REPORT DOCUMENTATION PAGE

AFRL-SR-AR-TR-05-

0462

1. REPORT DATE (DD-MM-YYYY)	2. REPORT TYPE	3. DATES COVERED (From - To)
10/14/05	Final Technical	05/01/01 - 05/31/05

4. TITLE AND SUBTITLE
Optically Addressed Nanostructures for High Density Data Storage

5a. CONTRACT NUMBER

5b. GRANT NUMBER
F49620-01-1-0363

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Dr. Mark A. Neifeld and Richard W. Ziolkowski

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Department of Electrial and
Computer Engineering
College of Engineering
University of Arizona
Tucson, AZ 85721

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
USAF, AFRL, AF Office of Scientific Research NE
4015 Wilson Blvd, Room 713
Arlington, VA 22203-1954

10. SPONSOR/MONITOR'S ACRONYM(S)
U

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release - Distribution is Unlimited

13. SUPPLEMENTARY NOTES

A memory is any physical system with a large number of distinguishable states. Through the work undertaken in this program we have shown that an optical field can be used to distinguish among N different physical configurations of a nanostructure whose volume is comparable to a cubic wavelength, thus achieving $\log_2(N)$ bits of capacity within a single resolution element. We have studied the use of near-field optical measurement together with spatially- and spectrally-selective defects to define and readout sub-resolution nanostructure storage configurations. We have successfully identified, studied, and characterized nanostructure configurations that provide optically distinguishable states with large interstate distances. We have focused on sub-resolution surface-relief structures combined with near-field detector arrays to demonstrate approximate storage densities of 25 bits/μm^2. We extended this work to so-called nano-structured voxels in which we employed sub-resolution volumetric degrees of freedom and near-field detector arrays. We designed the data carrying volumetric defects so as to exploit a plasmon resonance in their metallic nanoshells, and we predicted significantly improved storage densities of nearly 300 bits/μm^2. This work has resulted in an improved understanding of the interaction between space-time electromagnetic fields and various (imperfect) nanostructure volumes, new near-field optical characterization methods to determine the sub-resolution 3D configuration of artificial nanostructures, and new bounds on the abilities of optical fields to probe physical processes on length scales below the optical wavelength.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			19b. TELEPHONE NUMBER (include area code)
Unclassified	Unclassified	Unclassified			520-621-5864/6173

2005 AFOSR Final Progress Report

University of Arizona
Optically Addressed Nanostructures for High Density Data Storage
M. A. Neifeld and R.W. Ziolkowski
AFOSR # F49620-01-1-0363

I. Program Overview

A memory is any physical system with a large number of distinguishable states. Through the work undertaken in this program we have shown that an optical field can be used to distinguish among N different physical configurations of a nanostructure whose volume is comparable to a cubic wavelength, thus achieving $\log_2(N)$ bits of capacity within a single resolution element. We have studied the use of near-field optical measurement together with spatially- and spectrally-selective defects to define and readout sub-resolution nanostructure storage configurations. We have successfully identified, studied, and characterized nanostructure configurations that provide optically distinguishable states with large interstate distances. During year 1 we focused on sub-resolution surface-relief structures combined with near-field detector arrays to demonstrate approximate storage densities of 25 bits/μm^2. During year 2 we extended this work to so-called nano-structured voxels in which we employ sub-resolution *volumetric* degrees of freedom and near-field detector arrays. We design the data carrying volumetric defects so as to exploit a plasmon resonance in their metallic nanoshells and we predict significantly improved storage densities of nearly 300 bits/μm^2. This work has resulted in an improved understanding of the interaction between space-time electromagnetic fields and various (imperfect) nanostructure volumes, new near-field optical characterization methods to determine the sub-resolution 3D configuration of artificial nanostructures, and new bounds on the abilities of optical fields to probe physical processes on length scales below the optical wavelength.

One unanticipated outcome of the work outlined above was the creation of a novel method of designing nonlinear codes. These codes were necessary in order to tolerate the crosstalk that characterizes nano-structured optical memories. We invented these codes for use in the memory application and later found them to be valuable for use in fiber communications.

Our research activities within this program have progressed along two distinct lines of inquiry. The original program vision was exclusively concerned with static optical memory structures. Within this component of the program we have analyzed and demonstrated the utilization of sub-resolution degrees of freedom to achieve room-temperature high-density optical storage as outlined above. A complementary component of our activity grew out of some early success and gave rise to a parallel research path. This parallel path has been concerned with the use of sub-resolution volumetric degrees of freedom to control light in the time domain. This form of control facilitates simple delay-based storage and can be important in pulse shaping and optical buffering applications. During year 2 we pursued a theoretical understanding of information propagation in such systems. We also experimentally demonstrated that the information velocity in so-called fast-light media remains below the limit defined by relativistic causality. Our year 3 activities continued along these lines. We extended our theoretical understanding of information velocity into the domain of slow-light media and we have experimentally demonstrated the surprising result that even for systems in which the group velocity is much less than the speed of light in vacuum the *information* velocity remains close to c. The no-cost-extension of this work has concluded with a novel method of compensating distortion in slow-light optical delay lines.

This report is organized as follows. Our research is conveniently described in terms of 8 different projects. A brief summary of each project and its outcomes is provided below. For work that has already resulted in one or more peer-reviewed journal articles these are included as a detailed description of the relevant technical details.

II. Surface-Relief Nano-Structures

Our main project component has dealt with the study of surface relief structures. Within this component we have made contributions in three areas: analysis and simulation, planar memory concepts, and detector designs and realizations. We continued and extended the development of our FDTD simulators to study the memory density of nanoscale surface-relief structures. We have advanced several approaches to calculate the clique-based capacity of these structures. We have considered and bounded statistical cross-talk between different relief configurations. We have also considered a number of new planar memory concepts. These have included the study of relief layers consisting of polymer, a low (n=1.8) contrast medium, versus semiconductor, a high (n=3.5) contrast medium; structures that represent binary versus non-binary states; and relief structures that emphasize hole versus backfill configurations. Parameterization of a variety of the relief parameters including the gap size between the cells of interest, the base height, and the relief heights has been accomplished to determine their influence on the number of distinguishable states and, hence, the capacity. More details about this work can be found in Appendix 1.

III. Nano-Structured Voxels

Our second main project component has dealt with the study of volumetric memory structures. Within this component we have made contributions in two main areas: analysis and simulation of nano-structured voxels (NSVs) defined with dielectric, metallic, and/or plasmonic elements and the use of optical tweezers for novel fabrication methods to realize these volumetric structures. We have used our FDTD simulator to find which nano-structured configurations can be used to realize high-density memory storage devices. Detector arrays were placed appropriately under the scattering elements. Very distinctive states have been realized with the metallic and plasmonic scattering elements. We have found that the dielectric, metallic, and/or plasmonic elements could be positioned in the requisite matrix using optical tweezer technologies. This fabrication approach is being analyzed further to understand its limitations and practicality.

IV. Nonlinear Coding for Fiber Communications

The nonlinear codes developed for use in nano-structured storage application proved to be valuable for mitigating the distortion associated with nonlinear fiber propagation. We have considered a single wavelength channel in a nonlinear dispersive single-mode fiber in the normal dispersion regime. We study the propagation of binary sequences using OOK RZ modulation at a bit rate of 10 Gb/s, under different dispersion and input power conditions. We determine the impact that dispersion and nonlinearity have on the pulse sequences by measuring the Euclidian distances among all pairs of received sequences. We then model the set of Euclidian distances as a fully connected graph and by means of a clique-finding algorithm we search for subsets of sequences that form codes matched to this nonlinear channel, resulting in significant bit error rate improvement.We present a numerical analysis of the effects that GVD and SPM have on the Euclidian distances among pulse sequences at the receiver in a single channel of a WDM system. We observe that nonlinearity induces a decrease in the minimum Euclidian distance within the normal dispersion regime and that this decrease is stronger at high power levels. We model the set of Euclidian distances as a fully connected graph, where each sequence is a node and the Euclidian distance between each pair of sequences labels the corresponding edge. We then remove all edges with a Euclidian distance smaller than an arbitrarily chosen value. Using a Clique finding algorithm, we search for subsets of nodes that form fully connected subgraphs of this pruned graph. From these subsets we obtain nonlinear codes that are well matched to this nonlinear fiber channel, giving a significant reduction in bit error rate compared to linear codes of the same length. More details about this work can be found in Appendix 2.

V. Information Propagation in a Fast-Light Medium

Projects II and III are concerned with the use of nanostructures to control light in the space-domain. It is also possible to employ nanostructures to control light in the time-domain. One manifestation of this control is the engineering of materials with large anomalous dispersion. The special theory of relativity states that the velocity of information propagation is limited to the speed of light in vacuum (c). It is known

however, that optical pulses traveling in a medium with large anomalous dispersion have faster-than-*c* group velocities. The key to resolving this controversy is to quantify how information is (a) encoded on an optical pulse and (b) affected by the propagation medium. We have transmitted two distinct symbols through a 'fast light' medium and quantify both the propagation time and the detection latency. The former allows us to bound the information velocity; while, the latter reconciles the apparent discrepancy with group velocity. In our experiment we find that although the group velocity vastly exceeds *c*, the time required to detect information propagating through the medium is slightly longer than the time required to detect the same information travelling through vacuum. Our work highlights the issues that must be addressed in designing devices that operate at the ultimate limit imposed by the special theory of relativity. More details about this work can be found in Appendix 3 and related work on causality in metamaterials in Appendix 6.

VI. Information Propagation in a Slow-Light Medium

Through manipulation/engineering of materials at the nano-scale it is possible to create an optical medium in which the group velocity (v_g) is a small fraction of c. These systems represent a type of short-term memory that can be valuable in optical communication, computing, and processing applications. In our experiments we have $v_g = c/100$. It is often assumed that the velocity of information propagation in such a medium is given by the group velocity whenever $v_g<c$. During this part of our year 3 activities we sought to experimentally measure the velocity of information propagation in such a medium. By imposing a point of non-analyticity on the propagating waveform, we create a well-defined time at which information is launched. By estimating the time at which this information can first be measured we are able to obtain an estimate of the information velocity (v_i). Figure 1 shows some experimental results. Visually we can see that the time at which information (i.e., the discontinuity) appears at the detector is nearly the same for the signal propagating through vacuum and the slow-light medium. A bit error rate analysis of this data reveals that v_i is approximately 0.6c. This surprising result reveals that (a) the group velocity is not always a good measure of information velocity, (b) information can be carried in the pre-cursor fields and propagates at nearly c, and (c) these pre-cursor fields can be large enough to provide reliable measurements of the associated information. We are still working to understand the implications of these results; however, it is already clear that the design of optical pulse shapes should be conducted with awareness of the nature of information propagation in the system under study. More details about this work can be found in Appendix 4.

VII. Distortion Compensation in a Slow-Light Delay Line

A significant limitation to current methods of achieving slow-light pulse delay arises from the distortion that accompanies this delay. We have developed a new method through which large pulse delay may be achieved under the constraint that the distortion does not exceed a particular limit. Our technique is based on combining multiple dispersive devices in order to engineer the dispersion of the overall system. We have analyzed and experimentally demonstrated the case of a gain-doublet in which a pair of Lorentzian gain lines are used to achieve significant pulse delay relative to a single-line system. A factor of 6.25 improvement in delay and a factor of 2 improvement in pulse bandwidth can be achieved by use of our new method of jointly optimizing the overall system parameters. The optimization of more complex distortion-compensated systems based on resonators, interferometers, and PBGs is currently underway. More details about this work can be found in Appendix 5.

VIII. Pulse Shaping from a Spatially Varying Group Index

The new technique that we've developed is a spatio-temporal technique. It involves the spatial modulation of the pulse delay created by an optical medium, as shown in Figure 1. As the pulse propagates through the medium, some spatial components of the pulse are delayed more than others. This pulse "shear" can be used to control the temporal pulse shape. As the pulse is recombined, for example by refocusing or injecting the pulse into a fiber, the temporal intensity profile is determined by the spatial pulse delay profile of the medium. As in other spatio-temporal techniques, the pulse delay profile need only be controlled on the timescale that a single shape is used. This control can either be optical or electronic depending on the

pulse delay mechanism employed. Advantages of this technique over other spatio-temporal techniques are that it may be more easily extended to a wide range of pulse lengths (using media with different pulse delays) and it can be very compact. Figure 2 shows preliminary results demonstrating simple pulse-shaping of a pulse propagating through a BaTiO3 crystal with two regions of different pulse delays. These regions were created by controlling the intensity of a pump beam used to generate the pulse delay in each region. Because the response time of BaTiO3 is very slow, this material is best suited for shaping long pulses, whereas short pulses can be shaped using the same technique with faster-response-time materials such as nonlinear crystals, atomic systems, or "dispersion engineered" materials, possibly allowing electronic control with the use of EO inclusions.

The use of a medium with spatially varying optical properties also raises the possibility of diffraction. To our knowledge, the effect of diffraction from spatial varying group delay has not been previously studied. We are continuing both theoretical and experimental activities on this topic.

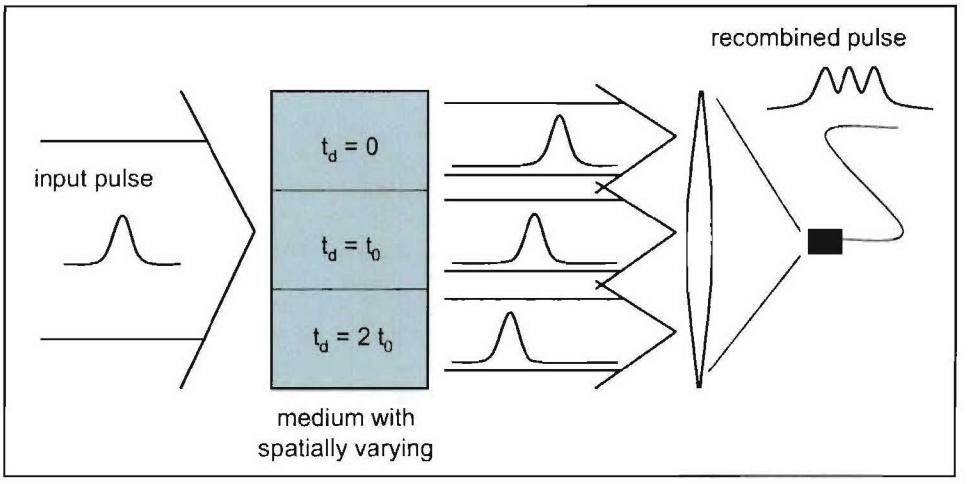

Figure 1: Medium with a spatially varying group index can be used as a pulse shaper.

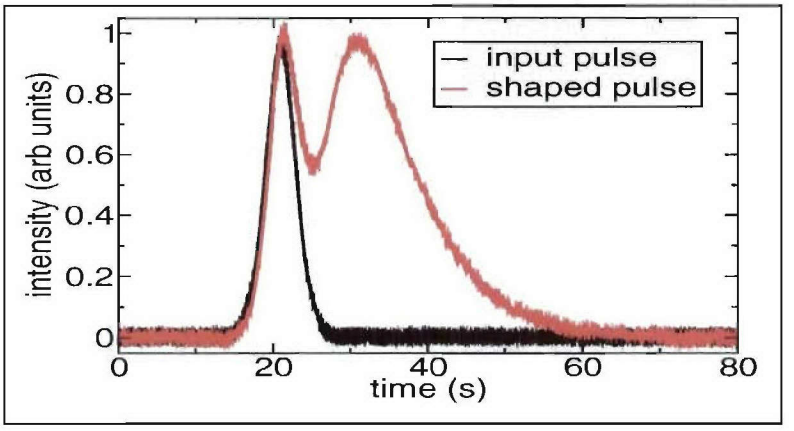

Figure 2: Initial results demonstrating the operation of our pulse shaper.

IX. Gaussian Beam Scattering from Nanostructures

We have continued to examine the scattering from nanostructures in the presence of complex material environments. Particular emphasis has been given to nanorods and nanospheres with plasmonic coatings. The enhanced scattering from these plasmonic nanostructures provides enhanced discrimination between scattering states for our optical data storage application. This effect has been used effectively for our nano-voxel studies. Additionally, it also allows one to guide light along subwavelength dimensions to produce subwavelength field localization. An example is shown below. A transmission line of nano-rods provides a means of capturing light and guiding it. The output spot is significantly subwavelength. More details about this work can be found in Appendix 7. This was a paper presented at a Special Session on *Advances in Plasmonic and Other Electromagnetic Phenomena* Co-Organized by R. W. Ziolkowski and N. Engheta and that appears in the Proceedings of the IEEE International Symposium on Antennas and Propagation, Washington DC, July 2005

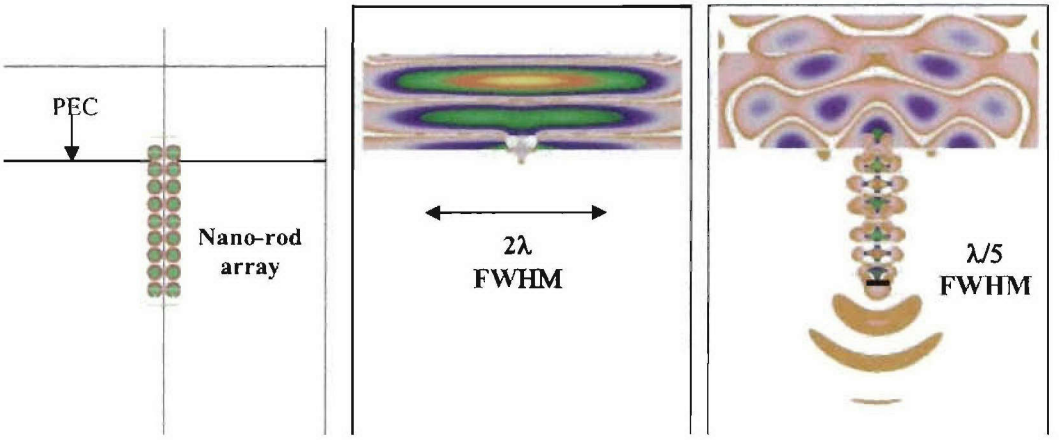

Figure 3: A nano-rod array transmission line is used to guide and to localize 500nm light acquired from an incident Gaussian beam to sub-wavelength resolutions.

X. Refereed Journal Publications

1. M. D. Stenner, D. J. Gauthier, and M. A. Neifeld, "The speed of information in a fast-light optical medium," *Nature*, Volume 425, pp.695-698, October 2003.
2. R. W. Ziolkowski and A. D. Kipple, "Causality and double-negative metamaterials," Physical Review E, Vol. 68, 2003.
3. J. Anguita and M. A. Neifeld, "Nonlinear codes for dispersive nonlinear fibers," *IEEE Photonics Technology Letters*, Vol.16, pp.2622-2624. 2004.
4. M. E. Potter, K. Gross, M. A. Neifeld, and R. W. Ziolkowski, "Nanostructure surface relief profiles for high-density optical data storage," *Optics Communications*, Vol.253, pp.56-69, 2005.
5. M. D. Stenner, D. J. Gauthier, and M. A. Neifeld, "Fast causal information transmission in a medium with a slow group velocity," *Physical Review Letters*, Vol.94, February 2005.
6. M. D. Stenner, M. A. Neifeld, Z. Zhu, A. Dawes, and D. Gauthier, "Distortion-management in slow-light pulse delay," submitted to *Optics Express*, October 2005.
7. R. W. Ziolkowski and M. A. Neifeld, "Nanostructured voxels for high-density optical memory," in preparation for submission to *Applied Optics*.

Appendix 1

Surface Relief Nano-Structures

Available online at www.sciencedirect.com
SCIENCE DIRECT®
Optics Communications 253 (2005) 56–69
OPTICS COMMUNICATIONS
www.elsevier.com/locate/optcom

Nanostructure surface relief profiles for high-density optical data storage

Mike E. Potter [a,*], Kelly Goss [a], Mark A. Neifeld [b], Richard W. Ziolkowski [b]

[a] *Department of Electrical and Computer Engineering, University of Calgary, 2500 University Drive NW, Calgary, AB, Canada T2N 1N4*
[b] *Department of Electrical and Computer Engineering, The University of Arizona, P.O. Box 210104, 1230 E. Speedway Blvd., Tucson, AZ 85721-0104, United States*

Received 6 January 2005; received in revised form 24 March 2005; accepted 19 April 2005

Abstract

A new optical data storage structure is proposed that utilizes surface relief profiles at sub-wavelength scales to distinguish memory states in a read-only memory system. Groupings of power sensors directly beneath the features read light transmitted through the structure from above. These are used to distinguish states from one another, where different surface profiles lead to different states. Two-dimensional simulations are performed using the finite-difference time-domain (FDTD) method to optimize parameters of the system, and to determine the data capacity and density of optimal systems. Two optimal systems are studied in detail: one intended as a polymer construction, the other as silicon. Techniques are developed to estimate capacity and density from subsets of simulations. Simulations estimate that densities of 3.17 GBits/cm^2 are possible with this type of system.

PACS: 42.79.Vb

Keywords: Optical storage; Memory systems

1. Motivation

Any system with a large number of distinguishable states can act as a memory (storage) device. Current generation high-capacity memory devices (archival rather than e.g. flash) generally rely on one of two technologies: optical or magnetic. Optical systems include CD-ROMs, DVDs, and the next generation of Blu-Ray discs. For all three, states are created and distinguished from each other by surface features (pits and ridges); *each* feature represents a single binary state. CDs are limited by classical resolution limits of approximately one bit per square wavelength. DVDs and

* Corresponding author. Tel.: +1 403 220 4129; fax: +1 403 282 6855.
E-mail address: mpotter@ucalgary.ca (M.E. Potter).

0030-4018/$ - see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.optcom.2005.04.050

Blu-Ray discs use multiple layers and can achieve four to eight bits per square wavelength. In the case of magnetic storage media, data is encoded in the magnetic polarity of small sub-domains of the medium. Scanning a probe over the structure reads the polarity in given regions. Demonstrations by Kong et al. [1] have shown magnetic storage densities on the order of 10 Gbits/in^2 (1.55 Gbits/cm^2). Sub-wavelength topographical features created with lithography can also be used to imprint data [2]. Features are then probed using atomic force microscopy techniques, since lasers cannot be effectively used to probe each *individual* feature. Impressive storage densities on the order of 400 Gbits/in^2 (62 GBits/cm^2) have been demonstrated [3]. One major limitation with disc-based systems is that each feature is read serially, and the time-scale of the reading process introduces latency effects.

An alternate optical storage methodology that shows promise is holography. The advantage of holography is that multiple data "pages" can be stored in the same medium simply by altering the angle at which the hologram is recorded in the medium. This is a form of parallel data addressing since an entire page can be recovered at once. Impressive storage densities of 10–40 Gbits/cm^2 have been predicted for such systems, and demonstrations with holographic RAM have indicated faster access times than hard discs for computers [4–6].

The focus of our research is to examine systems that will capitalize on parallel, near-field readout. Our intended application for this system would be analogous with CD-ROM or DVD uses, or any usage that requires multiple replicas of the same data. A target application would be for systems requiring low latency, such as video games. Other applications would be for mass distribution of high content media such as movies, games, and reference materials. Specifically, we postulate a read-only memory system that rivals magnetic storage technologies for capacity/density, but utilizing optical readout at a wavelength of $\lambda = 500$ nm. States will be defined by groupings of sub-wavelength surface features, which are identified by reading the signal (in the near-field) transmitted through the structure. The aim is to achieve a density of 16 Gbits/cm^2 in order to approach the same densities as magnetic, while surpassing DVD and Blu-Ray densities. This paper outlines the details of the system in Section 2, and the modeling technique used is introduced in Section 3. Results from initial simulations, discussion, and conclusions are then presented in Sections 4–6.

2. System description

2.1. Intended implementation

The proposed system consists of a series of wavelength-scale nanostructures (cells), with designed sub-wavelength features which will be distinguished using near-field optical measurements. The features are formed by introducing dielectric inhomogeneity. For the purposes of this study, the system is considered in two dimensions. The entire system consists of a row of cells on a flat surface, of which three cells are shown in Fig. 1 (in the third dimension, multiple rows would be added). The entire system is illuminated from above by an optical source at a wavelength of $\lambda = 500$ nm. Sensors directly underneath each cell record optical power measurements in parallel. These measurements are used to identify the profile of the sub-wavelength features, which is then mapped to the stream of bits it is meant to represent. Since the entire system is illuminated at once, the system does not require moving components and the data stored in the system is read in parallel at the physical level.

Each cell is one wavelength wide, although the entire system will consist of many such cells on

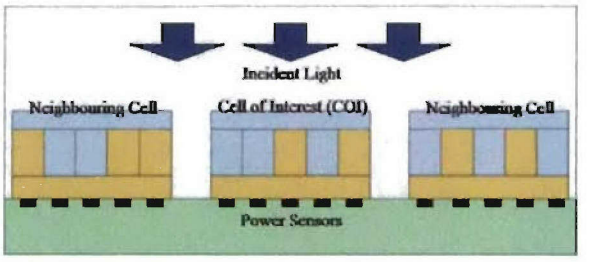

Fig. 1. System layout.

something similar to a compact disc. For the purposes of analysis, we focus on one cell of interest (COI) and two neighbouring cells. This allows us to analyze the uncertainty of the measurements introduced by the proximity of neighbouring cells via crosstalk phenomena. All the cells in a system have the same dimensions as described by the set of parameters indicated in Fig. 2. Three horizontal layers, namely the fill, feature, and base layers, make up a cell. The fill and base layers are homogeneous materials. The feature layer is divided into different sections allowing for either the fill or base material to be present, creating a non-homogeneous layer and therefore the desired sub-wavelength features.

The system is intended to be used as a ROM device, whose low latency and high density may justify the cost premium associated with silicon fabrication processes. In fact, we anticipate that the surface relief structures can be stamped, and thus may benefit from the economies of scale possible from mass production technologies such as for CD-ROMs and DVDs.

We consider that the fabrication of the detector layer will be asynchronous with that of the feature layers, and as such there will have to be a calibration procedure to tune for alignment tolerance. For instance, a set of calibration elements would be fabricated at higher tolerances and analyzed. Each storage "disc" would contain a few fiducial elements, and the readings from these compared to the calibrated standards would allow us to select the correct decoder to extract the stored data on the rest of the "disc." The detectors themselves will not require individual sense amps, but rather we anticipate a design that would be a matrix-addressed photo-conductor array requiring signal conditioning only for each row and column.

2.2. Distinguishing memory states and capacity

The current research focuses on determining the capacity for a two-dimensional nanostructured system with the aim to map as many features per COI as possible (n_f). Thus, for example, with $n_f = 5$ features per cell there are $2^5 = 32$ different possible COI configurations (states). Parameterization was performed to find systems that allow maximal distinguishability amongst the states. Parameters that were varied include: the feature height, the base height, the gap size, and the index of refraction of the base and feature layers.

Different states are distinguished using near-field optical power measurements from sensors located directly underneath each COI. Each sensor produces a power measurement, P_i, and each cell has an array of n_f power sensors resulting in a power vector, $P = (P_1, P_2, \ldots, P_{n_f})$. By altering the configuration of features for a given COI, there are 2^{n_f} possible power vectors that may be used to represent different states. Note that we do not seek a one-to-one mapping between features and representative bits: any state vector may be used to represent any bit stream we wish. In spite of the desired scatter from features within a cell, there will also be undesirable scattering from adjacent cell structures (neighbours) which will introduce crosstalk. This crosstalk will manifest itself in uncertainty in the measurement of the state vector for a given cell. This uncertainty may mean that some state vectors could become indistinguishable and, hence, they may not be used. Thus, the loss of distinguishability reduces the capacity of that system parameterization. Obviously the goal is to find the system parameterization that provides the maximum capacity. This is accomplished by finding all the state vectors that are distinguishable from all other state vectors.

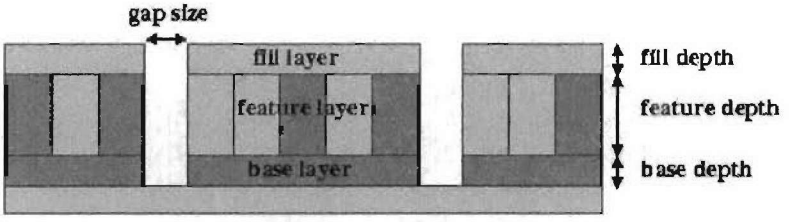

Fig. 2. Geometric parameters of the system.

It should be repeated that the capacity is limited only by the number of distinguishable patterns that can exist in the multi-dimensional space of the power vector. This allows us to take advantage of many other degrees of freedom such as source wavelength, feature heights, etc. For example, the height of the feature above each detector may be four-tiered instead of two-tiered, giving the possibility of $4^5 = 1024$ possible states, and hence the potential for more than one bit per detector.

3. Modeling methodology

3.1. Methodology

To investigate different parameterizations of our proposed system, it was necessary to simulate different systems with changing physical dimensions and dielectric properties. To meet our goal of 16 Gbits/cm^2, a system with 5–7 features/cell is required. Even at the lower limit of 5 features/cell, to obtain an exhaustive characterization (to examine cross-talk for a single system) would require us to simulate each possible COI configuration ($2^5 = 32$), and for every possible neighbour configuration ($2^{10} = 1024$) requires approximately 32,000 simulations. Fully investigating several different systems quickly becomes prohibitively time-consuming.

In the first stage of our methodology, COIs are analyzed and parameterized in isolation. Then full sets of simulation for systems with 3 features/cell (512 simulations/system) were performed to see if we could extrapolate what fraction of the five-feature simulations would be necessary for reasonable estimates of capacity. This analysis method is described in further detail in Section 4. A further benefit of this three-feature analysis is that it allows us to visualize the clustering of power measurements in three-dimensional visual mappings.

3.2. Modeling tool

The modeling tool must be well-developed and able to handle complex geometries, generating reliable results so that we can concentrate on the analysis and statistics. For these reasons a two-dimensional (TM polarization) finite-difference time-domain (FDTD) method was chosen due to its flexibility in materials and geometries. It is a reliable and time-proven method, and is particularly well suited for heterogeneous structures. Maxwell's equations in differential form are discretized in time and space, and explicit expressions determine the electric and magnetic fields at future time steps based solely on their values at previous time steps and on the local electromagnetic properties of materials. More details can be found in [7].

The simulation space was discretized at a resolution of $\Delta x = \Delta z = 5$ nm (100 cells per free space wavelength) which – because of stability conditions – results in a time discretization of $\Delta t = 11.55$ as. The absorbing boundary condition at the edge of the simulation space was a L2TDLM (see [8]) of 11 layers, and the simulations were allowed to progress for 3000 timesteps – approximately 20 cycles of the source.

The optical source was a continuous wave Gaussian beam with TM polarization and a wavelength of $\lambda = 500$ nm. For simulation purposes, the beam is generated by invoking the equivalence principle at a line lying along the waist of the beam (see, for example, the total/scattered field formulation in [7]). The waist of the beam was three wavelengths in width in our simulations, and was approximately two wavelengths above the structure. Observation tasks consisted of producing time histories of the electric and magnetic fields at each of the sensor locations, from which the Poynting vector and the associated power measurements were obtained in a post-processing phase. Note that reported power measures are in arbitrary units, which will scale accordingly as the fields are scaled. The source used is identical for all simulations reported, so that comparisons can be made. In this sense, we are simulating that one detector lies directly beneath each feature in the COI (i.e. one detector per feature), and that the fill-factor of the detector is 100% – that is, we assume that each detector completely covers the plane beneath each feature.

3.3. Structures modeled

Two possibilities for systems were envisaged: one based on polymers that can be easily stamped to

create the features, and one based on silicon for which lithographic processes will probably be necessary to create the features. To account for these two cases in the FDTD simulations, only two refractive index profiles were used. For the former, the base and feature layers had an index of refraction of $n_{\text{polymer}} = 1.8$; and for the latter the base and feature layers had an index of refraction of $n_{\text{silicon}} = 3.5$. In both cases a fill layer with an index of refraction of $n_{\text{fill}} = 1.2$ was used, as well as a substrate with an index of refraction of $n_{\text{substrate}} = 3.5$.

4. Analysis and results

To analyze capacity, it is necessary to determine the number of distinguishable states in a given system. For a given COI, reflecting and diffracting fields from neighbouring cells interact to produce interference in the COI power measurement. If neighbouring cells were infinitely far away, then there would be no interference and no ambiguity in the assignment of a particular state to a particular power measure (Fig. 3(a)). When neighbouring cells are close, the effect of the interference is to cause a slight shift in the measured power vector (Fig. 3(b)), and uncertainty is introduced in assignment of the state since different neighbour configurations will affect it to differing degrees. When the regions of uncertainties for two different states are close, there is a greater chance for mis-assignment of a state. The goal of our analysis was to establish how many states for a given system are distinguishable, and hence determine the system capacity.

Analysis consisted of an initial analysis to determine systems that provide maximum separation of state vectors in isolation (i.e. no crosstalk). Good candidates were then rigorously analyzed with neighbouring cells, to determine measures of uncertainty and thus capacity.

4.1. Initial analysis

In the first stage, COIs were analyzed in isolation to get a initial estimation of which parameters provide the greatest distinguishability in state vectors. Because there are no neighbours, all state vectors are unique. Consequently, the aim was to identify parameterizations that provide maximum separation (Euclidean) distance between state vectors. Intuitively we should expect that systems with more separation will provide better distinguishability.

To understand the metrics we define for evaluating systems, first consider a set of n power vectors, with an individual power vector being described by

$$\mathbf{P}_i = (P_{i1}, P_{i2}, \ldots, P_{in_f}), \quad 1 \leqslant i \leqslant n.$$

Then we define the vector describing the centroid of the distribution

$$\mathbf{P}_c = \frac{1}{n}\left(\sum_{i=1}^{n} P_{i1}, \sum_{i=1}^{n} P_{i2}, \ldots, \sum_{i=1}^{n} P_{in_f}\right).$$

The distance from one vector to another vector is (in the Euclidean sense) $d_{ij} = |\mathbf{P}_i - \mathbf{P}_j|$, and the distance from a vector to the centroid is $d_{ic} = |\mathbf{P}_i - \mathbf{P}_c|$.

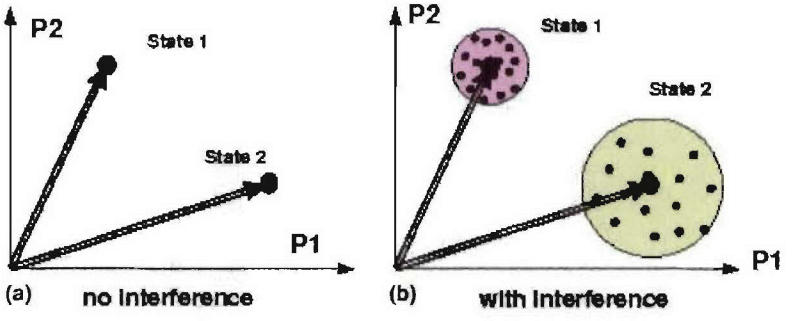

Fig. 3. The effect of interference from neighbouring COIs for a simplified system (two features). In (a), for COIs in isolation (i.e. neighbouring cells at infinity) would exhibit no uncertainty in identifying two different states – power measure would be exact. In (b), neighbouring cells are close, and for a given COI state, changes in the neighbouring cells will cause interference and hence uncertainty in the state power measurement.

The average distance from the centre is then given by $\tilde{d}_c = \frac{1}{n}\sum_i d_{ic}$. We define several variables to be able to assess the merit of a system:

$diststd = \frac{1}{n-1}\sqrt{\sum_{i=1}^{n}(d_{ic} - \tilde{d}_c)^2}$ the standard deviation in distance of vectors from the average vector location,

$avgpdist = \langle d_{ij}\rangle = \langle|\mathbf{P}_i - \mathbf{P}_j|\rangle$ the average in distance between vectors,

$pdiststd = \sqrt{\langle d_{ij}^2\rangle} = \sqrt{\langle|\mathbf{P}_i - \mathbf{P}_j|^2\rangle}$ the standard deviation in distance between vectors.

The average and standard deviation in power for each detector were also identified and compared graphically. Again, if the power measurements were distributed more widely throughout five-space, we would expect better distinguishability.

For the geometry of the structures in the simulations, the fill depth was kept constant at $d_{\text{fill}} = 100$ nm, and the width of each feature was always $w_{\text{feature}} = 100$ nm. For parameterization, the height of the features is varied between $h = 50$ and $h = 250$ nm, and the depth of the base layer is varied between $d = 10$ and $d = 500$ nm.

Results are found in Table 1. In general, the higher refractive index contrast provides better separation between state vectors. This can be seen better by observing Fig. 4. The power at each detector for all 32 configurations is plotted as points, the mean power at each detector is plotted as the solid line, and error bars indicate two standard deviations in the measured power from the mean at each detector. Examples of poor systems and good systems are shown for both cases of refractive index profiles. Though there are probably several choices of systems that could be made for more detailed studies, we will focus on two – one $n = 1.8$ (polymer) and one $n = 3.5$ (silicon) – that show the best promise. Systems are designated with the refractive index n, base depth d, and feature height h, as $nXXdYYhZZ$. Using this naming convention, the chosen best polymer system is $n1.8d500h250$ since it excels in all three measures compared to the others. The chosen best silicon system is $n3.5d10h150$; although the $n3.5d500h150$ is similar in measures, we have opted to look at the former because it would require less material for fabrication. The chosen systems are indicated in Fig. 5, and are the basis for all reported results from this point on.

4.2. Distinguishable states and capacity measures

To determine how well candidate systems perform in the presence of crosstalk (i.e. with neighbouring cells), we first had to define metrics for this crosstalk. With a full set of simulations for every possible COI and neighbour configuration, this allows us to identify distinguishable states, and therefore capacity.

For the purposes of this study, we designate a region of uncertainty for a given COI by a power vector cloud, where all of the possible interference conditions lie. This region is then designated by a centre C and radius R, as demonstrated in Fig. 6 for an example two-feature system. The centre is defined by the arithmetic mean of all the possible vectors in the region, and the radius is defined as the scalar distance between this centre and the furthest outlying measurement. For the three-feature system, regions are described by spheres, and beyond three features by hyperspheres of the appropriate dimension.

For a COI to be distinguishable, it is necessary that its region of uncertainty be distinct from all other regions for all other COIs, otherwise those states can not be used. This allows us to determine the capacity of the system. As an example, in Fig. 7 a two-feature system is shown with the regions of uncertainty for each of the four possible COI configurations. In this case, two of the states are indistinguishable from each other in the presence of crosstalk and therefore can not be used. Thus this system has a storage capacity of 2 states/COI (1 bit/COI).

In point of fact, this method of determining capacity assumes that if two states are overlapping, then they are both discarded. In reality, we only need discard one of the two. The choice of which state to discard is complicated however, because this depends on its relationship with all of the other states. The choice of which states to keep and which to discard is actually an NP-hard problem. For instance, even with 32 possible states for our COI, there are 2^{32} possible subsets of states

Table 1
Initial analysis of several systems in isolation

Refractive index of base layer	Depth of base d (nm)	Height of feature h (nm)	*diststd*	*avgpdist*	*pdiststd*
1.8		50	10.85	120.18	30.55
	10	150	46.16	286.96	86.70
		250	150.54	577.22	223.86
		50	11.47	80.55	22.71
	100	150	120.28	397.07	171.30
		250	186.69	598.97	264.03
		50	4.80	139.74	70.47
	500	150	137.94	426.24	210.88
		250	198.38	777.97	352.26
3.5		50	58.71	211.09	85.31
	10	150	155.07	848.95	275.54
		250	247.19	583.26	310.94
		50	156.19	469.51	230.78
	100	150	230.63	753.23	330.48
		250	153.69	627.51	228.06
		50	133.01	560.83	233.23
	500	150	206.51	836.29	285.81
		250	253.70	782.23	342.44

that would need investigating. We are currently investigating optimization and clique-finding methods for this NP-hard problem, to be reported at a future date.

For our system to rival current magnetic storage densities, we require at least 5 features per wavelength. As mentioned before, an exhaustive set of simulations for one such system would require over 30,000 simulations, which with the given choice of simulation parameters and the available computer resources would take several months to complete. In order to reduce this requirement, we assume that a random subset of neighbour configurations will allow us to quantify the system – the centres, radii, and hence capacity – within a certain measure of error. In order to determine what fraction of the full set is necessary, we performed simulations on the candidate systems with three features instead of five (512 instead of over 30,000 simulations). This then gives us a full set to compare measures against as we increase the fraction of neighbours used. Our intention is then to use this knowledge to estimate the fraction necessary for five-feature simulations.

The capacity (and density) will be affected by the spacing between cells (the gap size); as the spacing gets larger and larger, we should expect better and better distinguishability, and hence better capacity. However, the density (capacity per linear measure) should peak at some optimal spacing. We wished to make this spacing a constant in all future simulations, and so before proceeding with full simulations for capacity, we did some nominal simulations with a very small subset of neighbours, but increasing gap size to get the density trend. Fig. 8 shows the results of these initial simulations as the gap size increases from 50 to 500 nm. We see that the turnover point for capacity vs. density lies in approximately the 250–400 nm, depending on whether we are using the silicon or polymer system. In order to aim for maximum density, we thus choose the 250 nm measure as our gap size for all future simulations.

4.3. Three-feature measures

Full sets of simulations were performed for the two candidate systems indicated in Fig. 5, which

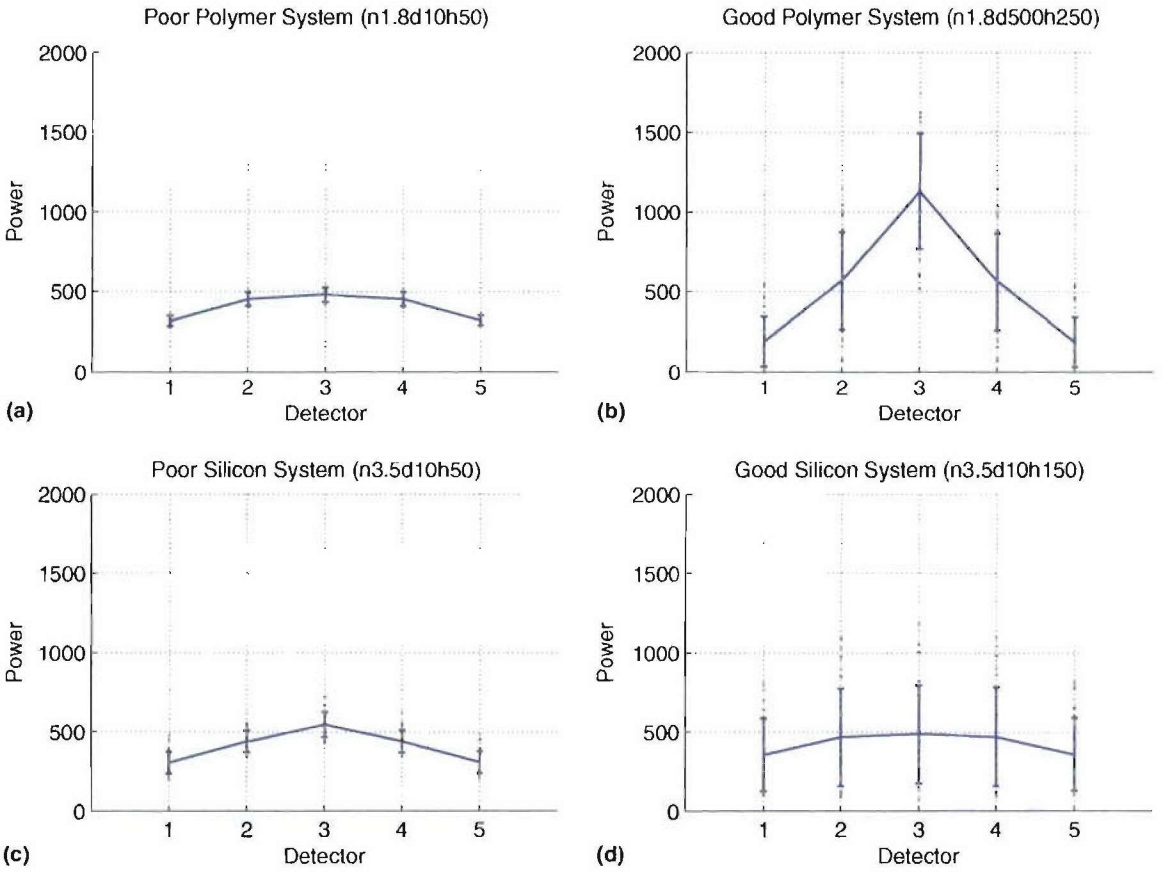

Fig. 4. Example candidate systems in polymer ((a) and (b)) and silicon ((c) and (d)). The power at each detector for all 32 configurations is plotted as points; mean power at each detector is plotted as a solid line; error bars indicate two standard deviations in the measured power from the mean. Clearly there is a marked difference in the separation of power measures for poor and good candidates.

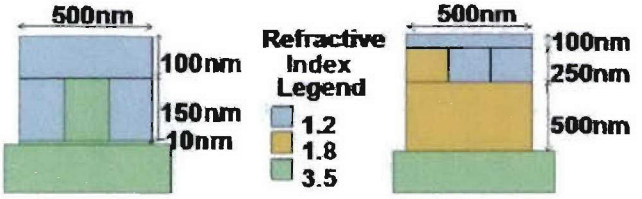

Fig. 5. Candidate system parameters in both silicon (left) and polymer (right) that show promise for good capacity.

included the COI along with two neighbouring cells. The cells were separated from each other by 250 nm. Other modeling parameters were as discussed in Section 3. The main output of these full sets was a complete set of power vectors for each and every possible configuration of COI and neighbour.

The full set of power vectors determined our comparison benchmark for the actual centres and radii of the power vector clouds. We then performed statistical sub-samplings at certain fractions of the full set of neighbours (in this case, 64 is a full set), and determined the centre and radii with these random sub-samples. This was

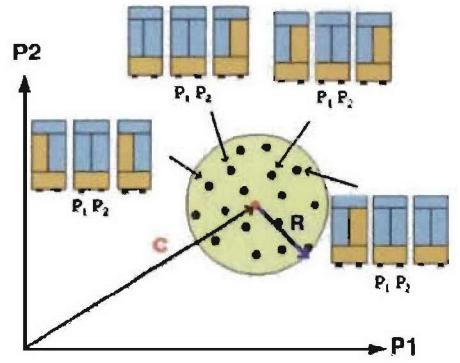

Fig. 6. For a given COI, as the neighbour configuration changes, the measured power vector P will vary slightly. The variance is represented by an uncertainty cloud with a centre C and a radius R.

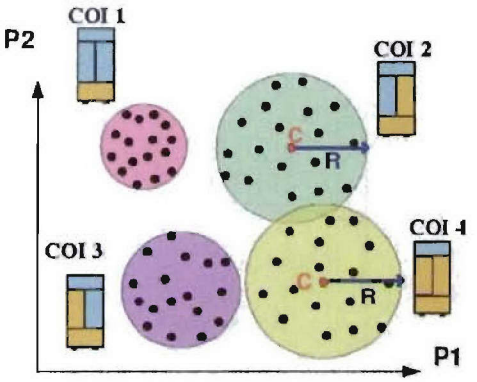

Fig. 7. A two-element system, indicating a graph of the power vectors. In this case, COIs #2 and #4 overlap, and so are indistinguishable from each other.

performed 100 times at each fraction ranging from 3-of-64 to 64-of-64, and at each fraction the RMS error of the 100 random sub-samples (relative to the actual radius from the full set) was found for the centre and radius. The results of this fractional randomization are displayed in Figs. 9 and 10.

Note that, as expected, as the fraction taken increases, the RMS error decreases from around 50% to zero at the full fraction. The error in both measures show a similar trend in this regard, although there is less error in the estimate of the centre. The utility in such a graph comes from attempting to infer what a necessary fraction is for a tolerable level of error. For instance, if an error of 10% is allowable in this measure, then these

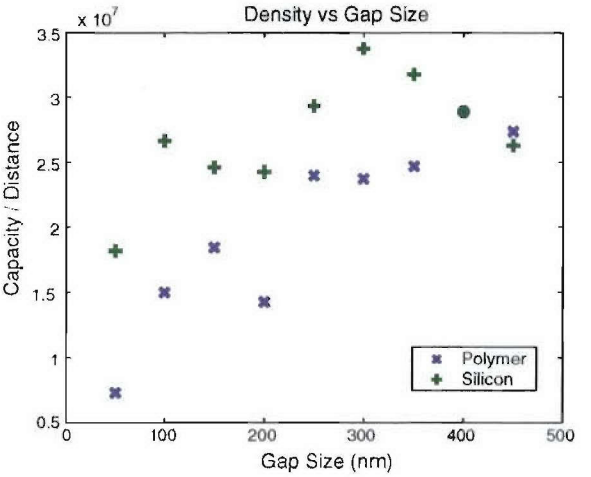

Fig. 8. Graph showing the trend in density as the spacing between cells increases, for both polymer and silicon systems.

graphs would indicate that a fraction of approximately 40–60% of the full set is required. This inference means that, in the five-feature case, that we should expect to have to perform simulations for every COI and a random subset of around 400–600 of the 1024 possible neighbour configurations. Unfortunately this is not the substantial reduction in the simulation time that we were hoping for.

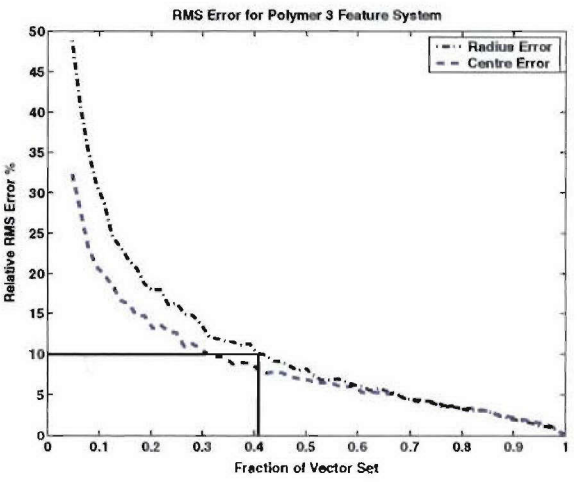

Fig. 9. The RMS error in the centre and radius for the three-feature polymer system.

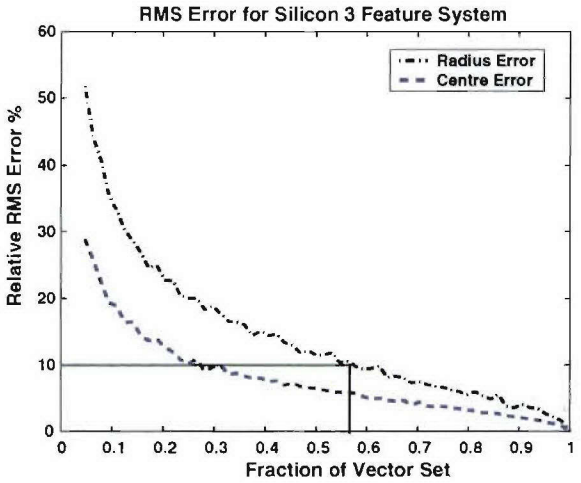

Fig. 10. The RMS error in the centre and radius for the three-feature silicon system.

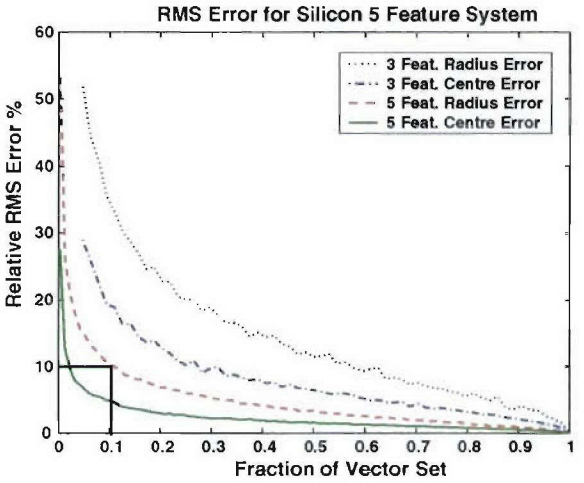

Fig. 11. The RMS error in the centre and radius for the five-feature silicon system, overlaid on the silicon three-feature system previously shown in Fig. 10.

4.4. Five-feature measures

In order to verify our hypothesis about whether the inference of the necessary fraction from the three-feature studies can be carried over to the five-feature case (and perhaps to the n-feature case), we completed one exhaustive full set of simulations for the five-feature silicon case. Simulation parameters were identical to the three-feature case, except that the cell of one wavelength in size was divided in five instead of three.

Using this full set of power vectors, we then performed the same statistical sub-sampling as previously mentioned, over the range of fractions spanning from 0 to 1. The results of this are shown in Fig. 11, overlaid on the previous results for the three-feature in silicon. It is interesting to note that in this case the error starts at the same levels, but decreases much more rapidly than in the three-feature case. In this case, if an error of 10% is allowable we only need to take a fraction of approximately 0.1, corresponding to a substantial reduction in simulation time.

Since we had a full set of data, we were then able to find our estimate of the number of distinguishable states, and hence the capacity. For this particular system, the capacity was determined to be 13 of 32 states, using the algorithm described at the beginning of this section. The final check in the methodology was to see at what fraction the capacity is well predicted. Fig. 12 shows the estimated capacity as the fraction of neighbours taken is increased. From this graph, it appears that the capacity can be estimated well at a fraction of around 0.2. From the previous graph, this indicates that an error of around 7% is tolerable for our estimates of capacity.

4.5. Ellipsoid algorithm

One of the benefits of performing the three-feature simulations is that we are able to visualize the distribution of the state vectors in three-dimensional space (a luxury not afforded to us for greater levels of features). For instance, the distribution of the states (the end-point of their vectors) for one particular system are indicated in Fig. 13. From this figure, one can see that the distributions for particular COIs tend to be quite elongated in particular directions. This indicates that spheres are not necessarily the best way of representing the distributions, as they tend to over-predict the volume occupied by the distribution. It would seem that ellipsoids could be better predictors of the occupied volume of states, and using these ellipsoids would provide better estimations of the capacities of systems.

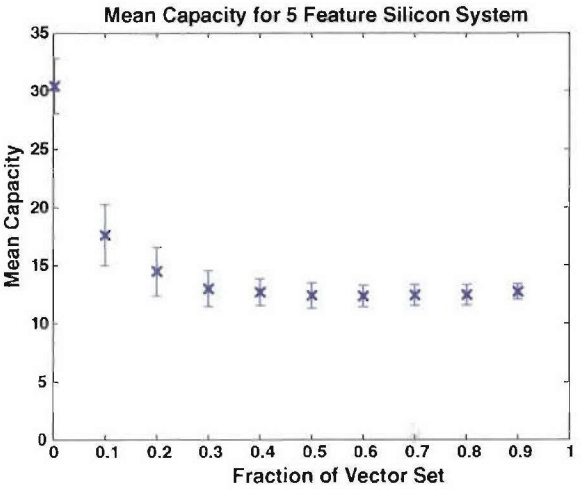

Fig. 12. The estimated capacity as a function of the fraction of the neighbours taken. At each fraction, 100 random neighbours were chosen and the capacity calculated. The error bars indicate one standard deviation on either side of the mean.

To test this methodology, we developed an algorithm which defines the enclosing ellipsoid for a given COI's distribution as follows. First, the centre of the distribution is found. We then find the furthest outlying state point, and define the vector between the centre and this point as the major axis of the ellipsoid. All state points are then projected onto the plane perpendicular to this axis. The next largest axis of the ellipse is then defined in a similar manner, by finding the furthest outlier on this plane. The process is repeated until all axes are found – three for the three-feature case, or five for the five-feature case (thus defining a hyperellipsoid).

From these defining ellipsoids, we can then determine the overlap between two different states. Because the ellipsoids provide a better indication of the states' defining volumes, we expect the capacity estimates for systems to increase from previous estimates done with defining spheres. For example, in Figs. 14 and 15 the same state distributions are shown with defining spheres and defining ellipsoids. Clearly there is interference in the former case, but no interference in the latter. Using the ellipsoids to define the state volumes, we performed capacity analyses on the previous three systems. The results of these measures are tabulated in Table 2, where we can see that the estimated capacity of the five-feature silicon system has nearly doubled to having 27 of 32 states distinguishable.

5. Discussion

Using the analyses in the previous sections, we were able to use the knowledge from the three-feature systems to establish the number of simulations needed for higher orders of features. This has allowed us to reduce the simulation times necessary for investigating other systems. Using one full set of simulations for a five-feature system, we were able to show how the necessary fraction of neighbour simulations decreases rapidly as the number of features increases. More importantly, the use of ellipsoids rather than spheres enables a more accurate method of distinguishing states, and can dramatically improve the predictions on capacity.

To compare the performance of this alternative system with existing technologies, it was necessary to estimate the achievable system density. From the simulations, it is possible to calculate a linear density, where we have to take into account the spacing of 250 nm between cells. Assuming that the same linear density can be achieved in the other planar dimension, an areal density can be determined. With this assumption, the maximum achieved density is predicted to be 3.17 GBits/cm^2 for the silicon five-feature system. Although this is much less than our system goal, it is still five times better than current DVD densities.

Initial simulations with other degrees of freedom provide some evidence indicating even higher densities are achievable. By allowing each feature to take on more than two possible levels, we can create even more possible states. For instance, a five-feature, four-level system would have 1024 possible configurations of the COI (rather than the 32 for the two-level system). A larger percentage of states will be indistinguishable, but it could still be that, for instance, 128 are distinguishable – a capacity of 7 bits/λ. We have performed a very preliminary analysis of such a system – using the silicon template that has been used throughout

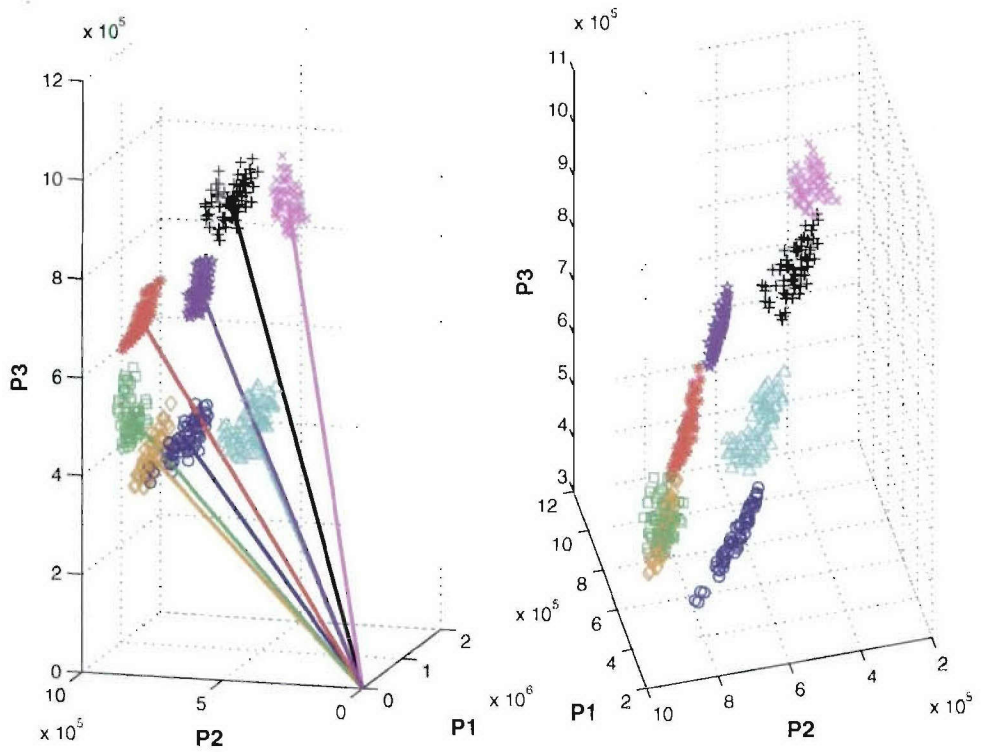

Fig. 13. The figure on the left shows the plot of the state vector distributions, for every COI, for the silicon three-feature system, along with the vector indicating the centre of the distribution. The figure on the right is an expanded view, without the centre vectors, demonstrating how the distributions are, in general, elongated in shape.

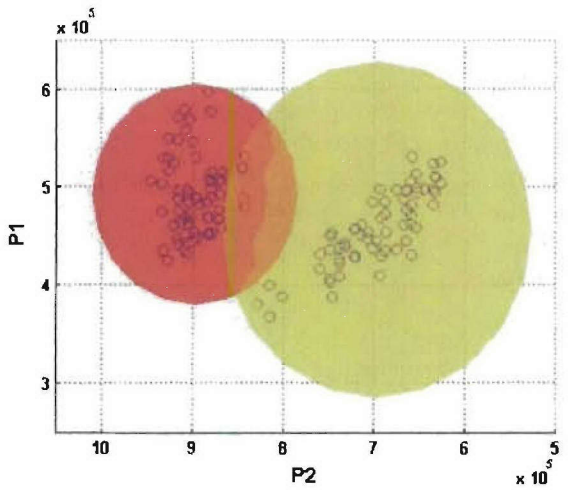

Fig. 14. The distribution for two COIs of a system, and the spheres that have been calculated to indicate the extent of the distribution. Note that in this case, to determine capacity, both of these states would be discarded because of the overlap in their spheres.

the last section – with a very small number of neighbour COIs. Using methods that over-estimate the uncertainty sphere radii, and hence provide a conservative estimate of the capacity, we find that at least 54 of the 1024 states are distinguishable. This would provide roughly double the linear density of the best system discussed so far in this paper.

Although larger densities will be required to surpass current magnetic storage technology, there are still a many possibilities to explore to increase the capacity of the proposed nanostructure optical data storage method. For instance, other parameters can be exploited in order to distinguish states, such as the wavelength and polarization of the source. These other degrees of freedom could feasibly dramatically increase the capacity of the system, but are left for future research.

Besides utilizing methods to increase the degrees of freedom, there are many avenues to be

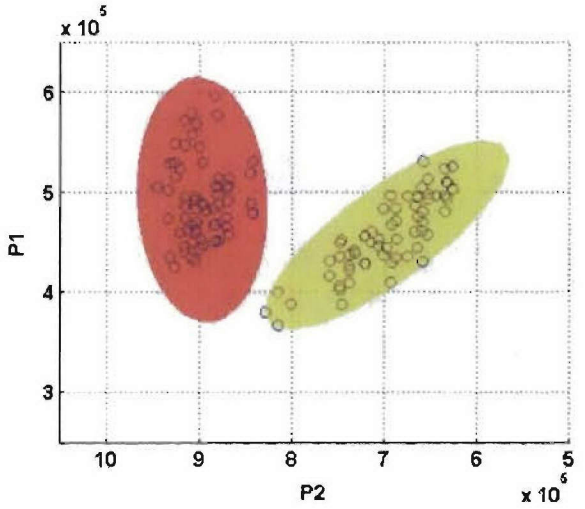

Fig. 15. The same two distributions as in Fig. 14, except with calculated ellipsoids rather than spheres to represent their distributions. In the capacity measure for this system, these two states would now be valid (distinguishable) because there is no overlap.

Table 2
A comparison of the capacity of the three different systems, using spheres or ellipsoids to define the state distributions

Capacity (# states/λ)	Three-feature		Five-feature
	Polymer	Silicon	Silicon
Using spheres	3 of 8	4 of 8	13 of 32
Using ellipsoids	6 of 8	8 of 8	27 of 32

explored for improving the implementation of the system. For instance, the spatially distributed near-field power data could be converted to spectral data for measurement in the far-field. One could imagine utilizing a set of patterned array of organic fluorophores as the detector level for just such a thing. The relative spectral intensities of the fluorophores measured in the far-field would then provide the data measure for a COI.

6. Conclusion

A read-only memory system was proposed that uses near-field optical power measurements to distinguish between designed deformities in nanoscale structured cells. Differing deformities or features affect the power measurements allowing for different states to be stored and read. The capacity (and hence the density) of a given system is governed by the crosstalk between cells, causing some states to be indistinguishable.

Measurements and methods of analysis were developed using simple simulations of three-feature systems, from which inferences were made on the fraction of simulations necessary at higher orders of features. Trends showed that as the number of features increases, the fraction of simulations necessary to get an accurate estimate of capacity decreases. In the example of a five-feature system that was presented, approximately 20% of the full set of simulations was enough to provide an accurate estimate of the capacity. Simulations for three-feature systems also provided insight on the shape of the power distributions, which allowed better estimates of capacity and hence density. Densities of 3.17 GBits/cm^2 have been estimated that greatly surpass current optical storage techniques.

Although the densities of magnetic storage have not yet been achieved, the optical system shows great potential. The third dimension is the main factor that provides hopes for larger capacity with an increased number of degrees of freedom, such as different modes of polarization. Another possibility is investigating the effects of the wavelength of the source to provide additional data to distinguish the different states. Future work will simulate the system in three dimensions to investigate these additional avenues available to increase capacity.

Acknowledgements

This work was supported through the generosity of the Natural Sciences and Engineering Research Council (NSERC) of Canada, and supported in part by AFOSR under contract number F49620-01-1-0363 #5. We would also like to express our gratitude to the anonymous reviewer for their comments on implementation, and in

particular for the use of space-to-spectrum conversion for far field sensing.

References

[1] L. Kong, L. Zhuang, M. Li, B. Cui, S.Y. Chou, Jpn. J. Appl. Phys. Part I 37 (1998) 5973.
[2] S.Y Chou, P.R. Krauss, W. Zhang, L. Guo, L. Zhuang, J. Vac. Sci. Technol. B 15 (1997) 2897.
[3] P.R. Krauss, S.Y. Chou, Appl. Phys. Lett. 71 (1997) 3174.
[4] G. Zhou, X. An, A. Pu, D. Psaltis, F. Mok, in: Proc. 1999 Advanced Optical Data Storage: Materials, Systems, and Interfaces to Computers – SPIE 3802 (1999) 2–5.
[5] W. Liu, D. Psaltis, 2000 IEEE Aerospace Conf. Proc. (2000) 391–397.
[6] G. Zhou, D. Psaltis, F. Mok, Opt. Quantum Electron. 32 (2003) 405.
[7] A. Taflove, Computational Electordynamics: the Finite-difference Time-domain Method, Artech House, Boston, MA, 1995.
[8] D.C. Wittwer, R.W. Ziolkowski, IEEE Trans. Antenn. Propag. 48 (2000) 192.

Appendix 2

Nonlinear Coding for Fiber Communications

Nonlinear Codes for Dispersive Nonlinear Fibers

Jaime A. Anguita, *Student Member, IEEE,* and Mark A. Neifeld, *Member, IEEE*

***Abstract*—We consider a single wavelength channel in a nonlinear dispersive single-mode fiber. We simulate the propagation of binary sequences using ON–OFF keying return-to-zero modulation at a bit rate of 10 Gb/s, under different dispersion and input power conditions. We determine the impact that dispersion and nonlinearity have on the pulse sequences by measuring the Euclidian distances (EDs) among all pairs of received sequences. We then model the set of EDs as a fully connected graph and by means of a clique-finding algorithm we search for subsets of sequences that form codes matched to this nonlinear channel, resulting in significant bit-error-rate improvement.**

***Index Terms*—Dispersive nonlinear fiber, fiber-optic communications, forward error correction, nonlinear codes.**

I. Introduction

GROUP-VELOCITY dispersion (GVD) is known to limit the performance of fiber-optic communication systems by causing intersymbol interference (ISI) that increases with propagation distance [1]. Moreover, in single-mode fibers, the pulse propagation is power-dependent due to Kerr nonlinearity. This nonlinearity interacts with dispersion, further increasing the pulsewidth under normal dispersion and also producing spectral broadening [2]. In dispersion-compensated systems, intrachannel pulse interaction causes distortion and jitter that increases with power and propagation length [3].

In this letter, we present a numerical analysis of the effects that GVD and nonlinearity have on the Euclidian distances (EDs) among pulse sequences at the receiver in a single channel of a wavelength-division-multiplexing (WDM) system. We observe that nonlinearity and dispersion induce a significant decrease in the minimum ED while the overall ED distributions are not considerably changed. We describe a graph theory-based methodology with which we obtain codes that are well matched to this nonlinear fiber channel, giving a significant reduction in bit-error rate (BER) compared to linear codes of the same length.

II. Description of the Simulations

The propagation of optical pulses is governed by the well-known nonlinear Schrödinger (NLS) equation [4]

$$i\frac{\partial A}{\partial z} + i\frac{\alpha}{2}A - \frac{\beta_2}{2}\frac{\partial^2 A}{\partial \tau^2} - i\frac{\beta_3}{6}\frac{\partial^3 A}{\partial \tau^3} + \gamma|A|^2 A = 0 \qquad (1)$$

where A is the slowly varying amplitude of the pulse, z is the propagation distance, and τ is the time in a reference frame that travels at the pulse group velocity. GVD, third-order dispersion, nonlinearity, and losses are represented by the parameters β_2, β_3, γ, and α, respectively. In (1), we have not included higher order nonlinear terms like those that account for stimulated Raman and Brillouin scattering, as they are negligible within the range of power and pulsewidth we consider in this letter. To solve the NLS equation given in (1), we use the symmetrized split-step Fourier method [5], [6] with constant step size.

We simulate the propagation of 10-bit sequences using Gaussian pulses and ON–OFF keying return-to-zero modulation at a bit rate of 10 Gb/s, along a fiber of length 50 km. The sequences comprise equally spaced bit slots, with each pulse centered in its bit slot. Also, sequences are padded with 2.5 bit slots of zeros at the beginning and at the end to avoid energy spilling at the boundaries. Sequences are necessary to account for the memory in the channel, manifested through the broadening and distortion of the pulses, which can extend to several adjacent bit slots. We find that using 10 bits is enough to accurately account for the spreading of the pulses, while maintaining a reasonable processing time, as the complexity of our procedure grows exponentially with the number of bits. The power P_0 applied to the pulses ranges from 1 to 100 mW.

The GVD parameter β_2 is varied from 2 to 30 ps^2/km, in the normal dispersion regime and we fix the third-order dispersion at $\beta_3 = 0.1$ ps^3/km. The nonlinear parameter γ is set to a fixed value of 1.3 W^{-1}km^{-1}. We assume no fiber losses in order to clearly show the effects of nonlinearity in the ED measurements, although we do consider losses when evaluating the codes as we want to use them in a realistic environment.

In order to analyze the impact of the pulsewidth for a given data rate, four different duty cycles are studied, namely, 25%, 33%, 40%, and 50%. We define duty cycle as the full-width at half-maximum (FWHM) of the pulse intensity over the bit slot width. For a given peak power P_0, the average power applied to a pulse clearly depends on the duty cycle. Therefore, to make a fair comparison of performance among different duty cycles, we apply a power penalty by scaling the measured ED by the average intensity of each pulse. This scaling produces a normalized ED (NED) that also allows us to compare the effect of nonlinearity at different input power levels, namely 1, 10, and 100 mW. For our system, we assume incoherent and direct detection. Hence, ED is computed in the intensity domain

$$\mathrm{ED}_{p,q} = \sqrt{\sum_i (I_{p,i} - I_{q,i})^2} \qquad (2)$$

where $I_{p,i}$ and $I_{q,i}$ are the measured intensities for sequences p and q, respectively, at the discrete time step i. The summation is over all samples in the sequence.

Manuscript received March 11, 2004; revised July 27, 2004. This work was supported in part by AFOSR under Contract Number F49620-01-1-0363.

J. A. Anguita is with the Electrical and Computer Engineering Department, University of Arizona, Tucson, AZ 85721 USA (e-mail: janguita@ece.arizona.edu).

M. A. Neifeld is with the Electrical and Computer Engineering Department, and the Optical Sciences Center, University of Arizona, Tucson, AZ 85721 USA (e-mail: neifeld@ece.arizona.edu).

Digital Object Identifier 10.1109/LPT.2004.836754

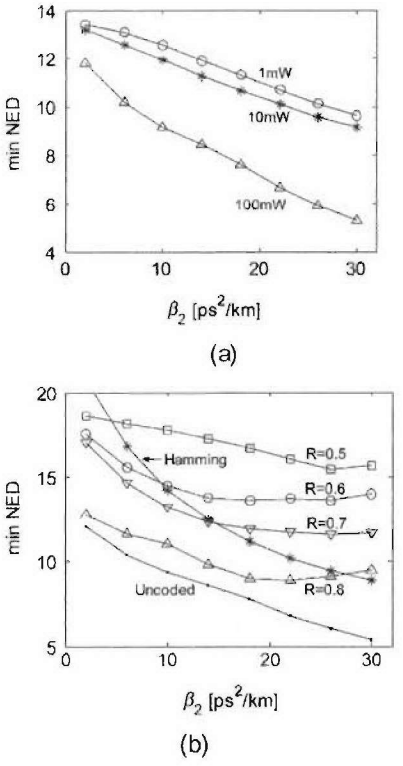

Fig. 1. (a) Minimum NED versus dispersion at $P_0 = 1$, 10, and 100 mW for 50% duty cycle. (b) Minimum NED versus β_2 at $P_0 = 100$ mW for several nonlinear codes using a 50% duty cycle. Hamming (7, 4) is also included for comparison.

Fig. 1(a) shows the minimum NED among pairs of sequences, versus β_2, at an input peak power P_0 of 1, 10, and 100 mW, and a duty cycle of 50%. Within all the duty cycles considered in our simulations, the performance at 50% duty cycle consistently showed to have the largest minimum NED. We observe in Fig. 1(a) that the minimum NED decreases monotonically as dispersion increases. As the power is increased, the minimum NED decreases more rapidly with dispersion, as an effect of nonlinearity. At 100 mW and 30 ps^2/km, the minimum NED has decreased to almost a half of the normalized value found at 1 mW.

III. Nonlinear Codes

For a given set of fiber parameters we obtain an ED map, the set of EDs among all pairs of 10-bit sequences. Within the ED map, we search for subsets of sequences with large minimum distance. This map can be thought of as an interconnected graph where each binary sequence is a node and every ED is an edge. This graph is fully connected, because there is a nonzero distance between every pair of nodes. We would like to remove those edges with an ED smaller than some minimum preferred value. This value will be chosen according to the code rate we want. Since our modulation scheme is binary, we are interested in finding codes with a number of codewords that is a power of two. Since it is not possible to predict the minimum ED required to produce a code with a previously chosen number of codewords, we have to do an iterative search. After removing all edges with ED smaller than the preferred value, we seek the largest subset of nodes that are fully connected, i.e., where there exits an edge between every pair of nodes in the subset. This subset of nodes forms a code. This is known as the clique finding problem [7], [8], and only an exhaustive search can guarantee the optimal result. There is certain symmetry in the ED map that facilitates the search for a good (nonoptimal) solution. This solution will have the best performance under the fiber parameters in which it was found, but, as we will show, it can perform well under other conditions.

TABLE I
Full Set of Codewords of the Code NL1(10, 6)

0000000000	0111100101	1011110110	1110010100
0000000111	1010101001	1011111010	1110100111
0000011001	1010101111	1011111111	1110101110
0000011111	1010110111	1100000011	1110111001
0001011101	1010111000	1100011010	1110111111
0001101101	1010111011	1100101100	1111000011
0001110000	1010111101	1100110011	1111011001
0001110110	1011000000	1100111110	1111011111
0001111010	1011000111	1101011111	1111100000
0001111111	1011001101	1101100001	1111100111
0010110101	1011011000	1101100111	1111101100
0011010011	1011011011	1101101101	1111101111
0101001011	1011011110	1101110110	1111110100
0101110011	1011101011	1101111010	1111110111
0101111101	1011101110	1101111111	1111111011
0111010110	1011110000	1110001101	1111111101

A property of linear codes is that the addition of any two codewords of the code produces a codeword that is part of the code. A nonlinear code is a more general set of codewords that does not have this property. As a consequence, nonlinear codes cannot be generated by a matrix, but instead by a lookup table. In practice, codes with the best distance properties will likely to be of the nonlinear kind due to the nature of this method, and also because, under ISI, ED does not necessarily grow monotonically with binary (Hamming) distance.

Fig. 1(b) shows the minimum NED versus β_2 using a few of these nonlinear codes, for $P_0 = 100$ mW and 50% duty cycle. These codes are 10 bit long and have various numbers of codewords according to their code rates (the rate of a code is $R = \log_2$ (#codewords/10). They are all optimized for $\beta_2 = 30$ ps^2/km. We observe that there is a significant increase of the minimum NED for the entire range of dispersion compared to the minimum NED of the uncoded sequences, also shown in the plot. It is also worthwhile to include the minimum NED performance of a Hamming (7, 4) code, shown by the curve with an x marker in Fig. 1(b), as this shows that the minimum NED of the set of sequences defined by this code decreases significantly as dispersion increases. This implies that the code does not have good performance under ISI. Table I shows the binary codewords of a rate 0.6 nonlinear code [which we label NL1(10, 6)], optimized for $P_0 = 100$ mW, $\beta_2 = 30$ ps^2/km, $L = 50$ km, and 16.7 Gb/s (such that the information rate is 10 Gb/s). Under these fiber conditions, the overall minimum NED, the maximum NED, and the chosen minimum NED to get the code NL1(10, 6) are 0.0526, 1, and 0.3409, respectively (all values normalized to the maximum NED). We measure the BER performance of code NL1(10,6) in Section IV.

IV. BER Performance

Fig. 2 shows the BER performance of the code NL1(10, 6) at a bit rate of 10 Gb/s, with parameters $\beta_2 = 30$ ps^2/km, $\beta_3 = 0.1$ ps^3/km, $P_0 = 100$ mW, $\alpha = 0.2$ dB/km, $\gamma = 1.3$ W^{-1}km^{-1},

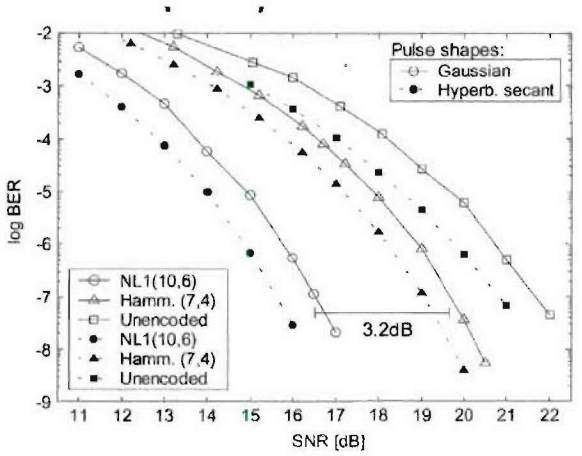

Fig. 2. BER performance of code NL1(10, 6). $L = 50$ km, $P_0 = 100$ mW, $\beta_2 = 30$ ps^2/km, using Gaussian pulses (hollow markers) and hyperbolic secant pulses (filled markers).

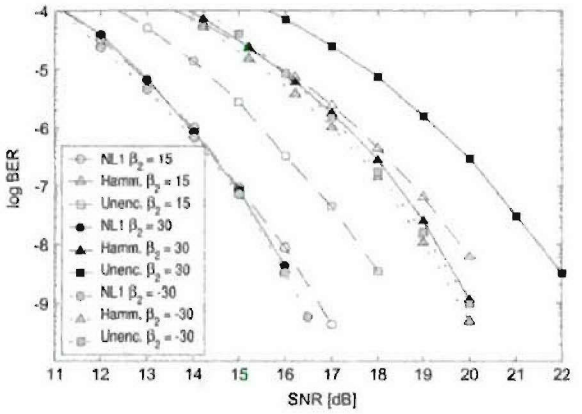

Fig. 3. BER performance of code NL1(10, 6) on different channel conditions, using Gaussian pulses. $P_0 = 20$ mW, $L = 100$ km, $\beta_2 = 15, 30$, and -30 ps^2/km.

and $L = 50$ km, using Gaussian (hollow markers) and hyperbolic secant pulses (filled markers). We assume additive white Gaussian noise and a maximum-likelihood (ML) receiver with a bandwidth of 20 GHz. The performances of a Hamming (7, 4) code and that of the uncoded set are also included. A power penalty has been applied to each curve to account for the different code rates, and the coded sequences were simulated at 16.7 Gb/s so that the information bit rate was maintained at 10 Gb/s. We see in Fig. 2 that at a BER $= 10^{-7}$, the coding gain of this code is about 5.1 dB over the uncoded set and about 3.2 dB over the Hamming (7, 4) code, which has a comparable rate of 0.57.

Fig. 3 shows the BER performance of the code NL1(10, 6) under various other settings. For $P_0 = 20$ mW, $\beta_2 = 15$ ps^2/km, and $L = 100$ km (bit rate and other parameters kept the same as in previous case), the coding gain is 1.5 dB over the uncoded set and about 3.8 dB over the Hamming code, at a BER $= 10^{-7}$ (dashed lines with hollow markers). Note that the Hamming code performance is worse than that of the uncoded sequences. This occurs because the Hamming code cannot overcome the increased ISI that the pulses suffer as they propagate at a higher bit rate (16.7 GB/s) in order to maintain the information bit rate at 10 Gb/s. Under the same fiber parameters but with a dispersion of 30 ps^2/km (continuous lines with black markers), the results of the code NL1(10, 6) are excellent too: At a BER $= 10^{-7}$, the coding gain is 5.8 dB over the uncoded bits, and 3.8 dB over the Hamming code.

This code shows to give good results in the anomalous dispersion regime as well. For $P_0 = 20$ mW and $\beta_2 = -30$ ps^2/km (maintaining all other parameters as in the previous case), the coding gains are 3.4 and 3.6 dB over the Hamming code and the uncoded sequences, respectively (dotted lines with gray markers).

V. Discussion and Conclusion

Using an approach from graph theory, we have found nonlinear codes from the ED maps that outperform linear codes of similar length in both the normal and anomalous dispersion regime, as these nonlinear codes account for the nonlinearity and dispersion present in the optical fiber.

We have determined as a rule of thumb that good performance can be achieved with code NL1(10, 6) if the ratio of dispersion length $L_D = T^2_{\mathrm{FWHM}}/|\beta_2|$ over propagation length L is $L_D/L < 2.0$ approximately. Also, an ML receiver is required. Other codes can be found to provide good performance under lower dispersion and/or bit rates.

These codes can be used as an alternative to dispersion compensation and can be used in WDM systems. This methodology could still be used on systems that already incorporate dispersion compensation, and this is under study.

By operating at high power, like 100 mW, spectral efficiency is reduced by the effects of nonlinearity, as it causes the spectrum to broaden. In practice, spectral broadening is not conspicuous below 40 mW in the presence of typical fiber losses and within the dispersion range in which this code is design to give good performance. Nevertheless, code NL1(10, 6) performs very well at lower power, so it is not essential to incur in a spectral efficiency penalty to obtain good results.

We are working on increasing the code length by extracting the salient features of the nonlinear codes and forming linear codes that contain these features.

References

[1] R. J. Nuyts, Y. K. Park, and P. Gallion, "Performance improvement of 10 Gb/s standard fiber transmission systems by using the SPM effect in the dispersion compensating fiber," *IEEE Photon. Technol. Lett.*, vol. 8, pp. 1406–1408, Oct. 1996.

[2] M. Potasek, G. Agrawal, and S. Pinault, "Analytic and numerical study of pulse broadening in nonlinear dispersive fibers," *J. Opt. Soc. Amer. B*, vol. 3, pp. 205–211, 1986.

[3] A. Mecozzi, C. Clausen, and M. Shtaif, "Analysis of intrachannel nonlinear effects in highly dispersed optical pulse transmission," *IEEE Photon. Technol. Lett.*, vol. 12, pp. 392–394, Apr. 2000.

[4] G. P. Agrawal, *Nonlinear Fiber Optics*. San Diego, CA: Academic, 2001.

[5] ——, *Fiber-Optic Communication Systems*. New York: Wiley, 1997.

[6] T. R. Taha and M. J. Ablowitz, "Analytical and numerical aspects of certain nonlinear evolution equations," *J. Comput. Phys.*, vol. 55, pp. 203–230, 1984.

[7] M. R. Garey and D. S. Johnson, "Complexity of near-optimal graph coloring," *J. ACM*, vol. 23, no. 1, pp. 43–49, 1976.

[8] C. Bron and J. Kerbosh, "Algorithm 457. Finding all cliques of an undirected graph," *Commun. ACM*, vol. 16, no. 9, pp. 575–577, 1973.

Appendix 3

Information Propagation in a Fast-Light Medium

The speed of information in a 'fast-light' optical medium

Michael D. Stenner[1], Daniel J. Gauthier[1] & Mark A. Neifeld[2]

[1]*Duke University, Department of Physics, and The Fitzpatrick Center for Photonics and Communication Systems, Durham, North Carolina 27708, USA*
[2]*Department of Electrical and Computer Engineering, The Optical Sciences Center, University of Arizona, Tucson, Arizona 85721, USA*

One consequence of the special theory of relativity is that no signal can cause an effect outside the source light cone, the space-time surface on which light rays emanate from the source[1]. Violation of this principle of relativistic causality leads to paradoxes, such as that of an effect preceding its cause[2]. Recent experiments on optical pulse propagation in so-called 'fast-light' media—which are characterized by a wave group velocity v_g exceeding the vacuum speed of light c or taking on negative values[3]—have led to renewed debate about the definition of the information velocity v_i. One view is that $v_i = v_g$ (ref. 4), which would violate causality, while another is that $v_i = c$ in all situations[5], which would preserve causality. Here we find that the time to detect information propagating through a fast-light medium is slightly longer than the time required to detect the same information travelling through a vacuum, even though v_g in the medium vastly exceeds c. Our observations are therefore consistent with relativistic causality and help to resolve the controversies surrounding superluminal pulse propagation.

The speed of a light pulse travelling through an optical material is not precisely defined, because any pulse comprises a collection of elementary sinusoidal waveforms, each with a distinct frequency ω. Each constituent sinusoid travels at a well-defined velocity known as the phase velocity $v_p = c/n(\omega)$, where $n(\omega)$ is the refractive index of the optical material. Approximate theories of optical pulse propagation predict that the peak travels at the group velocity $v_g = c/(n + \omega dn/d\omega|_{\omega=\omega_o}) = c/n_g$, where n_g is the group index and ω_o is the central frequency of the wavepacket[6].

We refer to the quantity $dn/d\omega$ as the dispersion of an optical material. For typical optical materials, there exist narrow spectral regions where $n(\omega)$ is a decreasing function of frequency (that is, $dn/d\omega < 0$), resulting in a condition known as anomalous dispersion[7]. When ω_o is within such a region, n_g can be less than one and can even become negative when the anomalous dispersion is large. This results in 'fast light', for which it is possible that the peak of a light pulse may exit the optical material before it passes through the entrance face[8]. The amount of fast-light pulse advancement is largest when v_g is negative and near zero (n_g large and negative).

The possibility of superluminal group velocities ($v_g > c$ or $v_g < 0$) was such a concern to researchers around 1910 that several conference sessions were devoted to the topic[9]. Based on these discussions, Sommerfeld demonstrated theoretically that the velocity of the front of a square-shaped pulse propagating through any medium is identically equal to c and hence relativistic causality is preserved[10]. In a follow-up study, Brillouin suggested that the group velocity is not physically meaningful when the dispersion is anomalous because the pulse becomes severely distorted[11]. More recent research investigating the propagation of smooth-shaped pulses has shown that this conclusion is not justified, leading to renewed controversy[8,12–20].

Another outcome of the discussions in the early 1900s, as recounted in the preface and first chapter of the book by Brillouin[9], was a reformulation of the fundamental postulate of the special theory of relativity. This reformulation states that, rather than limiting the speed of an 'object', it is the information velocity v_i that is limited by c. Unfortunately, there is no agreed-upon definition of the information velocity[2].

In our experiment, we use a fast-light medium that exploits the spectral region of anomalous dispersion between two closely spaced amplifying resonances[15,20] realized by creating large atomic coherence[21] in a laser-driven potassium vapour[22], as shown in Fig. 1a. We obtain larger pulse advancement for a smooth gaussian-shaped pulse, as shown in Fig. 1b, in comparison to the experiment of ref. 15, by increasing the gain and hence the size of the anomalous dispersion. The larger advancement relative to the pulse width obtained in our experiment makes it easier to distinguish the different velocities describing pulse propagation. From this data, we infer that $n_g = -19.6 \pm 0.8$, indicating that we are operating in the highly superluminal regime.

Measuring v_i requires an understanding of the fundamental mechanism for information encoding and detection. Garrison *et al.*[2] propose that new information is encoded on an optical pulse by creating a point that is non-analytic (for example, a discontinuity in the pulse amplitude or its derivatives) and that

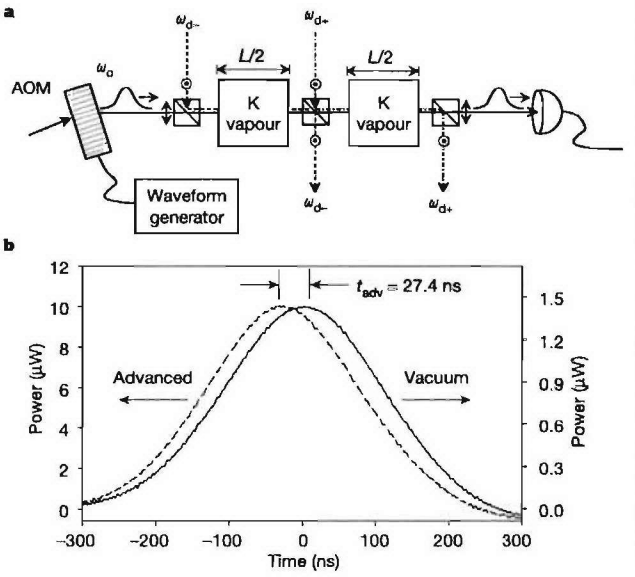

Figure 1 Fast-light pulse propagation. **a**, Experimental set-up. The potassium vapours are contained in two uncoated Pyrex cells of length $L/2 = 20$ cm (to suppress unwanted parametric instabilities[22]) and heated to obtain an atomic number density of 4.5×10^{11} atoms cm^{-3}. Linearly polarized coherence-preparation laser beams (frequencies ω_d and ω_{d+}) are combined with the linear and orthogonally polarized pulses using polarizing beam splitters. The pulses are detected by an avalanche photodiode with a 25 kHz–125 MHz bandwidth. The coherence preparation beams are adjusted with ω_d set at 1.36 GHz to the high-frequency side of the centre of the potassium $4P_{1/2} \leftrightarrow 4P_{1/2}$ transition and $\omega_{d+} - \omega_{d-} = 23$ MHz, chosen to optimize the pulse advancement using procedures similar to those discussed in refs 15, 18 and 22. The pulses are generated by passing a continuous-wave laser beam through an acousto-optic modulator (AOM) driven by a computer-controlled arbitrary waveform generator. The time origin has been set arbitrarily to coincide with the peak of this pulse. **b**, The solid line shows the temporal evolution of a 263.4-ns-long (full-width at half-maximum) pulse propagating through the cells when the lasers are tuned far from the atomic resonance and hence the vapour-cell portion of the path is equivalent to vacuum. The dashed line shows the observed fast-light pulse advancement for a smooth pulse shape when the coherence-preparation laser is tuned near the atomic resonance and ω_o is set between the gain resonances. The peak of the pulse is advanced by $t_{adv} = 27.4$ ns $\pm$ 1.1 ns, corresponding to a relative pulse advancement of 10.4%. Using $t_{adv} = L/c - L/v_g$ with $L/c = 1.3$ ns, we find $v_g/c = -0.051 \pm 0.002$. Careful inspection of the fast-light pulse reveals that it has been compressed by 1.9%, which is due primarily to the frequency dependence of the gain[18,19].

letters to nature

this point always travels at c regardless of the value of the other velocities associated with the pulse[2,5]. Essentially, they have generalized Sommerfeld's concept of the front velocity to a non-analytic point of the pulse amplitude, where the front of a square-shaped pulse is an example of a point of non-analyticity. These workers[2,5] suggest that the point of non-analyticity is the only part of the pulse representing new information because measurements of the early part of the pulse cannot be used to predict anything about the part of the pulse arriving after the point of non-analyticity, and hence v_i equals the speed of a point of non-analyticity. For counter-views, see refs 2, 4, 23 and 24. We note that some aspects of their proposal have been verified using electronic circuits where no propagating waves are involved, and hence only issues of causality, but not relativistic causality, can be tested[25–27].

To enhance our ability to estimate the location of this non-analytic point in the presence of noise, we use two optical pulses that are initially identically gaussian-shaped, which allows us smoothly to turn on the pulse amplitude to a level above the noise floor of our detection electronics and to monitor the fast-light pulse advancement. Near the peak of the gaussian function and at the same moment for both symbols, we switch the amplitude of the gaussian function to a high (1) or low (0) value for the remainder of the pulse. The moment when a decision is made to switch between the symbols corresponds to the point of non-analyticity. Note that this transition is smoothed out by the finite response time of the optical switch.

The location of the point of non-analyticity is determined by detecting the arrival of new information using a receiver that can distinguish between symbols to a desired level of certainty, characterized by the bit error rate (BER). Before the arrival of the point of non-analyticity at the detector, we expect no detected information, corresponding to a BER of 1/2. Once the point of non-analyticity propagates past the detector, the received information will grow smoothly from zero and the BER drops. A symbol is considered to be detected when the BER falls below some threshold. Hence, the detection time of information is later than the time when information is first available at the detector, even for pulses propagating through vacuum. This detection latency Δt depends on the characteristics of the medium through which the pulses propagate, the shape of the symbols, the detection algorithm, noise[24] in the detection process (including quantum noise[28,29]), and the BER threshold. It increases as the signal-to-noise ratio decreases because it takes longer for the receiver to achieve the same BER. Achieving the limit $\Delta t \to 0$ requires the use of optimal symbol shapes and detection algorithms, and infinite energy in the optical pulse so that the signal-to-noise ratio of the detected waveform approaches infinity. Although it is possible to estimate Δt for a specific experimental apparatus, it cannot be measured directly because it requires measuring the point of non-analyticity. A crucial aspect of our experiment is to make Δt as small as possible and to make it as similar as possible for both vacuum and advanced pulses.

Figure 2 shows the propagation of both symbols through the fast-light medium and vacuum. From a simple visual inspection of the

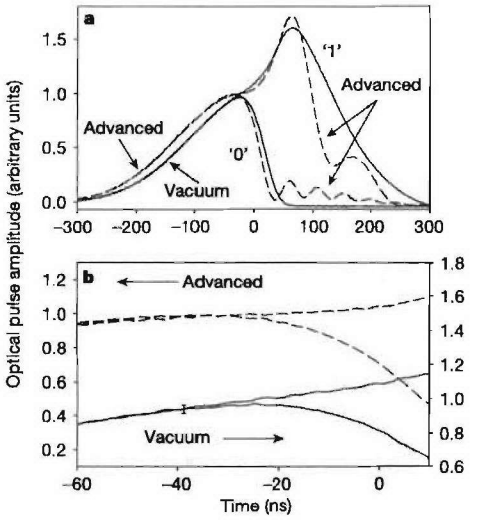

Figure 2 Transmitting information-encoded optical pulses through a fast-light medium. **a**, Transmitting '0' and '1' through the fast-light medium (dashed line) and vacuum (solid line). Each symbol is transmitted separately through the medium and vacuum, where each curve is an average of 50 pulses. **b**, High-resolution plot of part of **a** with an offset for clarity. The amplitude of the advanced and vacuum pulses have been scaled so that their heights would be the same if a gaussian pulse propagated through the system, as in Fig. 1b. The error bar indicates the typical standard deviation of the pulse amplitudes. From Fig. 2a, it is seen that the fast-light medium advances the early part of the pulses during the smooth turn on, identically to that observed for the full gaussian-shaped pulses shown in Fig. 1b. Most important is the observation that both symbols are the same for early times so that it is not possible to distinguish between them. Hence, no information can yet be conveyed to a receiving party at the end of the communication channel.

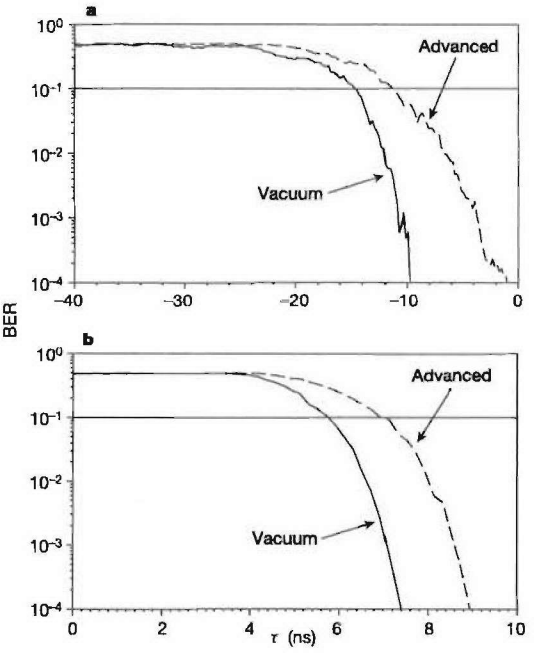

Figure 3 Detecting the arrival of new information. Shown is the BER as a function of the upper limit of the integration time τ for the vacuum (solid line) and advanced (dashed line) pulses. The horizontal line indicates the detection threshold. **a**, Experimental observations. The time origin has been selected arbitrarily. **b**, Theoretical predictions based on Maxwell's equations. The time origin corresponds to the moment when the point of non-analyticity first arrives at the detector. The BER is determined using the following matched-filter procedure. For each symbol (0 and 1), there are 50 pulse waveforms. Reference waveforms for each symbol, denoted by $R_0(t)$ and $R_1(t)$, are generated by averaging 49 pulse waveforms for each symbol and the integral $D(\tau) = I_0(\tau) - I_1(\tau)$ is determined for each pulse, where the waveform being detected is not included in the respective reference waveform. Here: $I_j(\tau) = \int_{t_s}^{t_s+\tau} x(t)R_j(t)\mathrm{d}t/\alpha_j(\tau_\alpha)N_j(\tau)$ (for $j = 0$, 1), where $x(t)$ is an individual pulse waveform, t_s is an integration start time chosen arbitrarily at a time in the early beginning of the pulse, $N_j(\tau) = \int_{t_s}^{t_s+\tau} R_j^2(t)\mathrm{d}t$, $\alpha_j(\tau_\alpha) = \int_{t_s}^{t_s+\tau_\alpha} x^2(t)\mathrm{d}t/N_j(\tau_\alpha)$, and τ_α is a normalization integration time chosen arbitrarily on the leading edge of the gaussian pulse before the point where the symbols separate. For each symbol type, the probability density of D is fitted to a gaussian distribution normalized to an area of 1/2 and the overlapping area of these two gaussian distributions is the BER.

data, we see that the time where it is possible to first distinguish between the two symbols for the advanced pulses is slightly later than the time where it is possible to first distinguish between the same two symbols propagating through vacuum. In addition, it is seen that the manner in which the average symbol waveforms separate for the vacuum and advanced case are only slightly different, so that the detection latency times (denoted by Δt_{vac} and Δt_{adv} respectively) should be similar.

To quantify our results, we determine the BER for the vacuum (Fig. 3a, solid line) and advanced (dashed line) pulse pairs using an integrate-and-dump matched filter technique. The BER is high for final observation times in the range between -40 and -25 ns, during which the pulse amplitudes are large (see Fig. 2a). Hence, even though the signal-to-noise ratio for a single pulse is high at these times, the pulses are not yet distinguishable and no information is detected. Placing the detection threshold at BER = 0.1, chosen to keep Δt_{vac} and Δt_{adv} small, we determine the detection time for vacuum (advanced) pulse pairs T_{vac} (T_{adv}) and the difference in detection times $T_i = T_{adv} - T_{vac}$. The time difference is approximately constant for BER values around 0.1; its average value in the range of BERs between 0.08 and 0.2 is equal to 3.2 ± 1.5 ns. Our observations demonstrate that the information detection time for pulses propagating through the fast-light medium is longer than the detection time for the same information propagating through vacuum, even though the group velocity is in the highly superluminal regime for the fast-light medium.

From our direct measurement of T_i, we do not know whether an observed difference between the detection times is due to changes in the detection latencies or differences in the information velocities for vacuum ($v_{i,vac}$) and the fast-light medium ($v_{i,adv}$). The relation among these quantities is given by:

$$T_i = (L/v_{i,adv} - L/v_{i,vac}) + (\Delta_{radv} - \Delta_{rvac}) \quad (1)$$

To gain some insight about the importance of detection latency, we analyse a mathematical model, based on Maxwell's equations, that describes approximately the generation, propagation, and detection of our symbols. Consistent with previous research[2,5], this model predicts that $v_{i,adv} = v_{i,vac} = c$, and hence T_i is completely determined by $(\Delta t_{adv} - \Delta t_{vac})$. Using the same matched-filtering approach, we determine the predicted BER as shown in Fig. 3b. We see that information is detected later for the advanced pulses than for the vacuum pulses, qualitatively similar to the experimental observations. We find that $T_i = 1.5 \pm 0.5$ ns, where the error only accounts for statistical uncertainty in the BER determination. The fact that $T_i \neq 0$ demonstrates that subtle changes in the shape of the symbols after information has been encoded give rise to substantial changes in the detection latency. The predicted value of T_i might be smaller than the observed time owing to our assumption that the fast-light medium does not change the noise properties of the optical pulses[28].

Using the model prediction for $(\Delta t_{adv} - \Delta t_{vac})$ in equation (1) and taking $v_{i,vac} = c$, we find that $v_{i,adv} = (0.4 \pm 0.5)c$. Thus, our observations are consistent with the special theory of relativity even for a medium where v_g is highly superluminal, demonstrating that the peak of the advanced pulse at the exit face of the medium (see Fig. 1b) is not causally connected to the peak at the entrance face[2].

Because our analysis makes no assumptions about the sources of noise in the encoding, transmission and decoding process, our general experimental approach and conclusions should hold even in the limit where quantum fluctuations are dominant[3,28,29]. □

Received 11 July; accepted 20 August 2003; doi:10.1038/nature02016.

1. Jackson, J. D. *Classical Electrodynamics*, 3rd edn Sec. 11.3.C (John Wiley & Sons, New York, 1999).
2. Garrison, J. C., Mitchell, M. W., Chiao, R. Y. & Bolda, E. L. Superluminal signals: causal loop paradoxes revisited. *Phys. Lett. A* **245,** 19–25 (1998).
3. Boyd, R. W. & Gauthier, D. J. in *Progress in Optics* (ed. Wolf, E.) Vol. 43, 497–530, Ch. 6 (Elsevier, Amsterdam, 2002).
4. Nimtz, G. & Gaibel, A. Basics of superluminal signals. *Ann. Phys.* **11,** 163–171 (2002).
5. Chiao, R. Y. & Steinberg, A. M. in *Progress in Optics* (ed. Wolf, E.) Vol. 37, 345–405, Ch. 6 (Elsevier Science, Amsterdam, 1997).
6. Agrawal, G. P. *Nonlinear Fiber Optics* 3rd edn, Ch. 1 (Academic, San Diego, 2001).
7. Born, M. & Wolf, E. *Principles of Optics*, 7th edn, Sec. 2.3.4 (Cambridge Univ. Press, Cambridge, 1999).
8. Garrett, C. G. B. & McCumber, D. E. Propagation of a Gaussian light pulse through an anomalous dispersion medium. *Phys. Rev. A.* **1,** 305–313 (1970).
9. Brillouin, L. *Wave Propagation and Group Velocity* Ch. 1 (Academic, New York, 1960).
10. Sommerfeld, A. Über die fortpflanzung des lichtes in dispergierenden medien. *Ann. Phys.* **44,** 177–202 (1914); English translation available in Brillouin, L. *Wave Propagation and Group Velocity* Ch. II (Academic, New York, 1960).
11. Brillouin, L. Über die fortpflanzung des lichtes in dispergierenden medien. *Ann. Phys.* **44,** 203–240 (1914); English translation available in Brillouin, L. *Wave Propagation and Group Velocity* Ch. III (Academic, New York, 1960).
12. Chu, C. & Wong, S. Linear pulse propagation in an absorbing medium. *Phys. Rev. Lett.* **48,** 738–741 (1982).
13. Ségard, B. & Macke, B. Observation of negative velocity pulse propagation. *Phys. Lett.* **109,** 213–216 (1985).
14. Peatross, J., Glasgow, S. A. & Ware, M. Average energy flow of optical pulses in dispersive media. *Phys. Rev. Lett.* **84,** 2370–2373 (2000).
15. Wang, L. J., Kuzmich, A. & Dogariu, A. Gain-assisted superluminal light propagation. *Nature* **406,** 277–279 (2000).
16. Akulshin, A. M., Cimmino, A. & Opat, G. I. Negative group velocity of a light pulse in cesium vapor. *Quantum Electron.* **32,** 567–569 (2002).
17. Bigelow, M. S., Lepeshkin, N. N. & Boyd, R. W. Superluminal and slow light propagation in a room-temperature solid. *Science* **301,** 200–202 (2003).
18. Cao, H., Dogariu, A. & Wang, L. J. Negative group delay and pulse compression in superluminal pulse propagation. *IEEE J. Sel. Top. Quantum Electron.* **9,** 52–58 (2003).
19. Macke, B. & Ségard, B. Propagation of light-pulses at a negative group velocity. *Eur. Phys. J. D* **23,** 125–141 (2003).
20. Steinberg, A. M. & Chiao, R. Y. Dispersionless, highly superluminal propagation in a medium with a gain doublet. *Phys. Rev. A.* **49,** 2071–2075 (1994).
21. Harris, S. E. Electromagnetically induced transparency. *Phys. Today* **50,** 36–42 (1997).
22. Stenner, M. D. & Gauthier, D. J. Pump-beam-instability limits to Raman-gain-doublet "fast-light" pulse propagation. *Phys. Rev. A.* **67,** 063801 (2003).
23. Diener, G. Superluminal group velocities and information transfer. *Phys. Lett. A* **223,** 327–331 (1996).
24. Wynne, K. Causality and the nature of information. *Opt. Commun.* **209,** 85–100 (2002).
25. Mitchell, M. W. & Chiao, R. Y. Negative group delay and "fronts" in a causal system: An experiment with very low frequency bandpass filters. *Phys. Lett. A* **230,** 133–138 (1999).
26. Nakanishi, T., Sugiyama, K. & Kitano, K. Demonstration of negative group delays in a simple electronic circuit. *Am. J. Phys.* **70,** 1117–1121 (2002).
27. Solli, D., Chiao, R. Y. & Hickmann, J. M. Superluminal effects and negative group delays in electronics, and their applications. *Phys. Rev. E* **66,** 056601 (2002).
28. Kuzmich, A., Dogariu, A., Wang, L. J., Milonni, P. W. & Chiao, R. Y. Signal velocity, causality, and quantum noise in superluminal light pulse propagation. *Phys. Rev. Lett.* **86,** 3925–3929 (2001).
29. Caves, C. M. & Drummond, P. Quantum limits on bosonic communication rates. *Rev. Mod. Phys.* **66,** 481–537 (1994).

Acknowledgements M.D.S. and D.J.G. acknowledge discussions with M. Gehm and J. Thomas, the loan of an argon-ion pump laser from J. Thomas, and the financial support of the US National Science Foundation.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to D.J.G. (gauthier@phy.duke.edu).

Q2

Appendix 4

Information Propagation in a Slow-Light Medium

Fast (but causal) information trasmission in an optical material with a slow group velocity

Michael D. Stenner,[1] Daniel J. Gauthier,[1*] Mark A. Neifeld[2]

[1]Duke University, Department of Physics, and The Fitzpatrick Center for Photonics and Communications Systems, Durham, North Carolina, 27708 USA

[2]Department of Electrical and Computer Engineering, The Optical Sciences Center, University of Arizona, Tucson, Arizona, 85721 USA

*To whom correspondence should be addressed; E-mail: gauthier@phy.duke.edu.

We demonstrate transmission of information encoded on optical pulses at a velocity 60 times faster than the group velocity v_g, which describes the speed of the peak of the pulses. Our experiments are conducted using a laser-driven potassium vapor with $v_g \approx 0.01c$, where c is the speed of light in vacuum. We measure the velocity of information v_i by transmitting different pulse shapes through the vapor and measuring the time required to distinguish the different shapes. Our observation that $v_i \gg v_g$ directly contradicts the conventional wisdom that $v_i = v_g$ when $0 < v_g < c$.

One of the most fundamental characteristics of a pulse of electromagnetic radiation is the speed at which it propagates. In vacuum, the pulse travels at speed c and does not change its shape. On the other hand, a pulse traveling through a dispersive material

(such as an optical fiber or a gas of atoms) experiences some degree of pulse reshaping due to the frequency dependence of the refractive index $n(\omega)$. The reshaping arises from the fact that a pulse is a coherent superposition of component sinusoidal waves (*1*), each of which travels through the medium at a different speed known as the *phase velocity* $v_p = c/n(\omega)$.

In general, there is no simple way to completely characterize the speed of a pulse due to the reshaping effect. To address this problem, it is customary to introduce several characteristic velocities that describe different aspects of the reshaping process. For example, it is well-known that the peak of the pulse travels approximately at the *group velocity* $v_g = c/(n + \omega dn/d\omega|_{\omega=\omega_0})$, where ω_0 is the central frequency of the pulse (*1, 2*). This velocity describes the lowest-order effect of the reshaping process: a translation in space-time of the pulse in comparison to the same pulse traveling in vacuum.

Recent research has shown that it is possible to make dramatic changes in the group velocity of a gas of atoms by driving it with an intense laser beam, thereby creating large atomic coherence, which in turn induces large changes in $dn/d\omega$ (*3*). For example, by inducing large normal dispersion (positive $dn/d\omega$) or anomalous dispersion (negative $dn/d\omega$), it is possible to slow light to less than 20 m/s (*4*), stop or freeze light (*5–7*), or even make superluminal pulses (*8, 9*).

Superluminal pulses, with $v_g > c$ or $v_g < 0$, have created considerable controversy because their existence might suggest that information can travel faster than c, which would violate relativistic causality (*10*). This controversy has been addressed in a recent experiment where it was found that the information velocity in fast-light media is indeed less than c (*11*).

In contrast, there has been little research to directly measure the information velocity in a normally-dispersive, slow-light medium. This situation is perhaps understand-

able: many textbooks on electromagnetism demonstrate theoretically that, in a normally-dispersive material, v_g is approximately equal to the *signal velocity* v_s as defined by Sommerfeld and Brillouin nearly a century ago (*1*, *2*). The name "signal velocity" implies that it is related to the transmission of information, a concept that is echoed in many texts and research articles (*12*, *13*). As a result, it is commonly believed that $v_i = v_s \approx v_g$ for normally-dispersive media.

Chiao and collaborators offer a different view. They show that v_s can exceed c, suggesting that v_s cannot be the true velocity of information (*10*). Also, the intensity threshold used to determine v_s is arbitrary (*2*), allowing for different values of v_s for the same pulse. As an alternative, they propose that information is contained only in points of non-analyticity on electromagnetic waveforms, and therefore that v_i is equal to the speed with which these points propagate (*10*, *14*). An example of such a non-analytic point is a discontinuity in the waveform or one of its derivatives. Their proposal is based on the notion that measurements of a waveform on one side of a non-analytic point cannot be used to infer anything about the waveform on the other side of the point. Hence, the arrival of a non-analytic point brings new information. Perhaps the most elegant aspect of their proposal is that it can be viewed as an extension of Sommerfeld and Brillouin's earlier research where they showed that the *front velocity* v_f (the velocity at which the leading edge of a waveform propagates) is always equal to c regardless of the properties of the medium (*2*). In linear media, one can always view a waveform with multiple points of non-analyticity as a superposition of multiple waveforms that are each analytic after an initial front. Therefore, according to the theory of Chiao and collaborators, $v_i = v_f = c$.

In order to help resolve the conflict between Chiao's new theory and the common belief that $v_s = v_i$, we observe the propagation of information in a medium where $v_s \approx v_g \ll c$. We prepare this "slow-light" medium by creating a single amplifying resonance, realized

by creating large atomic coherence in a laser driven potassium vapor (Fig. 1a). In the spectral region of the amplifying resonance, the group velocity is very slow, and pulses injected with center frequencies in this region experience large delays in comparison to identical pulses traveling through vacuum. Figure 1b shows pulses propagating through vacuum (solid line) and through the slow-light medium (dashed line). We find that the peak of the pulse is delayed by $t_{del} = 67.5 \pm 2$ ns, corresponding to a relative pulse delay of 25%. Using $t_{del} = L/v_g - L/c$ with $L/c = 0.66$ ns, we infer that $v_g/c = 0.01 \pm 0.0003$.

To enhance our ability to estimate the location of a point of non-analyticity in the presence of noise, we use two optical pulse shapes (our "symbols") that have identically Gaussian-shaped leading edges. This allows us to smoothly turn on the pulse amplitude to a level above the noise floor of our detection electronics and to monitor the pulse delay. Near the peak of the Gaussian function and at the same moment for both symbols, we switch the amplitude of the Gaussian function to a high ("1") or low ("0") value for the remainder of the pulse. The moment when a decision is made to switch between the symbols corresponds to the point of non-analyticity. Note that this point is smoothed out by the finite rise time of the optical switch.

Figure 2 shows the propagation of both symbols through the slow-light medium and vacuum (2a) along with an enlargement (2b) of the diagram in the vicinity of the transition between the symbols. From Fig. 2a, it is seen that the slow-light medium delays the early part of the pulses during the smooth turn on, identically to that observed for the full Gaussian-shaped pulses shown in Fig. 1b. Most important is the observation that both symbols are the same for early times so that it is not possible to distinguish between them. Hence, no information can yet be conveyed to a receiving party at the end of the communication channel. From a simple visual inspection of the data, we see that the time when it is possible to first distinguish between the two symbols for the delayed pulses is

nearly the same as the time when it is possible to first distinguish between the same two symbols propagating through vacuum.

The arrival time of the information is determined by observing the pulses with a receiver that can distinguish between the two symbols to a particular level of certainty, characterized by the bit-error-rate (BER). Before the arrival of the point of non-analyticity at the detector, we expect no detected information, corresponding to a BER of 1/2. Once the point of non-analyticity propagates past the detector, the BER drops as the received information grows smoothly from zero. A symbol is considered to be detected when the BER falls below some threshold. Hence, the detection time of information is later than the time when information is first available at the detector, even for pulses propagating through vacuum. This detection latency Δt depends on the characteristics of the medium through which the pulses propagate, the shape of the symbols, the detection algorithm, noise in the detection process (including quantum noise (*15, 16*)), and the BER threshold. It increases as the signal-to-noise ratio decreases because it takes longer for the receiver to achieve the same BER. Achieving the limit $\Delta t \to 0$ requires the use of optimal symbol shapes and detection algorithms, and infinite energy in the optical pulse so that the signal-to-noise ratio of the detected waveform approaches infinity. Although it is possible to estimate Δt for a specific experimental apparatus, it cannot be measured directly because it requires measuring the point of non-analyticity. Making Δt very small and very similar for both vacuum and delayed pulses is a primary consideration in our choice of symbol shapes. In Fig. 3b, we see that the manner in which the average symbol waveforms separate for the vacuum and delayed case are only slightly different so that the detection latency times (denoted by Δt_{vac} and Δt_{del}, respectively) should be similar.

To quantify our results, we determine the BER for the vacuum (Fig. 2a. solid line) and delayed (dashed line) pulse pairs using an integrate-and-dump matched filter technique

(*17*). It is seen that the BER is high for final observation times in the range between −40 and −20 ns during which the pulse amplitudes are large but not yet distinct within the noise limit (see Fig. 2a). Even though the signal-to-noise ratio for a single pulse is high at these times, the pulses are not yet distinguishable and therefore no information is detected. Placing the detection threshold at BER = 0.1, chosen to keep Δt_{vac} and Δt_{del} small, we determine the detection time for vacuum (delayed) pulse pairs T_{vac} (T_{del}) and the difference in detection times $T_i = T_{del} - T_{vac}$. The time difference is approximately constant for BER values around 0.1; its average value in the range of BERs between 0.08 and 0.2 is 8.5 ± 0.5 ns.

Based only on a direct measurement of T_i, one cannot determine whether an observed difference between the detection times is due to changes in the detection latencies or differences in the information velocities for vacuum ($v_{i,vac}$) and the slow-light medium ($v_{i,del}$). The relation among these quantities is given by

$$T_i = (L/v_{i,del} - L/v_{i,vac}) + (\Delta t_{del} - \Delta t_{vac}). \qquad (1)$$

To gain some insight about the importance of detection latency, we analyze a mathematical model, based on Maxwell's equations, that describes approximately the generation, propagation, and detection of our symbols. Consistent with previous research, this model predicts that $v_{i,del} = v_{i,vac} = c$ (*2*, *10*, *12*), and hence T_i is completely determined by $(\Delta t_{del} - \Delta t_{vac})$. Using the same matched filtering approach, we determine the predicted BER as shown in Fig. 3b. We see that information is detected later for the delayed pulses than for the vacuum pulses, qualitatively similar to the experimental observations. We find that $T_i = 8 \pm 2$ ns, where the error arises from uncertainty in the similarity between the model and physical experiment. The fact that $T_i \neq 0$ demonstrates that subtle changes in the shape of the symbols after information has been encoded give rise

to substantial changes in the detection latency.

Using the model prediction for $(\Delta t_{del} - \Delta t_{vac})$ in Eq. 1 and taking $v_{i,vac} = c$, our best estimate (*18*) is that $v_{i,del} = 0.6c$, which is 60 times faster than v_g. Thus, even in normally-dispersive slow-light media, information does not propagate at the group velocity. This result is in direct contradiction to the conventional wisdom regarding propagation of information on optical pulses and hopefully helps resolve a century-old confusion. Additionally, this result serves as a demonstration of a new technique for measuring v_i. Future refinements of this technique will allow precise examination of Chiao's proposal that the information velocity is always equal to c.

References and Notes

1. J. D. Jackson, *Electrodynamics, Third Edition* (Wiley, New York, 1999), pp. 322–339.

2. L. Brillouin, *Wave Propagation and Group Velocity* (Academic Press, New York, 1960).

3. R. W. Boyd, D. J. Gauthier, *Progress in Optics, Vol. 43* (Elseivier, Amsterdam, 2002), pp. 497–530.

4. L. V. Hau, S. E. Harris, Z. Dutton, C. H. Behroozi, *Nature* **397**, 594 (1999).

5. A. B. Matsko, *et al.*, *Advances in Atomic, Molecular and Optical Physics* (Academic, San Diego, 2001), pp. 191–242.

6. C. Liu, Z. Dutton, C. H. Behroozi, L. V. Hau, *Nature* **409**, 490 (2001).

7. D. F. Phillips, A. Fleischhauer, A. Mair, R. L. Walsworth, M. D. Lukin, *Phys. Rev. Lett.* **86**, 783 (2001).

8. L. J. Wang, A. Kuzmich, A. Dogariu, *Nature* **406**, 277 (2000).

9. M. S. Bigelow, N. N. Lepeshkin, R. W. Boyd, *Science* **301**, 200 (2003).

10. R. Y. Chiao, A. M. Steinberg, *Progress in Optics XXXVII* (Elsevier, Amsterdam, 1997), pp. 345–405.

11. M. D. Stenner, D. J. Gauthier, M. A. Neifeld, *Nature* (2003). In press.

12. G. Diener, *Phys. Lett. A* **223**, 327 (1996).

13. E. Hecht, *Optics, Fourth Edition* (Addison Wesley, San Francisco, 2002), pp. 299–302.

14. J. C. Garrison, M. W. Mitchell, R. Y. Chiao, E. L. Bolda, *Phys. Lett. A* **245**, 19 (1998).

15. A. Kuzmich, A. Dogariu, L. J. Wang, P. W. Milonni, R. Y. Chiao, *Phys. Rev. Lett.* **86**, 3925 (2001).

16. C. M. Caves, P. Drummond, *Rev. Mod. Phys.* **66**, 481 (1994).

17. Materials and methods are available as supporting material on *Science* Online.

18. Propagating the uncertainty leads to values ranging from $0.2c$ through $\pm\infty$ to $-0.5c$. Recall that negative velocities are very fast in this context.

19. MDS and DJG gratefully acknowledge discussions of this research with Michael Gehm and John Thomas, the loan of an argon-ion pump laser from John Thomas, and the financial support of the U.S. National Science Foundation Grant #PHY-0139991.

Supporting Online Material

www.sciencemag.org

Materials and Methods

Fig. S1

Fig. 1. Slow light pulse propagation. (a) Experimental setup (*17*). (b) Temporal evolution of 265-ns-long (full width at half maximum) Gaussian-shaped pulses propagating through the slow light medium (dashed line) and vacuum (solid line), demonstrating slow-light pulse delay. The time origin has been set arbitrarily to coincide with the peak of the vacuum pulse. Careful inspection of the delayed pulse reveals that it has been expanded to a pulse width of 311 ns (17% expansion), which is due primarily to the frequency dependence of the gain.

Fig. 2. (a) Transmitting "0" and "1" through the slow light medium (dashed line) and vacuum (solid line). Each symbol is transmitted separately through the medium and vacuum, and each curve is an average of 50 pulses. (b) High-resolution plot of (a). The amplitude of the advanced and vacuum pulses have been scaled so that their heights would be the same if a Gaussian pulse propagated through the system, as in Fig. 1b. The error bar indicates the typical standard deviation of the pulse amplitudes.

Fig. 3. Detecting the arrival of new information: the BER as a function of the upper limit of the integration time for the vacuum (solid line) and advanced (dashed line) pulses. The horizontal line indicates the detection threshold. (a) Experimental observations. The time origin is the same as in Fig. 1b. (b) Theoretical predictions based on Maxwell's equations. The time origin corresponds to the moment when the point of non-analyticity first arrives at the detector.

Supporting Online Material

Preparation of the slow-light medium

The potassium vapor is contained in an uncoated pyrex cell of length $L = 20$ cm and heated to obtain an atomic number density of 3.5×10^{12} atoms/cm^3. A linearly-polarized coherence-preparation laser beam (frequency ω_d) is combined with the linear and orthogonally-polarized pulses (center frequency $\omega_0 \approx \omega_d + 462$ MHz) using a polarizing beam splitter. The pulses are generated by passing a continuous-wave laser beam through an acousto-optic modulator (AOM) driven by a computer-controlled arbitrary waveform generator. The pulses are detected by a photoreceiver with a 25 kHz - 125 MHz bandwidth.

The vacuum pulses are observed by tuning the laser very far from the atomic resonance so that neither the coherence-preparation laser beam nor the pulses interact with the atoms. In this configuration, the potassium vapor is equivalent to vacuum.

To observe slow light pulse propagation, the coherence preparation beam is adjusted with ω_d set at 2.61 GHz to the high-frequency side of the center of the potassium $4S_{1/2} \rightarrow 4P_{1/2}$ transition. This creates an amplifying resonance at frequency $\omega_d + 462$ MHz, the ground-state splitting of ^{39}KS. We then set the pulse frequency to the center frequency of the generated amplifying resonance, where the group velocity is minimized. Figure S1 shows the gain exponent and group index as functions of frequency.

BER Calculation

The bit-error-rate (BER) is determined using the following integrate-and-dump matched-filter procedure. For each symbol (0 and 1), there are 50 test pulse waveforms. For each test waveform, two reference waveforms $R_0(t)$ and $R_1(t)$ are generated. Each reference

waveform $R_j(t)$ is an average of 49 pulse waveforms of symbol type j and does not include the test waveform.

For each $R_j(t)$, the normalization integral

$$N_j(\tau) = \int_{t_s}^{t_s+\tau} R_j^2(t)dt \tag{1}$$

is calculated, where t_s is a time chosen arbitrarily on the leading edge of the pulse. Next, the normalization constants

$$\alpha_j(\tau_a) = \frac{1}{N_j(\tau_a)} \int_{t_s}^{t_s+\tau_a} x^2(t)dt \tag{2}$$

are determined, where $x(t)$ is the test pulse waveform and τ_a is a normalization integration time chosen arbitrarily on the leading edge of the Gaussian pulse before the point where the symbols separate. The integral $D(\tau) = I_0(\tau) - I_1(\tau)$ is determined for each test waveform, where

$$I_j(\tau) = \frac{1}{\alpha_j(\tau_a)N_j(\tau)} \int_{t_s}^{t_s+\tau} x(t)R_j(t)dt \qquad (j = 0, 1). \tag{3}$$

Each integral $I_j(\tau)$ is a measure of of the similarity between the test waveform $x(t)$ and the reference waveform $R_j(t)$ over the time interval from t_s to $(t_s + \tau)$.

For each symbol type, the 50 test pulse waveforms are used to generate a probability density of $D(\tau)$, which is then fit to a Gaussian distribution normalized to an area of $1/2$. The overlapping area of these two Gaussian distributions is the BER.

Fig. S1. Gain and group index. (a) The dotted line shows the measured gain exponent (gL) and the solid line shows the Lorentzian least-squares fit. (b) The theoretically predicted group index n_g corresponding to the Lorentzian gain resonance shown in (a).

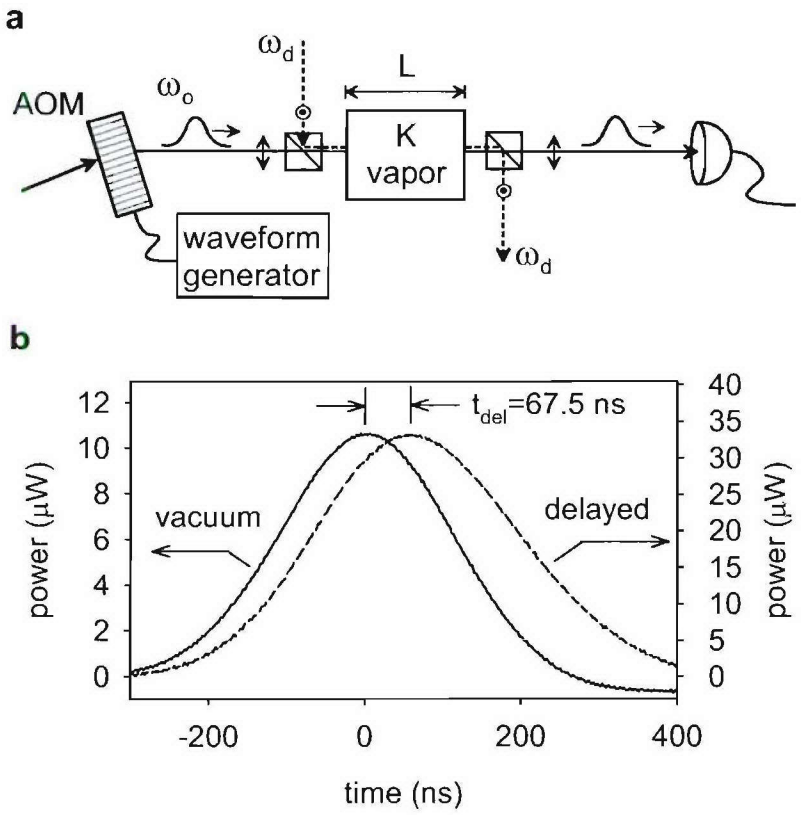

Figure 1 Stenner

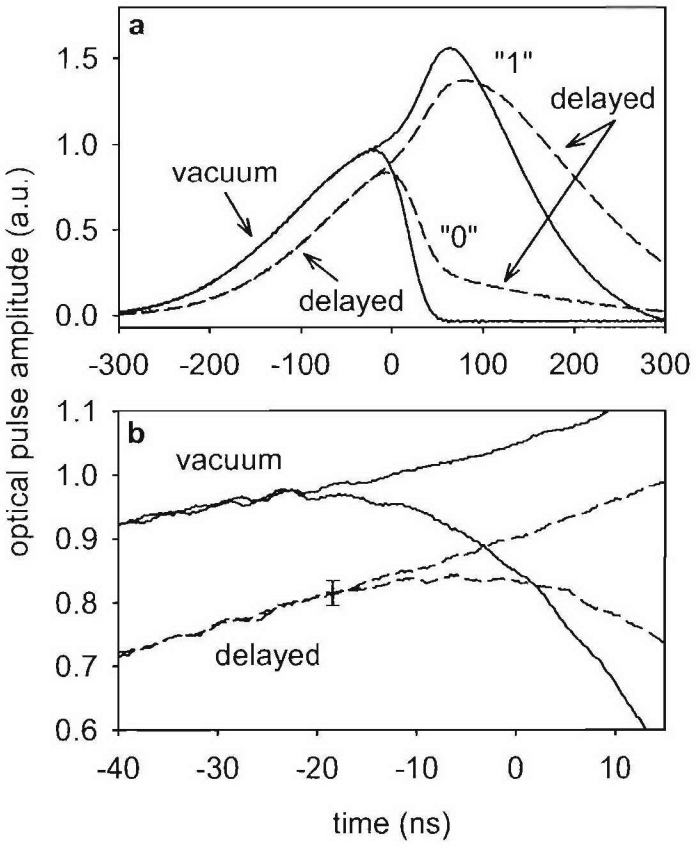

Figure 2 Stenner

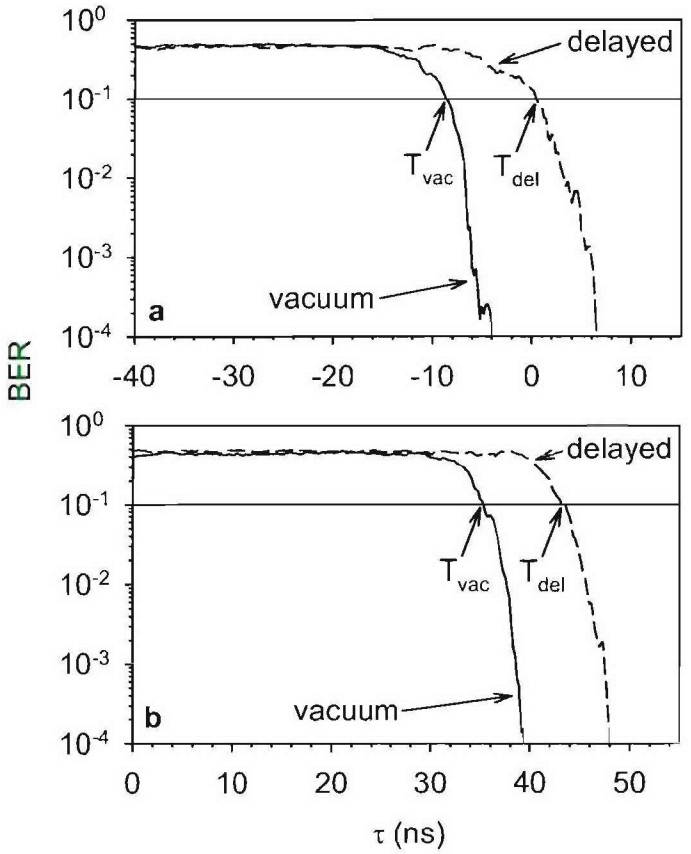

Figure 3 Stenner

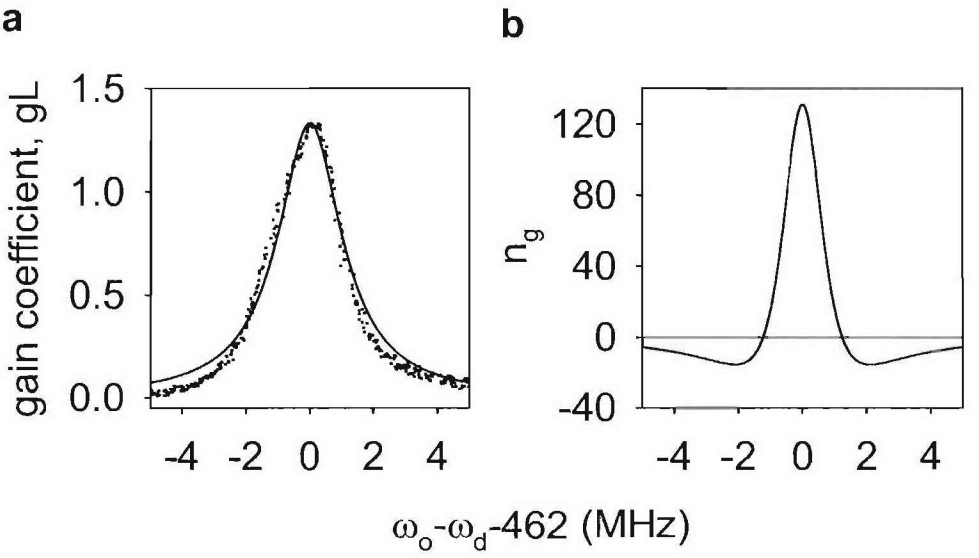

Figure S.1 (Supplemental Material) Stenner

Appendix 5

Distortion Compensation in a Slow-Light Delay Line

Distortion management in slow-light pulse delay

Michael D. Stenner and Mark A. Neifeld
University of Arizona, Department of Electrical and Computer Engineering and The Optical Sciences Center, Tucson, Arizona, 85721 USA
mstenner@ece.arizona.edu

Zhaoming Zhu, Andrew M. C. Dawes, and Daniel J. Gauthier
Duke University, Department of Physics, and The Fitzpatrick Center for Photonics and Communications Systems, Durham, North Carolina, 27708 USA

Abstract: We describe a methodology to maximize slow-light pulse delay subject to a constraint on the allowable pulse distortion. We show that optimizing over a larger number of physical variables can increase the distortion-constrained delay. We demonstrate these concepts by comparing the optimum slow-light pulse delay achievable using a single Lorentzian gain line with that achievable using a pair of closely-spaced gain lines. We predict that distortion management using a gain doublet can provide over a factor of 6 increase in slow-light pulse delay as compared with the optimum single-line delay. Experimental results employing Brillouin gain in optical fiber confirm our theoretical predictions.

OCIS codes: (060.5530) Pulse propagation and solitons, (290.5900) Scattering, stimulated Brillouin

References and links

1. R. W. Boyd and D. J. Gauthier, "'Slow' and 'Fast' Light," in *Progress in Optics, Vol. 43*, E. Wolf, ed. (Elsevier, Amsterdam, 2002), pp. 497–530.
2. Y. A. Vlasov, S. Petit, G. Klein, B. Hönerlage, and C. Hirlimann, "Femtosecond measurements of the time of flight of photons in a three-dimensional photonic crystal," Phys. Rev. E **60**, 1030–1035 (1999).
3. J. E. Heebner, R. W. Boyd, and Q. Park, "Slow light, induced dispersion, enhanced nonlinearity, and optical solitons in a resonator-array waveguide," Phys. Rev. E **65**, 036619 (2002).
4. Y. Xu, R. K. Lee, and A. Yariv, "Scattering theory analysis of waveguide-resonator coupling," Phys. Rev. E **62**, 7389–7404 (2000).
5. A. Yariv, Y. Xu, R. K. Lee, and A. Scherer, "Coupled resonator optical waveguide: a proposal and analysis," Opt. Lett. **24**, 711–713 (1999).
6. R. W. Boyd, D. J. Gauthier, A. L. Gaeta, and A. E. Willner, "Maximum time delay achievable on propagation through a slow-light medium," Phys. Rev. A **71**, 023801 (2005).
7. H. Cao, A. Dogariu, and L. J. Wang, "Negative group delay and pulse compression in superluminal pulse propagation," IEEE J. Sel. Top. Quantum Electron. **9**, 52–58 (2003).
8. B. Macke and B. Ségard, "Propagation of light-pulses at a negative group-velocity," European Phys. J. D **23**, 125–141 (2003).
9. M. Bashkansky, G. Beadie, Z. Dutton, F. K. Fatemi, J. Reintjes, and M. Steiner, "Slow-light dynamics of large-bandwidth pulses in warm rubidium vapor," Phys. Rev. A **72**, 033819 (2005).
10. Y. Okawachi, M. S. Bigelow, J. E. Sharping, Z. M. Zhu, A. Schweinsberg, D. J. Gauthier, R. W. Boyd, and A. L. Gaeta, "Tunable all-optical delays via Brillouin slow light in an optical fiber," Phys. Rev. Lett. **94**, 153902 (2005).
11. K. Y. Song, M. G. Herráez, and L. Thévenaz, "Observation of pulse delaying and advancement in optical fibers using stimulated Brillouin scattering," Opt. Express **13**, 82–88 (2005).
12. K. Y. Song, M. G. Herráez, and L. Thévenaz, "Long optically controlled delays in optical fibers," Opt. Lett. **30**, 1782–1784 (2005).

13. Z. Zhu, D.J. Gauthier, Y. Okawachi, J.E. Sharping, A.L. Gaeta, R.W. Boyd, and A.E. Willner, "Numerical study of all-optical slow-light delays via stimulated Brillouin scattering in an optical fiber," to appear in J. Opt. Soc. Am. B **22** (2005).
14. A. V. Oppenheim and A. S. Willsky, *Signals and Systems, 2nd Ed.* (Prentice Hall, Upper Saddle River, 1997).
15. M. Nikles, L. Thévenaz, and P. A. Robert, "Brillouin gain spectrum characterization in single-mode optical fibers," J. Lightwave Tech. **15**, 1842–1851 (1997).
16. Z. Dutton, M. Bashkansky, M. Steiner, and J. Reintjes, "Channelization architecture for wide-band slow light in atomic vapors," SPIE **5735**, 115–129 (2005).
17. Q. Sun, Y. V. Rostovtsev, J. P. Dowling, M. O. Scully, and M. S. Zhubairy, "Optically controlled delays for broadband pulses," Phys. Rev. A **72** 031802(R) (2005).

Over the last decade, there has been great progress in devising new methods for tailoring the dispersion of optical materials, such as electromagnetically induced transparency [1], photonic crystals [2], and nano-optic resonators [3–5]. This work has been motivated by the need for electronically or optically controllable pulse delays for applications such as optical buffers, optical memories, and signal processing. In these applications, the primary requirements for slow-light pulse delay are that the temporal pulse delay t_d be large relative to the pulse width τ and that the pulse not be substantially distorted (defined below). These two requirements largely oppose each other, with large delay coming at the cost of greater distortion. These tradeoffs have been studied in simple Lorentzian systems [6–8] and in Doppler-broadened media [9] with encouraging results, but distortion remains a major limitation to the usefulness of slow-light pulse delay.

In this paper, we present a method for achieving large slow-light pulse delay under the constraint that the distortion does not exceed a particular limit. We also present a simple distortion-managed medium based on dual Lorentzian gain lines. The dispersive properties of this medium are tailored by adjusting the spacing between two adjacent gain lines. We predict a factor of 6.25 improvement in the relative slow-light pulse delay for the two-gain-line case in comparison to the standard single-gain-line case for the same distortion constraint in both cases. We demonstrate dispersion-managed pulse delay in an experiment where the slow-light effect arises from the dispersion associated with stimulated Brillouin scattering (SBS) resonances in a pumped optical fiber [10–13].

For pulses propagating through linear optical systems, the output pulse amplitude $A(\omega,z)$ in the frequency domain can be related to the input pulse amplitude $A(\omega,0)$ by

$$A(\omega,z) = A(\omega,0)e^{ik(\omega)z}, \tag{1}$$

where z is the length of the medium and $k(\omega)$ is the complex wavenumber as a function of frequency ω. A pulse propagates undistorted through a dispersive material when $k(\omega)$ takes the form

$$k(\omega) = k_0 + k_1(\omega - \omega_c), \tag{2}$$

where ω_c is the carrier frequency of the pulse and k_1 is real. That is, the pulse shape remains unchanged and the only effects of propagation are delay, an overall phase shift, and gain or attenuation. In this ideal case, the pulse delay t_d is equal to the group delay $t_g = z(k_1 - 1/c)$, where c is the speed of light in vacuum. In practice, dispersive media do not satisfy Eq. (2) precisely, but have higher-order terms in the Taylor expansion $k(\omega) = \sum_{j=0}^{\infty} k_j(\omega - \omega_c)^j/j!$, where $k_j \equiv d^j k(\omega)/d\omega^j$. These higher-order terms can lead to pulse distortion, or a change in the pulse shape. One can always use pulses with narrower bandwidth, and thereby improve the linearity of $k(\omega)$ over the pulse bandwidth, but this results in longer pulses without increasing the delay. It is sometimes possible to simultaneously increase k_1 and decrease the pulse bandwidth, which can result in larger relative delays, but may result in unacceptably large gain or absorption [6].

One method for creating delayed pulses with minimal distortion is to reduce the effects of the higher-order terms in the Taylor series expansion of $k(\omega)$. The effects of these terms have been studied in simple systems [6–9, 13], but additional degrees of freedom can be used to eliminate or balance them. We propose using custom slow-light media designed to minimize the effects of distortion. Creating a custom dispersion profile $k(\omega)$ is in general quite difficult. However, one can instead create a custom system by combining multiple simple systems. We consider here a system composed of two Lorentzian gain lines; a gain doublet, as shown in Fig. 1. The wavenumber of such a system is given by

$$k(\omega) = \frac{\omega}{c} n_0 + \frac{g_0}{z}\left(\frac{\gamma}{\omega - (\omega_0 - \delta) + i\gamma} + \frac{\gamma}{\omega - (\omega_0 + \delta) + i\gamma}\right), \tag{3}$$

where n_0 is the background refractive index, g_0/z is the line-center amplitude gain coefficient for each line, γ is the linewidth, and 2δ is the separation between the lines. In many slow-light media, the dominant source of distortion is the second-order term in the expansion of $k(\omega)$, k_2 [6]. For the gain doublet, we have

$$k_2 = -\frac{4ig_0\gamma^2}{z}\frac{3\delta^2 - \gamma^2}{(\delta^2 + \gamma^2)^3}. \tag{4}$$

As can be seen from Eq. (4), $k_2 = 0$ for a separation of $\delta = \gamma/\sqrt{3}$, thereby eliminating the lowest-order distortion-causing term. For a doublet in this configuration, the group delay is given by

$$t_g = z(k_1 - 1/c) = \frac{z}{c}(n_0 - 1) + \frac{3}{4}\frac{g_0}{\gamma}, \tag{5}$$

which should be compared to the delay for a single line

$$t_g = \frac{z}{c}(n_0 - 1) + \frac{g_0}{\gamma}. \tag{6}$$

By comparing Eqs. 5 and 6, it is seen that the group delays for these two systems are very similar. While this result is interesting and illuminating, it is of limited practical value; it eliminates only second-order distortion, ignoring all higher-order terms.

A more practical approach to distortion management is to measure distortion as the deviation of the medium from the ideal one described in Eq. (2). At this point, it becomes convenient to describe the medium in terms of its transfer function $H(\omega)$, which is easily related to $k(\omega)$ by $H(\omega) = \exp(ik(\omega)z)$. An ideal medium has a transfer function with constant amplitude $|H(\omega)| = H_0$ and a phase that varies linearly with frequency $\angle H(\omega) = t_p\omega$, where t_p is the total propagation time of the pulse. Relating these to Eq. (2) yields $H_0 = \exp[i(k_0 - k_1\omega_0)z]$ and $t_p = k_1 z$. Any deviation from this behavior leads to pulse distortion [14].

There are many ways to quantify the deviation of the transfer function from ideal. One very simple way is to calculate the infinity-norm—the maximum magnitude—of the amplitude and phase deviation from ideal. This leads to two distortion metrics, one for the amplitude variation (D_a) and one for the phase variation (D_p). The amplitude distortion is given by

$$D_a = \frac{H_{\max} - H_{\min}}{H_{\max} + H_{\min}}, \tag{7}$$

where $H_{\min}$ and $H_{\max}$ are the minimum and maximum values of $|H(\omega)|$ over the frequency range $(\omega_0 - \Delta_b, \omega_0 + \Delta_b)$. Similarly, we define the phase distortion as

$$D_p = \frac{1}{2\pi}\max[|\angle H(\omega) - (t_p\omega + \phi_0)|]_{\omega_0 - \Delta_b}^{\omega_0 + \Delta_b}, \tag{8}$$

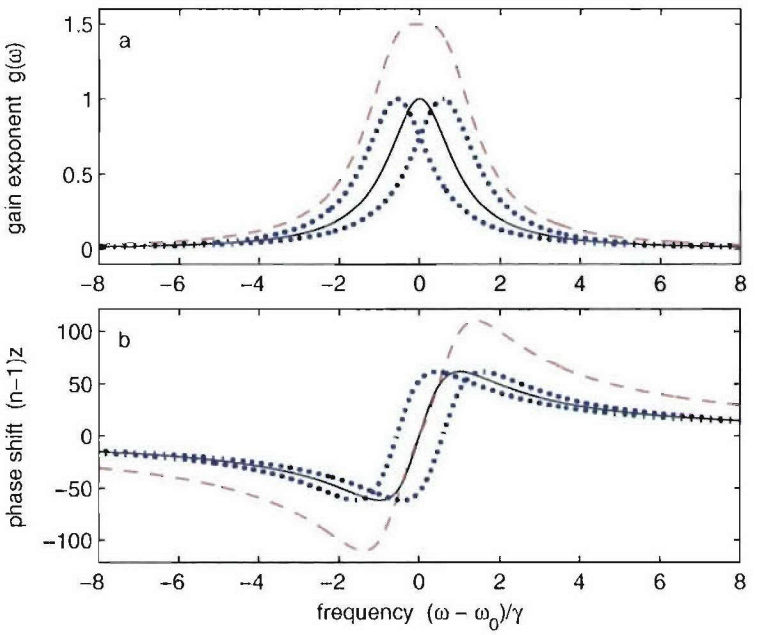

Fig. 1. Gain and dispersion for a single Lorentzian line (solid) and a doublet with separation $\delta = \gamma/\sqrt{3}$ (dashed). Also shown are the two constituent lines (dotted) that make up the doublet. (a) The gain exponent has a broad flat top (a result of setting $k_2 = 0$), although it is larger for the same value of the individual gain coefficients. (b) The dispersion is very similar for both situations near the center of the lines, but extends farther for the gain-doublet case.

where t_p and ϕ_0 are chosen to minimize D_p. This provides a very conservative distortion measure for pulses whose power falls entirely within the frequency range $(\omega_0 - \Delta_b, \omega_0 + \Delta_b)$. The calculation of D_p also serves to define an effective propagation time t_p as the propagation time of a pulse through the ideal medium most closely approximated by the real medium. A delay defined in terms of this propagation time $t_d = t_p - n_0 z/c$ is similar in concept to the group delay, except that it is based on the material dispersion over the entire bandwidth of interest rather than just at the carrier frequency. In fact, the pulse delay t_d is exactly equal to the group delay of the best-fit ideal medium. Because this delay includes dispersive effects over the entire bandwidth of interest, it provides a better prediction of actual pulse delay (peak delay, for example) than the group delay. As is common in communication systems [14], pulses with power extending outside this range can still be used, but the result is then approximate. These distortion metrics are reasonable in the case where the signal spectrum is approximately limited to the given bandwidth window but is otherwise unknown. If more is known about the signal spectrum, then more complicated distortion measures can be used that consider a fixed spectrum and the form of the material deviation from ideal.

Having defined these distortion metrics, we can use them to explore the gain-doublet slow-light media discussed above. Let us first consider the simple Lorentzian gain line, whose trans-

fer function can be written as

$$H_1(\omega) = \exp\left(izn_0\frac{\omega}{c}\right)\times\exp(g_1(\omega)) = \exp\left(izn_0\frac{\omega}{c}+g_{01}\frac{i\gamma}{(\omega-\omega_0)+i\gamma}\right). \quad (9)$$

Holding γ fixed, we find that for each Δ_b there is an optimum value of g_{01} that provides maximum delay t_d subject to gain and distortion constraints. We have chosen to use an amplitude gain constraint of $g_{01} \leq 2.5$ so that the system can be reasonably implemented experimentally. Larger gain is often accompanied by nonlinear optical effects.

The communication system of interest will define both the appropriate form of the relevant distortion and its maximum allowed value. In conjunction with the infinity norm, we select a conservative value of maximum allowed distortion $D_a < 0.05$ and $D_p < 0.05$. This value is somewhat arbitrary and results in pulse distortion that is just noticeable upon visual inspection.

Figure 2(a) shows the maximum relative delay $t_d\Delta_b$ as a function of the relative bandwidth Δ_b/γ. Figure 2(b) shows the Lorentzian gain exponent g_{01}. At each value of Δ_b, the gain is chosen to maximize the delay subject to the gain and distortion constraints described above. By comparing the solid lines in Figs. 2(a) and 2(b), we find that the delay is limited by the distortion constraint for large relative bandwidths—$\Delta_b/\gamma > 0.2$; driving the gain higher will produce too much distortion. For small bandwidths, the delay is limited by the gain constraint. The maximum relative delay is achieved at some intermediate bandwidth—$\Delta_b/\gamma \approx 0.2$—for which both distortion and gain limits are met simultaneously. At this value, we find a maximum relative delay of $t_d\Delta_b \approx 0.5$ under the constraint that neither D_a nor D_p exceed 0.05.

We now consider a distortion-managed system constructed using two nearby Lorentzian gain lines. The transfer function of this system is

$$H_2(\omega) = \exp\left(izn_0\frac{\omega}{c}\right)\times\exp(g_2(\omega)) = \exp\left(izn_0\frac{\omega}{c}+g_{02}\frac{i\gamma}{(\omega-\omega_0-\delta)+i\gamma}+g_{02}\frac{i\gamma}{(\omega-\omega_0+\delta)+i\gamma}\right). \quad (10)$$

As above, we can optimize over the free parameters—now g_{02} and δ—to maximize the delay at each Δ_b subject to the same constraints, $D_a \leq 0.05$, $D_p \leq 0.05$, and $\max[|g_2(\omega)|] \leq 2.5$. The dashed line in Fig. 2(a) shows distortion- and gain-limited relative delay as a function of bandwidth for this doublet system.

As described above, the second order distortion—dominant in many low-distortion cases—can be eliminated entirely ($k_2 = 0$) by setting $\delta = \gamma/\sqrt{3} \rightarrow \delta/\gamma \approx 0.58$. One might expect this result to provide the optimal line separation. However, as shown in Fig. 2(c), delay can be increased at large bandwidths by using a larger separation because, although there is a gain dip at the carrier frequency, the gain excursion over the full bandwidth can be minimized. For small bandwidths, a smaller separation is desirable because the delay is limited by gain, not distortion, and so it is more useful to make the separation small, creating larger dispersion for the same maximum gain.

There are three distinct frequency regions on these plots. The left-most region ($\Delta_b/\gamma < 0.25$) corresponds to gain-limited configurations where the second line provides no additional benefit. The right-most region ($\Delta_b/\gamma > 0.8$) corresponds to the distortion-limited regime. In the central region, both constraints are hit simultaneously. In this gain-doublet medium, we observe a maximum relative delay of 0.9 at a bandwidth of $\Delta_b/\gamma = 0.8$, an improvement by a factor of 6.25 over the single line at that bandwidth and nearly twice the best single-line relative delay.

To test our distortion-management concept, we propagate optical pulses through a SBS slow-light medium [10–13]. A gain-doublet with an adjustable frequency separation 2δ is realized

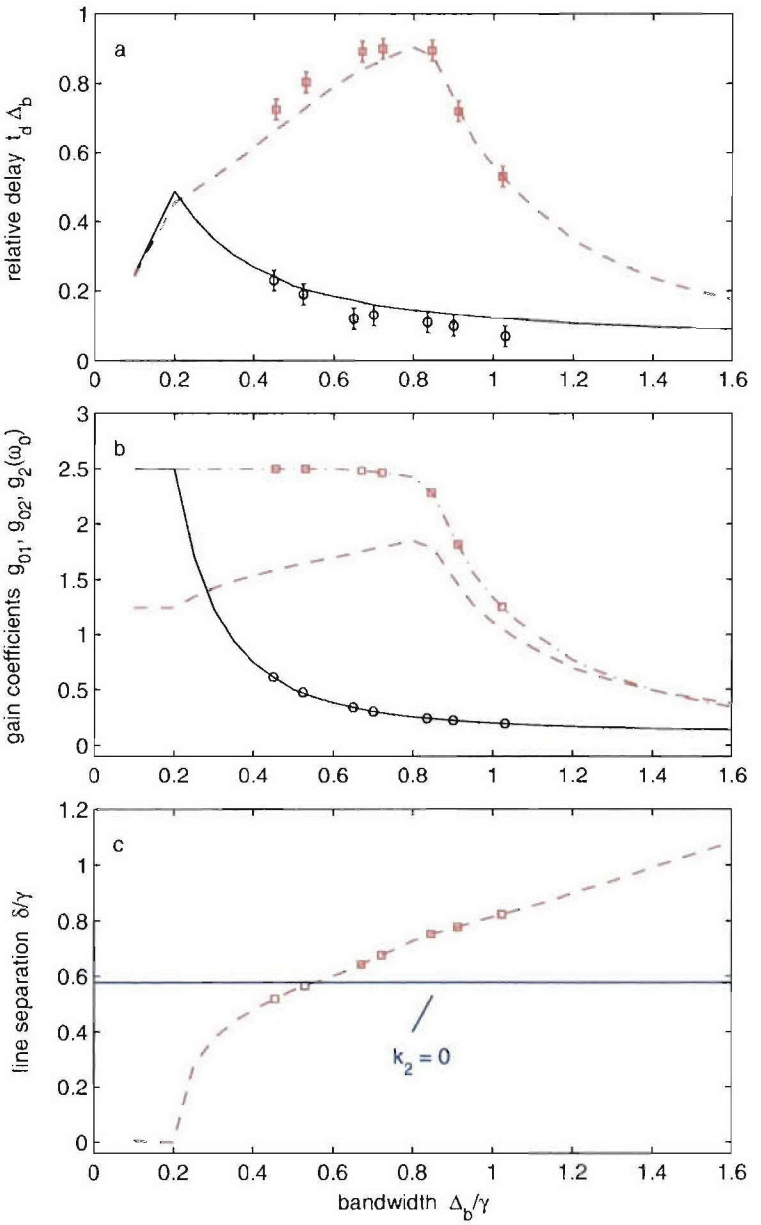

Fig. 2. Simulation and experimental results for both the single Lorentzian line (solid lines for simulation, circles for experimental results) and the doublet (dashed lines for simulation, squares for experimental results). (a) Relative delay. (b) Lorentzian line-center amplitude gain coefficients g_01 and g_02. Also shown is the center-frequency gain coefficient for the doublet $g_2(\omega_0)$ (dot-dashed). (c) Line separation for the doublet. The horizontal line indicates the value of δ/γ that leads to $k_2 = 0$

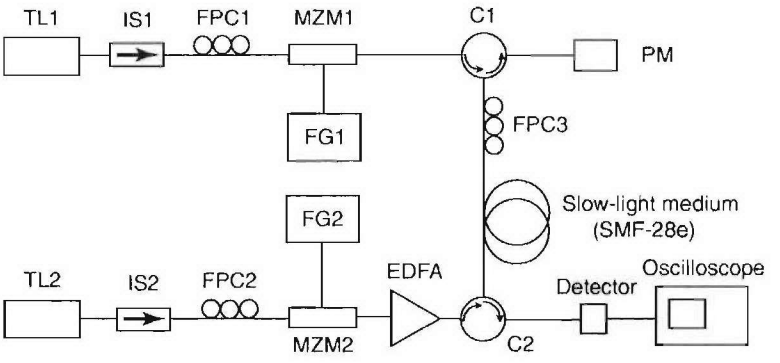

Fig. 3. Experiment setup based on a fiber Brillouin amplifier. TL1, TL2: tunable lasers; IS1, IS2: isolators; FPC1, FPC2, FPC3: fiber polarization controllers; MZM1, MZM2: Mach-Zehnder modulators; FG1, FG2: function generators; EDFA: Erbium-doped fiber amplifier; C1, C2: circulators; SMF-28e: 500-m-long SMF-28e fiber (the SBS amplifier); PM: power meter.

by pumping a standard room-temperature telecommunications optical fiber with a bichromatic laser field, which is obtained via carrier-frequency-suppression modulation of a continuous-wave (CW) single-frequency laser field [15]. The doublet spacing 2δ is adjusted by changing the frequency of the voltage driving the modulator and the gain coefficient (and hence the slow-light delay) is adjusted by changing the pump power. Either of these adjustments can be made very quickly, limited by the transit time of the pump field through the optical fiber.

The transfer function $H_2(\omega)$ (Eq. 10) can be realized by setting the carrier frequency of the pulses close to the SBS amplifying resonances (the so-called Stokes resonances) and keeping the pulse intensity low enough so that the response is linear. For comparison, we also propagate pulses through a standard slow-light medium with a single gain resonance, corresponding to $H_1(\omega)$ (Eq. 9) and realized using a single-frequency pump field [10–13].

The experimental setup is shown schematically in Fig. 3. Two 1550-nm narrow-linewidth tunable lasers (TL1 and TL2) are used to produce Stokes signal pulses and the pump beams, respectively. A bichromatic pump beam is produced by passing the single-frequency CW laser beam generated by TL2 (angular frequency ω_p) through modulator MZM2, which is driven by a sinusoidal voltage of angular frequency δ. The bias voltage applied to MZM2 is chosen to suppress the carrier frequency, resulting in a beam with frequencies $(\omega_p \pm \delta)$. This beam is amplified by an erbium-doped fiber amplifier (EDFA) and routed via circulator C2 to pump a 500-m-long SMF-28e fiber (the slow-light medium). For this fiber, we measure an SBS linewidth (full-width at half-maximum) of $\gamma/\pi \simeq 35$ MHz. After passing through the slow-light medium, the beam is routed out of the system via circulator C1 and monitored by a power meter.

The signal pulses are produced from a laser beam generated by TL1. The beam passes through an isolator, a fiber polarization controller, and modulator MZM1, producing pulses with carrier frequency ω_c. They are approximately Gaussian-shaped with an intensity envelope of the form $I(t) = I_0 \exp[-(t/\tau)^2]$, where the spectrum $S(\omega) \propto \exp[-\tau^2(\omega - \omega_c)^2]$ has a $1/e$ half width of $\Delta_0 = 1/\tau$. The pulses enter the 500-m-long SMF-28e fiber via circulator C1 and counterpropagate with respect to the bichromatic pump beam in the fiber. The slow-light-delayed and amplified pulses are routed out of the system via C2, sensed by a fast photodetector, and displayed on a digital oscilloscope. The frequency difference between TL1 and TL2 is tuned so that the pulse carrier frequency is set precisely to the center of the SBS amplifying resonances, $\omega_c = \omega_0$ where $\omega_0 = \omega_p - \Omega_B$, and $\Omega_B/2\pi = 12.5\,\text{GHz}$ is the Brillouin frequency shift

for the SMF-28e fiber. Fiber polarization controllers FPC1 and FPC2 are used to maximize the transmissions through the Mach-Zehnder modulators, and FPC3 is used to maximize the SBS slow-light delay experienced by the Stokes pulses.

In the experiments with the distortion-managed slow-light medium, we select a pulse bandwidth Δ_b (using the assumption that $\Delta_b = \Delta_0$), set the resonance half-separation δ according to Fig. 2(c), set the central gain $g_2(\omega_0)$ according to Fig. 2(b) by adjusting the pump power, and measure the delay of the peak of the pulses induced by the slow-light medium (which is approximately equal to t_d). From Fig. 2(a), it is seen that the agreement with the theoretical predictions is very good. Also shown in Fig. 2(a) are our observations for a single-gain line (the non-distortion-managed slow-light medium), where the agreement between the observations and the predictions is excellent.

The data shown in Fig. 2(a) demonstrate that the distortion-managed slow-light medium vastly outperforms the standard slow-light medium under conditions of constant distortion. Around the optimum normalized bandwidth of ~ 0.84 ($\tau \sim 11$ ns), we measure an improvement in relative pulse delay of a factor of 8.1 ± 2.6. Thus, the distortion-managed approach substantially increases the usable bandwidth of a slow-light medium that is much easier to implement in comparison to previously suggested methods [16,17].

In conclusion, we have shown that, although distortion limits the slow-light delay achievable in simple systems [6–8], it is possible to achieve better distortion- and gain-limited delay using custom composite media. Dramatic improvement can be achieved over a single Lorentzian line with a system as simple as a Lorentzian doublet, suggesting that even greater improvement may be achievable using more flexible media.

We gratefully acknowledge the financial support of the DARPA DSO Slow-Light Program.

Appendix 6

Causality and Double Negative Metamaterials

Causality and double-negative metamaterials

Richard W. Ziolkowski* and Allison D. Kipple†
Department of Electrical and Computer Engineering, The University of Arizona, 1230 E. Speedway Boulevard, Tucson, Arizona 85721-0104, USA
(Received 9 May 2003; published 29 August 2003)

The causality of waves propagating in a double-negative (DNG) metamaterial ($\varepsilon_r<0$ and $\mu_r<0$) has been investigated both analytically and numerically. By considering the one-dimensional electromagnetic problem of a pulsed current sheet radiating into a DNG medium, it is shown that causality is maintained in the presence of a negative index of refraction only if the DNG medium is dispersive. A Drude model DNG medium is used in this study. Spectrograms of the wave phenomena in the dispersive DNG medium show that the higher frequency components, which create the leading edge of the electromagnetic signals and see a double positive (DPS) medium ($\varepsilon_r>0$ and $\mu_r>0$), arrive causally before the negative index effects germinate completely. Comparisons with approximate analytical results demonstrate the presence of the negative index of refraction properties in the continuous wave portion of the signals. This dynamic pulse reshaping between the positive and negative index of refraction wave components causes an apparent delay in the realization of the negative index of refraction properties. Pulse broadening of the signal tails is associated with both dispersion and a larger negative index of refraction seen by the associated wave components.

DOI: 10.1103/PhysRevE.68.026615 PACS number(s): 41.20.Jb, 42.25.Bs, 81.05.Zx, 03.50.De

I. INTRODUCTION

The index of refraction of a double-negative (DNG) metamaterial, i.e., a material with both negative permittivity and negative permeability (see, for instance, [1–4]), has been shown to be negative (see, for instance, [5–7]). There are now several theoretical and experimental studies that have been reported confirming this negative index of refraction (NIR) property and applications derived from it, such as phase compensation and electrically small resonators [8], negative angles of refraction [8–13], enhanced focusing [14–17], backward wave antennas [18], Čerenkov radiation [19], photon tunneling [20,21], and enhanced electrically small antennas [22]. These studies rely heavily on the concept that a continuous wave (CW) excitation of a DNG medium will lead to a NIR and, hence, to negative phase terms.

If one considers carefully the ramifications of a homogeneous, nondispersive DNG medium and the resulting NIR, one immediately encounters a paradox in the time domain. A source in such a DNG medium will generate a signal that propagates noncausally. On the other hand, many of the NIR studies have been based on numerical simulations in the time domain of dispersive DNG metamaterials. In fact, as shown in [5], realistic DNG metamaterials must be dispersive. How then does one reconcile the CW NIR properties associated with a nondispersive DNG medium with causal pulse propagation in a dispersive DNG medium? Can one rely on designs of NIR applications that rely on the CW behavior of the corresponding nondispersive DNG medium?

In this paper, the issue of causality in DNG metamaterials is considered by studying the one-dimensional electromagnetic problem of a pulsed source in a DNG medium. A lossy, dispersive Drude model is used to create the DNG medium. A finite difference time domain (FDTD) numerical solution of the problem is obtained. The source is driven by an excitation pulse that is based on a sinusoid and a smooth envelope. It is demonstrated that both the expected causal propagation and the NIR effects are obtained. Other recent considerations of causality (e.g., [13,23]) have, in effect, been limited to narrow bandwidth considerations. The corresponding nondispersive DNG medium problem is then considered. It is demonstrated analytically that only a noncausal exact solution exists. The noncausal analytical DNG solution and the causal FDTD DNG results are then reconciled by the presence of dispersion. It is shown that the leading edge of the pulse, which is associated with high frequencies, propagates causally in a double-positive (DPS) medium (i.e., a normal medium that has both positive permittivity and positive permeability), while the CW portion of the pulse senses the DNG medium and exhibits the NIR effects. The slower speed components appear to accumulate in the trailing edge. Both the leading and trailing edges of the pulse are shown to be strongly affected by the presence of dispersion. Using the nondispersive results to guide the construction of approximate solutions to the dispersive DNG medium problem, it is shown that the NIR effects obtained for the dispersive case coincide with those predicted by the nondispersive case. It is also shown that observed time delays in the realization of the NIR properties are directly associated with the transition from the causal DPS response to the DNG one. These observations reinforce the fact that DNG metamaterials and the associated NIR results are physical and that their experimental confirmation is consistent with the underlying wave propagation physics.

II. 1D-FDTD SIMULATOR

A one-dimensional FDTD simulation environment was utilized to study the characteristics of signals produced by a pulsed source in a DNG medium. This one-dimensional en-

*Email address: ziolkowski@ece.arizona.edu
†Email address: kipplea@ece.arizona.edu

vironment was convenient and incorporated all of the necessary physics. The field components were assumed to be E_x and H_y with the direction of propagation taken along the z axis. An electric current plane-wave source was located in the plane $z=0$, at the center of a very large DNG slab. The DNG slab was surrounded by free space and was centered in the FDTD domain. This slab geometry was used in order that standard, exact absorbing boundary conditions could be applied at the edges of the FDTD simulation space. The problem geometry is shown in Fig. 1.

As in [6], lossy Drude polarization and magnetization models were used to simulate the DNG medium. In the frequency domain, assuming an $\exp(-i\omega t)$ time dependence, this means the permittivity and permeability were described as

$$\varepsilon(\omega)=\varepsilon_0\left(1-\frac{\omega_{pe}^2}{\omega(\omega+i\Gamma_e)}\right),$$

$$\mu(\omega)=\mu_0\left(1-\frac{\omega_{pm}^2}{\omega(\omega+i\Gamma_m)}\right). \tag{1}$$

The corresponding time-domain equations for the polarization, P_x, and the normalized magnetization, $M_{ny}=M_y/\mu_0$, fields were

$$\partial_t^2 P_x+\Gamma_e\partial_t P_x=\varepsilon_0\omega_{pe}^2 E_x,$$

$$\partial_t^2 M_{ny}+\Gamma_m\partial_t M_{ny}=\mu_0\omega_{pm}^2 H_y. \tag{2}$$

The normalized magnetization was introduced to make the electric and magnetic field equations completely symmetric. By introducing the induced electric and magnetic currents

$$J_x=\partial_t P_x,$$

$$K_y=\partial_t M_{ny}, \tag{3}$$

and the source current, J_s, the field and current equations used to model the waves generated by a pulsed source in a DNG medium became

$$\partial_t E_x=-\frac{1}{\varepsilon_0}(\partial_z H_y+J_x+J_s),$$

$$\partial_t J_x+\Gamma_e J_x=\varepsilon_0\omega_{pe}^2 E_x,$$

$$\partial_t H_y=-\frac{1}{\mu_0}(\partial_z E_x+K_y),$$

$$\partial_t K_y+\Gamma_m K_y=\mu_0\omega_{pm}^2 H_y. \tag{4}$$

A lossless, nondispersive DPS medium was used as a comparison case. In such a DPS medium, the corresponding field and current equation set was simply

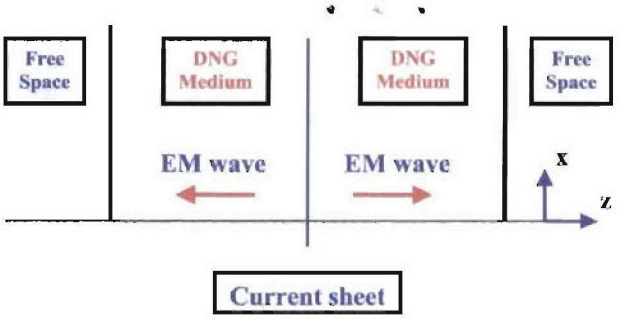

FIG. 1. (Color online) Configuration of the one-dimensional electromagnetic current sheet problem.

$$\partial_t E_x=-\frac{1}{\varepsilon_0}(\partial_z H_y+J_s),$$

$$\partial_t H_y=-\frac{1}{\mu_0}\partial_z E_x. \tag{5}$$

Equation sets (4) and (5) were solved self-consistently and numerically with the FDTD approach [24,25], i.e., these equations were discretized with a standard leap-frog in time, staggered grid approach. The electric field component was taken at the cell edges for integer time steps; the magnetic field components were taken at the cell centers for half-integer time steps. The electric and magnetic currents were located in space at the cell centers, and their time assignments were opposite to the corresponding electric and magnetic field components, i.e., the magnetic current components were sampled at integer time steps and the electric current component was sampled at half-integer time steps. This allowed a FDTD stencil that properly simulated matched medium conditions. The time step was set at 0.95 of the Courant value, i.e., $\Delta t=0.95\Delta z/c$.

In all cases, the center frequency of interest was chosen to be $f_0\equiv\omega_0/2\pi=30$ GHz, corresponding to a free-space wavelength $\lambda_0=1.0$ cm. Only matched media at f_0 were considered, i.e.,

$$Z(\omega_0)=\sqrt{\frac{\mu(\omega_0)}{\varepsilon(\omega_0)}}=\sqrt{\frac{\mu_0}{\varepsilon_0}}=Z_0. \tag{6}$$

Matching was obtained by setting the parameters for the electric and magnetic Drude models to be identical, i.e., $\omega_{pe}=\omega_{pm}=\omega_p$ and $\Gamma_e=\Gamma_m=\Gamma$. In all cases, only low loss values were considered by setting $\Gamma=10^{+8}=5.31\times10^{-4}\omega_0$. This value was selected to connect these results to those presented in [6] and [12]. A DNG medium matched to free space at $f_0=30$ GHz was considered, i.e.,

$$n(\omega_0)=\sqrt{\frac{\varepsilon(\omega_0)}{\varepsilon_0}}\sqrt{\frac{\mu(\omega_0)}{\mu_0}}=\sqrt{\varepsilon_r(\omega_0)}\sqrt{\mu_r(\omega_0)}=-1. \tag{7}$$

This required $\omega_p=2\pi\sqrt{2}f_0=2.66573\times10^{11}$ and, hence, $\Gamma=3.75\times10^{-4}\omega_p$.

The source was given by the expression

$$\vec{J}(z)=J_s(z)\hat{x}=-\delta(z)S(t)\hat{x}. \tag{8}$$

It was modeled in the FDTD simulation space by introducing the equivalent source $J_s\hat{x}=\hat{z}\times H_y\hat{y}$, i.e., $H_{y,\text{source}}=S(t)$ at the source plane. The input time signal was causal and was defined by

$$S(t)=g(t)\sin(\omega_0 t), \tag{9}$$

where the multiple cycle m-n-m envelope is given by the expression

$$g(t)=\begin{cases} 0 & \text{for } t<0 \\ g_{\text{on}}(t) & \text{for } 0\leq t<mT_p \\ 1 & \text{for } mT_p\leq t<(m+n)T_p \\ g_{\text{off}}(t) & \text{for } (m+n)T_p\leq t<(m+n+m)T_p \\ 0 & \text{for } (m+n+m)T_p\leq t, \end{cases} \tag{10}$$

where the period $T_p=1/f_0$. With the terms $x_{\text{on}}(t)=t/(mT_p)$ and $x_{\text{off}}(t)=[t-(m+n)T_p]/(mT_p)$, the continuous, two derivative smooth functions g_{on} and g_{off} can be written as

$$g_{\text{on}}(t)=10x_{\text{on}}^3(t)-15x_{\text{on}}^4(t)+6x_{\text{on}}^5(t),$$

$$g_{\text{off}}(t)=1-[10x_{\text{off}}^3(t)-15x_{\text{off}}^4(t)+6x_{\text{off}}^5(t)]. \tag{11}$$

The function $g_{\text{on}}(t)$ goes smoothly from 0 to 1 in m periods; the function $g_{\text{off}}(t)$ goes smoothly from 1 to 0 in m periods. The function $f(t)$ thus turns on smoothly in m periods, maintains a constant amplitude for n periods, and turns off smoothly in m periods. The more cycles in the center portion, i.e., the larger n is, the narrower the bandwidth and the closer the input pulse is to a CW signal. The shorter the turn-on and turn-off sections are, i.e., the smaller m is, the broader the bandwidth is and, in particular, the more high frequency content there is in the pulse.

The simulation space was discretized into cells with a length $\Delta z=\lambda_0/100=100\ \mu$m. The FDTD predicted electric field values at $z=100\Delta z$ and $z=120\Delta z$ for a 5-10-5 input pulse in the dispersive DNG medium are shown in Fig. 2(a). An expanded view of the leading edge of these pulses is shown in Fig. 2(b). The corresponding FDTD predicted electric field values for a lossless nondispersive DPS medium with $n=+1$ are shown in Figs. 3(a) and 3(b). Comparing these sets of figures, one can clearly see that, as in the DPS case, the fronts of the signals in the DNG medium arrive causally at the observation points. In the DPS case, the peaks of the central, CW portion of the pulse observed at the more distant *point 2* occur *later* in time than they do at *point 1*. In contrast, one can see that in the DNG case the NIR effects (e.g., negative phase and phase velocity) evolve as the CW portion of the signal is realized, e.g., the peaks of the CW portion of the pulse observed at the more distant *point 2* occur *earlier* in time than they do at *point 1*. Dispersive effects are also apparent in the DNG results as the observation points move farther away from the source. These results are consistent with those reported in [6] for a DNG slab.

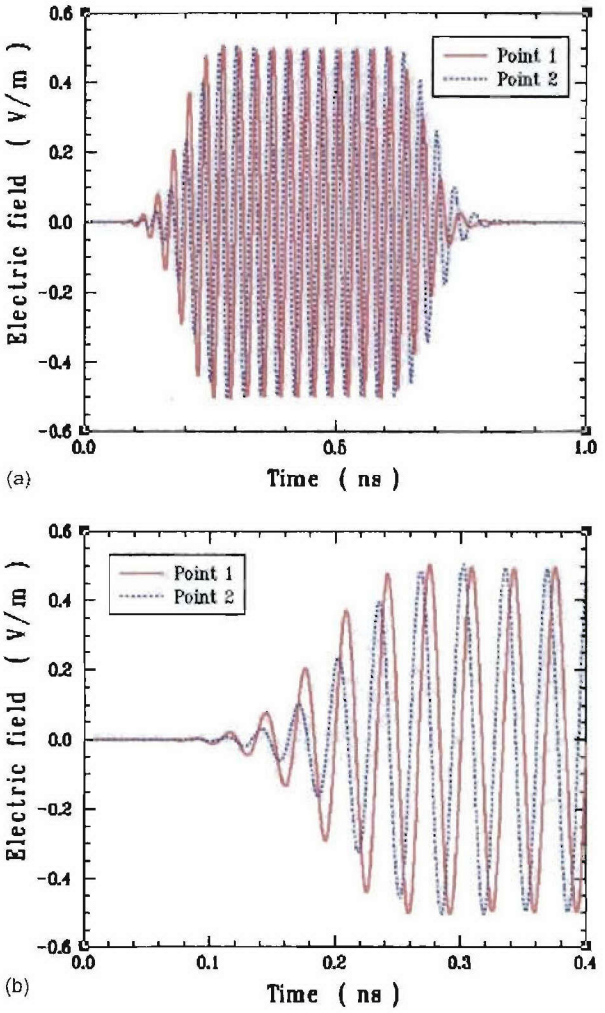

FIG. 2. (Color online) FDTD predicted electric field values in the DNG medium for the 5-10-5 input pulse (a) measured at the points $z=100\Delta z$ and $z=120\Delta z$ and (b) zoomed to the front of those pulses, where $\Delta z=100\ \mu$m.

III. ANALYTICAL CONSIDERATIONS

Consider now the plane-wave source in a hypothetical lossless, nondispersive DNG medium with $\varepsilon_r=\mu_r=-1$. The Maxwell equations become

$$\partial_t E_x=\frac{1}{|\varepsilon|}(\partial_z H_y+J_s),$$

$$\partial_t H_y=\frac{1}{|\mu|}\partial_z E_x. \tag{12}$$

First, let the excitation signal be given by the expression

$$S(t)=[H(t)-H(t-T)]\sin(\omega_0 t), \tag{13}$$

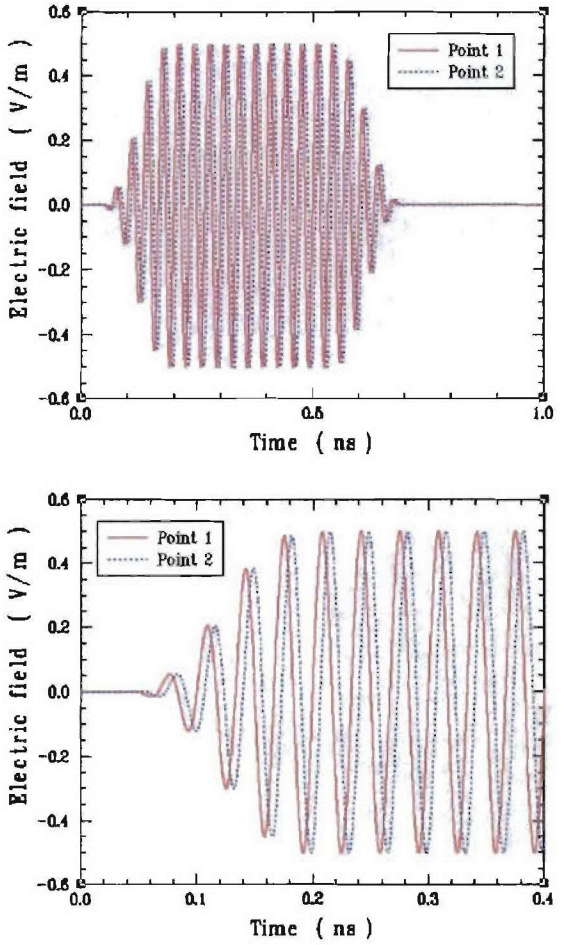

FIG. 3. (Color online) FDTD predicted electric field values in the DPS medium for the 5-10-5 input pulse (a) measured at the points $z=100\Delta z$ and $z=120\Delta z$ and (b) zoomed to the front of those pulses, where $\Delta z=100\ \mu$m.

where $T=(m+n+m)T_p$ and the Heaviside function

$$H(t)=\begin{cases}0 & \text{for } t<0\\ 1 & \text{for } t\geq 0.\end{cases} \tag{14}$$

Introducing the variable

$$t_0=|\varepsilon||\mu||z|, \tag{15}$$

it is readily shown that the solution to Eqs. (12) and (13) is

$$E_x(z,t)=\frac{1}{2}\frac{|\mu|^{1/2}}{|\varepsilon|^{1/2}}[H(t+t_0)-H(t+t_0-T)]\sin[\omega_0(t+t_0)],$$

$$H_y(z,t)=\frac{1}{2}\mathrm{sgn}(z)[H(t+t_0)-H(t+t_0-T)]\times\sin[\omega_0(t+t_0)], \tag{16}$$

where

$$\mathrm{sgn}(z)=\begin{cases}-1 & \text{for } z<0\\ +1 & \text{for } z\geq 0.\end{cases} \tag{17}$$

We note for completeness that $\partial_z|z|=\mathrm{sgn}(z)$ and $\partial_z\,\mathrm{sgn}(z)=+2\delta(z)$.

The corresponding DPS medium results with $\varepsilon_r=\mu_r=+1$ are described by the Maxwell equations

$$\partial_t E_x=-\frac{1}{|\varepsilon|}(\partial_z H_y+J_s),$$

$$\partial_t H_y=-\frac{1}{|\mu|}\partial_z E_x, \tag{18}$$

and their solutions

$$E_x(z,t)=\frac{1}{2}\frac{|\mu|^{1/2}}{|\varepsilon|^{1/2}}\{H(t-t_0)-H[t-(t_0+T)]\}\times\sin[\omega_0(t-t_0)],$$

$$H_y(z,t)=\frac{1}{2}\mathrm{sgn}(z)\{H(t-t_0)-H[t-(t_0+T)]\}\times\sin[\omega_0(t-t_0)]. \tag{19}$$

Comparing the two results, Eqs. (16) and (19), one can clearly see that the lossless, nondispersive DNG medium solution is noncausal. This observation is in complete agreement with those made in [5].

Similarly, with the m-n-m excitation pulse given by Eq. (9), the DNG medium solution is

$$E_x(z,t)=\frac{1}{2}\frac{|\mu|^{1/2}}{|\varepsilon|^{1/2}}g(t+t_0)\sin[\omega_0(t+t_0)],$$

$$H_y(z,t)=\frac{1}{2}\mathrm{sgn}(z)g(t+t_0)\sin[\omega_0(t+t_0)], \tag{20}$$

and the DPS medium solution is

$$E_x(z,t)=\frac{1}{2}\frac{|\mu|^{1/2}}{|\varepsilon|^{1/2}}g(t-t_0)\sin[\omega_0(t-t_0)],$$

$$H_y(z,t)=\frac{1}{2}\mathrm{sgn}(z)g(t-t_0)\sin[\omega_0(t-t_0)]. \tag{21}$$

The DNG medium result is again noncausal while the DPS medium result is causal. The FDTD results shown in Fig. 3 recover the DPS solution (21) to within 0.1%.

Led by the observed combination of causal envelopes and NIR effects of the central portions of the DNG FDTD results, we consider the following approximate analytical solutions to the fields produced by the abrupt and smoothed windowed sine waveforms, respectively:

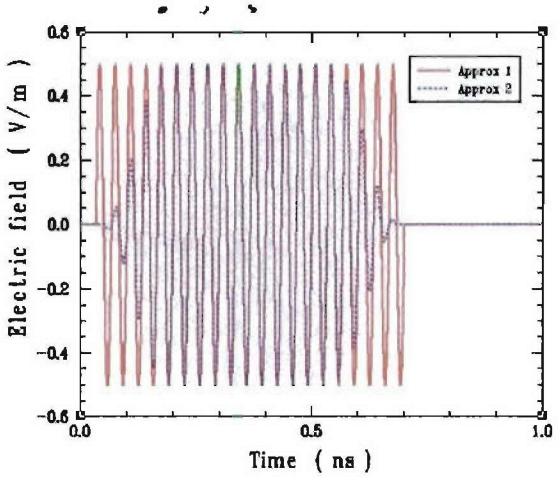

FIG. 4. (Color online) Approximate electric field values in the DNG medium for the 5-10-5 input pulse and $\Delta z = 100\ \mu$m measured at the point $z = 100\Delta z$.

$$E_x(z,t) = \frac{1}{2}\frac{|\mu|^{1/2}}{|\varepsilon|^{1/2}}\{H(t-t_0) - H[t-(t_0+T)]\} \times \sin[\omega_0(t+t_0)],$$

$$H_y(z,t) = \frac{1}{2}\mathrm{sgn}(z)\{H(t-t_0) - H[t-(t_0+T)]\} \times \sin[\omega_0(t+t_0)], \quad (22)$$

and

$$E_x(z,t) = \frac{1}{2}\frac{|\mu|^{1/2}}{|\varepsilon|^{1/2}} g(t-t_0)\sin[\omega_0(t+t_0)],$$

$$H_y(z,t) = \frac{1}{2}\mathrm{sgn}(z) g(t-t_0)\sin[\omega_0(t+t_0)]. \quad (23)$$

These solutions would allow for forward propagation in time while maintaining the solution's negative CW phase characteristics. Inserting the proposed solution (22) into Eqs. (12), one finds that it does not exactly solve them. The defects in this proposed solution occur as δ-function contributions at $t = t_0$ and $t = t_0 + T$. Similarly, inserting the proposed solution (23) into Eqs. (12), one finds that it also is not an exact solution and its failings as an exact solution occur as defects proportional to the derivatives of the turn-on and turn-off portions of the envelope function g. Note, however, that the CW portions of both of these proposed solutions do satisfy the Maxwell equations. Consequently, since the defects are highly localized, we will take Eqs. (22) and (23) as approximate solutions for the DNG case.

The electric field components of the approximate analytical solutions (22) and (23) for the 5-10-5 excitation pulse are compared at the point $z = 100\Delta z$ in Fig. 4. Approximation 1 denotes the electric field component given in Eqs. (22); approximation 2 denotes the electric field component given in Eqs. (23). Complete agreement of the CW portions of these results is observed, as expected. Their differences occur simply in how each of them is turned on and off. We note that the approximate solution (22) contains extremely high frequency content because of its abrupt turn-on and turn-off. The n-m-n form is more realistic and, as Fig. 4 shows, it contains the essential characteristics of this limiting case.

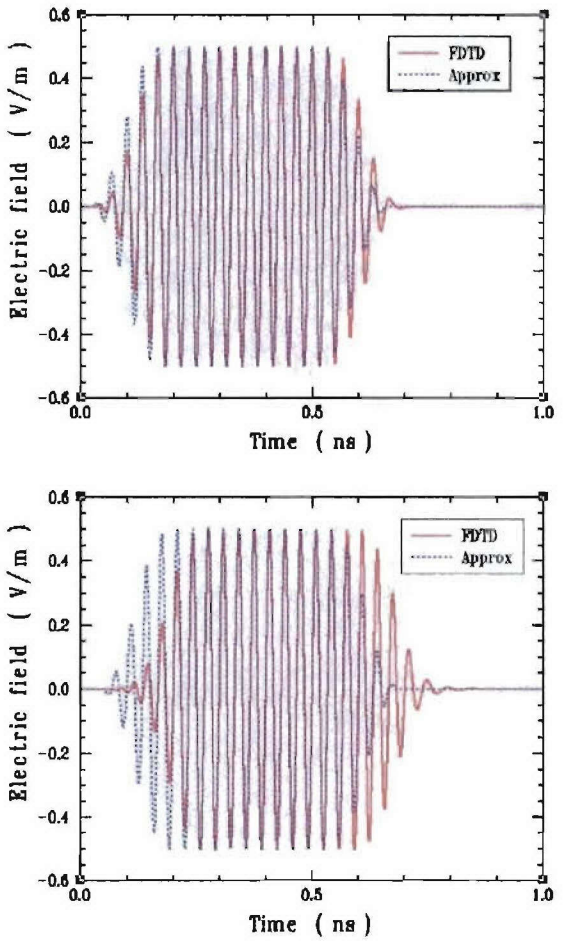

FIG. 5. (Color online) Comparison of the FDTD predicted and approximate analytical electric field values in the DNG medium for the 5-10-5 input pulse measured at the points (a) $z = 30\Delta z$ and (b) $z = 100\Delta z$, where $\Delta z = 100\ \mu$m.

The electric field component of the approximate solution given by Eqs. (23) for the 5-10-5 excitation pulse is compared to the corresponding FDTD results for the dispersive DNG medium at the points $z = 30\Delta z$ and at $z = 100\Delta z$ in Figs. 5(a) and 5(b), respectively. The corresponding results for a 2-16-2 excitation pulse were also obtained. Because the turn-on and turn-off portions of the 2-16-2 pulse are shorter, the time discretization was halved by decreasing the spatial discretization by a factor of 2 to $\Delta z' = 5.0 \times 10^{-5}$ m. The electric field component of the approximate solution given by Eqs. (23) for the 2-16-2 excitation pulse is compared to the FDTD result for the DNG medium at the points, $z = 60\Delta z'$ and $z = 200\Delta z'$, in Figs. 6(a) and 6(b), respectively. These observation points are located at the same distances away from the source, $z = 0.003$ m and $z = 0.01$ m, as they are for the 5-10-5 excitation pulse case.

From both sets of results, one can clearly discern the

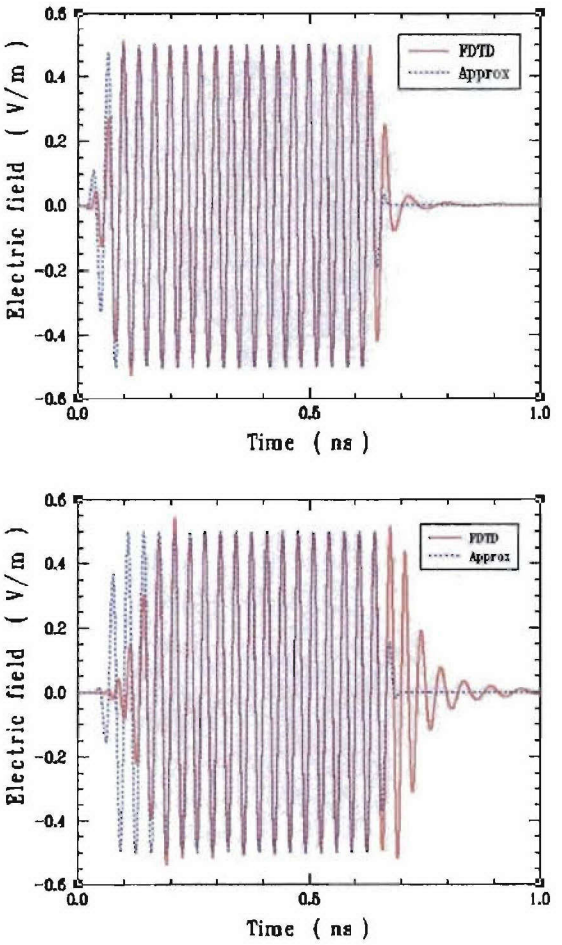

FIG. 6. (Color online) Comparison of the FDTD predicted and approximate analytical electric field values in the DNG medium for the 2-16-2 input pulse measured at the points (a) $z=60\Delta z'$ and (b) $z=200\Delta z'$, where $\Delta z'=50$ μm.

causal nature of the entire FDTD solution. One can also see that the NIR effects in the CW portions of the FDTD and the approximate results agree quite well. Thus, the NIR effects associated with a CW analysis are indeed recovered by the lossy, dispersive FDTD results. Clearly, if the loss were larger, the clarity of these CW components will be lost among the dispersive effects. This occurs even for the present cases as the observation points move much farther away from the source plane.

One can also immediately see from Figs. 5 and 6 that the main differences between the FDTD and the approximate results occur principally in the turn-on and turn-off portions of the envelope, as expected from the approximate solution analysis. The presence of dispersion is clearly seen in these regions. The faster turn-on envelope of the 2-16-2 pulse shows less dispersive effects than does the 5-10-5 excitation pulse. This might be expected since the turn-on portion of the envelope of the 2-16-2 pulse has higher frequency content than does the corresponding portions of the 5-10-5 envelope. Consequently, according to the Drude model, it sees a medium more like free space and less dispersion. On the other hand, there are larger differences in the turn-off portions of the 2-16-2 excitation pulse results than there are in those produced by the 5-10-5 excitation pulse. This trailing edge has more of its own higher and lower frequency components that are subject to dispersion; and it accumulates more of the slower moving, lower frequency components generated in the earlier portions of the pulse. Moreover, comparing the set of results at the point closer to the source with the one farther away from it, one observes that there are time delays between when the CW NIR effects are completely present and when they lose their dominance to the dispersive effects and that these delays increase as the observation point moves farther away from the source.

IV. SPECTROGRAM CONSIDERATIONS

How is it then that the dispersive effects reconcile the NIR and causal behaviors? To address this issue, we have studied spectrograms [26,27] of the source-generated pulse as it arrives at various observation points in the DNG medium. We recall that the magnitude of the phase and group speeds of the wave in the low loss limit are given, respectively, by the expressions

$$|v_p(\omega)|=\frac{c}{|\varepsilon_r|^{1/2}|\mu_r|^{1/2}}=\frac{c}{|1+\mathrm{Re}(\chi)|}\approx\frac{\omega^2}{|\omega^2-\omega_p^2|}c, \tag{24}$$

$$v_g(\omega)=\frac{c}{1+\mathrm{Re}(\chi)+\omega\partial_\omega[\mathrm{Re}(\chi)]}\approx\frac{\omega^2}{\omega^2+\omega_p^2}c, \tag{25}$$

where the real part of the susceptibility $\mathrm{Re}(\chi)=\mathrm{Re}[-\omega_p^2/\omega(\omega+i\Gamma)]\approx-\omega_p^2/\omega^2$. We also recall that the point in the frequency spectrum of the Drude model that isolates the DNG and the DPS regions occurs when $\varepsilon_r=\mu_r=0$, i.e., when $\omega=\omega_p$, hence when $f=\sqrt{2}f_0=42.42$ GHz. Well above this frequency, the permittivity and permeability are those of free space. Thus, the highest frequency components propagate causally in a medium that looks like free space. In particular, these components are the ones mainly responsible for the turn-on and turn-off portions of the envelope. The lower frequencies propagate with speeds slower than their value at the CW portion of the pulse, ω_0, i.e., $|v_p(\omega_0)|\approx c$.

To illustrate these effects, the spectrograms of the 1-1-1, 2-16-2, and 5-10-5 generated electric field pulses at $z=0.01$ m in the dispersive DNG medium were obtained. They are plotted in Figs. 7–9, respectively. The spectrograms were generated by taking Blackman-windowed Fourier transforms at every fourth sample in the signal data [26,27]. The number of signal samples analyzed by each transform was determined such that, after the data selection was windowed, the selection would represent at least a few and at most several wavelengths, depending on its frequency components. As a result, 512 signal samples were used in the transforms of the 5-10-5 pulse results, and 1024 signal samples were used in transforms of the 1-1-1 and 2-16-2 pulse results. Note that the spectrograms produced using a smaller signal slice in the Fourier transform became too noisy for interpretation, and larger signal slices resulted in

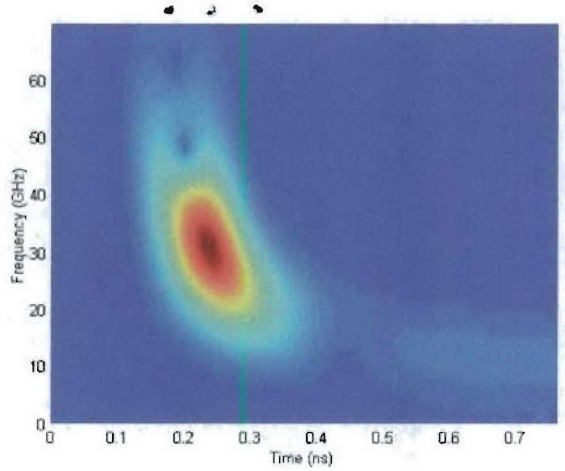

FIG. 7. (Color online) Spectrogram of the FDTD predicted field values in the DNG medium measured at the point $z=200\Delta z'$ for the 1-1-1 input pulse, where $\Delta z'=50$ μm.

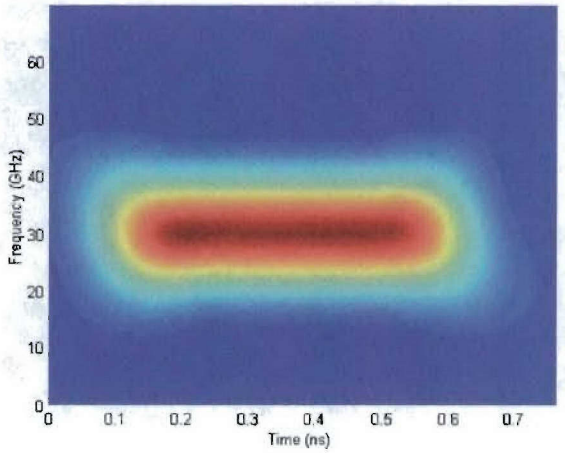

FIG. 9. (Color online) Spectrogram of the FDTD predicted field values in the DNG medium measured at the point $z=100\Delta z$ for the 5-10-5 input pulse, where $\Delta z=100$ μm.

broad areas in the spectrogram where the entire signal spectrum was displayed. The use of larger signal slices also shifted the estimates for the arrival times of various frequencies. The Fourier transforms were zero-padded to a size of 4096 to obtain sufficient resolution (<1 GHz sample spacing) near the center frequency (30 GHz) of the signals.

The resulting spectrograms shown in Figs. 7–9 then indicate, respectively, a good approximation of the frequency components present at the location $z=0.01$ m at various points in time for the 1-1-1, 2-16-2, and 5-10-5 cases. The three-cycle, 1-1-1 pulse spectrogram in Fig. 7 shows a negative slope. The higher frequencies propagate faster, hence they arrive sooner than the lower frequency components. While this behavior is present in the turn-on and turn-off portions of the results shown in Figs. 8 and 9, it is clearer in Fig. 8 since the 2-16-2 pulse has faster rates of change that are more closely related to the 1-1-1 case.

A plot of the Fourier transform output for data sections near the leading edge, central portion, and trailing edge of the FDTD DNG results for the 2-16-2 input pulse is shown in Fig. 10. All of the spectra are normalized to unity for comparison purposes. This figure clearly demonstrates that the leading edge of the pulse has significantly higher frequency content, the central portion of the pulse is dominated by the CW frequency, and the trailing edge of the pulse has significantly lower frequency content. Correlating this with Fig. 8, one then sees that those frequencies, hence the leading front of the pulse, arrive first to preserve causality.

To further investigate this dynamic pulse reshaping between the positive index of refraction (PIR) high frequency

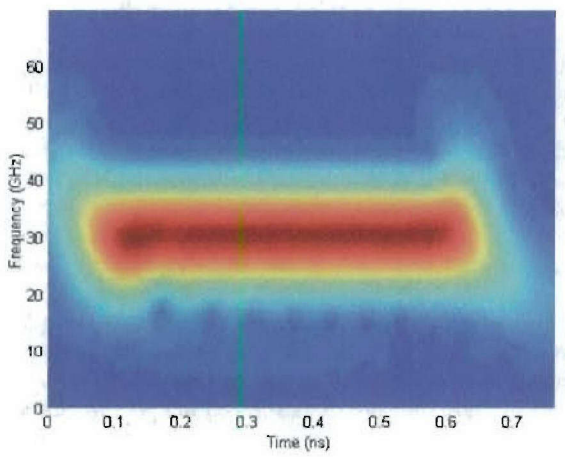

FIG. 8. (Color online) Spectrogram of the FDTD predicted field values in the DNG medium measured at the point $z=200\Delta z'$ for the 2-16-2 input pulse, where $\Delta z'=50$ μm.

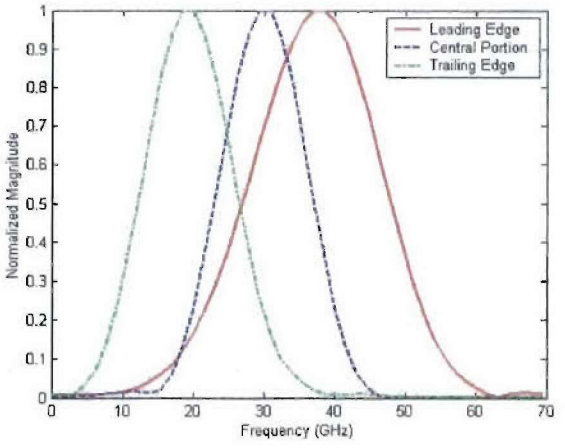

FIG. 10. (Color online) Normalized frequency composition of the (1) leading edge, (2) central portion, and (3) trailing edge of the DNG FDTD result measured at the point $z=200\Delta z'$ for the 2-16-2 input pulse, where $\Delta z'=50$ μm.

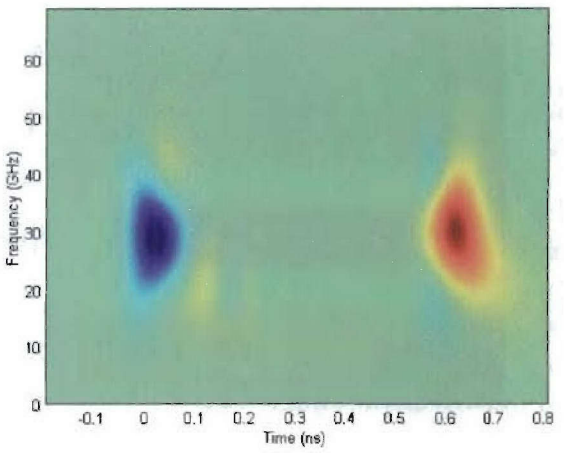

FIG. 11. (Color online) Spectrogram of the difference between the dispersive DNG 2-16-2 FDTD result shown in Fig. 8 and the corresponding approximate analytical solution. Red/dark region on the right (blue/dark region on the left) represents larger FDTD (approximate solution) values.

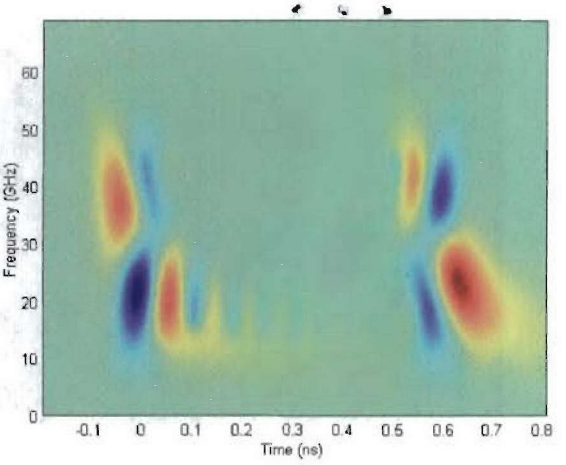

FIG. 12. (Color online) Spectrogram of the difference between the dispersive DNG 2-16-2 FDTD result shown in Fig. 8 and the corresponding approximate analytical solution when their CW portions are aligned. Red/dark region in the left-upper and right-lower portions of the left side and in the left-upper and right-lower portions of the right side of the figure represents larger FDTD values. Blue/dark region in the left-lower and right-upper portions of the left side and in the left-lower and right-upper portions of the right side of the figure represents larger approximate solution values.

components and the NIR center frequency and low frequency components, the differences between the spectrograms of the FDTD dispersive DNG results and the approximate solutions (23) were obtained. These spectrograms were also zero-padded at negative times to clarify the frequency components in the earliest portions of the signals. The 2-16-2 results measured at $z=0.01$ m are shown in Fig. 11. Although the FDTD DNG and the approximate solution signal fronts arrive at the same time and their CW portions are in very good agreement, the difference spectrograms filter out most of the CW behavior and reveal the importance of the apparent time delay and dispersive properties of the leading and trailing edges of the FDTD results. One finds that the CW portions of the FDTD results are not fully germinated until a time later than with the approximate solution. Similarly, the trailing edge of the FDTD result contains information that is delayed relative to the approximate solution. These delays reinforce similar observations made with regard to Figs. 5 and 6.

By aligning the CW portions of the FDTD and the approximate solution results and then generating the difference spectrogram, one can filter out all of the CW behavior and highlight the remaining low and high frequency behavior. The spectrogram for the 2-16-2 input pulse results measured at $z=0.01$ m is shown in Fig. 12. The transition from the off state to the CW state in the FDTD DNG data shows the necessary presence of higher frequency components very early in time to reconstruct the pulse front. The lower frequencies of the leading edge needed to complete the response in the DNG medium arrive later. This explains the slower turn-on of the FDTD results in comparison to the approximate solution. The trailing edge of the DNG FDTD result does contain significantly more of the lower frequency components while the presence of higher frequency components earlier in time aids the transition from the CW portion to the off state. This behavior adds delay to that already incurred with the turn-on portion of the pulse and explains the apparent lag in the turn-off of the FDTD results.

V. CONCLUSIONS

The one-dimensional electromagnetic problem of a current sheet source in a dispersive DNG medium was considered. A lossy Drude model of the DNG medium was used. The solution was generated numerically with the FDTD method. The analogous problem in a nondispersive DNG medium was also considered. It was shown that the solution to this problem is not causal in agreement with similar observations given in [5]. Approximate solutions that combined causal envelopes and the NIR properties of the nondispersive DNG medium were constructed; they were shown to compare well to the FDTD results for the dispersive DNG medium. It was thus demonstrated that causal results do indeed require the presence of dispersion in DNG media. Spectrograms of the FDTD results for the dispersive DNG medium and of the differences between those FDTD results and the approximate analytical solution revealed that the highest frequency components, which experience a DPS (PIR) medium, arrive first and are responsible for generating the front edge of the dispersive DNG pulse results. They also revealed that the trailing edge contained significant lower frequency components that see a DNG medium and have much slower wave speeds. The apparent delays in the germination of the CW portion of the pulse and in its termination were correlated with these dispersive effects. Thus, the dynamic reshaping of the DNG pulses occurs because of the different arrival times for the higher and lower frequency components.

It was also demonstrated that the CW portions of the

pulses do obey all of the NIR effects expected from a time-harmonic analysis in a nondispersive DNG medium. Thus, CW analyses of DNG media are credible as long as very narrow bandwidth pulse trains are considered for any practical realizations. This has been the case in all of the experimental results reported to date. Moreover, time delays for the realization of the NIR effects are inherent in the processes dictated by the dispersive nature of the physics governing them.

Even though the results presented here were derived for the broad bandwidth DNG Drude model, it is anticipated that the dynamic reshaping due to dispersive effects will also play a similar role in current metamaterial realizations of DNG media. This is particularly true of the planar implementations [15–18] that realize the Drude behavior. It should also be true for the nonplanar metamaterial constructs that involve highly resonant elements that achieve DNG properties in narrow frequency bands. In fact, this dynamic reshaping will occur at a faster rate since changes in the medium parameters from the DPS to the DNG states will be much more rapid. Furthermore, the dynamic reshaping must also occur at the interfaces between DPS and DNG media. The delays in the reflected pulse formation observed in [12] and [13] occur as the dispersive effects sort out the frequency components for the wavefronts and the appropriate propagation directions for the reflected and transmitted waves. Since most of the current nonplanar DNG metamaterial experiments are of the slab variety, these dynamic reshaping effects would be measurable if the experimental results were resolved in time.

It is interesting to note that the DNG results presented here can be related to Feynman's description of the interactions of electrons and positrons [28]. In particular, if one views the causal envelope propagating forward in time as an electron and the NIR CW wave propagating backward in time as a positron, then the positronlike CW wave is annihilated by the electronlike envelope at the pulse front. The scattered particle is the resulting wave that contains the NIR properties but is causal and propagating forward in time. The vertex interaction is smeared in time by the dispersive effects; it funnels the highest frequencies into the pulse front and properly distributes the higher and lower frequencies to provide the transition from the off-state to the CW on-state.

ACKNOWLEDGMENTS

This work was supported in part by DARPA under Contract No. MDA972-03-100.

[1] V. G. Veselago, Phys. Usp. **10**, 509 (1968).
[2] D. R. Smith, W. J. Padilla, D. C. Vier, S. C. Nemat-Nasser, and S. Schultz, Phys. Rev. Lett. **84**, 4184 (2000).
[3] R. A. Shelby, D. R. Smith, S. C. Nemat-Nasser, and S. Schultz, Appl. Phys. Lett. **78**, 489 (2001).
[4] R. W. Ziolkowski, IEEE Trans. Antennas Propag. **51**, 1516 (2003).
[5] D. R. Smith and N. Kroll, Phys. Rev. Lett. **85**, 2933 (2000).
[6] R. W. Ziolkowski and E. Heyman, Phys. Rev. E **64**, 056625 (2001).
[7] C. Caloz, C.-C. Chang, and T. Itoh, J. Appl. Phys. **90**, 5483 (2001).
[8] N. Engheta, Antennas Wireless Propagat. Lett. **1**, 10 (2002).
[9] A. Shelby, D. R. Smith, and S. Schultz, Science **292**, 77 (2001).
[10] J. A. Kong, B.-I. Wu, and Y. Zhang, Microwave Opt. Technol. Lett. **33**, 136 (2002).
[11] P. Kolinko and D. R. Smith, Opt. Express **11**, 640 (2003).
[12] R. W. Ziolkowski, Opt. Express **11**, 662 (2003).
[13] S. Foteinopoulou, E. N. Economou, and C. M. Soukoulis, Phys. Rev. Lett. **90**, 107402 (2003).
[14] J. B. Pendry, Phys. Rev. Lett. **85**, 3966 (2000).
[15] G. V. Eleftheriades, A. K. Iyer, and P. C. Kremer, IEEE Trans. Microwave Theory Tech. **50**, 2702 (2002).
[16] A. K. Iyer, P. C. Kremer, and G. V. Eleftheriades, Opt. Express **11**, 696 (2003).
[17] K. G. Balmain, A. A. E. Lüttgen, and P. C. Kremer, IEEE Antennas Wireless Propagat. Lett. **1**, 146 (2002).
[18] A. Grbic and G. V. Eleftheriades, J. Appl. Phys. **92**, 5930 (2002).
[19] J. Lu, T. M. Grzegorczyk, Y. Zhang, J. Pacheco, Jr., B.-I. Wu, J. A. Knog, and M. Chen, Opt. Express **11**, 723 (2003).
[20] Z. M. Zhang and C. J. Fu, Appl. Phys. Lett. **80**, 1097 (2002).
[21] L. Wu, S. He, and L. Chen, Opt. Express **11**, 1283 (2003).
[22] R. W. Ziolkowski and A. D. Kipple, IEEE Trans. Antennas Propagat., Special Issue on Metamaterials (to be published).
[23] J. Pacheco, Jr., T. M. Grzegorczyk, B.-I. Wu, Y. Zhang, and J. A. Kong, Phys. Rev. Lett. **89**, 257401 (2002).
[24] A. Taflove, *Computational Electrodynamics: The Finite-Difference Time-Domain Method* (Artech House, Inc., Norwood, MA, 1995).
[25] *Advances in Computational Electrodynamics: The Finite-Difference Time-Domain Method*, edited by A. Taflove (Artech House, Inc., Norwood, MA, 1998).
[26] R. S. Massey, S. O. Knox, R. C. Franz, D. N. Holden, and C. T. Rhodes, Radio Sci. **33**, 1739 (1998).
[27] A. V. Oppenheim and R. W. Schafer, *Discrete-Time Signal Processing*, 2nd ed. (Prentice Hall, Upper Saddle River, NJ, 1999).
[28] J. D. Bjorken and S. D. Drell, *Relativistic Quantum Mechanics* (McGraw Hill, New York, 1964), Chap. 5, pp. 63–75.

Appendix 7

Efficient Localization of Electromagnetic Fields Using a Plasmonic Nano-Cylinder Waveguide

EFFICIENT LOCALIZATION OF ELECTROMAGNETIC FIELDS USING A PLASMONIC NANO-CYLINDER WAVEGUIDE

Richard W. Ziolkowski

Department of Electrical and Computer Engineering, University of Arizona,
Tucson, Arizona, USA, ziolkows@ece.arizona.edu

The nanotechnology area holds much promise for the realization of ultra-small and ultra-fast devices with a variety of interesting applications. The optical sector is an immediate beneficiary of nanotechnologies. Consider, for instance, a coated nano-cylinder or nano-sphere whose outer radius is $a = 50\,nm$. At an excitation wavelength of $\lambda_0 = 500\,nm$, its radius is only $a = \lambda_0 / 10$. One could use arrays of such nano-cylinders or nano-spheres as inclusions in a substrate to realize optical metamaterials. On the other hand, because of their subwavelength sizes, such arrays could be used for several nano-photonic circuit concepts.

We have been studying analytically and with the finite difference time domain (FDTD) approach, the scattering of Gaussian beams from sets of nano-cylinders and nano-spheres. We selected the FDTD approach because of its versatility in the choices and configurations of materials and structures that can modeled. This versatility makes it an excellent candidate for studying the behavior of these ultra-small systems. One strong design goal has been to determine whether or not an array of plasmonic nano-cylinders could be used as an effective waveguide to localize efficiently an electromagnetic field to subwavelength dimensions.

It has been found theoretically and experimentally that by properly taking into account the actual material characteristics at optical frequencies and by properly designing the nano-structures, one can achieve enhanced scattering. In particular, the permittivity of metals such as gold and silver at visible (optical) frequencies can be described effectively by a lossy Drude (cold plasma) model. One finds that the real part of the permittivity of metals is in fact generally negative at those frequencies. On the other hand, it is readily shown from canonical plane wave scattering problems that the total scattering cross-section of an electrically small dielectric cylinder (sphere) (i.e., Rayleigh scatterers that are much smaller than the wavelength) is resonant for $\varepsilon = -1.0\varepsilon_0$ $(\varepsilon = -2.0\varepsilon_0)$. One typically realizes these *plasmon* resonance effects at optical frequencies and at nano-meter length scales with cylindrical (spherical) dielectric scatterers that are coated with metals whose permittivity is $\varepsilon_r = -1.0$ $(\varepsilon_r = -2.0)$ at the wavelength of interest,

We have used a finite difference time domain (FDTD) simulator to study the enhanced scattering from plasmonic cylinders [1]. Here, each cylinder is defined by an interior free space (and other) dielectric cylinder whose radius $r_1 = 40\,nm$ that is coated with a Drude metal shell whose outer radius $r_2 = 50\,nm$. The Drude metal coating has $\varepsilon_r = -1.0$ at $\lambda_0 = 500\,nm$, the wavelength of the optical field under consideration. We have combined together an array of these cylinders to form a nanometer waveguide environment that can capture a significant amount of power from an incident optical beam and can then transform it into a localized field with a highly subwavelength spot at the output of the waveguide. This geometry is shown in Fig. 1a. A parallel polarized Gaussian beam (electric field in the plane of the figure, magnetic field orthogonal to it), whose

waist is $2\lambda_0 = 1000\,nm$, is incident normally on the end of an 18 element, two column array of these plasmonic nano-cylinders. The longitudinal distance between the centers of the cylinders is $110\,nm$ giving a $10\,nm$ separation between them along the beam axis. The entire waveguide is thus slightly longer than $2\lambda_0$. The transverse distance between the centers of the cylinders is $120\,nm$ so that the channel formed between the edges of the cylinders is $20\,nm$. The distance between the outer edges of the cylinders is thus $220\,nm$. To isolate the input end of the waveguide and the incident beam from its output end, the cylinders are fed through a slot in a $10\,nm$ -thick perfect electric conductor (PEC) ground plane. The slot width is $170\,nm$. Detector arrays (*11* elements, each element being *6 nm* wide) are located $2r_2 = 100\,nm$ away from the ends of the waveguide. To further localize the field at the output end, the last two nano-cylinders have $10\,nm$ slots in their metallic shells [2, 3].

The incident beam is launched from a total-field/scattered field boundary into the FDTD simulation region. The beam scatters from and excites the input end of the waveguide. We are investigating the waveguide output fields as functions of all of the geometry and material factors. Snapshots in time of the magnetic field intensity distribution for the indicated plasmonic nano-cylinder waveguide configuration are shown in Fig. 1. The incident field is clearly coupled to the waveguide and the transmitted field is basically confined to the regions between the nano-cylinders. Each nano-cylinder is also seen to be strongly coupled to its nearest neighbors. Transverse cross-sections of the magnetic field intensity at various locations along the beam axis are shown in Fig. 2. The cross-sections before and just into the waveguide are given in Fig. 2a; those just after the output of the waveguide are given in Fig. 2b. The full width at half maximum of the intensity of the input beam is $235\,nm$. It is found that the intensity distribution across the front detector array has a $145\,nm$ full width at half maximum while it has a $120\,nm$ full width at half maximum across the back detector array. The output beam is seen to be highly localized near the waveguide. It has a $80\,nm$ full width at half maximum across the ouput face of the waveguide. This spot size is 6.25 time smaller than the wavelength. From the time histories obtained at the center of the front and back detectors it is found that the amplitude of the magnetic field at the back detector array is about 20% of its value at the front detector array and that it takes only about 15 cycles for this level to be reached. The plasmonic nano-cylinder waveguide array has thus localized the incident field with a reasonable efficiency to subwavelength spot sizes. These localized fields could be used for many nanoscale optical probe and photonic circuit applications.

REFERENCES

[1] R. W. Ziolkowski, "FDTD modeling of Gaussian beam interactions with metallic and dieletric mano-structures," *Proceedings of the 2004 URSI International Symposium on Electromagnetic Theory*, pp. 27-29, Pisa, Italy, May 23-27, 2004.

[2] Ziolkowski, R. W., W. A. Johnson, and K. F. Casey, "Applications of Riemann-Hilbert problem techniques to electromagnetic coupling through apertures," *Radio Sci.*, Vol. 19, pp. 1425-1431, Nov.-Dec.1984.

[3] Ziolkowski, R. W., and W. A. Johnson, "Electromagnetic scattering of an arbitrary plane wave from a spherical shell with a circular aperture," *J. Math. Phys.*, Vol. 26, pp.1293-1314, June 1987.

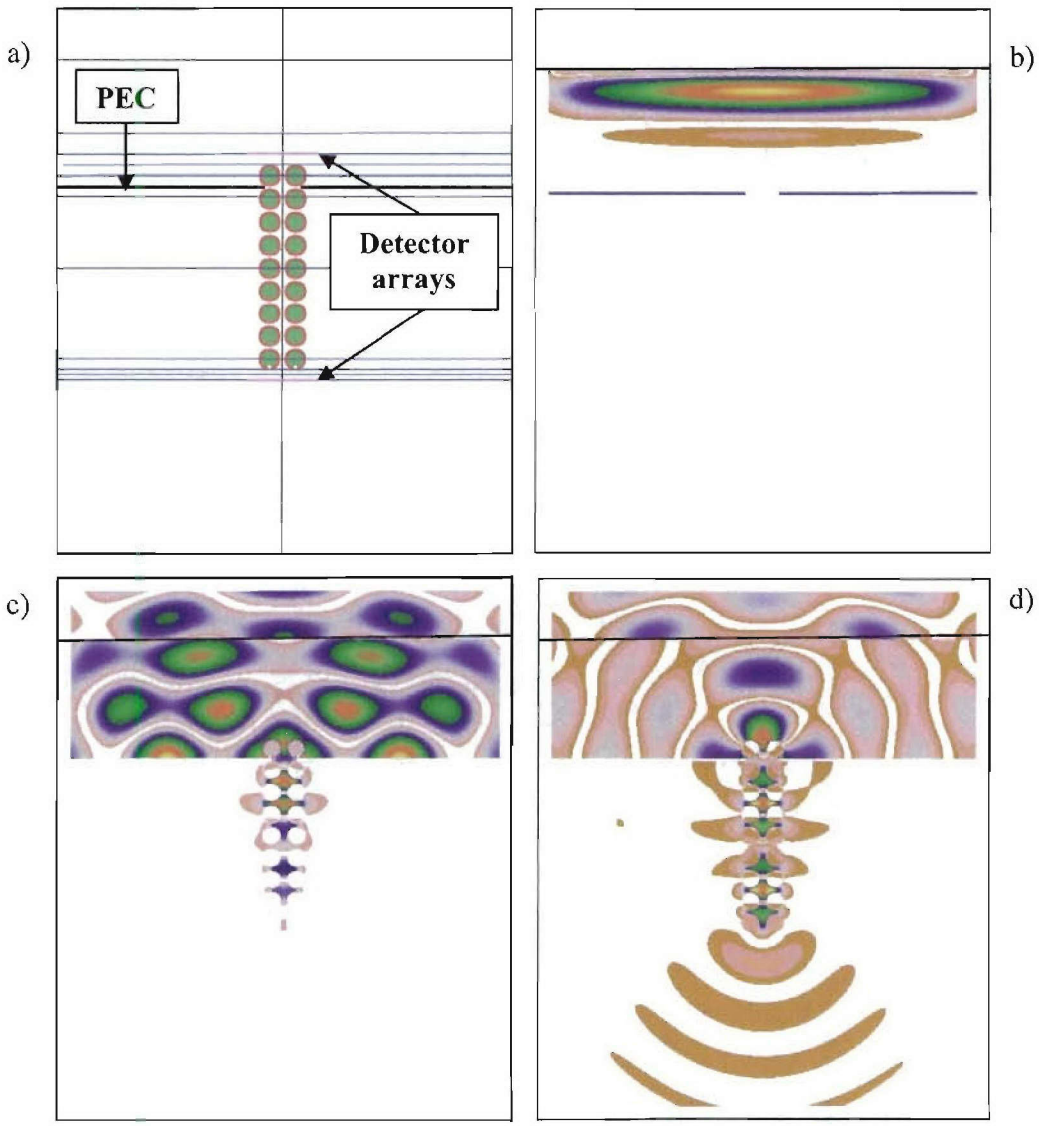

Fig. 1. Snapshots in time of the magnetic field intensity: a) $t = 0\ \Delta t$, b) $t = 150\ \Delta t$, c) $t = 1200\ \Delta t$, and d) $t = 3200\ \Delta t$. The field translated along the plasmonic nano-cylinder waveguide produces a localized distribution at the output face.

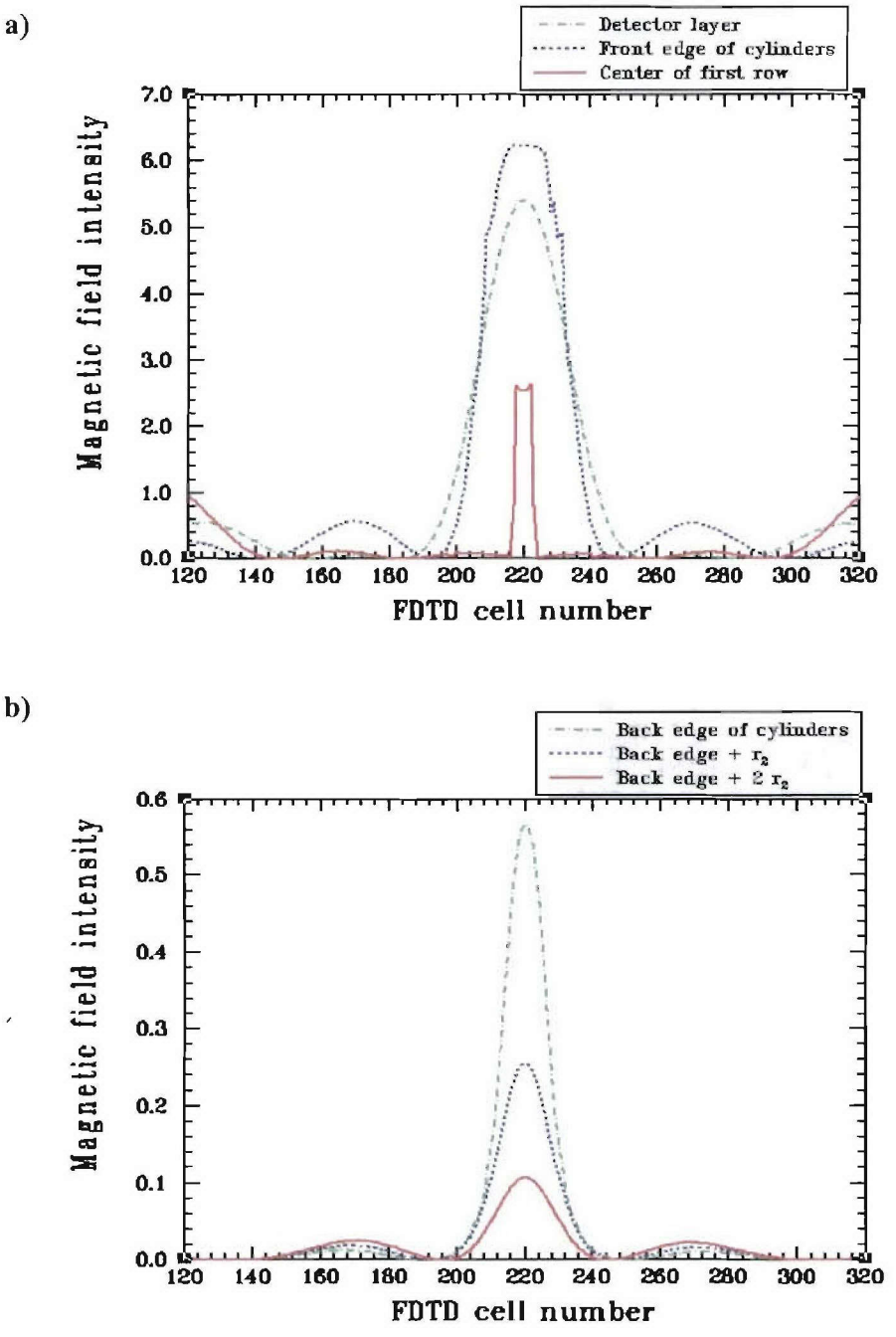

Fig. 2. Transverse cross-sections of the magnetic field intensity distribution at the (a) input and (b) output ends of the plasmonic nano-cylinder waveguide.