams. A description of the H-G beams is given in Appendix A.

Because the described optimization criterion is solely a function of the LSF, and the fact that the two-dimensional H-G functions are separable in rectangular coordinates, we only need to consider a one-dimensional expansion. We will restrict the generated beam to certain desirable symmetries appropriate to real-world applications. The first is that the LSF of the beams are symmetrical about $x = 0$. This condition is guaranteed by using only even-order H-G modes in the beam expansion. The second symmetry is that the beam propagates symmetrically about $z = 0$. This restricts the expansion coefficients to be real, and guarantees maximum depth of field for a chosen u_0.

The optimizing approach is based on representing the desired beam as a sum of H-G beams with unknown coefficients. In general, an infinite sum is required to represent an arbitrary beam. We will show that in practical applications, the series needs to include only a relative small number of N terms. This leaves N unknown parameters to be optimized. A finite H-G expansion of a one-dimensional beam (within a constant factor) using N even-order modes is given by

$$g_N(x,z) = u_0(x,z;b_0) + \sum_{n=1}^{N-1} A_n u_{2n}(x,z;b_0), \tag{8}$$

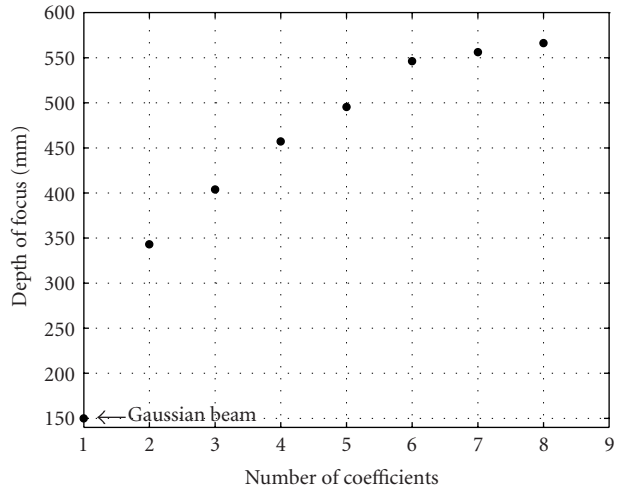


Figure 3: Depth of focus versus number of H-G coefficients.

where the N unknown parameters to be optimized are $[b_0, A_1, A_2, \ldots, A_{N-1}]$. The parameter b_0 is the confocal parameter and is shared by all the modes (see Appendix A). It should be noted that for Gaussian beams that correspond to the lowest order of the H-G family of beams, this parameter is known as the "Rayleigh distance." The LSF of the beam at a distance z_0 is then given by the beam intensity expressed as

$$s(x,z_0) = g_N(x,z_0) \cdot {g_N}^*(x,z_0). \tag{9}$$

As more terms of the series are included in the optimization, the depth of focus of the beam increases.

Optimization of the beam parameters was performed with computer search techniques using a sequential series of optimizations for increasing values of N (see Matlab Optimization Toolbox at http://www.mathworks.com/products/optimization). In particular, the multidimensional simplex search algorithm of Nelder and Mead was used due to the difficulty of obtaining accurate derivative information for the calculated depth of field.

The optimization for $N = 1$ (Gaussian beam) was analytically derived in Section 3. For $N > 1$, the critical initial guess required to seed the numerical multidimensional optimization of all N parameters was supplied by the previous $(N-1)$-term optimization. In addition, it was found useful to refine the initial guess by first performing a suboptimization using only the newly added coefficient together with the confocal parameter. This procedure greatly enhanced the convergence to the appropriate solution. It is important to note that all frequencies from 0 to u_0 must be checked at every z to determine if the beam is within focus as defined in (3).

The depth of focus achievable as a function of the number of terms in the series is shown in Figure 3 for a maximum spatial frequency of 3.9 lp/mm (e.g., bar code with a 5-mil narrow element) and a minimum contrast of 10%. The numerical values of the optimized parameters are given in Table 1.

Table 1: H-G coefficients.

Number of terms N	b_0	A_1	A_2	A_3	A_4	A_5	A_6	A_7
1	0.07	—	—	—	—	—	—	—
2	0.24	−1.19	—	—	—	—	—	—
3	0.33	−1.08	0.55	—	—	—	—	—
4	0.53	−0.9	0.78	−0.49	—	—	—	—
5	0.85	−0.87	0.78	−0.57	0.70	—	—	—
6	1.08	−1.03	0.91	−0.83	0.77	−0.7	—	—
7	1.36	−1.0	1.0	−0.88	0.74	−0.73	0.91	—
8	1.60	−0.96	1.0	−0.96	0.78	−0.67	0.80	−0.75

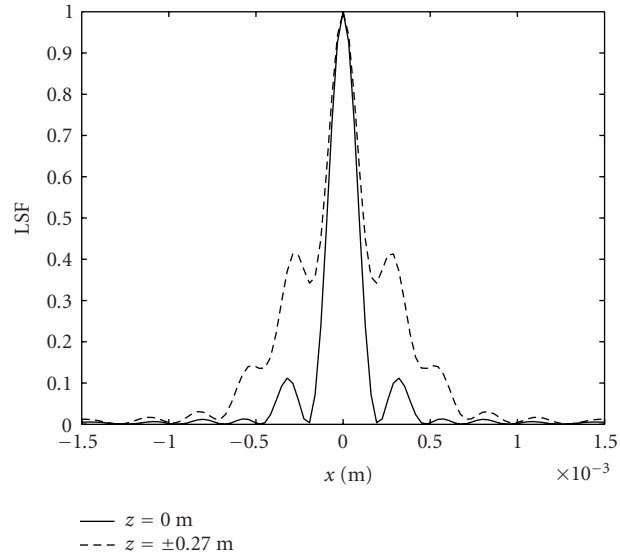


Figure 4: LSF curves (H-G).

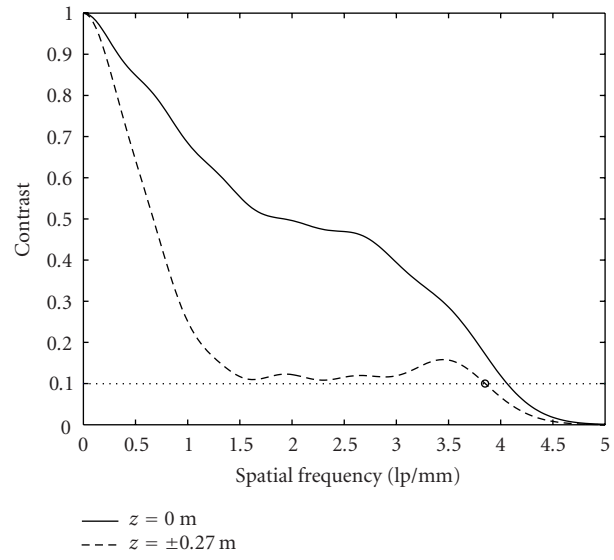


Figure 5: MTF curves (H-G).

Examining Figure 3 shows that simply adding a second term to the series (Gaussian beam and the next lowest even-order mode) doubles the depth of focus. As more terms are added, the rate of improvement decreases. In particular, there is little benefit after six terms. Figures 4 and 5 show the resulting LSF and MTF, respectively, of a beam that uses six terms. To physically realize the beam, diffractive optical elements can be employed [9], with higher-order beams requiring more spatial resolution.

5. OPTIMIZATION OF THE LAGUERRE-GAUSSIAN EXPANSION

In the previous section, we used the essentially one-dimensional H-G expansion to optimize optical beams used to read conventional bar codes. In the case of a two-dimensional bar code, a different optimization criterion is needed, since elongated optical beams cause severe inter-row interference. In this case, it is advantageous to use optical beams with circular symmetry. In this section, we will analyze one example based on the Laguerre-Gaussian series [10].

In spherical coordinates, the family of beams that form a complete set of solutions to the paraxial wave equation are the Laguerre-Gaussian (L-G) beams. A description of the L-G beams are given in Appendix B. In this case, the LSF can be expressed as

$$s(x, z_0) = \int_{-\infty}^{\infty} I\left(\sqrt{x^2 + y^2}, z_0\right) dy. \tag{10}$$

In a fashion similar to the optimization of the H-G beams, the coefficients of a six-term L-G expansion were optimized to guarantee a contrast level of 10% at 3.94 lp/mm. The resulting LSF and MTF curves are shown in Figures 6 and 7, respectively. The optimized parameters are given in Table 2.

It should be noted that while both the circularly symmetric optimization (L-G) and the 1D (H-G) optimization maintain a minimum contrast level, the circular beam's LSF undergoes significantly less variation over the depth of focus compared to the 1D beam, as evidenced in the curves presented in Figure 4 compared to those of Figure 6. This is due to the existence of a nondiffracting beam solution in spherical coordinates while no such beam exists in rectangular coordinates [2].

Table 2: L-G coefficients.

Number of terms N	b_0	A_1	A_2	A_3	A_4	A_5
1	0.07	—	—	—	—	—
2	0.21	0.70	—	—	—	—
3	0.64	1.58	1.29	—	—	—
4	0.82	1.35	1.66	1.46	—	—
5	0.98	1.25	1.60	1.87	1.17	—
6	1.11	1.17	2.07	3.35	2.03	1.17

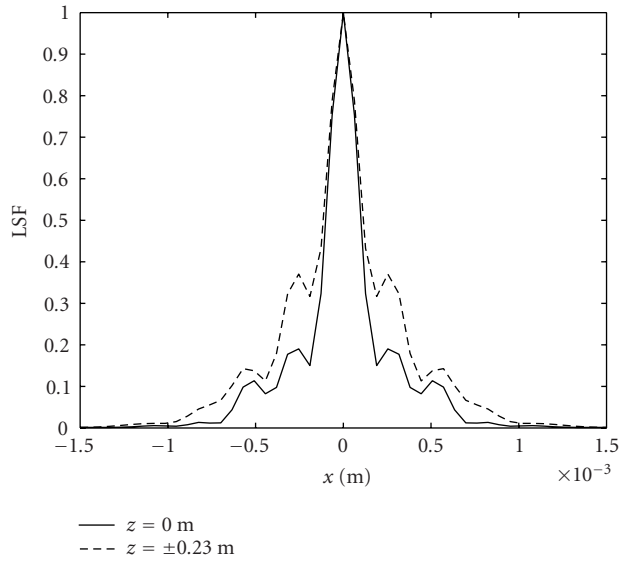


Figure 6: LSF curves (L-G).

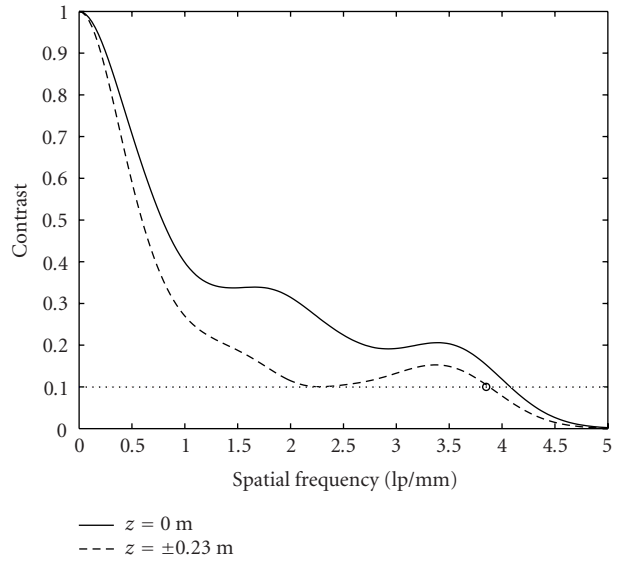


Figure 7: MTF curves (L-G).

6. SUMMARY AND CONCLUSIONS

Extended depth-of-focus laser beams have many practical applications. One such application is the scanning and decoding of printed bar codes over extended ranges. Previous work has concentrated on synthesizing such beams by approximating the nondiffracting Bessel beam solution to the wave equation [3, 4]. In this paper, a novel optimization technique was presented based on orthogonal beam expansions. One of the main advantages of this method is that it can generate noncircularly symmetrical beams which offer great advantage when scanning noisy one-dimensional patterns such as bar codes. Using this expansion, we have shown that with a relatively small number of terms, the laser beam operational depth of focus can be readily extended by more than threefold.

It is important to note that the overall performance of a bar reader depends not only on the optical properties of its optical beam, but also on the processing of the signal derived from the reflected beam.

When applying the beams described in this paper to bar code scanning applications, the price to be paid for the extended depth of focus is lower contrast (less optical power received) and waveform complexity resulting from the sidelobe structure of the beam. Higher-power lasers along with lower-noise electronics can address the contrast issue and more advanced signal processing techniques are required to remove the resulting distortion artifacts and to correctly reconstruct the original spatial pattern [11, 12].

APPENDICES

A. HERMITE-GAUSSIAN MODES

A Hermite-Gaussian function of order n is expressed as

$$\psi_n(\xi) = H_n(\xi)e^{-\xi^2/2}, \tag{A.1}$$

where $H_n(\xi)$ are Hermite polynomials. Some examples of low-order Hermite polynomials are

$$H_0(\xi) = 1, \qquad H_1(\xi) = 2\xi, \qquad H_2(\xi) = 4\xi^2 - 2. \tag{A.2}$$

The nth-order Hermite-Gaussian mode is defined as

$$u_n(x, z = 0, b_0) = C_n\psi_n\left(\frac{\sqrt{2}x}{\omega_0}\right) \tag{A.3}$$

with

$$\omega_0 = \sqrt{\frac{2b_0}{k}}, \tag{A.4}$$

where b_0 is the confocal parameter and k is the wave number. For Gaussian beams, the lowest order of the H-G beams, the confocal parameter b_0 is also known as the Rayleigh distance so that $\omega(b_0) = \sqrt{2}\omega_0$.

C_n is the energy normalization constant given by

$$C_n = \left(\frac{\sqrt{2}}{\omega_0 2^n n! \sqrt{\pi}}\right)^{1/2}. \tag{A.5}$$

For $z \neq 0$, the propagation of the Hermite-Gaussian modes are given by

$$u_n(x,z;b_0) = e^{j(kz-\pi/4)} \frac{C_n}{\left(1+z^2/{b_0}^2\right)^{1/4}} \psi_n\left(\frac{\sqrt{2}x}{\omega(z)}\right) \times e^{-jkx^2/2R(z)} e^{j(2n+1/2)\phi(z)}, \tag{A.6}$$

where

$$\omega(z) = \sqrt{\frac{2b_0}{k}\left(1+\frac{z^2}{{b_0}^2}\right)}, \qquad R(z) = \frac{z^2+{b_0}^2}{z}, \qquad \phi(z) = \tan^{-1}\left(\frac{z}{b_0}\right). \tag{A.7}$$

Consider an arbitrary one-dimensional beam with input amplitude function $h(x)$ at $z = 0$. The function $h(x)$ can be expanded as

$$h(x) = \sum_{n=0}^{\infty} A_n u_n(x, z=0; b_0), \tag{A.8}$$

where A_n are the expansion coefficients and b_0 is the confocal parameter shared by all the H-G modes. The propagation of the beam for $z \neq 0$ can then be expressed as a summation of the individual propagating beams given by

$$g(x,z) = \sum_{n=0}^{\infty} A_n u_n(x,z;b_0). \tag{A.9}$$

The intensity profile at a distance z_0 is given by

$$I(x,z_0) = g(x,z_0) g^*(x,z_0). \tag{A.10}$$

B. LAGUERRE-GAUSSIAN MODES

The nth-order Laguerre-Gaussian mode at $z = 0$ is defined as

$$u_k(r, z=0; b_0) = \frac{1}{\sqrt{\pi}} L_k\left(\frac{2r^2}{{\omega_0}^2}\right) e^{-r^2/{\omega_0}^2}, \tag{B.1}$$

where ω_0 is define in (A.4) and $l_k(\xi)$ are Laguerre polynomials. Some examples of low-order Laguerre polynomials are

$$l_0(\xi) = 1, \qquad l_1(\xi) = -\xi + 1, \qquad l_2(\xi) = \xi^2 - 4\xi + 2. \tag{B.2}$$

For $z \neq 0$, the propagation of the Laguerre-Gaussian modes is given by

$$u_k(x,z;b_0) = e^{j(kz)} \frac{1}{\sqrt{\pi}} \frac{1}{\left(1+z^2/{b_0}^2\right)^{1/2}} L_k\left(\frac{2r^2}{\omega^2(z)}\right) \times e^{-r^2/{\omega_0}^2} e^{-jkx^2/2R(z)} e^{j(2k+1)\phi(z)}, \tag{B.3}$$

where $\omega(z)$, $R(z)$, and $\phi(z)$ are defined in (A.7).

DISCLAIMER

This paper reflects the personal understandings and opinions of the authors only and is not intended in any way to convey any position, policy, or opinion of the author's employer, Symbol Technologies, Inc. The authors and Symbol Technologies, Inc. disclaim any liability for any errors or omissions in the paper.

ACKNOWLEDGMENT

The authors would like to acknowledge Emanuel Marom, Ed Barkan, and Vladimir Gurevich for useful technical discussions.

REFERENCES

[1] R. C. Palmer, *The Bar Code Book*, Helmers Publishing, Petersborough, NH, USA, 1995.

[2] J. Durnin, "Exact solutions for nondiffracting beams. I. The scalar theory," *Journal of the Optical Society of America* {*A*}, vol. 4, no. 4, pp. 651–654, 1987.

[3] Y. Li, V. Gurevich, M. Krichever, J. Katz, and E. Marom, "Propagation of anisotropic Bessel-Gaussian beams: sidelobe control, mode selection, and field depth," *Applied Optics*, vol. 40, no. 16, pp. 2709–2721, 2001.

[4] K. Tanaka, M. Tguchi, and T. Tanaka, "Quasi-diffraction-free beams," *Journal of the Optical Society of America* {*A*}, vol. 18, no. 7, pp. 1644–1649, 2001.

[5] R. Piestun and J. Shamir, "Synthesis of three-dimensional light fields and applications," *Proc. IEEE*, vol. 90, no. 2, pp. 222–244, 2002.

[6] U. Levy, D. Mendlovic, Z. Zalevsky, G. Shabtay, and E. Marom, "Iterative algorithm for determining optimal beam profiles in a three-dimensional space," *Applied Optics*, vol. 38, no. 32, pp. 6732–6736, 1999.

[7] A. Papoulis, *Systems and Transforms with Applications in Optics*, McGraw Hill Book Company, New York, NY, USA, 1968.

[8] H. A. Haus, *Waves and Fields in Optoelectronics*, Prentice Hall, Englewood Cliffs, NJ, USA, 1984.
[9] E. Marom, D. Mendlovic, N. Konforti, J. Katz, and C. Tan, "Diffractive optic elements for forming scanning beams", in *Diffractive and Holographic Device Technologies and Applications IV*, vol. 3010 of *Proceedings of SPIE*, San Jose, Calif, USA, February 1997.
[10] A. E. Siegman, *Lasers*, University Science Books, Mill Valley, Calif, USA, 1986.
[11] S. J. Shellhammer, D. P. Goren, and T. Pavlidis, "Novel signal-processing techniques in barcode scanning," *IEEE Robot. Automat. Mag.*, vol. 6, no. 1, pp. 57–65, 1999.
[12] S. J. Shellhammer and D. Goren, "Fuzzy logic bar code scanners," in *Proc. IEEE Workshop on Automatic Identification Advanced Technologies (WAIAT '97)*, pp. 49–52, Stony Brook, NY, USA, November 1997.

David P. Goren received his B.S. degree in electrical engineering from SUNY at Stony Brook in 1986. He received his M.S. degree in electrical engineering in 1991 from Polytechnic University, where he is currently pursuing a Ph.D. Since 1986, he has worked at Symbol Technologies in Holtsville, NY, and holds numerous patents in the bar code field.

Joseph Katz is a Vice President of Mitsubishi Electric Research Laboratories and Deputy Director of its Research Laboratory. In addition to his role as a member of MERL's management team, he pursues his research interests in optoelectronics, communications, and RFID. Until 2004 he was with Symbol Technologies, where as a Senior VP of R&D, he participated, initiated, and led projects in a wide range of technologies, including barcode/RFID data capture, optics, imaging, signal processing, computing, networking, security, biometrics, and communications. Prior to joining Symbol Inc., Joseph Katz worked at JPL/Caltech, where he conducted and led R&D efforts in optical communications and advanced optoelectronic materials and devices, and was a recipient of the JPL Director's Research Achievement Award and of 25 NASA awards for technical innovations. Joseph Katz authored/coauthored over 150 technical publications and holds over 120 US patents. He was elected as a Fellow of both the IEEE and the Optical Society of America.

Leonard Bergstein, a holder of a Ph.D. degree in optics/electro-optics in 1960 from Polytechnic University, Brooklyn, NY, is a Professor Emeritus of electro-optical sciences at Polytechnic University. He is a Fellow of the Optical Society of America and has been a Consultant on optics and electronics to various industrial and educational organizations. He is one of the inventors of and holds basic patents on zoom lenses and has written extensively on quantum electronics and electro-optics.