Shlomi Dolev
Mihai Oltean (Eds.)

LNCS 6748

Optical Supercomputing

Third International Workshop, OSC 2010
Bertinoro, Italy, November 2010
Revised Selected Papers

Springer

Lecture Notes in Computer Science 6748

Commenced Publication in 1973
Founding and Former Series Editors:
Gerhard Goos, Juris Hartmanis, and Jan van Leeuwen

Shlomi Dolev Mihai Oltean (Eds.)

Optical Supercomputing

Third International Workshop, OSC 2010
Bertinoro, Italy, November 17-19, 2010
Revised Selected Papers

Springer

Volume Editors

Shlomi Dolev
Ben Gurion University of the Negev
Beer-Sheva, Israel
E-mail: dolev@cs.bgu.ac.il

Mihai Oltean
Babeş-Bolyai University
Cluj-Napoca, Romania
E-mail: moltean@cs.ubbcluj.ro

ISSN 0302-9743 e-ISSN 1611-3349
ISBN 978-3-642-22493-5 e-ISBN 978-3-642-22494-2
DOI 10.1007/978-3-642-22494-2
Springer Heidelberg Dordrecht London New York

Library of Congress Control Number: 2011931465

CR Subject Classification (1998): F.1, B.4.3, C.5.1

LNCS Sublibrary: SL 1 – Theoretical Computer Science and General Issues

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

OCS, the International Workshop on Optical SuperComputing, is an annual forum for research presentations on all facets of optical computing for solving hard computation tasks. Optical computing devices have the potential to be the next computing infrastructure. The frequency limitations and cross-talk phenomena, as well as soft-errors of electronic devices on one hand, and the natural parallelism of optical computing devices, with the advance in fiber optics and optical switches on the other hand, make optical computing commercial-able. The focus of research is on the theory, design, specification, analysis, implementation, or application of optical supercomputers.

Topics of interest include, but are not limited to: designs or demonstrations of optical computing devices and systems; algorithmics and complexity issues of optical computing; computation representation by photons and holograms; neural and brain-inspired architectures; electro-optic devices for interacting with optical computing devices; practical implementations; analysis of existing devices and case studies; optical photonics and laser switching technologies; optical and photonic memories; optical signal processing subsystems; optical networks for high-performance computing; optical interconnections; quantum optical systems; applications and algorithms for optical devices; alpha particles; x-rays and nanotechnologies for optical computing.

The third OSC workshop was held during November 17–19, 2010, at the Bertinoro International Center for Informatics, BiCi, Italy.

This volume contains 13 contributions selected by the Program Committee. All submitted papers were read and evaluated by the Program Committee members. We are grateful to the EasyChair system for assisting the reviewing process.

OSC 2010 was organized in cooperation with OSA and Laserfest.

The support of Bertinoro International Center for Informatics, Ben-Gurion University and Babeş-Bolyai University are also gratefully acknowledged.

April 2011

Shlomi Dolev
Mihai Oltean

Organization

OSC, the International Workshop on Optical SuperComputing, is an annual forum for research presentations on all facets of optical computing. OSC 2010 was organized in cooperation with the OSA and Laserfest.

Steering Committee

H. John Caulfield	Fisk University, USA
Shlomi Dolev	Ben-Gurion University of the Negev, Israel
Yeshaiahu Fainman	University of California, San-Diego, USA
Tobias Haist	Universität Stuttgart, Germany
Mihai Oltean	Babeş-Bolyai University, Romania

Organizing Committee

Program Chair	Shlomi Dolev, Ben-Gurion University of the Negev, Israel

Program Committee

Hossin A. Abdeldayem	NASA, USA
George Barbastathis	MIT, USA
Yacob Ben-Aryeh	Technion Institute, Israel
Antonella Bogoni	CNIT, Italy
H.John Caulfield	Fisk University, USA
Shlomi Dolev	Ben-Gurion University, Israel
Yeshaiahu Fainman	University of California, USA
Dietmar Fey	University of Erlangen-Nürnberg, Germany
Debabrata Goswami	Indian Institute of Technology Kanpur, India
Tobias Haist	Universität Stuttgart, Germany
Zhanghua Han	University of Alberta, Canada
Jürgen Jahns	FU Hagen, Germany
Efstratios Kehayas	National Technical University of Athens, Greece
Boris Kryzhanovsky	Center of Optical Neural Technologies, SRISA RAS, Moscow
Shimon Levit	Weizmann Institute, Israel
Alastair McAulay	Lehigh University, USA
Kouichi Nitta	Kobe University, Japan
Mihai Oltean	Babeş-Bolyai University, Romania

Wolfgang Osten	Universität Stuttgart, Germany
Haldun Ozaktas	Bilkent University, Turkey
Joseph Rosen	Ben-Gurion University, Israel
Sukhdev Roy	Dayalbagh Educational Institute, India
Natan T. Shaked	Duke University, USA
Joseph Shamir	Technion Institute, Israel
Dan Tamir	Texas State University, USA
Kristof Vandoorne	Ghent University, Belgium
Damien Woods	Caltech, USA
Zeev Zalevsky	Bar-Ilan University, Israel
Xinliang Zhang	Huazhong University of Science and Technology, China

Table of Contents

Integrated Photonic Micro Logic Gate

Arkady Rudnitsky[1], Asaf Shahmoon[1], Menachem Nathan[2], Moshe Nazarathy[3], Bar Larom[3], Alexander Martucci[4], Luca Businaro[5], Annamaria Gerardino[6], and Zeev Zalevsky[1,*]

[1] School of engineering, Institute of Nanotechnology, Bar-Ilan University, Ramat-Gan, Israel
zalevsz@eng.biu.ac.il
[2] Faculty of Electrical Engineering, Tel-Aviv University, Israel
[3] Faculty of Electrical Engineering, Technion, Israel
[4] Università di Padova, Dipartimento Ingegneria Meccanica Set. Materiali, Padova, Italy
[5] LILIT Micro and nano fabrication Group, CNR-INFM TASC Laboratory, Trieste, Italy
[6] CNR – Inst. for photonics and nanotechnologies (IFM), Rome, Italy

Abstract. In this paper we present an approach for realizing an integrated all-optical logic gate. The basic principle is based upon stimulated emission process generated in an active gain medium while special interferometric photonic wave-guiding structure allows the realization of an integrated micro scale device. The operation rate of the proposed approach can theoretically reach tens of Tera-Hertz.

Keywords: All-optical devices; Photonic integrated circuits; Logic gates.

1 Introduction

Realizing optical integrated circuits is an important task due to both its scientific merit as well as its industrial applicability. The aim of this research is to generate all-optical integrated processing devices as modulators and logic gates capable of operating at high processing rates as well as being integrated with optics communication modules [1-6]. The ability to carry out concurrent large-scale on-chip information processing and multi-channel communication is valued in the information technology. All-optical systems may carry out these tasks at speeds or rates far exceeding electronic counterparts.

Typically, light beams which are to be processed by an all-optical device have a non-linearly interact with the electronic subsystem of the device's medium, so as to produce a certain non-linear effect (e.g. two and four waves mixing, frequency doubling, parametric oscillation, etc) on utilization of which the operation of the particular device is based. Accordingly, typical other approaches for realization of all-optical devices tend to use media with a large non-harmonicity of electronic oscillations. Typical all-optical devices also tend to require intense illumination and a large interaction length for proper operation [1-6].

* Corresponding author.

S. Dolev and M. Oltean (Eds.): OSC 2010, LNCS 6748, pp. 1–9, 2011.

In this paper we propose a new approach for an integrated all-optical logic gates capable of having nano and micro scale dimensions as well as ultra fast operation rates. The concept is based upon stimulated emission in pumped *gain medium* [7]. The technique utilizes the dependence of a signal propagating through a stimulated emission medium (herein called *gain medium*) on the gain. This dependence is typically non-linear with respect to the gain. The idea is that an input signals control the gain of the medium since the gain depends on the number of photons in the medium. Increasing the number of photons reduces the gain. Reference beam is passing through the gain medium and is coupled to the output of the device. Since the amplification of the medium is directly proportional to the logic inputs, a Boolean logic operation may be realized at the output of the device. Because a stimulated emission is involved the response rate of operation mechanism is very short (an immediate response).

The novelty of this paper includes also the design of special interferometric photonic wave-guiding structure which is a major component in the realization of an integrated micro scale logic device.

In section 2 we present the operation principle. Section 3 discuses the fabrication process of the device whose operation principle is being validated in section 4. The paper is concluded in section 5.

2 Operation Principle

The schematic sketch (see Fig. 1) for the basic all-optical logic gate contains a gain media which is pumped continuously from external source (not present in the Fig. 1). Reference beam is a continuously present beam. Output wave = Reference × G, where G is the gain factor of the gain media. There also two input signals to the logic gate. Those signals are used as the control signal for the gain media. If the control signal exists, the gain factor G is suppressed and the output beam exiting the gain module equals to zero. Only in the case when both of the control signals are zero (Input1 and Inout2 are both "0"), we have an amplified output beam coming from the output of the medium. This is a logical function "NOR". Although NOR is a universal Boolean operation in a sense that all other logic functions can be derived from it, the proposed concept can easily be used for realizing other logic functions as well. This to be further demonstrated in Section 4.

Mathematically the gain of a gain medium depends on the energy passing through it (in our case the pumping is always on) as follows:

$$\gamma = \frac{\gamma_0}{1 + \frac{I_v}{I_{sat}}} \tag{1}$$

where γ is the gain per unit length (overall gain G equals to $\exp(\gamma L)$ where L is the length of the gain medium), I_v is the energy in the gain medium and I_{sat} is the saturation energy level. From this equation it is simple to see that increasing the energy in the gain medium I_v, correspondingly reduces the overall gain.

The operation rate of the all-optical gate is very fast since, as seen in the rate equations for the amplification medium, the stimulated emission is a very fast

processes and basically depend only on the pumping power and the intensity developed in the medium (as presented in Eqs. 2 and 3) and not only on the spontaneous relaxation times:

$$\frac{dN_2}{dt} = R_p - \frac{N_2}{\tau_2} - (N_2 - N_1)W(\nu) \tag{2}$$

where N_2 and N_1 are the populations of the two levels between which the amplification is generated, R_p is the rate of pumping to the upper energetic level, τ_2 is the relaxation time related to the spontaneous processes, and W is related to the stimulated emission process and which is proportional to the intensity that is developed inside the medium:

$$W(\nu) = \frac{\kappa}{\nu^3} G(\nu) I_\nu \tag{3}$$

where κ is a constant, ν is the frequency, G is the gain spectral response and I_ν is the intensity in the gain medium.

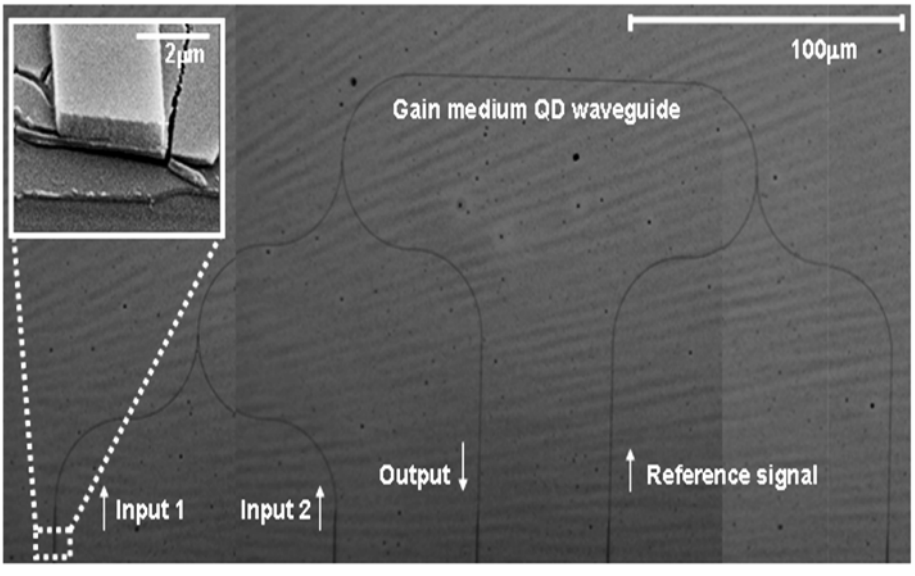


Fig. 1. Microscope image of the fabricated nano photonic device. On the upper left corner one may see the SEM image of the fabricated QD waveguide. The gain media is pumped continuously from external source which is not present in the figure.

3 Fabrication

The realization of an integrated all-optical circuit based upon gain medium can be obtained for instance by constructing optical waveguides in quantum dots (QD) photoresist [8,9]. The gain material is constantly pumped and thus following the rate equations of a gain medium, input signals passing through the medium at wavelength

corresponding to the gain/absorption spectral characteristics, will evoke immediate stimulated emission.

The microscope image of fabricated photonic chip following the above specified description may be seen in Fig. 1 (SEM image of the edge of the waveguide is seen in the upper left corner). The waveguide is made out of QD where on the left side we have the two logic inputs and on the right we input the reference beam. This reference is coupled to the right side of the chip where the output detector is located. The inputs affect the gain of the gain medium which controls the level of the reference beam that is going to the output. This yields, as previously explained, optical logic NOR gate.

The integrated chip option for realizing the proposed approach may include realization of the gain medium by fabricating ZrO_2 film (that was used as photo-resist) doped with CdSe@ZnS QDs which have a significant gain factor [8,9] over small interaction length. This fabrication process was applied to generate the chip seen in Fig. 1.

The X-ray diffraction (XRD) spectra of QD doped zirconia thin films at different annealing temperatures are seen in Fig. 2(a). A clear transition from amorphous to partially crystalline is observed above 400°C. Due to the low fill-factor of QDs in the sample (< 1% by volume) an effective medium model was not needed to account for the effect of the QD dispersion. Due to condensation of the matrix, a significant increase in the refractive index of the film and a concordant decrease in thickness were consistently observed with increasing annealing temperature. For example, following heat treatment at only 100°C, the average film refractive index n_g and thickness d (thin film of ZrO_2) were found to be 1.60 and 81nm respectively, while after annealing at 300°C these values changed to 1.75 (39 nm).

In addition to the usage of ZrO_2 thin films as the optical waveguide, there were attempts to fabricate waveguides out of TiO_2 film doped with CdSe@ZnS QD. For TiO_2 films following heat treatment at only 100 °C, the average film refractive index n_g and thickness d were found to be 1.69 and 54 nm respectively, while after annealing at 300 °C these values changed to 1.94 (31 nm). We found that in general n_g of TiO_2 films was 0.15 higher than values for ZrO_2 films at each studied annealing temperature. The XRD spectra of QD doped TiO_2 thin films at different annealing temperatures are seen in Fig. 2(b).

In Tables 1 and 2 one may find the summary of relevant parameters obtained for ZrO_2 and TiO_2 doped thin films respectively. In Tables 1and 2 n_d is the fully densified refractive index (anatase TiO_2 = 2.52, tetragonal ZrO_2 = 2.208) and n_g is the measured refractive index at the reference wavelength. The results in Tables 1 and 2 reveal that the overall film porosity significantly decreased for both matrices at higher annealing temperatures. This is expected due to the structural changes which occur during heating (i.e. removal of residual solvent and organics, to hydroxyl condensation and structural relaxation). The extent of densification of the ZrO_2 matrix was found to be very similar to that of the TiO_2 matrix at each studied annealing temperature. This similarity in densification rate led to the ZrO_2 matrix retaining a relatively high n_g value in comparison to TiO_2 at all the treatment temperatures.

From the fitted refractive index profiles and the film thickness, an estimation of the porosity for each sample was calculated using the Bruggeman model:

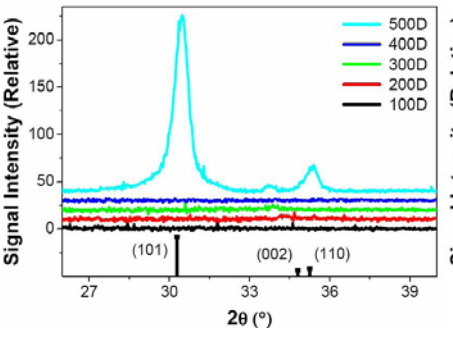
$$Porosity\,(\%) = \left(\left(\frac{n_d^{\,2} - n_g^{\,2}}{n_d^{\,2} + 2n_g^{\,2}} \right) \Big/ \left(\frac{n_d^{\,2} - n_g^{\,2}}{n_d^{\,2} + 2n_g^{\,2}} - \frac{1 - n_g^{\,2}}{1 + 2n_g^{\,2}} \right) \right) 100\% \tag{4}$$

Table 1. Summary of relevant parameters obtained for the ZrO_2 doped thin films

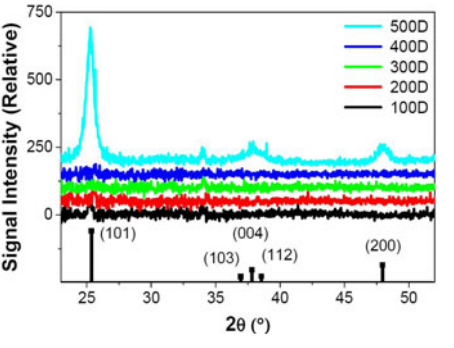
Matrix	n_e	Film Thickness (nm)	Porosity (%)	Min d/λ_0 Value for Guiding
ZrO_2-QD-100	1.601	81	47.53	0.234
ZrO_2-QD-200	1.683	56	41.35	0.165
ZrO_2-QD-300	1.753	39	36.05	0.132
ZrO_2-QD-400	1.794	30	32.93	0.118
ZrO_2-QD-500	1.829	27	30.26	0.108

Table 2. Summary of relevant parameters obtained for the TiO_2 doped thin films

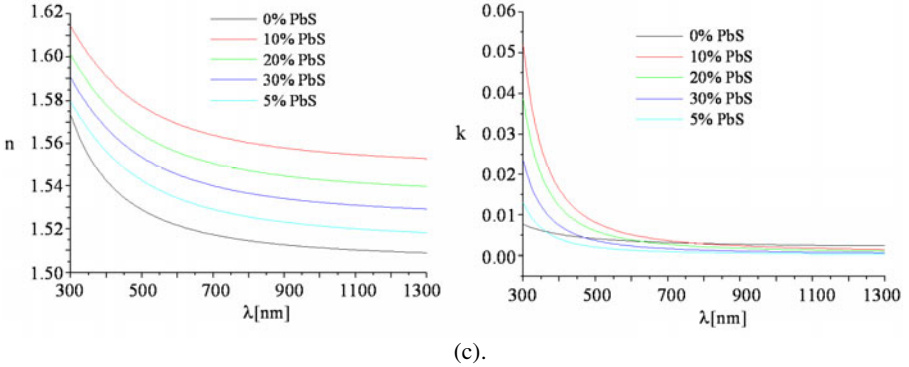
Matrix	n_e	Film Thickness (nm)	Porosity (%)	Min d/λ_0 Value for Guiding
TiO_2-QD-100	1.693	54	50.99	0.160
TiO_2-QD-200	1.849	37	41.87	0.103
TiO_2-QD-300	1.935	31	36.81	0.085
TiO_2-QD-400	1.963	27	35.15	0.081
TiO_2-QD-500	2.006	24	32.59	0.074



(a).



(b).



(c).

Fig. 2. (a). and (b). XRD spectra of ZrO_2 and TiO_2 films doped with CdSe@ZnS QDs annealed at 100, 200, 300, 400 and 500 °C, respectively. (c). Ellipsometry (real part n and imaginary part k of the refraction index) measurements on ZrO_2 film doped with PbS QDs annealed at 130 °C and containing 0, 5, 10, 20 and 30% of PbS.

In another fabrication attempt we used ZrO_2 film doped with PbS rather than CdSe@ZnS QD. The ellipsometry measurements (the real n and the imaginary k parts of the refraction index) on ZrO_2 film doped with PbS QD annealed at 130 °C and containing 0, 5, 10, 20 and 30% of PbS may be seen in Fig. 2(c). The introduction of lead sulphide increases both n (the real part of the refraction index) and k (the imaginary part of the refraction index).

4 Experimental Testing

Although we have fabricated the chip (Fig. 1), for the preliminary results of the experimental investigation we used the optical semiconductor amplifiers (SOA) [10,11] as the gain medium while the input signals are input through optical fibers. The bits of one of the input channels are delayed in comparison to the other. The purpose of the delay is to generate relative shift of half the pulse in order to demonstrate that even if the pulses are super imposed one on top of the other and the overall amplitude is increased, the output of the proposed device remains the same (due to the saturation of the gain medium). This is one of the very important advantages gained by the proposed approach. In the experimental investigation the modulation rate was 15MHz and the pulses width was 15nsec. As previously mentioned the SOA gain medium was constantly pumped.

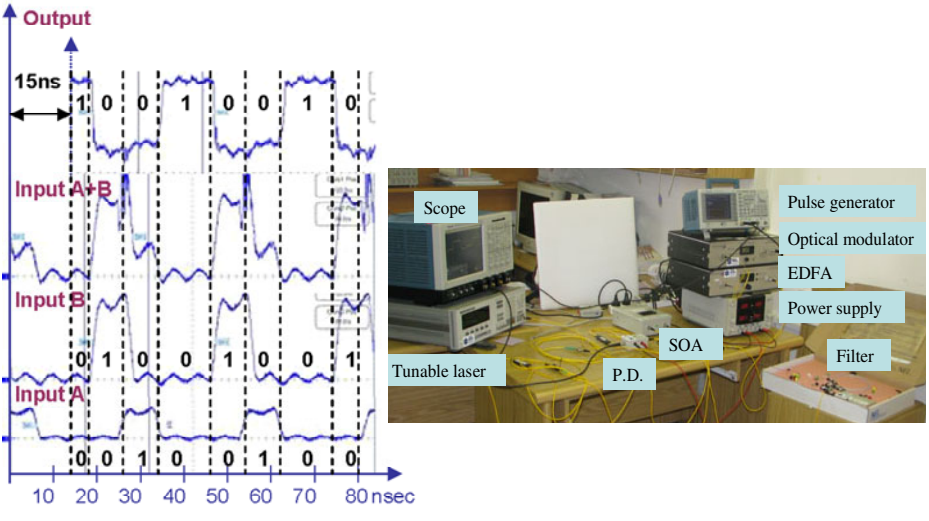


Fig. 3. Experimentally obtained preliminary results for all-optical logic NOR gate (using SOA). In the right part of the figure one may see image of the experimental setup as constructed in our laboratory.

In Fig. 3 one may see the measurements of input A, input B, both inputs together and the obtained output for that case respectively (going from the lower signals sequences presented in Fig. 3, upwards). From the results seen in Fig. 3 one may

conclude that only when both inputs A and B are zero the output is "1" otherwise it is "0". This is exactly the realization of the logic NOR gate operation. Note that the output channel as it is presented in Fig. 3 and which was captured by our scope was deliberately shifted in the time axis a shift of about 15nsec. We did this shift in the time axis in order to correct (or compensate) the temporal delay of about 15nsec (marked in the figure) that was generated due to the difference between the lengths of the cable that we used when connecting the input signals and the cable connecting the output of our device, to the scope channels.

In the right part of Fig. 3 we present the experimental setup that we used to produce the experimental results shown in the left part of the figure. The various electronic and optical components of the constructed measurements setup are indicated in the figure with proper labels.

Note that the proposed approach can be used for realizing different types of logic gates and not only the Boolean function of NOR. Figure 4(a) relates to the realization of a logic AND gate. The general structure for any logic function contains two parts: an integrated waveguide based interferometer and a gain medium. The output of the interferometer is input to the gain medium. In order to modify the Boolean function all that is required it to change the interferometer part while the gain medium module will remain unchanged. Thus, the difference between the AND gate in comparison to the previously discussed NOR gate is only in the interferometer part.

The interferometer part in the case of an AND gate includes three inputs: the two logic inputs (I_{in1} and I_{in2}) and a reference (I_{REF2}). The output of the interferometer is input to the gain medium. Another reference beam (I_{REF1}) is input to the gain medium and passed to the overall output of the logic gate (I_{out2}). The output of the

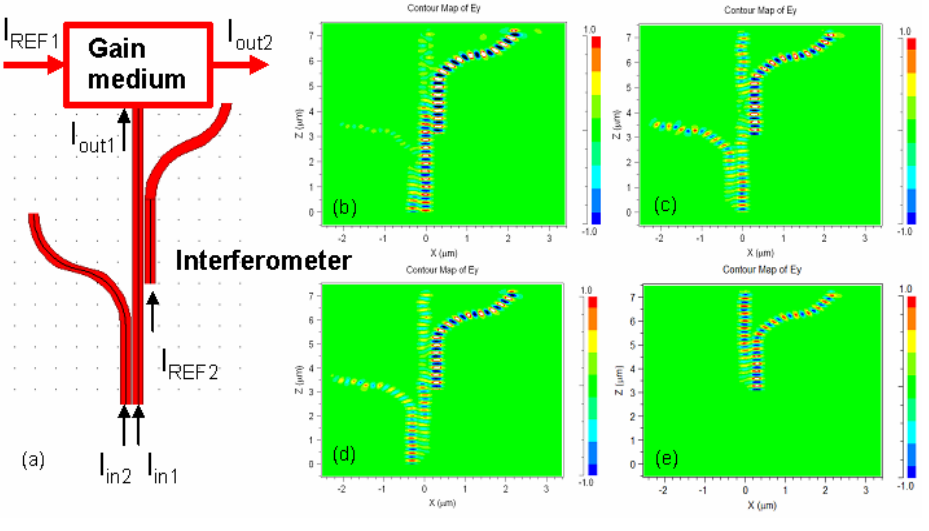


Fig. 4. (a). Schematic sketch of an AND logic gate. The interferometer which is later on simulated has inputs in its lower part and one output in its upper part. The right input is a reference. (b)-(e). Simulations of a nano metric interferometer. (b). I_{in1}= I_{in2}="1"; I_{out1}="0". (c). I_{in1}="1", I_{in2}="0"; I_{out1}="1". (d). I_{in1}="0", I_{in2}="1"; I_{out1}="1". (e). I_{in1}= I_{in2}="0"; Iout1="1".

Interferometer (I_{out1}) injects photons to the gain medium and therefore controls the gain level of the medium i.e. the gain applied over I_{REF1}. This results in obtaining the desired output from the device for proper combinations of the two logic inputs to the gate.

In Fig. 4(b)-4(e) we present the simulations of the interferometer device. All simulations presented in this section were done using numerical software of R-Soft using a Finite Difference Time Domain (FDTD) numerical approach in order to solve Maxwell's equations.

As seen from the simulations, the interferometer is designed such that its output equals to I_{REF2}- I_{in1}- I_{in2} (which is needed to obtain the overall AND functionality) while the energy of I_{REF2} is twice the energy of the two inputs such that only when both of them are present the output of the interferometer becomes zero (this by itself is similar to NAND functionality but when this logic combination illuminates the gain medium due to the fact that gain is reduced when the intensity of relevant photons is increased, the output of the overall device, i.e. the output of the gain medium, is logic one, i.e. a functionality of logic AND gate).

5 Conclusions

In this paper we have presented new concept of realizing ultra fast all-optical logic gates based upon the integration of a gain medium inducing stimulated emission and by that realizing the logic gate operation. Preliminary experimental results based upon SOA device as well as preliminary fabrication attempts of an integrated chip based upon of ZrO_2 and TiO_2 waveguides doped with quantum dots and their optical characterization were demonstrated.

References

1. Yabu, T., Geshiro, M., Kitamura, T., Nishida, K., Sawa, S.: All-optical logic gates containing a two-mode nonlinear waveguide. IEEE J. Quantu, Electron. 38, 37–46 (2002)
2. Yanik, M.F., Fan, S., Soljacic, M., Joannopoulos, J.D.: All-optical transistor action with bistable switching in a photonic crystal cross-waveguide geometry. Opt. Lett. 28, 2506–2508 (2003)
3. Buhl, L.L., Alferness, R.C.: Ti:LiNbO3 waveguide electro-optic beam combiner. Opt. Lett. 12, 778–780 (1987)
4. Lee, S.Y., Darmawan, S., Lee, C.W., Chin, M.K.: Transformation between directional couplers and multi-mode interferometers based on ridge waveguides. Opt. Exp. 12, 3079–3084 (2004)
5. Nagai, S., Morishima, G., Inayoshi, H., Utaka, K.: Multimode interference photonic switches. IEEE J. Lightwave Technol. 20, 675–681 (2002)
6. Zalevsky, Z., Rudnitsky, A., Nathan, M.: All-optical devices and methods for data processing. Patent application # 166810 (2005)
7. Zalevsky, Z., Rudnitsky, A.: Devices and methods for optical signal control. Patent application #12/401 779 (2007)

8. Jasieniak, J.J., Fortunati, I., Gardin, S., Signorini, R., Bozio, R., Martucci, A., Mulvaney, P.: Highly efficient amplified stimulated emission from CdSe-CdS-ZnS quantum dot doped waveguides with two-photon infrared optical pumping. Adv. Mater. 20, 69–73 (2007)
9. Jasieniak, J.J., Pacifico, J., Signorini, R., Chiasera, A., Ferrari, M., Martucci, A., Mulvaney, P.: Luminescence and amplified stimulated emission in CdSe-ZnS-nanocrystal-doped TiO2 and ZrO2 waveguides. Adv. Funct. Mater. 17, 1654–1662 (2007)
10. Kim, Y., Kim, J.H., Jeon, Y.M., Lee, S., Woo, D.H., Kim, S.H., Yoon, T.: All-optical flip-flop based on optical bistability in an integrated SOA/DFB-SOA. In: Sawchuk, A. (ed.) Optical Fiber Communications Conference. OSA Trends in Optics and Photonics (Optical Society of America, 2002), vol. 70 (2002) paper TuF5
11. Ono, H., Yamada, M., Shimizu, M.: S-Band EDFA with multistage configuration design, characterizations and gain tilt compensation. J. of Lightwave Tech. 21, 2240–2247 (2003)

An Optical System for Prime Factorization Based on Parallel Processing

Kouichi Nitta and Osamu Matoba

Department of Systems Science
Graduate of System Informatics, Kobe University
Rokkodai-cho 1-1, Nada-ku, Kobe, Hyogo, 657-8501, Japan
{nitta,matoba}@kobe-u.ac.jp

Abstract. A method for optical parallel processing in an algorithm for prime factorization is proposed as a brief announcement. In this method, amplitude modulation is utilized whereas our conventional method is based on phase modulation. The proposed method is suitable for the optical Fourier transform. This feature is considered to be useful in prime factorization. Basic concept of the proposed method is shown in this report.

Keywords: Prime factorization, Spatial parallel processing, Amplitude modulation.

1 Introduction

Some solutions for problems with computational hard is one of the main research topics in the field of optical supercomputing. As is well known, various methods have been reported for the travelling salesman problem, the Hamilton path problem, and so on [1-3].

We have also proposed a method for optical modulo operation [4]. Phase modulation of light wave is utilized in the proposed method. The method can treat modulo multiplication which is an important element in the algorithm for prime factorization. It also executes large scale parallel processing with simple optical hardware. Some optical implementations based on the proposed method have been demonstrated [5] and a system to solve prime factorization has been developed [6]. However, the method requires complex electronic processing to prepare a set of modulo exponentiation from results of optical processing.

In this presentation, we present a novel optical method for prime factorization. This method is based on optical amplitude modulation. Sequence of 4f optical systems is utilized for implementation. First, brief procedure for the prime factorization is described. Next we show simple numerical analysis to demonstrate the proposed method. From this analysis, it is shown that the proposed method is useful for prime factorization.

S. Dolev and M. Oltean (Eds.): OSC 2010, LNCS 6748, pp. 10–15, 2011.

2 Prime Factorization with Modulo Exponentiation

In our study, prime factorization is defined as the process to derive two different prime factors p and q from N (N=pq). The contents of an algorithm used in our method are briefly described as following. First, a specific operation called modulo exponentiation is obtained in $1 \leqq x \leqq 2N^2$-1. Modulo exponentiation $f(x)$ is given by Eq. (1)

$$f(x) = a^x \bmod N \tag{1}$$

In Eq. (1), a is a positive integer and satisfied with Eqs. (2) and (3).

$$1 < a < N \tag{2}$$

$$\gcd(a,N) = 1 \tag{3}$$

Here, $f(x)$ is known to be a periodic function. Also, the period is an integer. The period is given by analysis with the Fourier transform. Using the obtained period r, p and q can be derived as shown in Eqs. (4).

$$\begin{aligned} p &= \gcd(a^{r/2} - 1, N) \\ q &= \gcd(a^{r/2} + 1, N) \end{aligned} \tag{4}$$

This procedure is the same as that of the Shor's quantum algorithm [8]. In the algorithm, $f(x)$ is derived by quantum circuits for parallel processing. Many quantum gates are required in the circuits.

On the other hand, optical signal processing can achieve some complex operations. For example, the Fourier transform is executed by single lens. Based on the characteristics of optical processing, we have proposed parallel processing for modulo multiplication [7]. Modulo multiplication is important to obtain $f(x)$ as described in Sec. 3. In the conventional method, phase modulation of plane wave is utilized. Also, we have reported a system for prime factorization and some improvement of the system [4]. In our conventional method based on optical phase modulation, $f(x)$ is obtained with an optical set up for modulo multiplication. In this case, we have to derive a set of pixels corresponding x in preprocessing. Huge computational costs are required in the preprocessing.

3 Optical Parallel Processing with Amplitude Modulation

This section describes a novel optical parallel method. Let us consider a procedure to obtain a set of $f(x)$. As preprocessing, $l(i)$ defined as Eq. (5) is calculated in i=0, 1, 2, …, n-1.

$$l(i) = a^{2^i} \bmod N \tag{5}$$

Using $l(i)$, $f(x)$ is rewritten as the following equation.

$$f(x) = \prod_{i=0}^{n-1} l(i)^{x_i} \bmod N \tag{6}$$

Here, n is bit length of N. Also, x_i shows an i'th bit value of x. For example, (x_0, x_1, x_2, x_3, x_4) is (0, 1, 1, 0, 0) at x=6. Figure 1 shows principle of the procedure. As described in the figure, a set of $l(i)^{x_i}$ is prepared. And, inductive processing described in Eq. (7) is executed.

$$f_k(x) = \left\{ l(k-1)^{x_{k-1}} f_{k-1}(x) \right\} \bmod N \tag{7}$$

As results of n times iterations, $f(x)$ is obtained.

Our optical method is similar to the procedure. A schematic diagram of the method is shown in Fig. 2 (a). From the figure, optical hardware for the method consists of n'th sequence of 4-f optical systems. In this hardware, spatial amplitude modulators are put at all image planes. Fig. 2 (b) shows a diagram of distribution of amplitude transmittance on the modulators. These distributions in k'th modulator correspond with patterns of $l(k)^{x_k}$ shown in Fig. 1. Here $t(k)$ should be designed to be satisfied with Eq. (8).

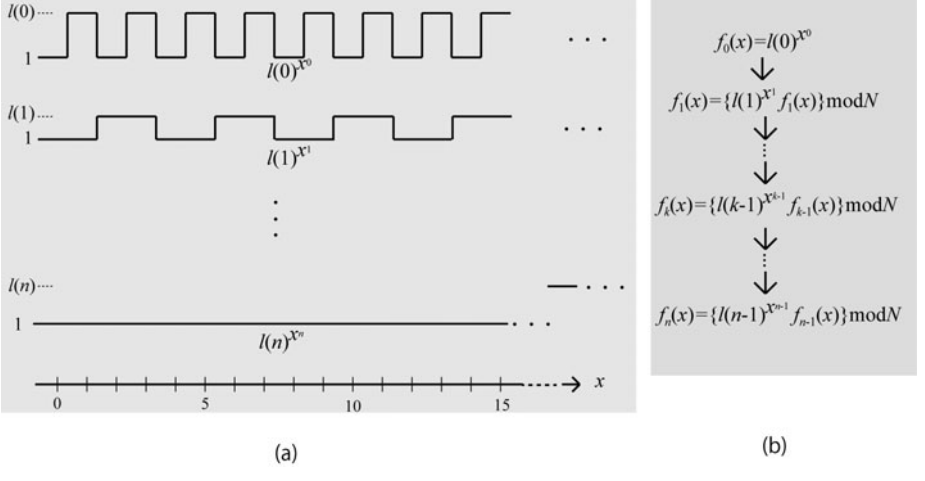


Fig. 1. (a) Structure of $l(i)^{x_i}$ and (b) the procedure to derive $f(x)$

$$t(0) - t_b : t(1) - t_b : \ldots : t(k) - t_b : \ldots : t(n-1) - t_b = l(0) : l(1) : \ldots : l(k) : \ldots : l(n-1) \tag{8}$$

Patterns measured at the detector plane correspond with distribution represented as Eq. (9).

$$f'(x) = \prod_{i=0}^{n-1} l(i)^{x_i} \tag{9}$$

In comparison with Eqs. (6) and (9), our optical method cannot implement modulo operations. However, our method seems to be useful to extracted the period of $f(x)$ as described in the next section.

4 Numerical Analysis

Usefulness of our proposed method is confirmed by numerical analysis. In the analysis, t_b is set to 0.5. $t(k)$ is determined in accordance with Eq. (10).

$$t(k) = \frac{l(k)}{\max\{l(i)\}} \times 0.5 + 0.5 \tag{10}$$

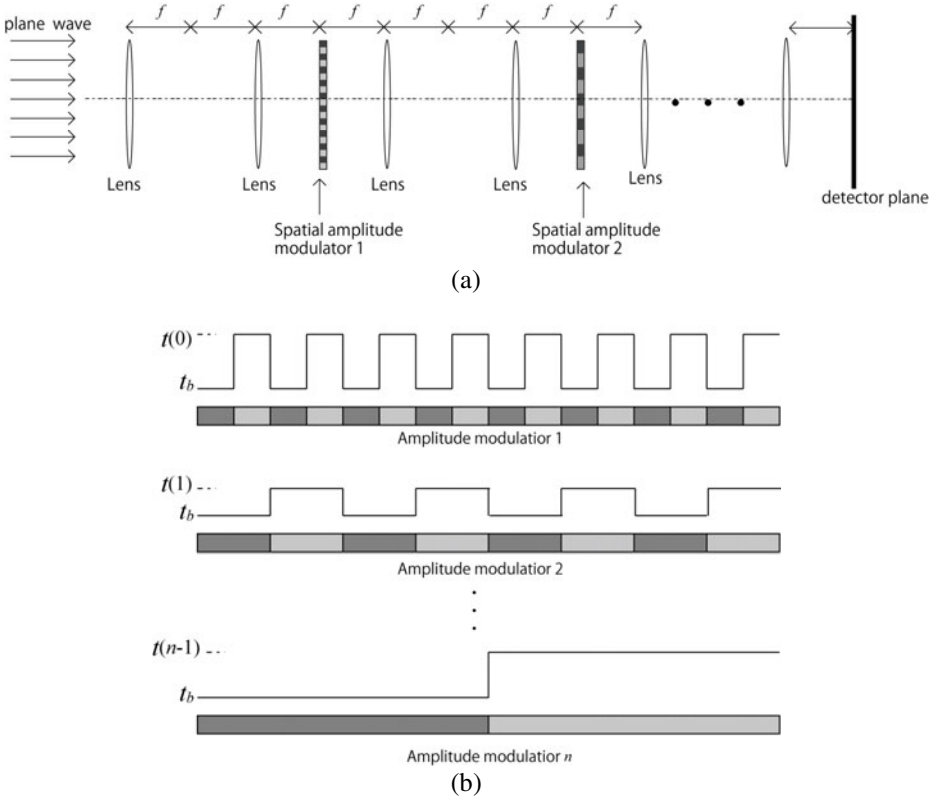
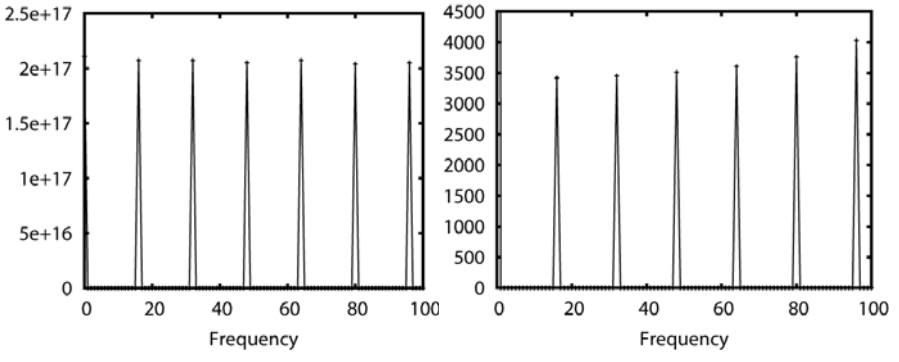


Fig. 2. Diagram of a scheme for of 2D binary images with image compression

Fig. 3 shows an example of results in the analysis. N and a are set to 203 and 106, respectively. Fig. 3 (a) is the part of the power spectrum of the result provided by the proposed procedure, and (b) is that of $f(x)$. From these figures, peak positions in the both profiles are the same coordinates in the horizontal direction. With analysis of

peak positions in the power spectrum, it is found that the value of r is 28. From the value, two prime numbers 7 and 29 are derived from Eq. (4). This analysis is quite same as the post processing in Shor's algorithm. Therefore, it is shown that the proposed procedure seems to be useful in large scale information processing for prime factorization. Note that aberration of lens and the point spread function of the optical system are not considered.



(a) (b)

Fig. 3. (a) Parts of power spectrum of a profile given by our method at (N, a)=(203, 106), and (b) that of $f(x)$

5 Discussions

As optical processing for factorization, the TWINKLE device is mentioned [8]. The TWINKLE gives two prime factors in accordance with the quadratic sieve (QS) algorithm and the number field sieve (NFS) one. These two algorithms are known to be practical for prime factorization in the present technology. Therefore, this device is considered to be suitable for high speed processing with optical and electronic implementations.

On the other hand, the purpose of Shor's quantum method is achievement of the system to solve prime factorization with polynomial time costs. The goal of our research is the same as that. Parallel processing is our optical system seems to be more practical than that based on quantum parallel processing. This is because physical implementation of massive data processing for quantum operations is difficult due to limitation of decoherence time.

Therefore, we should estimate computational costs of the proposed method. However, it is not found that our method can treat huge scale processing. Before the estimation, there are many issues in our research. Strictly speaking, the method cannot gives us $f(x)$ correctly. Also, influence on point spread function of imaging optics and that on unavoidable misalignment should be discussed. Based on the discussions, we would be able to clarify computational costs required in our optical procedure.

6 Comments and Summary

A novel optical method for the solution for prime factorization has been proposed as a brief announcement. By numerical analysis, it is demonstrated that the method can provide correct results. One of the features of this method is to implement simple optical set up. Another is suitability of optical Fourier transform. The latter is effective in comparison with our conventional method. Accuracy of our method should be estimated with detail analysis. Also, experimental demonstration is required.

References

1. Shaked, N.T., Messika, S., Dolev, S., Rosen, J.: Optical solution for bounded NP-complete problems. Appl. Opt. 46, 711–724 (2007)
2. Haist, T., Osten, W.: An Optical Solution For The Traveling Salesman Problem. Opt. Express 15, 10473–10482 (2007)
3. Oltean, M.: Solving the Hamiltonian path problem with a light-based computer. Natural Computing 7, 57–70 (2008)
4. Nitta, K., Matoba, O., Yoshimura, T.: Parallel processing for multiplication modulo by means of phase modulation. Appl. Opt. 47, 611–616 (2008)
5. Nitta, K., Katsuta, N., Matoba, O.: An Optical Parallel System for Prime Factorization. Jpn J. Appl. Phy. 48, 09LA02-1-5 (2009)
6. Nitta, K., Katsuta, N., Matoba, O.: Improvement of a system for prime factorization based on optical interferometer. In: Dolev, S., Oltean, M. (eds.) OSC 2009. LNCS, vol. 5882, pp. 124–129. Springer, Heidelberg (2009)
7. Shor, P.: Algorithms for quantum computation: Discrete logarithms and factoring Algorithms for quantum computation: Discrete logarithms and factoring. In: Proc. 35th Ann. Symp. on Foundations of Comput. Sci., vol. 1898, pp. 124–134 (1994)
8. Shamir, A.: Factoring Large Numbers with the TWINKLE Device. In: Koç, Ç.K., Paar, C. (eds.) CHES 1999. LNCS, vol. 1717, pp. 2–12. Springer, Heidelberg (1999)

Optical Graph 3-Colorability

Sama Goliaei and Saeed Jalili

SML Lab, Electrical and Computer Engineering Department,
Tarbiat Modares University, Tehran, Iran
{goliaei,sjalili}@modares.ac.ir

Abstract. The graph 3-colorability problem is a decision problem in graph theory which asks if it is possible to assign a color to each vertex of a given graph using at most three colors, satisfying the condition that every two adjacent vertices have different colors. It has been proved that the graph 3-colorability problem belongs to NP-complete class of problems which no polynomial resources solution is found for them yet.

In this paper, a novel optical solution to the graph 3-colorability problem is provided. In this solution, polynomial number of black filters are created in preprocessing phase each of which has exponential size and requires exponential time to be created. After preprocessing phase, the provided solution takes $O(n + m)$ time to decide if a given graph is 3-colorable or not, where the given graph has n vertices and m edges.

Keywords: Unconventional Computing, Optical Computing, NP-Complete, Graph Coloring, Graph 3-Colorability.

1 Introduction

Graph coloring problems is a class of problems in graph theory seeking to assign a color to each vertex of a given graph in such a way that every two adjacent vertices have different colors. Graph coloring problems arise from many real-world applications such as scheduling [1] making it necessary to find efficient solutions for these problems.

Graph 3-colorability problem is one of the graph coloring problems having specific applications in resource allocation and scheduling [2]. The graph 3-colorability problem asks if it is possible to assign a color to each vertex of a given graph in such a way that every two adjacent vertices (two vertices which are connected via an edge) have different colors, using at most three colors. It has been proved that 3-colorability problem is an NP-complete problem and like every other NP-complete problems, no polynomial resources solution is found yet.

Light is a natural phenomenon used in computation because of its special physical properties such as its parallel motion. Many NP-complete problems have recently investigated in optical computing such as the 3-SAT problem [3], the Hamiltonian path problem [4], the exact cover problem [5], the subset sum problem [6], the maximum clique problem, the vertex cover problem, the partition

S. Dolev and M. Oltean (Eds.): OSC 2010, LNCS 6748, pp. 16–22, 2011.

problem, the 3D-matching problem, the permanent problem, and the traveling salesman problem (TSP) [7].

New computational capabilities of optical computing in comparison to conventional computing have resulted to obtain more efficient solutions for these NP-complete problems, brings the idea that using optical computing to solve 3-colorability problem will also result to obtain more efficient solutions. Although the 3-colorability problem have been investigated in other branches of unconventional computing such as DNA computing [8] and quantum computing [9], but it seems that no optical solution for the graph 3-colorability problem (based on natural properties of light) is provided yet.

In this paper, a novel optical solution for the graph 3-colorability problem is provided. The solution takes just polynomial time to solve each problem instance, but exponential time in preprocessing phase. In the next section, the graph 3-colorability problem is defined. Provided optical solution and its complexity analysis are explained in sections 3 and 4, respectively. Finally, the conclusion of the paper is provided in section 5.

2 The Graph 3-Colorability Problem

In graph theory, graph coloring means to assign a color to each vertex of a graph. A proper graph coloring is assigning different colors to every two adjacent vertices (two vertices are adjacent if and only if they are connected via an edge). A proper 3-coloring for a given graph G, is a proper coloring for G using just three colors. A given graph G is 3-colorable, if and only if there exists at least one proper 3-coloring for G. Fig. 1 shows a 3-colorable graph and a proper 3-coloring solution. Fig. 2 shows a graph which is not 3-colorable.

The graph 3-colorability problem is a decision problem (requires answer "yes" or "no") which is looking to find if a given graph G is 3-colorable or not [10]. As the graph 3-colorability problem is a decision problem, it is not seeking to find a proper 3-coloring, but just requires answer "yes" if the given graph is 3-colorable, and answer "no" otherwise.

It has been proved that the graph 3-colorability problem is an NP-complete problem [10] and the best algorithms to solve this problem is exponential time in the conventional computers.

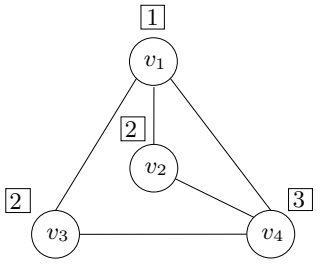


Fig. 1. Example of a 3-colorable graph and a possible proper 3-coloring solution. The colors assigned to the vertices are specified in the squares.

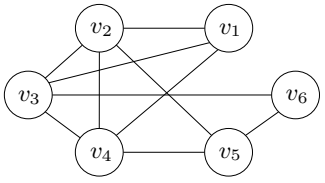


Fig. 2. Example of a graph which is not 3-colorable

3 The Optical Solution

In this section, the optical solution to the graph 3-colorability problem is explained. We first show how to represent a 3-coloring of a given graph having n vertices as a binary sequence having length $2n$. Then we explain how to use light and optical devices to generate all binary sequences having length $2n$ and filter improper sequences to obtain exactly those binary sequences representing a proper 3-coloring for the given graph in efficient time.

3.1 Graph 3-Colorings as Binary Sequences

It is possible to represent each 3-coloring of a graph G having n vertices $v_1, \cdots, v_n$ as a binary sequence having length $2n$ in the form of $a_1b_1a_2b_2 \cdots a_nb_n$ where $\overline{a_ib_i}$ represents the index of the color assigned to vertex v_i $(0 < i \leq n)$ in binary format. In the other words, $\overline{a_ib_i}$ is "01" if v_i is colored by the first color, "10" if it is colored by the second color, and "11" if v_i is colored by the third color. For example, the provided 3-coloring for the given graph in Fig. 1 is represented as "01101011".

Now, the question is how we can confirm if a given binary sequence $a_1b_1 \cdots a_nb_n$ represents a proper 3-coloring for a given graph G with n vertices $v_1 \cdots v_n$ and m edges $e_1, \cdots c_m$. Since $\overline{a_ib_i}$ $(0 < i \leq n)$ represent the color of v_i, two conditions should be satisfied (and it is also sufficient) in order the given binary sequence represents a proper 3-coloring for G:

- The color of each vertex should be one of three colors, so for each i $(0 < i \leq n)$, $\overline{a_ib_i}$ should be one of "01", "10", or "11". In the other words, for each i $(0 < i \leq n)$ it is needed $a_i \neq 0 \vee b_i \neq 0$.
- For each edge e of the given graph G connecting two vertices v_i and v_j, the color of v_i should be different from the color of v_j. So $\overline{a_ib_i}$ and $\overline{a_jb_j}$ should be different, or in the other words, it is needed $a_i \neq a_j \vee b_i \neq b_j$.

It can be easily seen that these two conditions are also sufficient for the given binary sequence having length $2n$ to represent a proper 3-coloring for G, as they guarantee a color is assigned to each vertex using at most three colors and every two adjacent vertices have different colors.

3.2 Finding Proper Binary Sequences

To solve the graph 3-colorability problem for a given graph G having n vertices $v_1, \cdots, v_n$, we are looking to find proper binary sequences representing proper 3-colorings for G among all binary sequences having length $2n$. The main idea is to divide a square of light into 2^{2n} square sections (number of all possible binary sequences with length $2n$) and consider each section as a binary sequence having length $2n$. Then we use some black filters which do not let light to pass from the sections not representing a proper 3-coloring for G.

Now we want to create some black filters which be punched in sections corresponding to binary sequences representing proper 3-colorings (In the other words, if a section is correspond to a binary sequence not representing a proper coloring, then the section is black in at least one filter). Note that a binary sequence in form of $a_1b_1 \cdots a_nb_n$ represents a proper 3-coloring for G if and only if it satisfies the two conditions explained in section 3.1. So we create two classes of filters, vertex filters to filter binary sequences not satisfying the first condition, and edge filters to filter binary sequences not satisfying the second condition.

Vertex Filters. The first condition described in section 3.1 expresses that the color of each vertex should be one of three colors 01, 10, or 11. Each vertex filter blocks those light rays which does not assign a valid color to one vertex. So we create n vertex filters denoted by $f_{v_1}, \cdots, f_{v_2}$, in such a way that for each vertex v_i $(0 < i \leq n)$, f_{v_i} is divided into 2^{2n} sections (representing all possible binary sequences having length $2n$ in form of $a_1b_1 \cdots a_nb_n$), and the filter is black in those sections where $a_i = 0 \wedge b_i = 0$. As f_{v_i} is punched in other sections, it filters those binary sequences which does not assign a valid color to v_i.

These vertex filters are independent from the structure of the given graph, so they can be created in preprocessing phase and be used to solve many problem instances.

Edge Filters. The edge filters are created to block light rays in the sections corresponding to binary sequences not satisfying the second condition described in section 3.1, which expresses that the color of each two adjacent vertices should be different. So for each edge $e_u = (v_i, v_j)$ $(0 < u \leq m)$ in G, we create an edge filter denoted by f_{e_u} and divide it into 2^{2n} sections (representing all possible binary sequences having length $2n$ in form of $a_1b_1 \cdots a_nb_n$), and make the filter black in those sections where $a_i = a_j \wedge b_i = b_j$ (the filter is punched in other sections). Hence f_{e_u} filters those binary sequences assigning the same color to v_i and v_j.

Note that these edge filters are dependent just on the corresponding edge and are independent from the whole structure of the graph, so they can be created in preprocessing phase and be used to solve many problem instances.

3.3 Solving the Problem Instances

To solve a problem instance for a given graph G having n vertices $v_1, \cdots, v_n$ and m edges $e_1, \cdots, e_m$, we place n vertex filters $f_{v_1}, \cdots, f_{v_n}$ and m edge filters

$f_{e_1}, \cdots, f_{e_m}$ one after each other (the order of placing filters is not important), and emit light to the filters as is shown in Fig. 3. Using a convex lens and an optical sensor, we examine if there are some light rays passing through the filters or not. If some light rays have been passed, so there are some sections which are punched in all $n + m$ filters, thus the binary sequences corresponding to these sections satisfy both conditions described in section 3.1. This means that there is at least one binary sequence representing a proper 3-colorings for G and hence G is a 3-colorable graph. If no light rays passes through the filters, no binary sequence with length $2n$ satisfies the conditions described in section 3.1 and G is not a 3-colorable graph.

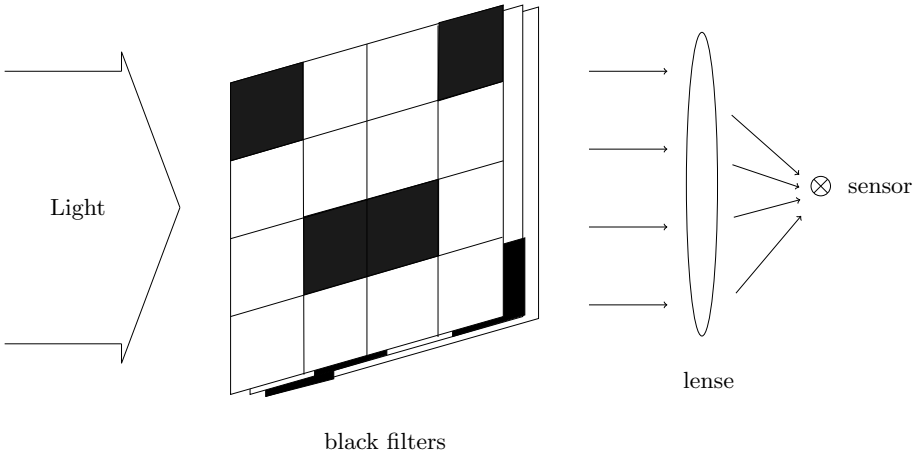


Fig. 3. Arrangement of optical devices in the provided optical solution for the 3-colorability problem

4 Complexity

In the provided optical solution, vertex filters and edge filters are created in preprocessing phase and are used to solve many problem instances. To solve a problem instance for graphs having at most n vertices, n vertex filters should be created, each of which has 2^{2n} sections and requires $O(2^{2n})$ time to be created.There are $\binom{n}{2}$ possible edges in graph having at most n vertices, so $\frac{n(n-1)}{2}$ edge filters are required to be created in preprocessing phase each of which requires $O(2^{2n})$ time to be created. Hence the provided solution requires $O(n^2 2^{2n})$ time in preprocessing phase to create $O(n^2)$ black filters (vertex and edge filters). As the shape of the filters is square and each filter has 2^{2n} sections, the side length of a filter is 2^n times side length of a section.

To solve each instance of the problem for a given graph having n vertices and m edges, $n+m$ black filters (n vertex filters and m edge filters) are placed next to each other and light is passed through them. So the provided solution takes $O(m+n)$ time to find if a given graph is 3-colorable or not.

Note that the exponential size of the black filters may cause some difficulties in implementation of the provided solution. Considering side length 30 micro meters for each section obtained by laser micro drilling technology, the side length of each black filter is 1.3 meters for $n = 15$ and 2.6 meters for $n = 16$. Hence, the provided solution is not practicable for large values of n.

5 Conclusion

In graph theory, the graph 3-colorability problem asks if it is possible to assign a color to each vertex of a given graph using just three colors in such a way that every two adjacent vertices have different colors. The graph 3-colorability problem arises in many real-world applications makes it necessary to find efficient solutions for this problem. It has been proved that the graph 3-colorability problem is an NP-complete problem, which no polynomial time and other resources solution is found yet.

In this paper, an optical solution to the graph 3-colorability problem is provided requiring polynomial time to solve each problem instance after an exponential time preprocessing phase. In the provided solution, each graph 3-coloring for a graph having n vertices is represented as a binary sequence with length $2n$. A square shape space is considered to be divided into 2^{2n} sections and each section is considered as a possible binary sequence having length $2n$. Black filters created in preprocessing phase are placed each after other in such a way that each section corresponding to a binary sequence not representing a proper 3-coloring for the given graph is black in at least one black filter. Passing light through the filters, if at least some light rays pass through all filters detecting by an optical sensor, the given graph is 3-colorable, and if no light ray passes through the filters, every sections are black in at least one black filter and hence the given graph is not a 3-colorable graph.

The black filters are created in preprocessing phase and may be used to solve many problem instances. To solve the problem instances for the graphs having at most n vertices, the provided solution requires to create $O(n^2)$ black filters each of which having exponential size and requires $O(2^{2n})$ time to be created. After preprocessing phase, the provided solution requires $O(m+n)$ time to solve each problem instance for a given graph having n vertices and m edges.

The exponential size of each black filter causes difficulties in implementation of the provided solution and the solution is not practicable for large values of n.

Acknowledgments. This work is partially supported by Iranian National Institute of Elites.

References

1. Myszkowski, P.B.: Solving scheduling problems by evolutionary algorithms for graph coloring problem. SCI, vol. 128, pp. 145–167 (2008)
2. Zweben, M., Fox, M.: Intelligent Scheduling. Morgan Kaufmann, San Francisco (1998)
3. Goliaei, S., Jalili, S.: An optical wavelength-based solution to the 3-SAT problem. In: Dolev, S., Oltean, M. (eds.) OSC 2009. LNCS, vol. 5882, pp. 77–85. Springer, Heidelberg (2009)
4. Oltean, M.: Solving the hamiltonian path problem with a light-based computer. Natural Computing: an International Journal 7(1), 57–70 (2008)
5. Oltean, M., Muntean, O.: Exact cover with light. New Generation Computing 26(4), 327–344 (2008)
6. Oltean, M., Muntean, O.: Solving the subset-sum problem with a light-based device. Natural Computing: an International Journal 8(2), 321–331 (2009)
7. Dolev, S., Fitoussi, H.: Masking traveling beams:optical solutions for np-complete problems, trading space for time. Theoretical Computer Science 411, 837–853 (2010)
8. Lin, M.: 3d dna self-assembly model for graph vertex coloring. Journal of Computational and Theoretical Nanoscience 7(1), 246–253 (2010)
9. DHondt, E.: Quantum approaches to graph colouring. Theoretical Computer Science 410(4-5), 302–309 (2009)
10. Goldreich, O.: Computational Complexity: a conceptual perspective. Cambridge University Press, Cambridge (2008)

Solving a Generalized Version of the Exact Cover Problem with a Light-Based Device

S.M. Shabab Hossain[1,2], Md. Mahmudur Rahman[1,2], and M. Sohel Rahman[1]

[1] AℓEDA Group
Department of CSE, BUET, Dhaka-1000, Bangladesh
{shabab,sajib}@cse.uiu.ac.bd
http://www.buet.ac.bd/cse
[2] Department of CSE, UIU, Dhaka-1209, Bangladesh
msrahman@cse.buet.ac.bd

Abstract. We propose a light-based device that is capable of solving a generalized version of the exact cover problem.

1 Introduction

In recent times, significant amount of attention has been given to explore different unconventional computing techniques to solve various computational problems. Different problems that are hard to solve efficiently in conventional methods, have received significant attention in the literature from this point of view. Optical computing is an exciting research avenue in this regard. Very recently,a number of researchers have suggested light-based devices to solve combinatorially interesting problems. Most of the problems handled in this way are hard to solve in the conventional computing paradigm and are categorized as NP-Complete or NP-Hard problems. For example, a system which solves the Hamiltonian Path problem (an NP-Complete problem [1]) using light and its properties has been proposed in [6]. Similar system was devised in [7] and [8] to solve the Exact Cover Problem and the Subset Sum problem respectively, both of which are NP-Complete problems [1]. Following up the work of [8], very recently, a generalized version of the subset sum problem was solved using a light-based device in [2]. In the above-mentioned work, only the decision versions of the corresponding problems have been considered. Finally, a way to compute an actual solution of the subset sum problem was presented in [3].

This paper follows up the work of [7] and [2]. In particular, we revisit the Exact Cover problem from optical computing point of view. We combine the work of [7] and [2] to solve a generalized version of the exact cover problem. Notably, a number of unresolved issues were highlighted in [7] as future research directions. By devising a light-based device capable of solving a generalized version of the problem, we are able to resolve some of these open problems.

S. Dolev and M. Oltean (Eds.): OSC 2010, LNCS 6748, pp. 23–31, 2011.

2 Exact Cover Problem

We start this section with a formal definition of the exact cover problem.

Problem 1. Exact Cover Problem

Instance: We are given a set $U = \{u_1, u_2, ..., u_n\}$ of elements and a set C of subsets of U.

Question: Is there a subset S of C such that every element in U is contained in exactly one set in S.

In the paper [7], the authors presented a light based device to solve the Exact Cover problem. Notably, the device presented in [7] only can solve the problem as a decision problem; it doesn't have the capability to provide a solution. Indeed, the author pointed out in the conclusion that designing a device with such capability should be explored as a future research. In this paper, we consider a more general problem than the problem defined above. In particular, we solve the question whether there is a subset S of C such that $|S|$ is less than or equal to a given integer m. Formally, the problem we handle is defined below:

Problem 2. Generalized Exact Cover Problem

Instance: We are given a set $U = \{u_1, u_2, ..., u_n\}$ of elements, a set C of subsets of U and a positive number m.

Question: Is there a subset S of C such that every element in U is contained in exactly one set in S and $|S| \leq m$.

Clearly, the Exact Cover Problem (Problem 1) is a restriction of the Generalized Exact Cover Problem (Problem 2), since the latter reduces to the former when $m = |U|$; therefore, the latter is also NP-Complete.

2.1 Example

$U = \{u_1, u_2, u_3, u_4\}$
$C = \{C_1, C_2, C_3\}$
$C1 = \{u_1, u_3\}$
$C2 = \{u_2, u_3, u_4\}$
$C3 = \{u_2, u_4\}$

An exact cover S of U is, $S = \{C_1, C_3\}$ also there is no solution set S such that $|S| \leq 1$.

3 The Device

We design our device in a step by step manner. We start from the device proposed in [7] see figure 1. We first give brief review of the device proposed in [7].

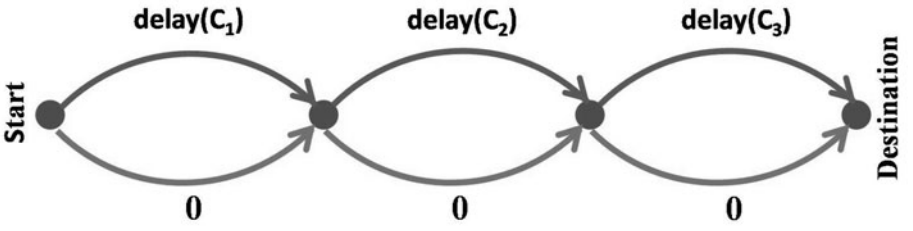


Fig. 1. previously proposed device

3.1 Previously Proposed Device

The previously proposed device in figure 1 is constructed for the problem definition given in 2.1. The device has a graph-like structure. The following two simple operations are performed by the device.

1. When passing through an arc the light ray is delayed by the amount of time assigned to the arc.
2. When passing through a node the light ray is divided into a number of rays equal to the external degree of that node.

Delay calculation. The $delay(C_i), 1 \leq i \leq |C|$ is computed as follows,

We assign each element of set $U = \{u_1, u_2, u_3, u_4\}$ a number $d_i (1 \leq i \leq n)$. These number have the following special property,

$$d_1 + d_2 + \ldots + d_n \neq a_1 \cdot d_1 + a_2 \cdot d_2 + \ldots + a_n \cdot d_n \tag{1}$$

where $a_i (1 \leq i \leq n)$ are natural numbers $(a \geq 0)$ and cannot be all 1 in the same time.

The integer number a_j tells us how many times a ray has passed through (covered) an element u_j. If value of a_j is equal to 1, we know that that particular element u_j has been covered (exactly once). Some example of numbers abiding the equation 1 is given in table 1. The numbers in this set are called Nialpdromes numbers (sequence A023758from The On-line Encyclopedia of Integer Numbers [9]).

These numbers have been used before to solve Hamiltonian path problem [6], Exact cover problem [7]. We can very easily generate these number for a given n by using the following formulae,

$$\begin{gathered} 2^n - 2^{n-1}, \\ 2^n - 2^{n-2}, \\ 2^n - 2^{n-3}, \\ \ldots \\ 2^n - 2^0 \end{gathered}$$

Although the numbers are very easy to generate, for a given n they are of exponential order, and thus have a adverse effect on the complexity of the device.

For each element $u_i \in U(1 \leq i \leq n)$ we assign the value $d_i(1 \leq i \leq n)$. for the problem definition in 2.1 each delay is computed as follow,

$delay(C_1) = d_1 + d_2$
$delay(C_2) = d_2 + d_3 + d_4$
$delay(C_1) = d_2 + d_4$

Because of the values assigned to each element $u_i(1 \leq i \leq n)$ and the way $delay(C_i)(1 \leq i \leq |C|)$ is computed we can identify whether a exact cover exist or not. We do so by waiting for a light signal at moment $d_1 + d_2 + \ldots + d_n$. if a light signal is present then a solution exists otherwise there is no solution.

3.2 The Basic Idea of the System

The basic idea of our device is that $delay(C_i), (1 \leq i \leq |C|)$ plus some fraction value represent the delays induced to the signals(light) that passes through our device. For instance the, if the sets C_1, C_3 produce a exact cover, then the total delay of the signal should be $delay(C_1) + delay(C_1) + (fractionalValue)$. The fractional value will help us determine the maximum number of sets needed for exact cover. If using light we can easily induce some delays by forcing the ray to pass through an optical cable of given length.

This is why we have designed our device as a directed graph. Arcs, which are implemented by using optical cables, are labeled with $delay(C_i), (1 \leq i \leq |C|)$ plus a fractional value which depends on the size of the set C.

As can be seen in figure 2 that each number plus fractional value f is assigned to arcs with an alternative path of delay equal to zero.

Property of the fraction f:

1. The f is equal for all arcs, so that we can identify the number of sets participating in the solution.
2. Sum of all f is less then 1. The significance of this property will be realized when we prove the correctness of this device.

We can easily calculate such fraction by using the formulae, $f = \frac{1}{(|C|+1)}$. So for the problem instance shown in the figure 2 with $|C| = 3, f = \frac{1}{(3+1)} = 0.25$ the signal passing through the arcs containing C_1, C_3 will reach the destination at

Table 1. Delay that are used for a given set U of size n

Problem Size, n	Labels for $C_j, d(C_j)$
1	1
2	2, 3
3	4, 6, 7
4	8, 12, 14, 15
5	16, 24, 28, 30, 31
6	32, 48, 56, 60, 62, 63

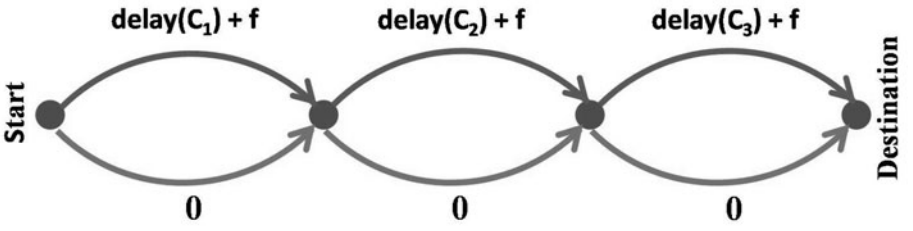


Fig. 2. First version of our device

time $delay(C_1) + delay(C_3) + 0.5$. So if we are interested in finding out whether there is a solution set S producing exact cover and $|S| = m$, then we wait for a signal at the moment $d_1 + d_2 + \ldots + d_n + \frac{m}{(|C|+1)}$.

Clearly, the above system is not capable to answer the question whether there is a solution subset S such that, $|S| \leq m$. In the rest of this section we present a novel system to achieve our goal.

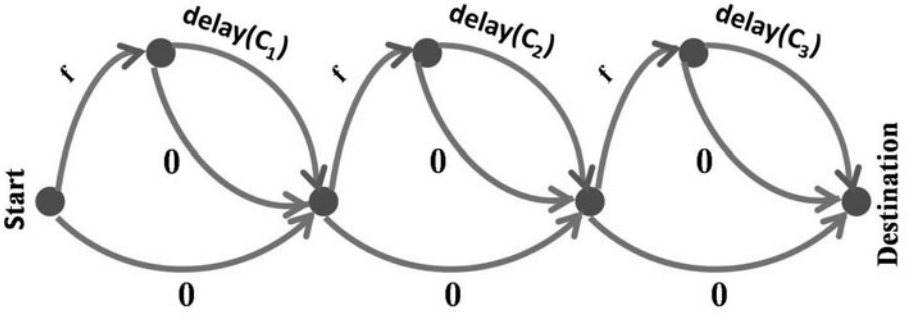


Fig. 3. Second version of our device

3.3 The Modified Device

We modify the device as shown in figure 3. This device ensures that if there is a light signal at moment $d_1 + d_2 + \ldots + d_n + \frac{\beta}{(|C|+1)}$, then there will be a light signal at moments $d_1 + d_2 + \ldots + d_n + \frac{\alpha}{(|C|+1)}$, where $\beta < \alpha \leq |C|$. For instance if $|C| = 3, f = 0.25$ the device in figure 3 will produce a signal which reaches destination at time $d_1 + d_2 + \ldots + d_4 + 0.5$ as before. But it will also produce a signal at time $d_1 + d_2 + \ldots + d_4 + 0.75$.

The final bit of modification needed is because of the presence of 0 arcs. Even if theoretically we could have arcs of length 0, we cannot have cables of length 0 in practice. We may use very short cables (let's say of length ϵ) for arcs which are supposed to have length 0. However, in that case we could obtain false signal

because of the cumulative effect of cables with length ϵ. Even if there is no exact cover of the set U, still there will be possible to have a signal at moment $d_1 + d_2 + \ldots + d_n + (fractionalValue)$ due to the situation presented above. For avoiding this situation we have added a constant k to the length of each alternative path of the light signal. The schematic view of this device is depicted in figure 4.

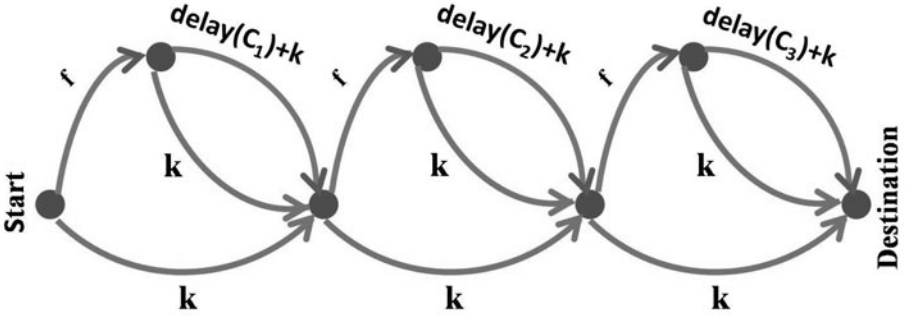


Fig. 4. Final version of our device considering implementation issues

We can see that each path from start to destination contains exactly $|C|$ time value k. Thus at the destination we will not wait anymore at moment $d_1 + d_2 + \ldots + d_n + \frac{m}{(|C|+1)}$, for a solution set S with $|S| \leq m$. Instead we will wait for a solution at moment $d_1 + d_2 + \ldots + d_n + \frac{m}{(|C|+1)} + |C| \cdot k$ Since all light signal will have the constant $|C| \cdot k$ added. Now, it is quite easy to realize that if there is a ray arriving to the destination (see figure 4) at moment $d_1 + d_2 + \ldots + d_n + \frac{m}{(|C|+1)} + |C| \cdot k$ then there is a subset S of C such that S contains every elements of U exactly once and $|S| \leq m$.

4 Physical Implementation

For implementing the proposed device the following components are required:

- A light source at the start node.
- Beam-splitters for splitting a light beam into two light rays. Half silvered mirror is a good example for beam-splitter
- Optical fibers of various length.
- Light sensor at the destination node, which converts optical pulses to electric signals. Photo diode is a good example for light sensor.
- Oscilloscope, a tool for detecting a fluctuation of power generated by photo diode.

5 Analysis of the Proposed Device

In this section, we analyze our proposed device following the framework for analysis presented in [6,8].

5.1 Precision

A problem is that we cannot measure the moment $d_1+d_2+\ldots+d_n+\frac{m}{(|C|+1)}+|C|\cdot k$ exactly. We can do this measurement only with a given precision which depends on the tools involved in the experiments. Actually it will depend on the response time of the photodiode and the rise time of the oscilloscope. The rise-time of the best oscilloscope available on the market is in the range of picoseconds (10^{-12} seconds). This means that each signal arriving at the destination at distinct moment must maintain above mentioned time interval to be recognized correctly. This can be ensured by setting a lower bound on the length of the cable as follows. Since the speed of light is 3×10^8, the minimum cable length must be 0.0003 meter.

This value is the minimal delay that should be introduced by an arc. Also, note that, all lengths must be integer multiples of 0.0003. We cannot allow to have cables whose lengths can be written as $p \times 0.0003 + q$, where p is an integer and q is a positive real number less than 0.0003, since by combining this kind of numbers we can have a signal in the above mentioned interval and that signal may not contain valid information.

Once we have the length for the minimal delay it is quite easy to compute the length of the other cables that are used in order to induce a certain delay as follows. We assign $f = 0.0003$ meter, where f stands for the fractional value. Then we define $unit = (1 + |C|) \cdot f = (1 + |C|) \cdot 0.0003$ meter Now for each $delay(C_i), (1 \leq i \leq |C|)$, we use arc of length $delay(C_i) \cdot unit$ where $1 \leq i \leq |C|$.

5.2 Hardware Complexity

The maximum value of delay of a particular arc in the device, is $O(2^n)$. So the maximum cable length needed for physical representation of a element is, $2^n \cdot unit = (|C| + 1) \cdot 2^n \cdot \Delta = O(|C| \cdot 2^n)$. So the total length of the wire is $O(|C|^2 \cdot 2^n)$.

5.3 Runtime Complexity

The ray corresponding to solution takes $O((d_1 + d_2 + \ldots + d_n) \cdot |C| + |C| \cdot k)$ time to reach destination node. So the time complexity is $O(|C| \cdot n \cdot 2^n + |C| \cdot k)$ or $O(|C| \cdot n \cdot 2^n)$ assuming $2^n \gg k$.

5.4 Problem Size

We are also interested to find the size of the instances that can be solved by our device. let, L be the maximum length of available cable. And Δ be the minimum length of cable (see Section 5.1). Then, we have $2^n \cdot (1 + |C|) \cdot \Delta = L$. This implies $2^n \cdot (1+2^n) \cdot \Delta = L$, [because $C = O(2^n)$]. Finally, we get $n = \frac{1}{2} \cdot lg \frac{L}{\Delta}$ (approximately).

5.5 Power Decrease

Beam splitters are used in our approach for dividing a ray in two subrays. Because of that, the intensity of the signal is decreasing. In our proposed device there $2 \cdot |C|$ nodes(we ignore the destination node, because there is no split there). At each node the power of the signal is reduced to half. So for detection of the light signal at the destination the input signal must have $2^{2 \cdot |C|}$ power.

5.6 Technical Difficulties

There are many technical difficulties that must be solved to implement the proposed device. Some of them are described bellow:

- Cutting the optical fiber with precision. Failing to do so can generate a fluctuation of light pulse at the time when the signal representing the solution is scheduled to reach the destination.
- Finding high precision oscilloscope. This is essential to measure the power at the accurate time.

5.7 Improving the Device

In our proposed device the delay a light signal incurs as it passes through an optical fiber is used to encode the elements of the given set. So the maximum number representable in the device depends upon the maximum delay that is practically possible to produce with an optical fiber. We can either increase the length of the cable used or decrease the speed of light to do so. Increasing the length of the cables will also increase the device size. So decreasing light speed is a preferable option.

The speed of light in the optical fiber is much less than the speed of light in void space (60% of original speed of light). Other methods for reducing the speed of light is also available.A very interesting solution was proposed in [4] which is able to reduce the speed of light by 7 order. This could help our mechanism significantly. But using this idea in our device remains an open question because of the complex equipment involved in those experiments [5] [4].

6 Conclusion and Future Work

In this paper, we have developed a light-based device that is capable of solving a generalized version of the exact cover problem. Further research directions can be focused on the followings:

- implementation of the proposed device.
- the device proposed is mechanical in nature. i.e. we have to change the cable length and number of beam splitters in order to solve different instances of the exact cover problem. So we need to automate the whole process.

References

1. Garey, M.R., Johnson, D.S.: Computers and Intractability: A Guide to the Theory of NP-Completeness. W.H. Freeman, New York (1979)
2. Hasan, M., Hossain, S.M.S., Rahman, M.M., Rahman, M.S.: Solving the generalized subset sum problem with a light based device. Natural Computing (2010)
3. Hasan, M. R., Rahman, M.S.: Computing a solution for the subset sum problem with a light based device. In: Dolev, S., Oltean, M. (eds.) OSC 2009. LNCS, vol. 5882, pp. 70–76. Springer, Heidelberg (2009)
4. Hau, L., Harris, S., Dutton, Z., Behroozi, C.: Light speed reduction to 17 meters per second in an ultracold atomic gas. Nature 397, 594–598 (1999)
5. Liu, C., Dutton, Z., Behroozi, C.H., Hau, L.V.: Observation of coherent optical information storage in an atomic medium using halted light pulses. Nature 409, 490–493 (2001)
6. Oltean, M.: Solving the hamiltonian path problem with a light-based computer. Natural Computing 7(1), 57–70 (2008)
7. Oltean, M., Muntean, O.: Exact cover with light. New Generation Comput. 26(4), 329–346 (2008)
8. Oltean, M., Muntean, O.: Solving the subset-sum problem with a light-based device. Natural Computing 8, 321–331 (2009)
9. Sloane, N.: The on-line encyclopedia of integer sequences, `http://www.research.att.com/~njas/sequences/A023758`

All-Optical Reconfigurable Logic Unit with Optically Controlled Microcavity Switches

Mohit Prasad and Sukhdev Roy*

Department of Physics and Computer Science, Dayalbagh Educational Institute
Dayalbagh, Agra 282 110 INDIA
sukhdevroy@dei.ac.in

Abstract. We propose an all-optical reconfigurable logic unit based on optically controlled microcavity switches that can execute different logic and arithmetic operations on the same hardware with different configuration. Theoretical designs considering bacteriorhodopsin protein-coated microcavities in a tree architecture have been presented. The combined advantages of high Q-factor, tunability, compactness, low-power control signals and flexibility of cascading switches to form circuits, makes the designs promising for practical applications.

Keywords: all-optical switching, microcavities, multiplexer, bacteriorhodopsin, optical comcputing.

1 Introduction

The prospect of light controlling light for all-optical information processing has received tremendous thrust as new optical and optoelectronic technologies become available [1]. The search for new optical materials, components, devices, algorithmic architectures as well as number systems which exploit the parallelism, non-interfering communication and wide bandwidth properties of optical systems is underway. There are several factors impeding the technology. The most important are cascadability and material development. One of the major impediments in realizing all-optical circuits is the absence of non-linear all-optical circuit elements preferably capable of fabrication in the form of 2D/3D arrays and that are small enough to have low switching energies and high speeds [2].

The basic requirement which photonic components should satisfy in order to achieve all-optical computing are high contrast, low-power operation, fast switching, steady-state bias, external address, cascadability, high fanout and high gain. The flexibility of electronic processing stems from its ability to perform non-linear operations such as thresholding. In optical processing, non-linear optical mechanisms play important role in ultra-fast optical switches and all-optical logic gates. The integration of photonic components is expected to increase significantly due to emerging novel photonic structures such as microresonators, photonic crystals and plasmonics.

* Corresponding author.

S. Dolev and M. Oltean (Eds.): OSC 2010, LNCS 6748, pp. 32–41, 2011.

Microcavities have emerged as extremely sensitive and versatile device configurations for a variety of operations due to their high Q-factor, low switching threshold and ultra compactness [3]. Optical microcavities confine light to small volumes by resonant circulation. A very wide range of microresonator shapes have been explored over the years for various applications. The most widely used are rotationally symmetric structures such as fabry-perot cavities, spheres, cylinders, disks, torroids, and photonic crystals which have been shown to support very high-Q whispering gallery (WG) modes, whose modal field intensity distribution is concentrated near the dielectric-air interface [3-7].

Devices based on microcavities are already indispensable for a wide range of applications and studies. By tailoring the shape, size or material composition, of a microcavity it can be tuned to support a spectrum of optical modes with required polarization, frequency and emission patterns [3-5]. Various all-optical logic operations have been shown using Si, GaAs and InGaAs microcavities [8, 9]. The nonlinear optical mechanism implemented in these gates is the change in the refractive index from free charge carriers generated by two photon absorption (TPA). Although the change in refractive index in these microcavities is high, absorption generates heat inside the microcavity as it is dependent on pump power. The pump and probe power is in milliwatts and Q-factor is of the order of ~10^4 [8, 9].

Silica microcavities have an inherent advantage of high Q-factor, relatively simple fabrication, possible on chip integration and control of the coupling efficiency through taper by the change of the fiber thickness [4, 6]. Fiber-optic tapers have been proposed as a means to couple quantum states to or from a resonator onto a fibre. Also the recent demonstration of a fiber-taper-coupled ultrahigh-Q microtoroid on a chip enables integration of wafer based functions with ultralow-loss fibre-coupled quantum devices. The bulk optical loss from silica is also exceptionally low and record Q factors of 8 * 10^9 (and finesse of 2.3 * 10^6) have been reported [3]. Ultra high-Q microtoroidal silica resonators represent a distinct class of optical microresonators with Q's in excess of 100 million [6]. Due to the high Q-factor and the small dimensions, switching at low power is feasible. Recently, large scale arrays of ultra high-Q coupled nanocavities have also been demonstrated in photonic crystals [10]. Moreover, coating the microcavity with a photosensitive material can further lead to switching at ultra-low powers. The photosensitive protein bacteriorhodopsin (BR) found in the purple membrane fragments of *Halobacterium halobium,* has emerged as an outstanding photonic material for practical applications due to its unique multifunctional photoresponse and properties [11-15].

1.1 All-Optical Bacteriorhodopsin Protein Coated Microcavity Switch

Recently, all-optical switching in the near-infrared with D96N-BR mutant coated silica microcavity has been reported, with a Q-factor on BR adsorption ~5 x 10^5 [11-13]. The state of BR in the optical microcavity is controlled by a low power (< 200 μW) continuous green pump laser coupled to the microsphere cavity using a tapered fiber [11-13]. We have recently demonstrated that faster switching (~μs) can be achieved by injecting a blue laser beam (405 nm) of same intensity in addition to the green laser beam (532 nm), which truncates the BR photocycle at the M-intermediate state [13]. Moreover, we have also shown that the same versatile BR-coated

microcavity configuration can be utilized for implementing conservative, reversible and universal all-optical Fredkin gate and for designing higher computing circuits such as full-adder, full-adder avoiding fan-out and demultiplexer/multiplexer circuits [14]. Thus, optically controlled microcavity switches in a suitable architecture can be useful to achieve all-optical computing. However, it would be more advantageous to design a general circuit which has the capability to reconfigure and realize different arithmetic and logic operations.

Cascading to integrate a large number of all-optical components such as switches, logic gates etc., is a complex problem and a major obstacle in the development of a complete all-optical computing system [2]. Branching of signals all-optically among various logic devices is a critical task. Theoretical schemes suggesting alternative solutions for parallel generation of logic gates have been recently reported in literature [15, 16]. An effective method for cascading involves the tree architecture [17]. It is a multiplying system of a single straight path into several distributed branches and sub-branch paths. Although, it is necessary to realize combinatorial logic through a multiplexer which is very important for computing operations, tree architectures have so far been utilized for designing logic operations and some all-optical circuits, such as half-adder/subtractor, full-adder/subtractor etc. Shen et al. demonstrated that re-configurability can be introduced into designs of all-optical logic circuits by electro-optic switches [18]. However, it would be more advantageous to use all-optical switches to achieve reconfigurability. Electronic reconfigurable systems often have stringent performance and power requirements, leading designers to incorporate special-purpose hardware into their designs [19-21]. In many embedded applications, such as multimedia, encryption, wireless communication, highly repetitive parallel computations well-suited to hardware implementation represent a significant fraction of the overall computation required by the system [19-22].

The key objective of this paper is to consider an optically controlled resonant coupler microcavity switch as the key element to propose general designs of an all-optical logic unit that can be configured to realize different logic and arithmetic operations in parallel, which is described in Section 2. In this paper, as an example we combine the advantages of BR-coated microcavity switches in a tree architecture to realize reconfigurable all-optical logic and arithmetic operations, as shown in Section 3. The proposed designs are general and can be realized by using any optically controlled microcavity with the possibility of fabrication on a chip in a array. Moreover, the horizontal and vertical extension of the designs can also be easily done. We discuss the merits and demerits of the proposed designs in Section 4 and present the conclusion in Section 5.

2 BR-Coated Microcavity Switch as a Four-Port Resonant Coupler

The basic configuration of an all-optical switch in the design has been considered to be the BR-coated microcavity switch [11-13]. The silica microcavity in contact between two tapered fibers serves as a four-port tunable resonant coupler. The switch shown in Fig.1 represents the simplified schematic representation of the resonant coupler. Here, port 1 serves for input and control signal, while port 2 and 3 act as

output ports. The various possible input and output states have been listed in Table 1. The switching of input signal operating at the infra-red (IR) wavelength of 1310 nm or 1550 nm between the output ports (2 and 3) is photo-induced by activating the WG mode of the resonator by a fiber-coupled green pump laser (at 532 nm) which controls the conformational state of the adsorbed BR. The molecularly functionalized microcavity thus redirects the flow of near-infrared light beam between the two optical fibers. With the pump OFF, the probing light from input port 1 is detuned from resonance and is directly transmitted into the output port 2. When the pump is ON, it evanescently excites WG modes propagating around the microsphere's equator, inducing photo-isomerization along their path. A low green cw laser (< 200 μW at 532 nm) is sufficient for this purpose as its effective absorption is resonantly enhanced. Isomerization reduces the retinal polarizability, tuning the peak/trough of the resonance to match the wavelength of the infrared probe which is then rerouted into the output port 3 [11-13]. The transmitted spectra has been reported to exhibit an extinction of -9.4 dB in port 2 and a 9.8 dB increase in transmission in port 3 [11].

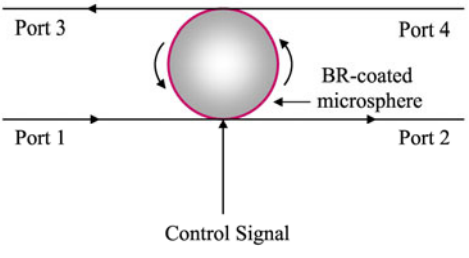


Fig. 1. Schematic of BR-coated microcavity switch

The four-port coupler configuration with an added control signal at one of the ports facilitates equal number of inputs (2 or 3) and output states. An equal number of input and outputs are essential for realizing reversible logic. This simple configuration has been shown to result in the universal all-optical Fredkin logic gate which has been utilized for realizing various Boolean functions and higher computing circuits [14]. In this paper, we use this versatile BR-coated microcavity switch to design an all-optical reconfigurable logic unit that can be configured for generating various logic and arithmetic functions.

Table 1. Truth Table of Four-Port Resonant Coupler of Fig. 1

Incoming Signal	Control Signal	Port 3	Port 2
0	0	0	0
0	1	0	0
1	0	0	1
1	1	1	0

3 Design of BR-Based All-Optical Reconfigurable Logic Unit

To design a BR-based all-optical reconfigurable logic unit (BR-AORLU), we consider the control signals at 532 nm to be the select lines that control the IR signal at 1310 or 1550 nm, which forms the output states of different switches at different ports, to realize the various arithmetic and logic operations in a tree architecture. The select lines decide the routing of the incoming signal to the desired output channel. Fig.2 (a) shows the architecture based on BR-coated microcavity switches S_1, S_2, and S_3 and Fig. 2 (b) shows the schematic diagram of the proposed design.

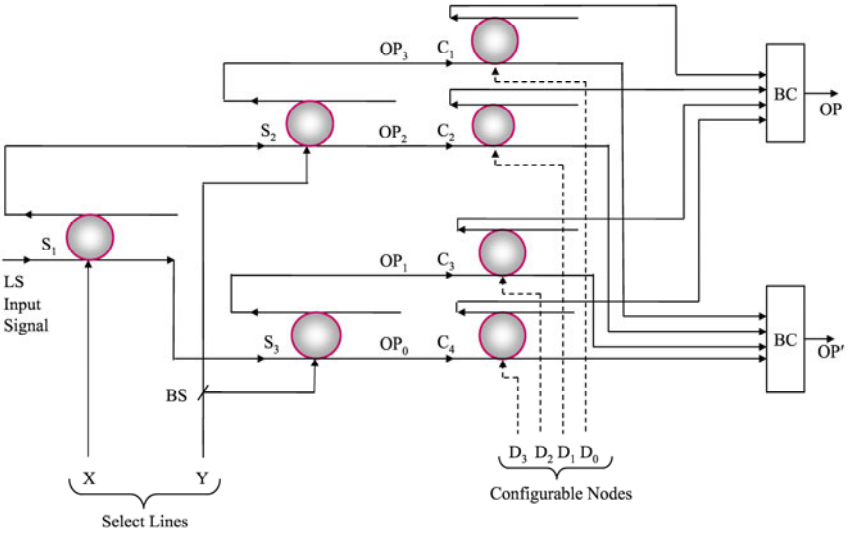


Fig. 2 (a). Architecture of two-variable BR-AORLU

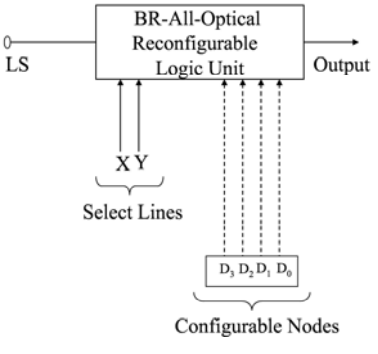


Fig. 2 (b). Schematic of two-variable BR-AORLU

A laser source (LS) is coupled at port 1 of switch S_1. Here X and Y act as select lines operating at wavelength 532 nm. The configurable switches C_1, C_2, C_3, and C_4 work as doors here. Electrodes D_0, D_1, D_2 and D_3 are the configurable nodes through which we can switch light signals ON or OFF in configurable switches C_1, C_2, C_3 and C_4. Various arithmetic and logic operations can be realized using this BR-AORLU. When the LS is switched OFF there is no light detected in any of the output port. When the LS is switched ON and both the optical control signals are zero i.e. X = 0 and Y = 0, no 532 nm light is incident on the BR-coated microcavity switches to induce photo-isomerization of BR and switch the input signal. As BR is in its ground state hence, the input signal by-passes switches S_1 and S_3 and emerges at output port OP_0. The logic generated at $OP_0 = X'Y'$, as shown in Table 2.

Now, when X = 0 and Y = 1, light is incident on BR-coated microcavity switches S_2 and S_3 and triggers the conformational changes in BR. The control signal Y activates S_2 and S_3. The input signal passes from switch S_1 to S_3 and gets redirected to OP_1. The logic generated at $OP_1 = X'Y$, as shown in Table 2. Considering the values taken by optical control signals as X = 1, Y = 0, now, the input signal gets switched at S_1 and emerges at OP_2 that corresponds to logic XY'. Similarly, when optical control signals are X = 1, and Y = 1, i.e. all the three switches are activated, the input signal emerges from OP_3 via switches S_1 and S_2. The logic generated at $OP_3 = XY$, as shown in Table 2. Logic operations generated at output ports OP_0, OP_1, OP_2 and OP_3 are $X'Y'$, $X'Y$, XY' and XY respectively as shown in Table 2. The output of these ports is fed into the switches C_1, C_2, C_3 and C_4 where electrodes D_0, D_1, D_2 and D_3 decide the routing of the incoming signals according to values configured at the electrodes. Signals from upper output ports of these configurable switches are combined using a beam combiner. Signals from lower output ports of the configurable switches C_1, C_2, C_3 and C_4 are combined using another beam combiner. Different logic operations can be derived by suitable selection of these electrodes. For e.g., if we combine signals of output ports OP_0 with OP_1, it results in X' logic operation, whereas combining OP_1 with OP_2 results in XOR. Sixteen different logic operations can be realized using this procedure. In order to achieve XOR operation, we just need to configure the electrodes and assign them the values 0 1 1 0. In this way, logic operations that can be generated are X, Y, X', Y', XOR′, XOR, X+Y′, X′+Y, X′+Y′, X+Y, X′Y, XY, X′Y′, XY′, T and F.

Table 2. State of different output ports for different values of X and Y in Fig. 2 (a)

Input		**Output at different Ports**			
X	**Y**	**OP_0 ($X'Y'$)**	**OP_1 ($X'Y$)**	**OP_2 (XY')**	**OP_3 (XY)**
0	0	1	0	0	0
0	1	0	1	0	0
1	0	0	0	1	0
1	1	0	0	0	1

We can use the two-variable-BR-AORLU to design all the two-variable logic operations. Table 3 lists the configuration of all 16 two-variable binary logic operations.

Table 3. Two-variable binary logic operations in Fig. 2 (a)

Configuration				Sum of Minterms	Logic Operation
D_3	D_2	D_1	D_0		
0	0	0	0	None	False
0	0	0	1	OP_3	XY
0	0	1	0	OP_2	XY′
0	0	1	1	OP_2+OP_3	X
0	1	0	0	OP_1	X′Y
0	1	0	1	OP_1+OP_3	Y
0	1	1	0	OP_1+OP_2	XOR
0	1	1	1	OP_1+OP_2+OP_3	X+Y
1	0	0	0	OP_0	X′Y′
1	0	0	1	OP_0+OP_3	XNOR
1	0	1	0	OP_0+OP_2	Y′
1	0	1	1	OP_0+OP_2+OP_3	X+Y′
1	1	0	0	OP_0+OP_1	X′
1	1	0	1	OP_0+OP_1+OP_3	X′+Y
1	1	1	0	OP_0+OP_1+OP_2	X′+Y′
1	1	1	1	OP_0+OP_1+OP_2+OP_3	True

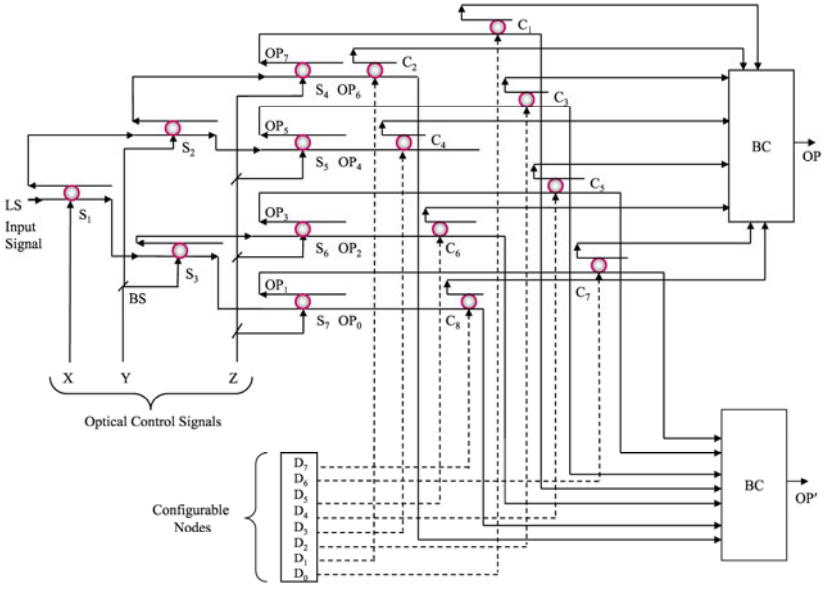


Fig. 3 (a). Architecture of an integrated all-optical three-variable BR-AORLU (BS = Beam Splitter, BC = Beam Combiner, OP = Output Port, LS = Laser Source)

To implement tree architecture for triple-input-binary logic using BR-AORLU we have to incorporate another four BR-coated microcavity optical switches S_4, S_5, S_6 and S_7 and eight configurable switches C_1, C_2, C_3, C_4, C_5, C_6, C_7 and C_8 controlled by eight electrodes D_0, D_1, D_2, D_3, D_4, D_5, D_6 and D_7 respectively as shown in Fig 3 (a). The schematic diagram for three variable BR-AORLU is shown in Fig. 3 (b). This BR-AORLU can work as full-adder/subtractor by just configuring the optical control signals.

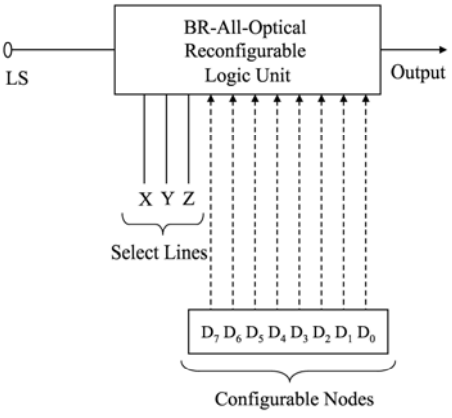


Fig. 3 (b). Schematic of an integrated all-optical three-variable BR-AORLU

4 Results and Discussion

All the designs presented in this paper are all-optical in nature. The proposed designs are simple and can easily implement multiple functions. The designs have good scalability. The probe operates at telecom wavelengths (1310 nm or 1550 nm) and pump operates at wavelength 532 nm, which makes them potentially useful in optical signal processing.

Coating the microcavities with a photosensitive material is of critical value, as the material should exhibit high sensitivity, high absorption, fast dynamics, high photo and thermal stability and flexibility to tailor its properties. BR-protein is a natural photochromic material that exhibits this unique combination of properties for practical realization. The proposed designs require optical excitation of BR-protein, which in principle requires some energy, which due to its high quantum efficiency, high absorption coefficient and high sensitivity is very low. The change in the refractive index and absorption in a BR-coated microcavity switch depends on the wavelength of the control beam making the switches tunable. For ultrafast operation, they can also be realized by coating the microcavity with other organic photochromic molecules such as diarylethenes, in which case the speed will be limited by the cavity photon lifetime.

All-optical switching in BR-coated microcavities has not been optimized for practical applications. At present, the energy/bit for switching that has been demonstrated corresponds to ~ nJ/bit [11-13]. Both the minimum switching energy/bit

and the maximum switching speed need to be investigated by exploiting the ultrafast initial intermediates.

In general, the BR-coated microcavity switch is operated with two different wavelengths. Cascading these switches can in general require output at one wavelength to form the input at the other wavelength, which would then require a wavelength converter. However, as shown in the proposed design, using one wavelength as the signal and the other as the control signal a versatile circuit can be judiciously be designed to circumvent this problem. Moreover the designs are reconfigurable, i.e., the same hardware realizes different logic functions by just configuring the switches at the configurable nodes.

Although the designs presented here have been based on the BR-coated microresonator switch, they can also be suitably adapted to any other optically controlled microcavity using semiconductor ring resonators or photonic crystals. Such planar configurations would be potentially useful for realizing large circuits and networks.

5 Conclusion

In this paper, we have presented a theoretical scheme with optically controlled microcavity switches to design all-optical reconfigurable circuits. As a specific case, we consider BR-coated microcavity switches in a tree architecture. This scheme can easily be extended and implemented for any higher number of input digits, by proper interconnection of microcavity switches using vertical and horizontal extension of the tree and by suitable branch selection. Due to the ultra-high-Q factor and small dimensions, along with the exceptional sensitivity of BR-protein, switching at low-power is possible with BR-coated microcavities. They represent an alternative to the waveguide based techniques providing exceptional sensitivity and straightforward optical integration on micron scales. The proposed designs can yield large computing circuits and networks within a mW-W power budget.

The combined advantages of high Q-factor, tunability, compactness and low power control signals of BR-coated microcavities along with the flexibility of cascading switches in tree architectures to form circuits, makes the designs promising for practical applications.The proposed schemes are general and can be implemented (i) in both fiber-optic and integrated optic formats, (ii) with any other coated photosensitive material or (iii) any externally controlled microresonator switch. The proposed designs provide a novel method for all-optical computing based on hybrid molecular photonic devices employing molecules to perform photonic functions.

References

1. Caulfield, H.J., Dolev, S.: Why future supercomputing requires optics. Nature Photon 4, 261–263 (2010)
2. Miller, D.A.B.: Are optical transistors the logical next step? Nature Photon 4, 3–5 (2010)
3. Vahala, K.J.: Optical microcavities. Nature 424, 839–845 (2003)

4. Benson, T.M., Boriskina, S.V., Sewell, P., Vukovic, A., Greedy, S.C., Nosich, A.I.: Micro-optical resonators for microlasers and integrated optoelectronics: Recent advances and future challenges. In: Janz, S., Ctyroky, J., Tanev, S. (eds.) Frontiers in Planar Lightwave Circuit Technology: Design Simulation and Fabrication, pp. 39–70. Springer, Dordrecht (2006)
5. Ilchenko, V.S., Matsko, A.B.: Optical resonators with whispering gallery modes-Part II: applications. IEEE J. of Select. Top. Quant. Electron. 12, 15–32 (2006)
6. Armani, D.K., Kippenberg, T.J., Spillane, S.M., Vahala, K.J.: Ultra-high-Q toroid microcavity on a chip. Nature 421, 925–928 (2003)
7. Manipatruni, S., Poitras, C.B., Xu, Q., Lipson, M.: High speed electro-optic tuning of the optical quality factor of a silicon micro-cavity. Opt. Lett. 33, 1644–1646 (2008)
8. Xu, Q., Lipson, M.: All-optical logic based on silicon micro-ring resonators. Opt. Exp. 15, 924–929 (2007)
9. Ibrahim, T.A., Amarnath, K., Kuo, L.C., Grover, R., Van, V., Ho, P.T.: Photonic logic NOR gate based on two symmetric microring resonators. Opt. Lett. 29, 2779–2781 (2004)
10. Notomi, M., Kuramochi, E., Tanabe, T.: Large-scale arrays of ultrahigh-Q coupled nanocavities. Nature Photon 2, 741–747 (2008)
11. Topolancik, J., Vollmer, F.: All-optical switching in the near infrared with bacteriorhodopsin–coated microcavities. Appl. Phys. Lett. 89, 1841031–1841033 (2006)
12. Topolancik, J., Vollmer, F.: Photoinduced transformations in bacteriorhodopsin membrane monitored with optical microcavities. Biophys. J. 92, 2223–2229 (2007)
13. Roy, S., Prasad, M., Topolancik, J., Vollmer, F.: All-optical switching with bacteriorhodopsin protein coated microcavities and its applications to low power computing circuits. J. Appl. Phys. 107, 53115 (2010)
14. Roy, S., Prasad, M.: Novel proposal for all-optical Fredkin gate based on bacteriorhodopsin coated microcavity and its applications. Opt. Engg. 49, 65201 (2010)
15. Sharma, P., Roy, S.: All-optical biomolecular parallel logic gates with bacteriorhodopsin. IEEE Trans. on Nanobiosci. 3, 129–136 (2004)
16. Roy, J.N., Gayen, D.K.: Integrated all-optical logic and arithmetic operations with the help of a TOAD-based interferometer device-alternative approach. Appl. Opt. 46, 5304–5310 (2007)
17. Mukhopadhyay, S.: An optical conversion system: from binary to decimal and decimal to binary. Opt. Commun. 76, 309–312 (1990)
18. Shen, Z.Y., Wu, L.L.: Reconfigurable optical logic unit with a tetrahertz optical asymmetric demultiplexer and electro-optic switches. Appl. Opt. 47, 3737–3742 (2008)
19. Garcia, P., Compton, K., Schulte, M., Blem, E., Fu, W.: An overview of reconfigurable hardware in embedded systems. EURASIP J. Embed. Syst. 56320, 1–19 (2006)
20. Todman, T.J., Constantinides, G.A., Wilton, S.J.E., Mencer, O., Luk, W., Cheung, P.Y.K.: Reconfigurable computing: architectures and design methods. In: IEE Proc.-Comput. Digit. Tech., vol. 152, pp. 193–207 (2005)
21. Dery, H., Dalal, P., Cywinski, L., Sham, L.J.: Spin-based logic in semiconductors for reconfigurable large scale circuits. Nature 447, 573–576 (2010)
22. Cardoso, J.M.P., Diniz, P.C., Weinhardt, M.: Compiling for reconfigurable computing: a survey. ACM Comp. Surv. 42, 1–65 (2010)

Simulation and Optimized Design of High Density Optical Crossconnect Systems for Massively Parallel Computing Architectures

Ulrich Lohmann[1,3], Jürgen Jahns[1], Steffen Limmer[2], and Dietmar Fey[2]

[1] Fernuniversität Hagen
Optical Information Technology
[2] Friedrich Alexander University Erlangen Nürnberg
Department of Computer Science 3
[3] Fernuniversität Hagen
Optische Nachrichtentechnik
Universitätsstrasse 27 /PRG, 58097 Hagen, Germany
Ulrich.lohmann@fernuni-hagen.de

Abstract. We demonstrate the simulation results of a skewless high density approach for a multi-channel optical cross connect using integrated free-space optics. Using the 3D nature of free-space optics, this approach is able to solve geometrical problems with static crossings of the signal paths that occur in waveguide optical and electrical interconnection, especially for large number of connections.

Keywords: High density I/O, high performance planar optical interconnection, Clos networks, modeling simulation and evaluation techniques, optical interconnections for parallel computing, parallelization of simulation, Evolutionary Algorithm, optimization, skewless parallel gigabit/s interconnections.

1 Introduction

Optical chip-level interconnections in the data-rate range of ≥ 10 *Gb/s* are becoming more and more interesting in the "short haul" range both for closely-coupled parallel computing and data center applications. In comparison to conventional electrical interconnections over metallic wires, optical connections have advantages with regards to the signal consistency in data rates up to several *Gb/s* [6,11].

Additionally, the problem of geometrical signal path crossing, which is essential in the case of Clos networks with large numbers of channels, can be circumvented by using 3D optical interconnection. Here, we consider in particular the implementation of the interconnection by using the concept of planar-integrated free-space optics (PIFSO) [9,10] combined with MT fiber-optics technology. The free-space approach offers the possibility of a high interconnection density and a significant reduction of crosstalk problems [2].

The goals that we pursue with this approach are: reduction of the physical dimension of a cross connection segment and elimination of skew. The focus of this

S. Dolev and M. Oltean (Eds.): OSC 2010, LNCS 6748, pp. 42–52, 2011.

work is to present a logical model for the cross interconnect and an optimization approach that allows easy implementation and management.

This holds e.g. for the topology of the links in Clos networks which are considered in this paper. In [1] we investigated the possibilities and benefits for a system architecture using wavelength-division-multiplexing techniques for optical Clos networks. In this paper we now move a step towards the hardware realization and present a solution for designing a concrete geometrical layout for the optical links.

The design of such an optical interconnection system on hardware layer is a complex procedure. Hardware parameters have to be optimized which are inversely correlated, i.e. optimizing one parameter can cause a degradation of another parameter.

The rest of the paper is structured as follows. Chapter 2 shows a brief review of Clos networks. Chapter 3 presents the mathematical modeling of Clos networks realized in PIFSO technology. This model is the base for the optimization of some PIFSO technology parameters for which we used an Evolutionary Algorithm (EA) due to the complexity of that task. This EA approach and the yielded design results are the topic of chapter 4. Finally chapter 5 closes the paper with an outlook.

2 Clos Networks

Clos networks are a form of multistage switching networks, first presented by Clos in 1953 ([3]). Clos networks are very important for parallel computer architectures. They have been used in former parallel architectures, e.g. in the MasPar [12], and they are found as well in the networks of current parallel computer, e.g. in InfiniBand [13] switches. A Clos network operates semantically like the well-known crossbar interconnect in the sense that both networks offer universality, i.e. all connection requirements between input and output ports can be switched conflict-free in principal. The advantage of a Clos network is that the number of required crosspoints can be much fewer compared to a network implemented with just one big crossbar.

An $n \times n$ Clos network has three stages (see Figure 1). The first stage consists of switches of size $r \times k$. Since we have n inputs in total, there must be n/r of those switches. In the second stage are k switches each of size $n/r \times n/r$. Every switch in stage two has a connection to every switch in stage one. The third stage has n/r switches of size $k \times r$. Again every switch in stage two has a connection to every switch in stage three. Between each of the subsequent stages (i.e. between the first and second stage, and the second and the third stage) the topology of the links is as follows. The i^{th} output port of the j^{th} switch in the input stage has a 1-to-1 link to the j^{th} input port of the i^{th} switch in the subsequent stage.

It can be shown that for $k \geq 2n-1$ the network is nonblocking and for $k \geq n$ it is rearrangeable nonblocking. That means, an unused input can always be connected to an unused output, but a rearrangement of the existing connections may be necessary. The number of crosspoints that can be saved compared to a network of just one crossbar depends on the exact choice of n and r.

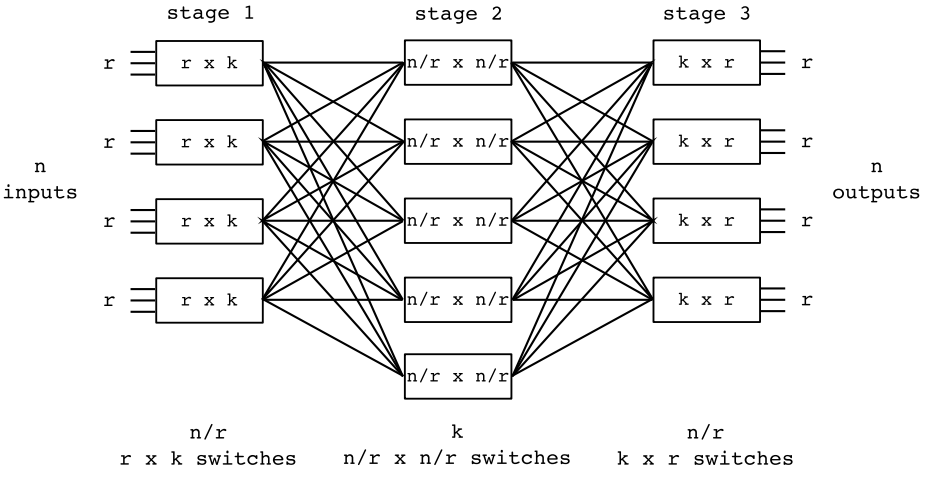


Fig. 1. Topology of an $n \times n$ Clos network

3 Mathematical Modeling of PIFSO $n \times n$ Cross Connects

Like described in Section 1 the realization of the cross interconnections between the stages of a Clos network in PIFSO technology can offer many advantages compared to classical fiber optical or electronic interconnections. The geometrical relations inside the optical plane are defined as shown in Figure 2. With the condition, that the angle of incident is the same as the angle of exit, we can write the total length of the needed interconnection from point A to point B as Equation 1.

$$L_{AB} = \sqrt{(\Delta x_{AB})^2 + (\Delta y_{AB})^2} = n_{FI} \bullet 2 \bullet \frac{H}{\tan \beta} \tag{1}$$

The terms under the square root are the geometric distances between the in- and out-coupling point at the plane $z = H$. The optical path length L_{opt} of a channel can be computed as shown in Equation 2. Equation 3 shows the ratio of L_{opt} to L_{AB}.

$$L_{opt} = n_{FI} \bullet 2 \bullet \frac{H}{\sin \beta} \tag{2}$$

$$\frac{L_{opt}}{L_{AB}} = \frac{1}{\cos \beta} \tag{3}$$

Prior to the realization of a Clos network in PIFSO technology it is necessary to determine concrete values for the deflection angle and the number of folding intervals for every connection. These values should be constituted in such a way, that

1. The light rays hit the output fibers with a minimal position aberration as possible. That means Equation 1 should be satisfied as best as possible.

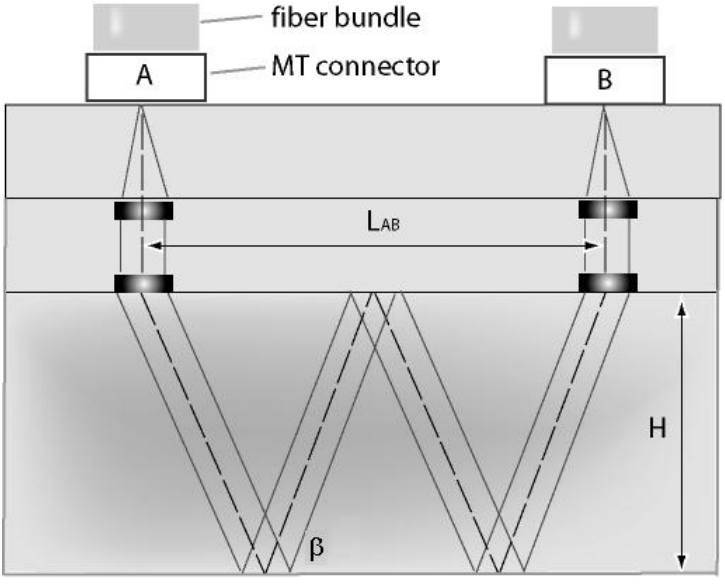


Fig. 2. Geometry of the PIFSO system. The light gray areas indicate optical substrate. The light paths are "folded" into the bottom substrate and propagate under an angle β relative to the substrate surface. n_{FI} denotes the number of double paths (also described here as "folding interval").

2. The aberrations of the optical path lengths L_{opt} of the rays from the average optical path length are as small as possible to minimize the skew effect for parallel lines. For example, a path length difference of 10 *mm* in SiO_2 causes a signal delay of 50 *ps*. This is a quarter of a 10 *Gb/s* digital signal pulse width.

3. The deflection angles lie in the range between 32° and 43° to insure total reflection.

The second condition is the reason for the zig-zag paths of the beams. In this way it is possible to adjust the optical path lengths by an appropriate choice of the deflection angle values and the number of folding intervals. Figure 3 shows a 4×4 Clos network with optimized values as described above.

If one has given the number of folding intervals for one connection, then all the other wanted values can be computed analytically, so that there are no position aberrations and the sum of the aberrations of the optical path lengths are minimal (under the condition that the position aberrations are zero). But with this approach the third condition is not regarded.

Since it is hard or even impossible to find an analytical approach for solving this problem with regard to all three conditions, we use Evolutionary Algorithms (EAs) for this task. The following section gives a short introduction to EAs.

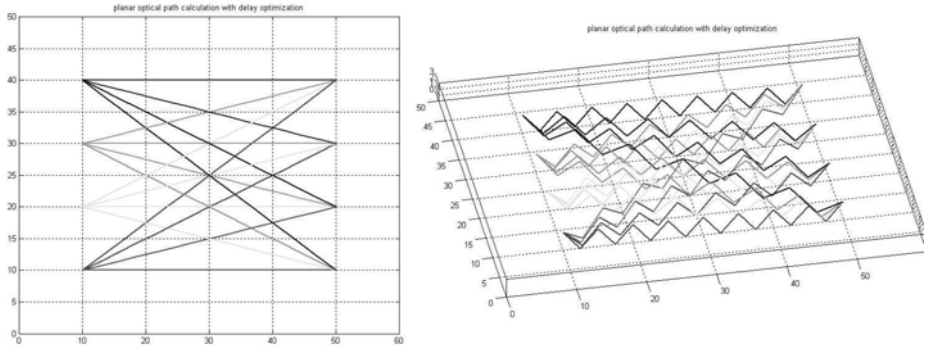


Fig. 3. Top and side view of a skewless optical *4* × *4* Cross interconnect as a part of a Clos network (Example from Figure 1). In the case of integrated free space optic, the path crossing problem is not relevant and allows high density network constellations with more channels.

4 Optimizing PIFSO Clos Networks with Evolutionary Algorithms

4.1 Evolutionary Algorithms

EAs are a popular form of iterative heuristic optimization algorithms ([4]). The process of an EA is oriented in biological evolution. The basis is always a (mostly randomly chosen) population of individuals, where an individual is a possible solution of the problem for which an optimum is searched. The individuals are encoded in an adequate form (in general number values). Such an encoded individual is called a *chromosome*. An individual is in general composed from several parameters. Such a parameter in encoded form is called a *gene*. Every individual will be associated with a real valued *fitness* which expresses how optimal the individual for the examined problem is.

The objective of the algorithm is to create new generations of populations, containing individuals with a higher or respectively smaller fitness and to find in this way an individual with a fitness as maximal or minimal as possible. For the creation of new populations, *genetic operators* are used. These operators are *selection*, *recombination* and *mutation*. Selection is the task of selecting a fixed number of so called parent individuals from the given population. This happens mostly randomized where individuals with a higher fitness have a higher chance to be chosen. Using recombination, offspring are generated from the parent individuals. Mostly one or more offspring are created from two parent individuals. It is desired that parents with a good fitness yield offspring with good fitness values. Subsequent mutation is carried out on the offspring, e.g. random modifications on genes of the offspring will be performed (with a small probability). After that the fitness values of the produced offspring will be computed and they are inserted in the old population according to a certain strategy. By repeating the described procedure continuously, new generations of populations are produced until a certain break condition is met.

4.2 Optimizing $n \times n$ Clos Networks

We implemented an EA that computes the optimal deflection angles and number of folding intervals. An individual consists of a sequence of n folding interval numbers n_{Fi} (one for each of the n rays) that are encoded as integers. The fitness of an individual is computed as the sum of the aberrations of the optical path lengths from the average optical path length and should be minimized in our case. The deflection angles for the different connections are computed directly from the number of folding intervals so that the position aberrations at the outputs are zero. Thus, the position aberrations do not have to be considered for the fitness evaluations. The algorithm turned out to be robust and fast. Robust in that sense, that it yields the same results in almost every run. So the number of runs we performed for obtaining the below mentioned results is not relevant. A mutation rate of *3/n* was used, so that in average three folding interval numbers of an individual are mutated. We used a population size of 100 and a crossover rate of 0.9. Several thousand generations were computed per run..

It is assumed that there are $\sqrt{n}$ in- as well as output MT connectors with $\sqrt{n}$ fibers each. The connectors are arranged on "top" of each other, as shown in Figure 4. Between two fibers of a connector is a pitch of 0.25 *mm* and between two connectors is a pitch of 4 *mm*. The length of the transmission distance is 100 *mm*. For the height of the PIFSO module we chose 1 *mm*.

For $n = 16$ the algorithm yielded a result where all delays of the optical paths lie in tolerable ranges (under 1 *mm*). However, for higher dimensions like $n = 256$ this is not the case. For $n = 256$ the result has a fitness of 1248.08 (the sum of the delays of all paths in *mm*) and there are several paths with a delay of the optical path larger than 1 *mm*, what is not acceptable.

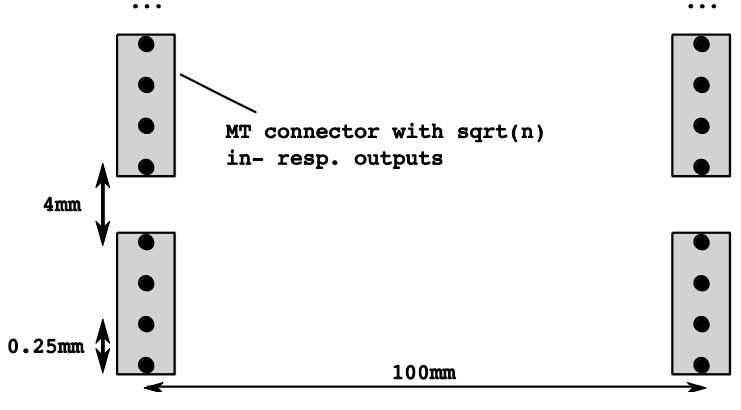


Fig. 4. Schematic illustration of the geometry of a 16 × 16 network

Better results can be found for other arrangements of the connectors. Figure 5 shows different grid-like arrangements of the 16 input connectors (output connectors are arranged in the same way) and the resulting fitness values.

We can see that the best result was achieved with an *8 × 2* arrangement, but here, too, the computed delays are not adequate, yet. From all results, we got so far, it is predictable that the connectors should be arranged as regular and near to each other as possible. Thus, "displacing" individual connectors in the grid-like arrangement is not promising. But by displacing all outputs (resp. analogue the inputs) in x- and/or y-direction (see Figure 6) the result can be further improved.

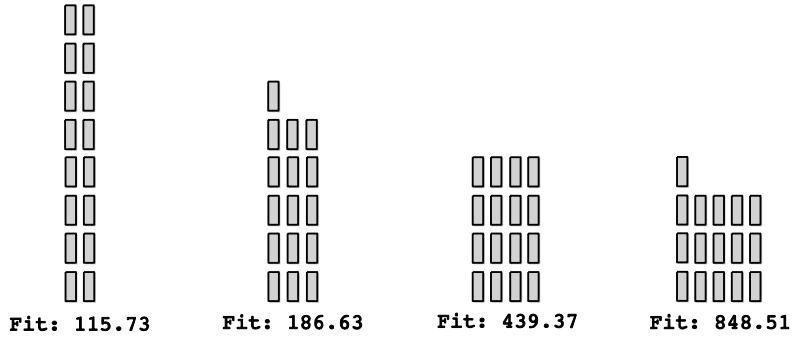


Fig. 5. Four different grid-like arrangements of 16 connectors and the fitness values that were achieved with them

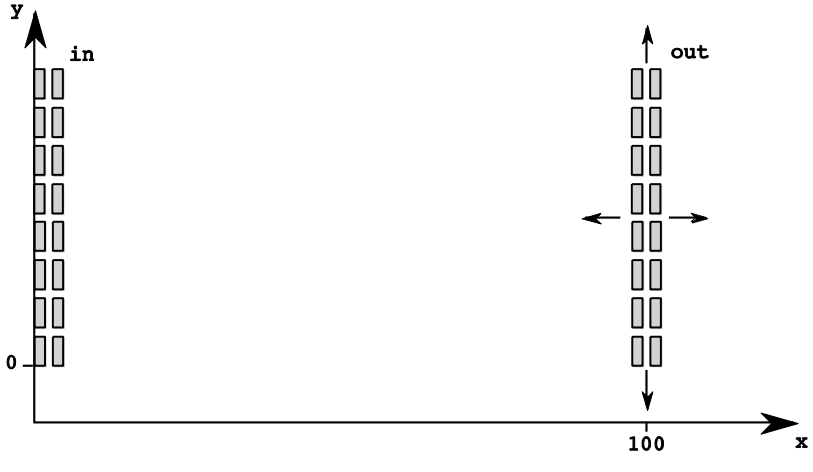


Fig. 6. Displacement of the output connectors from the initial position in x- and y-direction

Figure 7 shows the fitness values dependent on the displacement of the outputs in x-direction. A displacement by +7.1 *mm* yields a fitness of 88.225 and a displacement by +13.45 *mm* a fitness of 74.96. However, the resultant increase of the transmission distance may be problematic for the production. Via a displacement by -1.92 *mm* in x-direction and in addition by +0.2 *mm* in y-direction a fitness of 111.9995 can be

achieved, the best value for a transmission distance of max. 100 *mm*, so far. But the results can be further improved, namely by rotating the connectors by 90°. This yields a fitness of 91.284, which can be improved to even 75.88 via displacement of the outputs by -0.7 *mm* in x-direction and +4.1 *mm* in y-direction. Here the optical paths of only 4 out of the 256 connections are not in the desired range. That is the best result, we achieved so far.

Better results are possible for smaller pitches between the in- resp. output fibers, a bigger interval of permitted deflection angles or a longer transmission distance. But all this may cause problems in implementation. A possible alternative is the employment of other connectors, like MPO Connectors ([7]) that allow an arrangement of the fibers of the connector.

4.3 Conclusions for the Design

In Figure 8a-c the results of the performed optimization are shown. The best founded values for deflection angle and number of folding intervals generate the shown values for optical path length variances. In the diagrams all 256 optical channels and their deviation values are shown. This optimization is performed for the rotated MT-connector arrangement.

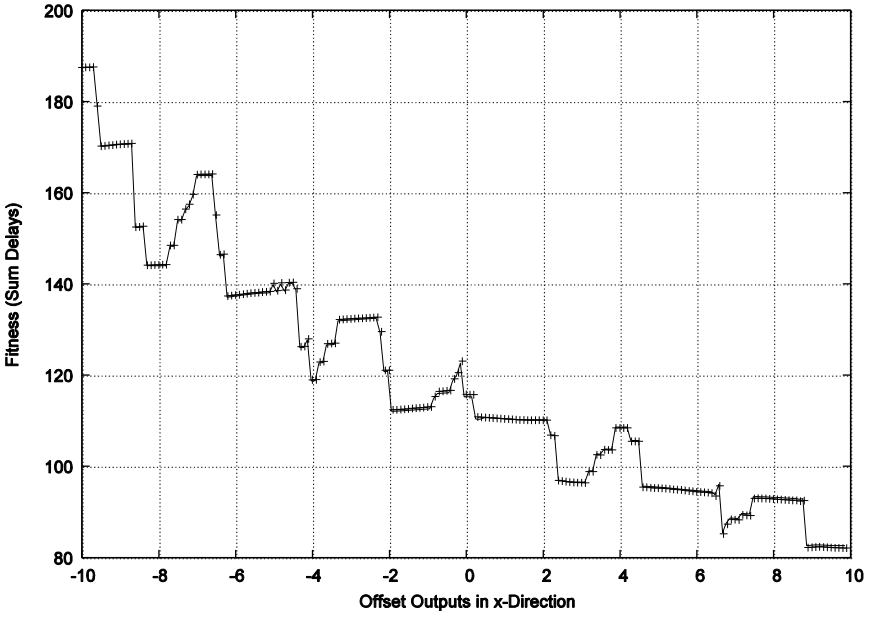


Fig. 7. Optimal fitness value subject to the displacement of the outputs in x-direction

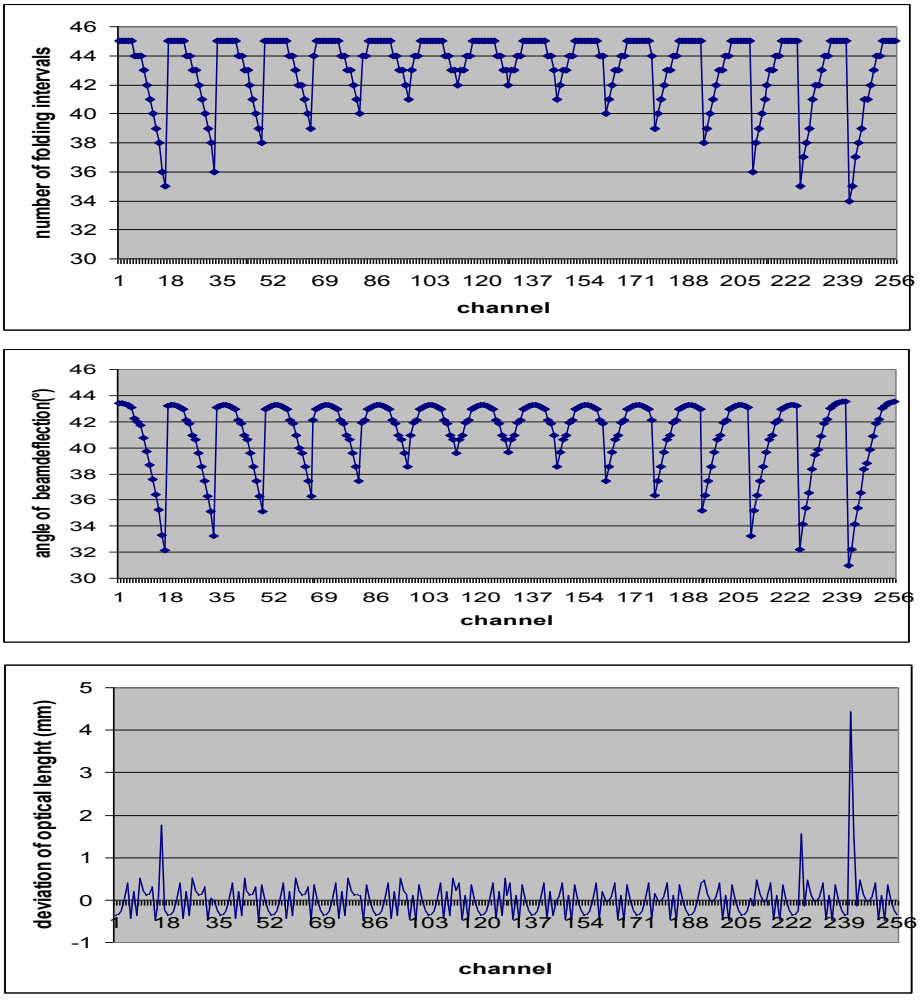


Fig. 8 a-c. The results of a) number of folding intervals, b) angle of beam deflection c) the deviation of the optical path length

5 Outlook

We have presented a novel approach to implement passive free-space optical cross interconnections to realize a high density Clos network. Although the results of the optimizations were not completely as desired, they can be used for an initial guess for design parameters.

In the next step we make use of the large number of individual mirror elements in a DMD array [8] to implement dynamic routing following a photonic network on chip. In Figure 10 the scheme of the experimental setup for the DMD approach is shown. The used optimization algorithm can be modified to calculate this DMD based dynamic routing with defined optical relations inside the PIFSO cuboid.

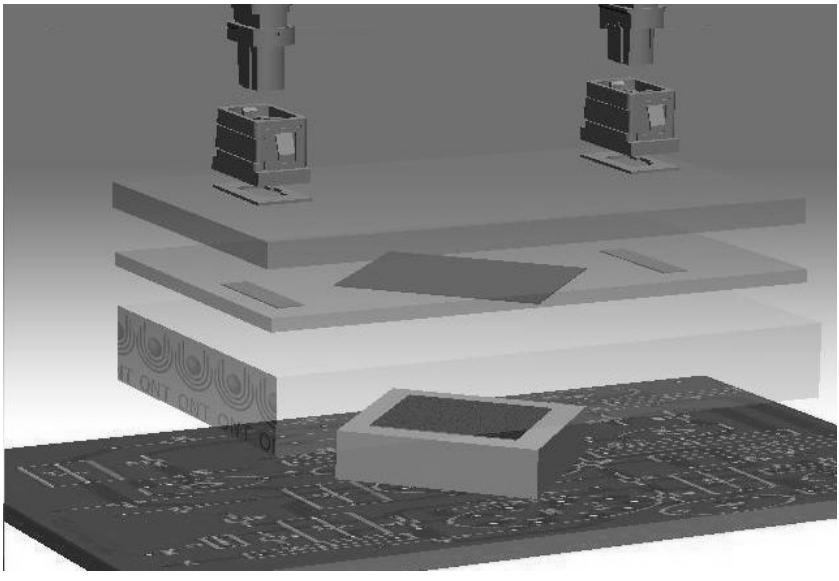


Fig. 9. Exemplary PIFSO & MEMS implementation of high density planar optical switch matrix Clos configurations (Photonic network on chip)

The results presented here, are first results that were obtained by using various simplifying assumptions. Next steps will include adapting the model step-by-step to realistic assumptions. Furthermore, our model and simulation tool will allow us to include modifications of the setup such as a variation of the pitch in the input and output fiber bundles and the geometrical relations for the optical beam deflection.

References

1. Fey, D., Schneider, M., Jahns, J., Knuppertz, H.: Optical Multiplexing Techniques for Photonic Clos Networks in High Performance Computing Architectures. In: Dolev, S., Oltean, M. (eds.) OSC 2009. LNCS, vol. 5882, pp. 110–123. Springer, Heidelberg (2009)
2. Jahns, J.: Digital Optical Computing and Interconnection. In: The Handbook of Photonics, 2nd edn. CRC Press, Boca Raton (2007)
3. Clos, C.: A Study of Non-Blocking Switching Networks. Bell Sys. Tech. J. 32, 406–424 (1953)
4. Goldberg, D.E.: Genetic Algorithms in Search, Optimization, and Machine Learning. Addison-Wesley, Reading (1989)
5. Miller, D.A.B.: Rationale and Challenges for Optical Interconnects to Electronic Chips. Proc. IEEE 88, 728–749 (2000)
6. Saleh, B.E.A., Teich, M.C.: Fundamentals of Photonics. Wiley, Chichester (2007)
7. Euromicron Corp. Datasheet: Specification of Standard MPO-Connectors. IEC (1996)
8. Texas Instrument Corp. Datasheet: Specification of Digital Mirror Devices (2009)
9. Jahns, J., Huang, A.: Planar integration of free-space optical components. Appl. Opt. 28, 1602–1605 (1989)

10. Gruber, M., Jahns, J., Sinzinger, S.: Planar-integrated optical vector-matrix multiplier. Appl. Opt. 39, 5367–5373 (2000)
11. Baudet, D., Braux, B., Prieur, O., Hughes, R., Wilkinson, M., Latunde-Dada, K., Jahns, J., Lohmann, U., Fey, D., Karafolas, N.: Innovative On Board Payload Optical Architecture for High Throughput Satellites. In: Proc. ICSO 2010 (2010)
12. El-Ghazawi, T.A.: Characteristics of the MasPar parallel I/O system, frontiers. In: Fifth Symposium on the Frontiers of Massively Parallel Computation (Frontiers 1995), pp. 265–272 (1995)
13. InfiniBand Trade Association, `http://www.infinibandta.org` (accessed November 2010)

An Optical Solution for the SAT Problem

Mihai Oltean and Oana Muntean

Department of Computer Science,
Faculty of Mathematics and Computer Science,
Babeş-Bolyai University, Kogălniceanu 1,
Cluj-Napoca, 400084, Romania
{moltean,oanamuntean}@cs.ubbcluj.ro
http://www.cs.ubbcluj.ro/∼moltean/optical

Abstract. We describe a delay-based optical device for solving the the Satisfiability problem. The device has a graph structure which is traversed by light in order to generate a solution. The device has 2 special nodes: a start node (where the initial pulse is sent) and a destination node (where the solution is read). Multiple signals are expected at the destination. Some of them contain valid solutions and others do not. A special, time filter, is proposed in order to detect a solution. Graphical simulations show how the device works.

Keywords: NP-complete, optical computing, SAT.

1 Introduction

Recently, an increased number of problems have been proposed to be solved using optical computers. Hamiltonian path [1, 11, 19], Travelling Salesman [1, 3, 4], subset sum [1, 12, 15, 17, 18, 25], exact cover [14], Diophantine equations [16], 3-SAT [1, 22, 23], SAT [7], prime factorization [20, 21, 9, 10], security [6], arithmetic operations [5], Boolean algebra [8] are just few of the problems whose solution can be found by using an optical device.

Here we show how to solve the Satisfiability (SAT) problem, which is an NP-complete problem [2]. The underlying mechanism is to use delays for encoding possible solutions. The problem was also attacked in [1] and in [22] in a different manner: by using masks and filters for wavelengths.

The paper is organized as follows: Properties of the signal useful for our research and the operations performed on that signal are described in section 3. Section 4 contains a brief description of the problem. Section 5 deeply describe the proposed device. How the system works is described in section 5.2. Reconstructing the solution is described in section 6. A short discussion of the weaknesses of this method is given in section 7. Section 8 concludes our paper.

2 Related Work

We are aware of three other optical approaches for solving the SAT problem:

S. Dolev and M. Oltean (Eds.): OSC 2010, LNCS 6748, pp. 53–62, 2011.

- Dolev et al. [1] constructed a tree arhitecture for each clause of the formula to be satisfied. Solutions are constructed by adding a new variable at each step. Transitions between states are implemented either by reducing the intensity of the signal or by inducing a delay. Beams which have a low intensity or which have been delay over a certain amount of time are filtered out. The results which have passed through each clause are combined using a lens and read with a detector.
- Goliaei et al. [22] proposed a device for the 3-SAT problem where the space of wavelengths is divided into 2^n zones, each of them encoding a possible solution to the problem. The first zone encode the assigments $x_1 = 0, x_2 = 0, ...x_n = 0$ and the last zone encode asignments $x_1 = 1, x_2 = 1, ...x_n = 1$. Filters for variables and clauses are constructed. These filters will let only some wavelengths which satisfy the solution to pass. If we have some wavelengths available at the output it means that we have a solution, otherwise we have no solution. Based on the zones that have passed all filters we may easily say which are the assignments which satisfy the formula. The weak point is that the space must be divided into an exponential number of zones.
- Tom Head [7] used tranparencies and a xerox machine to perform usefull computations. For the SAT problem with n variables, m clauses and at most k literals per clause, the table of all 2^n truth settings is constructed in n steps. Then each clause is evaluated in at most $2k$ copies of the truth table. Finally the solutions are combined in m steps. The drawback is related to the exponential increase in the information density.

3 Delay-Based Systems

Two properties of signal are of great interest for our research. Most types of signal that we know (light, sound, electric etc) have these properties.

- The speed of the signal has a limit. We can delay any signal by forcing it to pass through a cable of a certain length.
- The signal can be easily divided into multiple signals of smaller intensity/power.

The following manipulations of the signals are performed within the devices:

- When the signal passes through an arc it is delayed by the amount of time assigned to that arc.
- When the signal passes through a node it is divided into a number of signals equal to the out degree of that node. Every obtained signal is directed toward one of the nodes connected to the current node. In this way we add parallelism to our devices. This feature is actually the source for a major drawback: due to repetitive divisions the strength of the signal decreases exponentially and more and more powerful signals are required for larger and larger instances of the problems.

4 The SAT Problem

We have n Boolean variables (x_1, x_2, ... ,x_n) and m clauses of form $C = x_k \vee x_j \vee$ ..., where a variable can also appear negated. We are asking to find if there is an assignment for variables which satisfies a formula:

$F = C_1 \wedge C_2 \wedge ....C_m$

In the first step we are not interested in finding the value of the variables. Rather, we are interested in finding whether such assignment does exist. Only in section 6 we will show how a solution can be reconstructed.

5 The Device

Our device is represented as an incomplete matrix. On each column we have all the variables (as they appear) inside a clause. Thus, we have a matrix with m columns and max n elements on each row.

Example

Consider the formula:

$F = (x_1 \vee x_2 \vee \overline{x}_3) \wedge x_1 \wedge (\overline{x}_2 \vee \overline{x}_3) \wedge (x_1 \vee \overline{x}_3)$

We represent it as a matrix with 4 columns:

x_1	x_1	$\overline{x}_2$	x_1
x_2		$\overline{x}_3$	$\overline{x}_3$
$\overline{x}_3$			

Next we have to design a device where the signal will flow. For this purpose we do the followings:

- We place each element of this matrix in a Cartesian coordinates system.
- We connect each element with all elements from the next column.
- We also add 2 extra elements (nodes) called *Start* and *Destination*.

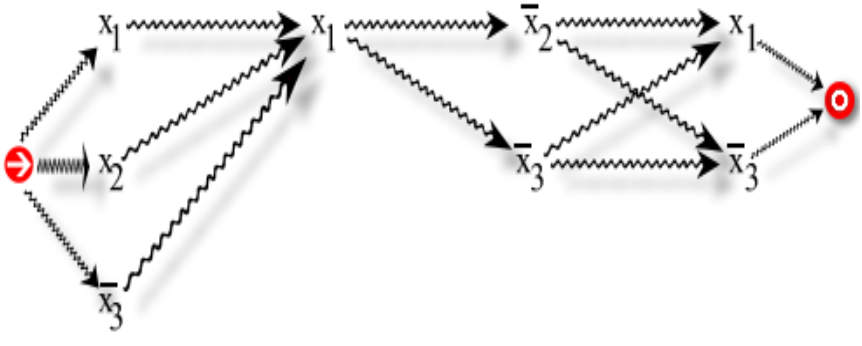


Fig. 1. Optical implementation for the SAT problem. On each column we have a clause. On the left side we have the *Start* node and on the right side we have the *Destination* node. Arcs connecting nodes have 0 delay.

The device is represented in Figure 1.

If we find a path that links *Start* with *Destination* and which does not contain both the variable and its negation it means that the formula is satisfiable. We call that path **satisfiable path**.

For our example, a satisfiable path is: $Start, x_1, x_1, \overline{x}_2, x_1, Destination$.

The big challenge is to identify such path.

5.1 The Delaying System

When solving the SAT problem with an optical system we assign delays to elements of the graph. Those delays will accumulate as the signal traverse the graph. At the end we will know that we had a solution only if we have received a signal having a certain (precomputed) delay.

Here we assign some delays to nodes in order to easily recognize the satisfiable paths.

To arcs connecting nodes we assign a negligible delay (very small - or even 0) compared to all other delays assigned to nodes.

5.1.1 The First Variable

Let us now focus on variable x_1 and its negation $\overline{x}_1$.

Suppose that we have a formula with 4 clauses. What are the minimal numbers that we can assign as delays for x_1 and $\overline{x}_1$ in order to safely decide that x_1 and its negation $\overline{x}_1$ do not (both) appear on a path from *Start* to *Destination*?

For instance, if we assign a delay of 1 for x_1 and a delay of 2 for $\overline{x}_1$ we get some trouble recognizing if the path whose delay is 4 is satisfiable or not. In this case, we can have a path containing 4 times x_1 (which is satisfiable) or a path containing 2 times x_1 and 1 time $\overline{x}_1$ (which is not satisfiable). Note that (having 4 clauses) x_1 can appear at most 4 times and $\overline{x}_1$ can appear at most (4 - number of appearances of x_1) times.

Let's try another arrangement: $delay(x_1) = 1$ and $delay(\overline{x}_1) = 4$. The number of clauses is again 4.

The analysis of paths' delay is given in Table 1.

If we are talking in the number of clauses (m) we have the following delays: $delay(x_1) = 1$ and $delay(\overline{x}_1) = m$. There is no problem if $delay(\overline{x}_1) = m * delay(x_1)$ because we have only m clauses and we cannot have both m times x_1 and 1 time $\overline{x}_1$.

In Table 2 we have the generalized values from Table 1.

If we look to the values in Table 2 we can see that only some delays encode a satisfiable path. More specific: only values $0...m$, $2m$, $3m$, ... m^2 encode a satifiable path. All other values encode either a path that is either not possible (more than m variables on it) or is not satisfiable (contains both x_1 and $\overline{x}_1$).

Having said that we have ended the discussion about the first variable.

5.1.2 The Second Variable

Let us move our attention to the next variable: x_2. We cannot assign delays less or equal to m^2 because this will lead to conflicts to the previous variable. What

Table 1. The delays induced by x_1 and $\overline{x}_1$ only. The *Start* and *Destination* have not been printed. Also, other variables have not been displayed. *Not possible* is because we have only 4 clauses, so the path cannot have more than 4 variables.

Delay	Content of path	Satisfiable?
0	neither x_1 or $\overline{x}_1$ appear	YES
1	x_1	YES
2	$x_1 + x_1$	YES
3	$x_1 + x_1 + x_1$	YES
4	$\overline{x}_1$	YES
5	$\overline{x}_1 + x_1$	NO
6	$\overline{x}_1 + x_1 + x_1$	NO
7	$\overline{x}_1 + x_1 + x_1 + x_1$	NO
8	$\overline{x}_1 + \overline{x}_1$	YES
9	$\overline{x}_1 + \overline{x}_1 + x_1$	NO
10	$\overline{x}_1 + \overline{x}_1 + x_1 + x_1$	NO
11	Not possible	-
12	$\overline{x}_1 + \overline{x}_1 + \overline{x}_1$	YES
13	$\overline{x}_1 + \overline{x}_1 + \overline{x}_1 + x_1$	NO
14	Not possible	-
15	Not possible	-
16	$\overline{x}_1 + \overline{x}_1 + \overline{x}_1 + \overline{x}_1$	YES

we can do is to assign: delay$(x_2) = m^2$ and delay$(\overline{x}_2) = m^3$ (obtained with the same reasoning as in the case of the first variable).

In this case the delays encoding a satisfiable path are: $m^2, 2m^2, 3m^2, ..., m^3$, $2m^3, 3m^3, ..., m^4$.

5.1.3 The Other Variables

For the next variable (x_3) we can assign: delay$(x_3) = m^4$ and delay$(\overline{x}_3) = m^5$ (obtained with the same reasoning as in the case of the first variable).

In this case the delays encoding a satisfiable path are: m^4, $2m^4$, $3m^4$, ..., m^5, $2m^5$, $3m^5$, ..., m^6.

For the k^{th} variable we can assign: delay$(x_k) = m^{2k-2}$ and delay$(\overline{x}_k) = m^{2k-1}$ (obtained with the same reasoning as in the case of the first variable).

In this case the delays encoding a satisfiable path are: m^{2k-2}, $2m^{2k-2}$, $3m^{2k-2}$, ..., m^{2k-1}, $2m^{2k-1}$, $3m^{2k-1}$, ..., m^{2k}.

5.1.4 Combining Delays

The SAT formula is made up of clauses containing multiple variables (not only single variables as we discussed so far in sections 5.1.1, 5.1.2 and 5.1.3).

What we have to do is to add the delays. If we take into account only variables x_1 and x_2 (and their negation) we have to add each delay for variable x_1 (or its negation) to each delay for variable x_2 (or its negation). In total are $4m^2$ moments when a signal traversing a satisfiable path arrives in the destination.

The moments when the signal traversing a satisfiable path can arrive in the destination are given in see Table 3. Note that some moments are not possible

Table 2. The generalized delays induced by a single variable (x_1 and $\overline{x}_1$). m is the number of clauses. Some delays are not possible (for instance $3m-1$ - see Table 1 for an example.)

Delay	Content of path	Satisfiable?
0	neither x_1 or $\overline{x}_1$ appear	YES
$1...m-1$	x_1 only	YES
m	$\overline{x}_1$ only	YES
$m+1...2m-1$	both $\overline{x}_1 + x_1$	NO
$2m$	$\overline{x}_1$ only	YES
...	...	...
m^2	only $\overline{x}_1$	YES

because more than m variables appear in path. For instance a delay of $(m-1)*m^2+m-1$ means that x_2 appears $m-1$ times and x_1 appears $m-1$ times, which is not possible.

Table 3. The delays induced by single variables (first 2 lines) and by 2 variables (third line). Some delays are not possible.

Involved variables	Delays
$\overline{x}_1, x_1$	$0...m,\ 2m,\ 3m,\ ...\ m^2$
$\overline{x}_2, x_2$	$m^2, 2m^2, 3m^2, ..., m^3,\ 2m^3, 3m^3, ..., m^4$
$\overline{x}_1, x_1, \overline{x}_2, x_2$	$0...m,\ 2m,\ 3m,\ ...$ $m^2,$ $m^2+1,\ m^2+2,\ m^2+3,\ ...,\ m^2+m,\ m^2+2m,\ m^2+2m,\ ...\ ,$ $2m^2,$ $2m^2+1,\ 2m^2+2,\ 2m^2+3,\ ...,\ 2m^2+m,\ 2m^2+2m,\ 2m^2+2m,$ $...\ ,\ 3m^2,$ .... $m^3+1,\ m^3+2,\ m^3+3,\ ...,\ m^3+m,\ m^3+2m,\ m^3+2m,\ ...\ ,$ $m^3+m^2,$ $2m^3,\ 2m^3+1,\ 2m^3+2,\ 2m^3+3,\ ...,\ 2m^3+m,\ 2m^3+2m,$ $2m^3+2m,\ ...\ ,\ 2m^3+m^2,$ ..., $m^4,\ m^4+1,\ m^4+2,\ m^4+3,\ m^4+m,\ m^4+2m,\ m^4+2m,\ ...\ ,$ $m^4+m^2.$

This set of delays is only for 2 variables (including their negation). For n variable we would have $(2m)^n$ possible moments for listening for a solution.

The main advantage is that these delays do not depend on the actual problem to be solved. These delays do not take into account what literals we have in each clause. This is very important because we can construct the arcs encoding delays and time filters once and we can use them again and again for different instances of the problem (with m clauses and n literals). This is a big advantage.

5.2 How the System Works

In the graph depicted in Figure 5.2 the light will enter in *Start* node. It will be divided into 3 subrays of smaller intensity. These 3 rays will arrive into the second node at moments 1, m^2 and m^5 units of time. Each of them will be further delayed with 1 unit of time and will arrive in the 3^{rd} node at moments 2, m^2+1 and m^5+1. The process goes on until all photons arrive in the destination node.

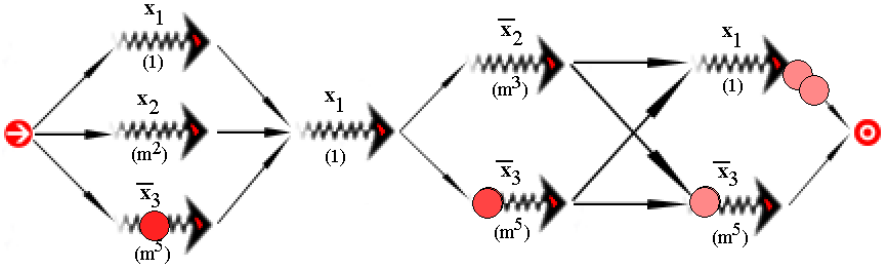


Fig. 2. A screenshot of the optical simulation for the SAT problem. Multiple photons are depicted running through device. Some of them are at the begining of their trip (left side of the picture) because they have been delayed more. Some others are already near the destination (right side of the device) because they have followed some paths with small delays. The intensity decrease can also be observed: the photons in the left side have a stronger color than those from the right side

In the destination node we will have multiple rays arriving at no more than $(2m)^n$ different moments (see section 5.1.4).

Because we are working with continuous signal we cannot expect to have discrete output at the destination node. This means that rays arrival is notified by fluctuations in the intensity of the light. These fluctuations will be transformed, by a photodiode, in fluctuations of the electric power which will be easily read by an oscilloscope.

A movie showing the entire simulation can be viewed here: `http://www.youtube.com/watch?v=_hdVKKVMocA`.

6 Reconstructing the Solution

Once we have a satisfiable delay lets see how to reconstruct which variables are inside it and what is their form (positive or negated).

First of all we discuss some special cases, where we have exactly delay of m or m^2 or m^3 etc units.

For instance if we have a delay of exact m it means that we have exactly m times x_1. We cannot have a delay of exact m which contains $\overline{x}_1$ because the path contains m nodes (excluding *Start* and *Destination*) and those nodes will increase the delay to more than m (which is already induced by $\overline{x}_1$).

If the total delay is m^3 it means that we have x_2 m times. It cannot contain $\overline{x}_2$ because the path must have other $m-1$ nodes which will add more delay.

Thus, for the particular cases when the total delay is of form m^k, $(1 \leq k \leq 2n+1)$ we know exactly what nodes the satisfiable path contains. These particular cases will be handled separately.

Lest see how we discover the content of the path for all other cases. First we divide the delay by : m^{2n} (this value corresponds to the delay of $\overline{x}_n$ - being the highest delay possible). We have 2 cases:

- If the quotient is 1 ... m it means that $\overline{x}_n$ belongs to the path. It means that variable x_n will be assigned value 0.
- If the quotient is 0 it means that $\overline{x}_n$ does not belong to the path.

Next we take the remainder of the previous division and we repeat the process with delay corresponding to the previous variable. The order is $\overline{x}_n$, x_n, $\overline{x}_{n-1}$, x_{n-1}, ... $\overline{x}_1$, x_1. If the quotient is greater than 0, it means that the variable (positive or negated) belongs to the path. We assign a value to the variable based on the form in which has appeared (0 if negated, 1 if positive). If a variable has not appeared in the path, it can take any value (0 or 1) because this has no influence over the value of formula F from section 4.

7 Weaknesses

The proposed device has some weaknesses:

- it contains exponential delays,
- it requires time-filters for events that appear in time,
- the number of possible moments when a solution can appear is exponential in the number of clauses. This is different from other problems (see [13]) where only one moment was enough for detecting the solution,
- no proof yet for the optimality of the delay system. Are shorter delays possible ?

8 Conclusion

We have shown how SAT can be solved with an optical device. The proposed device is based on the brute-force approach which generates all good or wrong solutions. Finally the good ones are selected with a time-filter. Optical implementation is natural option for this kind of approach because light can be easily divided into multiple rays which can follow different paths. Physical implementation could small or moderate instances.

Acknowledgment

A graphical simulation of the device can be viewed at: `http://www.youtube.com/watch?v=_hdVKKVMocA`. This research was supported form the CNCSIS grant IDEI 2412/2008.

References

[1] Dolev, S., Fitoussi, H.: The Traveling Beams Optical Solutions for Bounded NP-Complete Problems. In: Crescenzi, P., Prencipe, G., Pucci, G. (eds.) FUN 2007. LNCS, vol. 4475, pp. 120–134. Springer, Heidelberg (2007)

[2] Garey, M., Johnson, D.: Computers and intractability: A guide to NP-Completeness. Freeman & Co, San Francisco (1979)

[3] Haist, T., Osten, W.: An Optical Solution For The Traveling Salesman Problem. Opt. Express 15, 10473–10482 (2007)

[4] Haist, T., Osten, W.: An Optical Solution For The Traveling Salesman Problem:erratum. Opt. Express 15, 12627 (2007)

[5] Haist, T., Osten, W.: Ultrafast Digital-Optical Arithmetic Using Wave-Optical Computing. In: Dolev, S., Haist, T., Oltean, M. (eds.) OSC 2008. LNCS, vol. 5172, pp. 33–45. Springer, Heidelberg (2008)

[6] Haist, T., Osten, W.: Proposal for Secure Key Distribution Using Classical Optics. In: Dolev, S., Oltean, M. (eds.) OSC 2009. LNCS, vol. 5882, pp. 99–101. Springer, Heidelberg (2009)

[7] Head, T.: Parallel Computing by Xeroxing on Transparencies. Algorithmic Bioprocesses, part 9, 631–637 (2009)

[8] Head, T.: Using Light to Implement Parallel Boolean Algebra. In: Gao, Y., Lu, H., Seki, S., Yu, S. (eds.) DLT 2010. LNCS, vol. 6224, pp. 231–242. Springer, Heidelberg (2010)

[9] Nitta, K., Katsuta, N., Matoba, O.: Improvement of a System for Prime Factorization Based on Optical Interferometer. In: Dolev, S., Oltean, M. (eds.) OSC 2009. LNCS, vol. 5882, pp. 124–129. Springer, Heidelberg (2009)

[10] Nitta, K., Katsuta, N., Matoba, O.: A Method for Modulo Operation by Use of Spatial Parallelism. In: Dolev, S., Haist, T., Oltean, M. (eds.) OSC 2008. LNCS, vol. 5172, pp. 98–103. Springer, Heidelberg (2008)

[11] Oltean, M.: A Light-Based Device for Solving the Hamiltonian Path Problem. In: Calude, C.S., Dinneen, M.J., Păun, G., Rozenberg, G., Stepney, S. (eds.) UC 2006. LNCS, vol. 4135, pp. 217–227. Springer, Heidelberg (2006)

[12] Muntean, O.: Optical Solutions for NP-complete problems, graduation thesis, Faculty of Mathematics and Computer Science, Babes-Bolyai University, Cluj-Napoca, Romania, defended on July 3 (2007)

[13] Oltean, M., Muntean, O.: Solving NP-Complete Problems with Delayed Signals: An Overview of Current Research Directions. In: Dolev, S., Haist, T., Oltean, M. (eds.) OSC 2008. LNCS, vol. 5172, pp. 115–127. Springer, Heidelberg (2008)

[14] Oltean, M., Muntean, O.: Exact Cover with light. New Generation Computing 26(4) (2008)

[15] Muntean, O., Oltean, M.: Using light for solving the unbounded subset-sum problem. IJICIC 5(8), 2159–2167 (2009)

[16] Muntean, O., Oltean, O.: Deciding whether a linear Diophantine equation has solutions by using a light-based device. Journal of Optoelectronics and Advanced Materials 11(11), 1728–1734 (2009)

[17] Oltean, M., Muntean, O.: Solving the subset-sum problem with a light-based device. Natural Computing 8(2), 321–331 (2009)

[18] Hasan, M. R., Rahman, M.S.: Computing a Solution for the Subset Sum Problem with a Light Based Device. In: Dolev, S., Oltean, M. (eds.) OSC 2009. LNCS, vol. 5882, pp. 70–76. Springer, Heidelberg (2009)

[19] Shaked, N., Messika, S., Dolev, S., Rosen, J.: Optical solution for bounded NP-complete problems. Applied Optics 46, 711–724 (2007)
[20] Shamir, A.: Factoring large numbers with the TWINKLE device. In: Koç, Ç.K., Paar, C. (eds.) CHES 1999. LNCS, vol. 1717, pp. 2–12. Springer, Heidelberg (1999)
[21] Lenstra, A.K., Shamir, A.: Analysis and optimization of the TWINKLE factoring device. In: Preneel, B. (ed.) EUROCRYPT 2000. LNCS, vol. 1807, pp. 35–52. Springer, Heidelberg (2000)
[22] Goliaei, S., Jalili, S.: An Optical Wavelength-Based Solution to the 3-SAT Problem. In: Dolev, S., Oltean, M. (eds.) OSC 2009. LNCS, vol. 5882, pp. 77–85. Springer, Heidelberg (2009)
[23] Goliaei, S., Jalili S.: An Optical Solution to the 3-SAT Problem using Wave-length Based Selectors. International Journal of Supercomputing (in press)
[24] Woods, D., Naughton, T.J.: Parallel and Sequential Optical Computing. In: Dolev, S., Haist, T., Oltean, M. (eds.) OSC 2008. LNCS, vol. 5172, pp. 70–86. Springer, Heidelberg (2008)
[25] Optical simulation for the subset sum problem, `http://www.youtube.com/watch?v=moPGLwhm_XM` (last accessed on March 17, 2011)

Compressive Sensing of Object-Signature

Dan E. Tamir[1], Natan T. Shaked[2], Wilhelmus J. Geerts[3], and Shlomi Dolev[4]

[1] Department of Computer Science, Texas State University, San Marcos, Texas 78666, USA
dt19@txstate.edu
[2] Department of Electrical and Computer Engineering, Ben-Gurion University of the Negev, P.O. Box 653, Beer-Sheva 84105, Israel
natis@ee.bgu.ac.il
[3] Department of Physics, Texas State University, San Marcos, Texas 78666, USA
wg06@txstate.edu
[4] Department of Computer Science, Ben-Gurion University of the Negev, P.O. Box 653, Beer-Sheva 84105, Israel
dolev@cs.bgu.ac.il

Abstract Compressive sensing is a new framework for signal acquisition, compression, and processing. Of specific interest are two-dimensional signals such as images where an optical unit performs the acquisition and compression (i.e., compressive sensing or compressive imaging). The signal reconstruction and processing can be done by optical signal processing and/or digital signal processing. In this paper we review the theoretical basis of compressive sensing, present an optical implementation of image acquisition, and introduce a new application of compressive sensing where the actual signals used in the compressive sensing process are image object-signature (an object-signature is a specific representation of an object). We detail the application of compressive sensing to image object-signature and show the potential of compressive sensing to compress the data through analysis of several methods for obtaining signature and evaluation of the rate/distortions results of different compression methods including compressive sensing applied to object-signature.

Keywords: Digital Signal Processing, Compressive Sampling, Compressive Sensing, Compressive Imaging, Optical Super Computing.

1 Introduction

Compressive sensing (CS) is a new paradigm for data acquisition [1,2]. The method exploits the fact that in most cases, especially when lossy compression is involved, digital signals used in signal processing, image processing, multimedia applications, and data communication have structural information and redundancy [1,2]. While the structural information is not always evident in the time domain, current processing techniques, in specific numerous data compression procedures, exploit the structure and the implied redundancy through transform coding and quantization [3,4]. In general, the result of transform coding and quantization is a sparse signal. That is, a

S. Dolev and M. Oltean (Eds.): OSC 2010, LNCS 6748, pp. 63–77, 2011.

signal where many of the signal elements have a value of 0. For example, several compression-methods, such as the JPEG image compression; apply discrete cosine transform (DCT) and quantization to an image sub-block [3,5]. The result is a sparse signal that is efficiently encoded using lossless compression. The decompression module applies inverse DCT to the sparse DCT-domain sub-block. Depending on the quantization and the sparsity of the sub-block in the DCT-domain, JPEG can result in superb rate/distortion. That is, JPEG can produce high compression rate with low visual and/or mathematical based (e.g., minimum square error - MSE) reconstruction error. Note, however, that the JPEG process itself is not designed to minimize the l_2 error (MSE), yet low MSE is generally a by-product of the compression. While JPEG is embedded in almost every commercial grade digital camera, JPEG2K which uses wavelet transform in lieu of the DCT transform is less popular. Nevertheless, as a part of the inception of the JPEG2K standard it has been demonstrated that wavelet based compression can result in better rate/distortion at about the same complexity (computational complexity and/or hardware complexity) of the DCT based JPEG compression. Again, the application of wavelet along with quantization produces a sparse signal [3,5].

Under the traditional approach to digital signal processing, the signal is obtained through analog to digital conversion (ADC) [6]. The ADC process includes sampling and quantization of an analog signal. To maintain reconstruction quality the sampling is complying with the Nyquist bounds [6]. Theoretically the compliance with the Nyquist bounds guarantees perfect reconstruction of the underlying analog signal [6]. The quantization, which is a "destructive" process, can be done in numerous ways and is generally governed by the register length of the digital signal processing hardware [3,5,6]. For data compression, additional subsequent quantization, such as vector quantization, might be applied in order to reduce the data size [7,8]. Nevertheless, this quantization generally results in increased reconstruction distortion. The entire process of sampling quantization and reconstruction is referred to as the Shannon-Nyquist ADC process [3,6].

For numerous signals, in specific for sparse signals, compressive sampling can be used as an alternative to the ADC process [1,2]. The novelty of compressive sampling is the observation that, in many cases, analog and digital signals can be measured (rather than sampled) using a finite set of points and then reconstructed with very low (often 0) distortion. The measurement process, referred to as CS, involves obtaining a finite set of values of the signal, or the result of signal transform, provided that the signal/transformed-signal is sparse. In practice, the signal/transformed-signal is measured, calculated, or "sensed" only in the points of interest. In other word the transform is not applied to the entire signal and is only calculated for a finite set of points of the signal [1,2]. This observation has opened the door for numerous new and innovative methods for data acquisition, data compression, and signal processing. Of specific interest are optical methods for the acquisition and processing of data using the principles of compressive sampling [4]. In the context of image data these methods are referred to as compressive imaging (CI) [4].

Significant research on CS has started in the late 90's, and picked up following seminal papers by Candès and Donoho (2002 – 2006) [1,2]. Following these breakthrough research and publications, numerous research efforts summarized in a

large number of papers have emerged. Currently, there is an extensive research activity in the area. In essence the research has progressed from concentrating on the data acquisition to the examination of the combination of data acquisition, remote sensing, data compression, and signal processing [9-12].

One of the active areas of research in image processing and machine vision is shape-representation [13-15]. Under this discipline the pixels in the boundary of image-objects are represented using a set of descriptors. For Example, the signature of an object is a set of distances from a reference point, generally the object centroid, to boundary pixels. Baggs and Tamir (1995) as well as Keogh (2003) present several variants of the signature [14,15]. These representations are invariant to affine transformations ([14,15]) as well as several other non-linear transformations which result in warping of objects [14].

Despite the intensive exploration of the utility of CS to general image processing techniques there are no reports concerning the application of CS to image object-signature. Motivated by the importance of using shape descriptors, in specific object-signature, in the area of image alignment/registration, image recognition, and image compression, or more broadly, in the area of machine vision and machine learning, we have initiated research into several aspects of CS of object-signature. As far as we know, and based on an extensive literature review, this is the first report on the application of CS to object-signature applying the results to machine vision and machine learning.

This paper presents the application of CS to image object-signature. In section 2, we introduce the theoretical background of compressive sampling. Next, in section 3, we review the subject of CS concentrating on the optical computing aspects of this method. In section 4 we introduce the application of CS to object-signature. Section 5 presents experiments with CS of object-signature, and section 6 includes conclusions and proposals for further research.

2 Compressive Sampling

Compressive sampling is an alternative to the traditional Shannon-Nyquist based analog to digital conversion (ADC) [1,2,6]. Under this approach, rather than sampling an analog signal, measurements of a sparse signal or a sparse transform of the signal are obtained and quantized [5]. This sampling method has an advantage over the traditional sampling provided that it results in sufficient reconstruction accuracy and the number of samples is smaller than the number of samples required by the Nyquist theorem. Candès and Donoho, as well as many other researchers, have investigated numerous aspects of compressive sampling and their applications in the fields of sensing, imaging, and signal processing [9-12]. In this section we concentrate on compressive sampling which is the theoretical foundation of CS. We examine the requirements posed on a signal and the number of measurements required in order to obtain reconstruction with sufficient quality [1,2]. The theory of CS applies to analog signals. Nevertheless, without loss of generality and in order to simplify the discussion we assume that the signal is discrete.

Consider an N-dimensional vector (signal) $s \in R^N$. The goal is to obtain a representation of s in the form of $x \in R^M$ such that: 1) x is a set of linear measurements of s, 2) s can be reconstructed from x, and 3) $M \ll N$. Hence, the vector x is of the form:

$$x_i = \langle s, \varphi_i \rangle, \ i = 1, \ldots, M \text{ or } x = \Phi s \qquad \text{(eq. 1)}$$

That is, x is the result of correlating s with the waveforms φ_i. In other words, x, the representation of s, is obtained via sensing of s using M vectors of the form $\varphi_i \in R^N$. The theory of compressive sampling states the condition on s, x, and M as well as the conditions on the reconstruction method such that s can be reconstructed from x with acceptable quality. Next, we discuss these conditions.

Let $\|a\|_{l_1}$ stand for the l_1-norm of a. That is $\|a\|_{l_1} = \sum_i^N |a_i|$, and let $|a|$ stand for the cardinality of a. A signal s is S-sparse if the cardinality of its support complies with $|\{i: \ s_i \neq 0 \}| \leq S$.

According to Candès *et al*, one could almost always recover the signal s in an exact way from the set of measurements x by solving the convex program

$$\min_{s' \in R^N} \|s'\|_{l_1} \quad \text{subject to } \Phi s' = x \quad (x \in R^M) \qquad \text{(eq. 2)}$$

provided that $M \geq C \cdot S \cdot \log(N)$. Where C is a constant.

The theoretical value of C and bounds over this value has been investigated and documented in several research papers [1,2]. In practice, often the value $C = 4$ gives good signal to noise ratio (SNR) results [2].

Using the traditional sampling theory, N, the number of samples required for exact reconstruction of s, must comply with the Nyquist theorem. The novelty of compressive sampling is that the number of required measurements M is much smaller than the number of required samples N.

In practice, the vector s might be the coefficients of a signal $y \in R^N$ in an orthonormal basis Ψ, that is $s = \Psi y$, and $y = \Psi^* s$. The signal y is considered to be sparse and compressible in the Ψ-domain if it can be represented with a set of coefficients s such that the support of s is small, and s is concentrated in a small subset of the domain. In this case, we can obtain s via $x = \Phi' s$ where $\Phi' = \Phi\Psi^*$. Furthermore, we can recover y using:

$$\min \|s'\|_{l_1} \ \text{ subject to } \Phi' s' = x \quad (x \in R^M) \qquad \text{(eq. 3)}$$

Additional theoretical results and practical implementations of the theory of compressed sampling can be found in [1,2].

3 Compressive Sensing and Compressive Imaging

In the conventional sensing framework, a maximal amount of data is first collected and then redundant data is removed by digital compression methods. In contrast, CS tries to minimize the collection of redundant data in the data acquisition stage [2,16]. In CS, the data is measured indirectly, typically by non-conventional acquisition techniques, so that fewer samples have to be acquired. Suitable algorithms are then applied to convert the acquired data to a conventional form. CI is CS with visual data

(images) [4, 16]. In cases where regular image acquisition is time consuming, has high pixel cost, or when the system requirements cannot afford digital compression before data storage due to data transfer limitations or data processing power limitations, advantages can be gained by using CI. The cost of using CI, however, is special hardware design for the acquisition process, and dedicated digital post-processing (typically not performed during the acquisition or storage stages).

In general, CI refers to obtaining an image with a small number of sensors. Takhar *et al*, however, propose a single pixel representation [4]. This is the first CI system, in which digital mirror-array device is used to randomly project images of the recoded object on a single-pixel sensor in a successive manner [4]. Then, a non-linear numerical reconstruction algorithm is applied to calculate the visual image. Using this time multiplexing method, compression is obtained by requiring less sensor exposures than the number of pixels in the 2-D image. Terahertz imaging version of this camera is presented in [18,19].

Stern and Javidi present the CI-vector camera, which is a single-shot motion-free CI approach [20]. In this approach, a randomly-coded aperture is used to obtain the random projections. In Ref. [21], Stern proposes to use vector sensor that scans the field of view by rotational motion. The latter technique creates non-fully-random projections, and thus induces sub-optimal compression. In practice, the technique can obtain good compression ratios, with excellent acquisition time and complexity of the system, reconstruction algorithm, and storage requirements [22,23]. In Ref. [24], random aperture coding is implemented for spectral CI. In Ref. [25], random micro-lenses are used to obtain the random projections. Mahalanobis proposes to use spatially multichannel aperture coding [26].

Using a more general perspective on CI, Brady *et al* [27,28] suggest the compressive holography technique, showing that under certain assumptions, it is possible to obtain multidimensional images from lower dimensional data. Using coherent illumination, holography can capture the wave-front of the light that has interacted with the object. In general, however, it is not possible to obtain 2-D digital slices (2-D regular images in different depths of the object) from a single hologram. In contrast, Brady *et al.* show that for sparse objects, it is possible to obtain a sliced or sectioned 3-D image, although only a single 2-D hologram is acquired [28].

A typical setup for a single-pixel-sensor system using the principle of CI consists of a detector, two lenses, and a spatial modulator. Figure 1 shows a schematic diagram of this setup. The object is imaged on the spatial modulator by one of the lenses, while the second lens projects this image on the detector. The detector signal is sampled and digitized by a sample and hold and an ADC convertor. The signal reconstruction is performed by a DSP. The role of the spatial filter is to convert the spatial dependent intensity distribution into time-dependent intensity at the detector. In the simplest form the spatial filter could select one pixel at a time, scanning the light transmitted through, or reflected from, the sample; pixel by pixel. Takhar *et al* have shown that by an educated choice of the spatial filter's transmission, a high quality image can be obtained or a small amount of image samples (with the limit being one) are required [4].

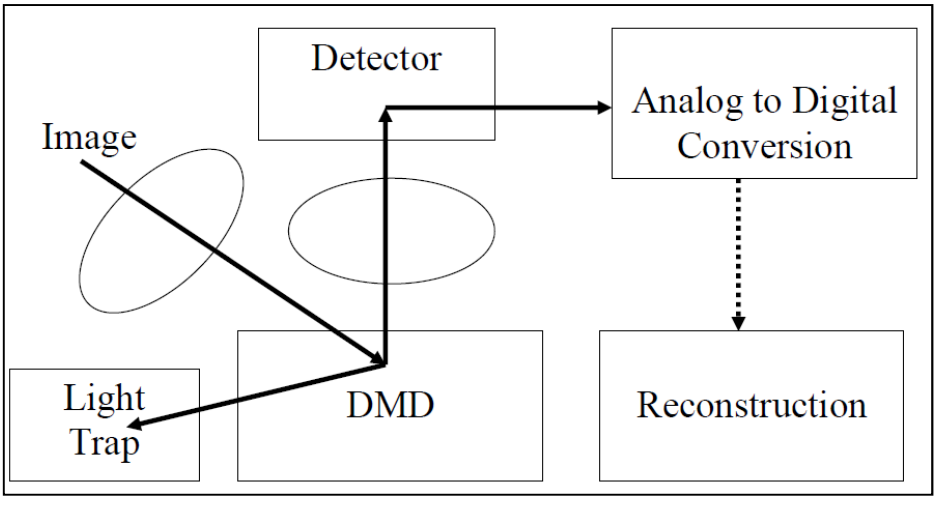


Fig.1. A Schematic Diagram of the Compressive Imaging Process

The spatial filter is normally implemented by a digital mirror device (DMD) consisting of an array of small mirrors. The orientation of the mirrors can be controlled by small electromagnet [29], piezo-electric [30], memory material based [31], or electrostatic driven actuators [32]. Most often used are the DMDs of Texas Instruments (TI); where individual mirrors can be set at 12 and -12 degrees. The optical design of the system is such that for one of the two angular positions the mirror reflects the light outside of the light path into a light trap while for the other angular position; the light beam is reflected to a lens that focuses it on the detector. The TI DMD is fast, but allows for only two stable orientations for each mirror. Large arrays over 1 megapixel are available for a rather low price. Other researchers have implemented the spatial filter by moving pre-fabricated masks in and out the light path with a step-motor [18]. Additional options for implementing the spatial filter include the application of a micro-electro-mechanical system (MEMS) driven shutter array [31,33] or a liquid crystal spatial filter [34]. Note that for some applications the spatial filter can be placed between the light source and the object.

For low light and low dose applications, the CI approach is generally a better choice than the scanning beam technique since a second sensor that measures the deflected beam is included, and all the photons emitted from the object can be measured. In addition, this approach enables imaging systems for technologies where a pixel array is expensive [4]. Further advantages are for low-power and/or low-processing-capability applications such as long space-missions.

4 Object-Signature

Object-signature is an effective and compact way to represent object shape. A set of experiments with CS of object-signature is reported in section 5. This section (section 4) provides the details of the object-signature shape-representation method.

One of the active areas of research in image processing is shape-representation [13-15]. Under this discipline the pixels in the boundary of image-objects are represented using a set of descriptors. In this context an object is a contiguous set of pixels with uniform gray level. That is, the variance of gray levels of object pixels is "small;" potentially 0. The representation can be used for object/image recognition, image alignment/registration, and image data compression. Generally, the procedure for shape-representation includes identifying image-objects, marking the pixels in the boundary of the object, and using a specific, compact and efficient, method for representing these pixels. Of interest are shape-representation techniques that are invariant to spatial transformations such as translation, rotation, and scaling [13]. Some of the commonly used methods for shape-representation include chain codes [5], Fourier descriptors [5], B-Splines [13], Moments [5,13] and several variants of the signature of the object [14]. The signature of an object is a set of distances from a reference point, generally the object centroid, to boundary pixels [14]. Baggs and Tamir as well as Keogh present several variants of the signature [14,15]. These representations are invariant to object warping in the form of affine transformations such as translation, rotation, scaling, and shearing. In addition, Baggs and Tamir demonstrate methods for dealing with non-linear warping of objects [14]. Furthermore the representations proposed can be efficiently used for object recognition, alignment/registration, and compression [14].

In order to obtain object-signature the image has to be segmented, objects have to be identified, and the pixels that reside on object contours have to be marked [13]. Image segmentation can be implemented in several different ways [5]. The method chosen in this research is the robust vector quantization (VQ) based image segmentation method described in [8]. Moreover, our implementation of the VQ algorithm is about four times faster than the commonly used Linde, Buzo, and Gray (LBG) algorithm [35]. This results in a computationally efficient segmentation method.

4.1 Shape-Representation Using Object-Signature

In general, the term object-signature relates to a set of measurements of the distances from a fixed point to object boundary elements. The term object-signature, however, is "overloaded." In a fixed angle signature, the distances from the object center to its boundary are measured in increments of equal angles. A second approach is to measure the distance from the center to every pixel on the contour of the object. A problem with the fixed angle method occurs with certain concave objects where the same angle might yield more than one distance to points on the contour. The "every-pixel" method removes the requirement for convexity (in practice it is enough to require that the shape is a star convex shape [36]) and can use an arbitrary center with arbitrary objects. On the other hand, the "every-pixel" approach generates a variable number of samples which depends on the object and is sensitive to object scaling. This might necessitate non-linear warping, referred to as dynamic space warping, of the signature [14,15]. This work uses a variant of the object-signature referred to as the "length-code" (LC) which is the cyclic auto correlation of the set of distances measured from the object centroid to its boundary. The signature generation method

used in our CS experiments is the equal angel. In addition, before obtaining the LC, the objects are normalized via scaling.

The cyclic auto correlation function of the signature sequence is rotational and translational invariant [14]. The first autocorrelation coefficient of the signature, R_0, is the sum of squares of the values of signature elements. Hence, it approximates the energy of the signal and is proportional to the area of the object. Thus, in order to normalize the sequence each autocorrelation coefficient R_i is divided by the coefficient R_0. The resultant sequence is scale invariant. It is referred to as the LC of an object. Note, that under our terminology the signature of an object is a set of distances, the LC of an object is the normalized cyclic autocorrelation of the signature.

Object signature and length-code have several advantages over other shape-representation methods with respect to object matching, reconstruction, and compression. The main advantage is that it is less sensitive to "noise". This is due to the fact that a small change in shape, which may be due to noise, causes minimal change to the distance between the centroid and contour points. In addition, length-coding is invariant to the set of affine transformations. Finally, this representation converts a two-dimensional signal into a one-dimensional signal with minimum loss of information and provides a computationally efficient framework.

4.2 Compression and Compressive Sensing of Object-Signature

The object-signature can be used for several applications including image alignment/registration, object/image recognition, and image data compression. Baggs *et al* report on the usage of the LC variant of object-signature for image alignment/registration [14], Keogh reports on the use of signature for object recognition [15]. Another application of object-signature is in image data compression. In this method a compact representation of an image is obtained via representing image-objects through their center, scaling factor, average gray level, and signature. The reconstruction stage reconstructs each object from its signature and places it with the right scale in the right location; thereby recovers the image from the object-signature. An interesting question is the resilience of these methods for object identification and recognition, image recognition and image compression, to compression applied to the object-signature. That is, rather than storing the signature, a compressed version of the signature is applied. We are currently performing experiments related to this question.

The experiments include recognition and reconstruction. In the recognition stage, a library of objects is constructed and each object is represented using its LC. Next, the LC is compressed. Hence, the library contains compressed LCs. The object recognition is implemented by comparing the LC of an inspected object to the LC of every object in the library. Under this process, in the first phase, the compressed LC of library objects is decompressed and the LC of the inspected object is obtained. Next, the LC of the inspected object is compared to each of the library LCs and the closest LC, provided it is not above a threshold, is used for object identification. In the image compression application, image-objects (which might be stored in a central library) are represented by compressed signatures. Hence, the image decompression involves decompression of signatures and reconstruction of the respective objects.

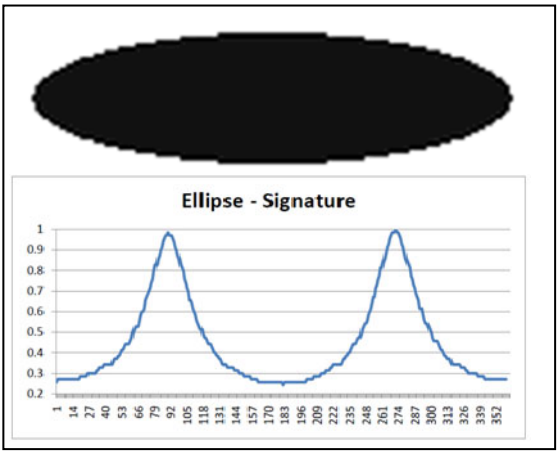


Fig. 2. An Ellipse and its Signature

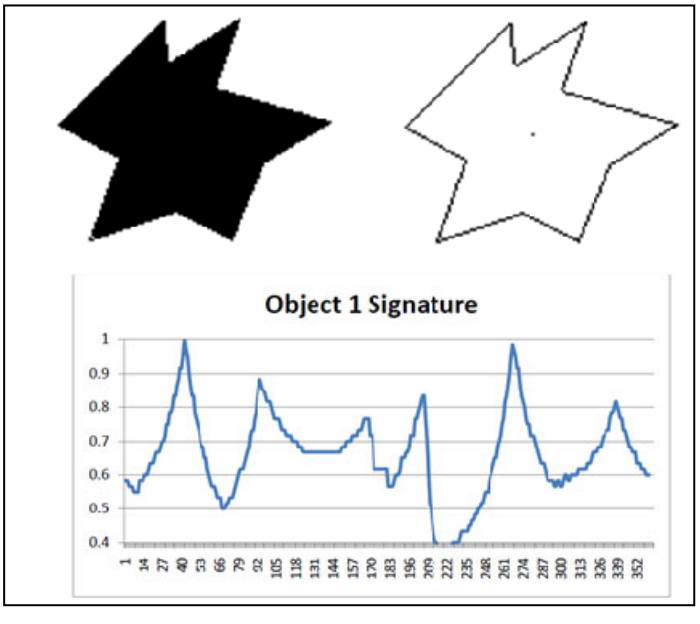


Fig. 3. A star convex Object, Contour, and Signature

Figure 2 shows an ellipse and its signature. Figure 3 shows a star convex object (referred to as Object 1), its contour and its signature. Note that in this case the signature is obtained form a point that can "view" the entire shape. Nevertheless, this point is not the centroid of the object. Figure 4 shows the flowchart for image compression and object recognition processes. In the figure, CCL stands for connected component labeling [13], and LC stands for length code [14].

The next section (section 5) reports on a set of experiments where lossless compression, DCT based compression, predictive/differential coding, and CS have been applied to a library of object LCs (for recognition) and signatures (for compression). Following the compression the LCs/signatures were decompressed and used for recognition/reconstruction. Current experiments use simple synthetic objects. Future experiments will include objects extracted from real images.

5 Experiments, Results, and Results Evaluation

The goal of the experiments is twofold. First, the experiments explore the impact of different compression methods applied to object-signature/LC on the object recognition/reconstruction accuracy. Next, the potential of CS as a method for economic extraction of compact representation of objects is evaluated and compared to other compression methods. Current experiments include:

1. Comparing the rate/distortion of several compression methods as well as CS applied to the signature/LC of simple and complex, synthetic and natural image-objects.
2. Comparing the recognition rate of simple and complex, synthetic and natural image-objects represented by LC and compressed via different compression methods as well as CS. These experiments include objects that appear in the object library and object that do not exist in the library as well as variants of these objects obtained through affine transformation.
3. Comparing the reconstruction accuracy of simple and complex, synthetic and natural image-objects represented by LC and compressed via different compression methods as well as CS.
4. Comparing the rate-distortion of image representation via compact (and compressed) object-signature/LC.

The work on these items is currently on-going. In this section, we only report on the result of experiments for comparing the rate/distortion of several compression methods as well as CS with respect to the signature of simple synthetic image-objects.

Experimental Set-up

A library of the signatures of several synthetic objects is obtained. All the objects used are star-convex objects [36], with the additional restriction that the centroid is in the star-convex kernel [36] and is not co-linear with any edge. This guarantees that an equal angle signature does not “hit” the object’s boundary more than one time for the same angle. First, each object is generated via a drawing program. Next, the image containing the object is segmented using the LBG algorithm with a code-book of 2 [7]. Following this stage, a connected component labeling (CCL) algorithm and a contour following algorithm are applied to the segmented image [13]. Finally, the set of distances from object-centroids to object-boundary are obtained using the equal-angle approach. An angle of 1 degree is used hence each object is represented with 360 signature elements. These elements are normalized to values between 0 and 1 via division by the maximal element and quantized to 16 bits.

Experiments

The experiments reported in this paper include evaluation of the rate/distortion of several methods for signature compression. The methods examined include:

1) Lossless Lempel Ziv (LZ) based compression [37,38],
2) Lossless chain-code encoding [14],
3) Differential Pulse code Modulation (DPCM) with a quantizer of 3 bits/sample [3],
4) Differential Linear Predictive Coding (DLPC) with 10 LPC coefficients and quantizer of 3 bits/sample [3],
5) DCT where only the first $k = 32$ coefficients are maintained
6) CS of the DCT transform of the signature with $k = 80$ measurements.

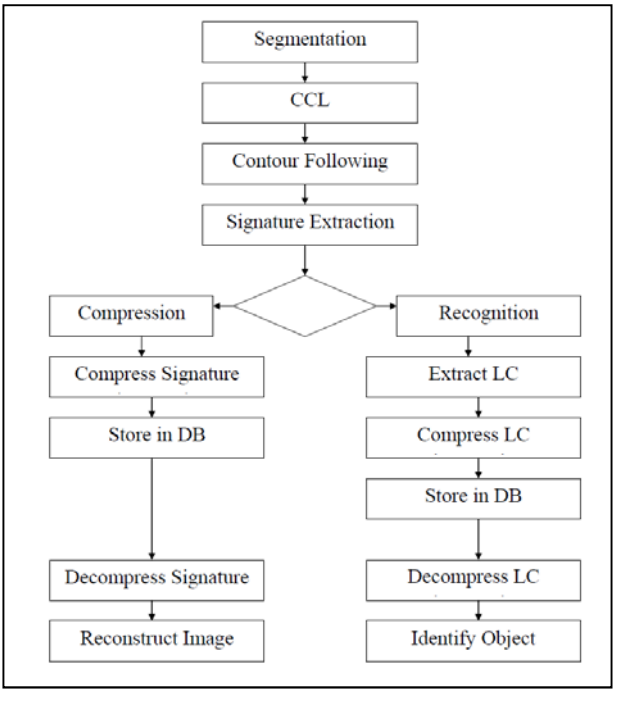


Fig. 4. Image Compression and Object Recognition Flowchart

Results

The experiments results can be represented via rate/distortion. The rate is specified in bits/sample, where a sample is a signature element. The distortion is specified in terms of signal to noise ratio (SNR) in db.

The LZ compression and the chain code representation yield a bit rate of about 0.2 with no distortion. Further lossless compression of the chain code has not been

done and is expected to improve the compression ratio without compromise in quality. The DPCM and the DLPC use 3 bits quantizer, hence, taking into account a small overhead, they yield compression rate of about 0.2 with high SNR (>50db). For the same SNR (>50db) the DCT yields a rate of less than 0.1 bits/pixel. Finally, the CS provides a bit rate of about 0.25 with an SNR of 40db-45db. Figure 5 shows the signature of the object depicted in Figure 3 along with the reconstructed signature obtained from compressive sensing of the original signature.

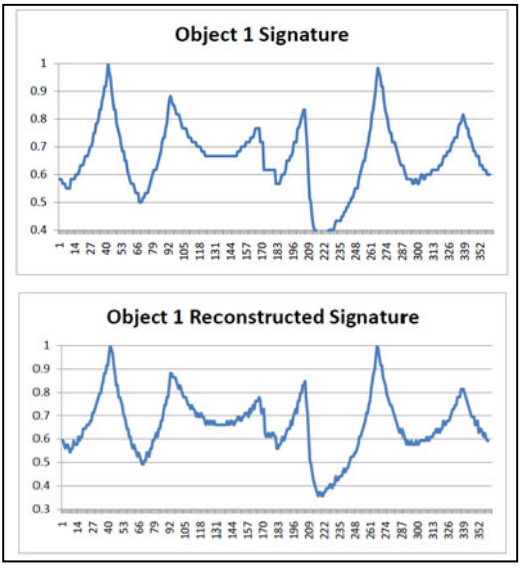


Fig. 5. Signature and reconstructed signature obtained from compressive sensing of the original signature

Result Evaluation

The lossless compression results show that there is significant redundancy in the signature elements. They provide high rate with no distortion. The DPCM and DLPC are lossy; they produce about the same bit rate as the lossless compression methods with non-zero distortion. Hence, they do not provide a benefit over lossless compression. Nevertheless, the distortion is small. Hence, it is conceivable that higher rate could yield competitive results. The DCT provides better rate than other methods with high SNR. Hence, it can be used as an alternative for the lossless compression methods. Finally, the CS provides competitive compression with about the same rate as lossless compression, DPCM, and DLPC and acceptable SNR of about 40db. The CS however, has a significant advantage over other methods in the fact that the compression is achieved via the use of a reduced amount of samples (80 instead of 360). Hence, a non-expensive compressing sensor would enable cost effective acquisition, storage, and processing of signature elements.

6 Conclusions and Proposals for Further Research

We have presented the basic concepts of compressive sampling, CS, and CI as well as the application of CS to image object-signature. The experiment results show that the CS of object-signature yields rate distortion results that are competitive yet somewhat inferior to generally used compression methods. On the other hand, the reduction in the number of required samples is a significant benefit of the CS.

Future work will include experiments with natural images as well as comparative experiments using the every-pixel approach. Additionally, there is a potential for addressing some of the challenges of the field of CS including the sampling and the computation via optical supercomputing (OSC) systems and this paper can serve as a starting point for research involving the utilization of OSC systems for CS. Finally, the results of the current research can be used to advance the application of CS in the field of machine vision and machine learning.

Acknowledgments. The last author is partially supported by the Rita Altura Trust in Computer Science.

References

1. Donoho, D.L.: Compressed Sensing. IEEE Transactions on Information Theory 52(4), 1289–1306 (2006)
2. Candès, E.J., Tao, T.: Near-Optimal Signal Recovery from Random Projections: Universal Encoding Strategies. IEEE Transactions on Information Theory 52(8), 5406–5425 (2004)
3. Sayood, K.: Introduction to Data Compression, 3rd edn. Morgan Kaufmann, NY (2006)
4. Takhar, D., Laska, J.N., Wakin, M.B., Durate, M.F., et al.: A New Compressive Imaging Camera Architecture using Optical-Domain Compression. In: Proceedings of Computational Imaging IV at SPIE Electronic Imaging, CA (2006)
5. Jain, K.A.: Fundamentals of Digital Image Processing. Prentice-Hall, NJ (1989)
6. Porat, B.: A Course in Digital Signal Processing. Wiley, NY (1997)
7. Linde, Y., Buzo, A., Gray, R.: An Algorithm for Vector Quantizer Design. IEEE Transactions on Communications 28(1), 84–95 (1980)
8. Coleman, G., Andrews, H.: Image Segmentation by Clustering. Proceedings of the IEEE, 773–785 (1979)
9. Lustig, M., Donoho, D.L., Pauly, J.M.: Sparse MRI: The Application of Compressed Sensing for Rapid MR Imaging. Magnetic Resonance in Medicine 58(6), 1182–1195 (2007)
10. Elad, M.: Optimized Projections for Compressed Sensing. IEEE Transactions. on Signal Processing 55(12), 5695–5702 (2007)
11. Cotter, F., Rao, B.D.: Sparse Channel Estimation via Matching Pursuit with Application to Equalization. IEEE Transactions on Communications 50(3) (2002)
12. Sen, P., Darabi, S.: Compressive Dual Photography. Computer Graphics Forum (2009)
13. Pavlidis, T.: Algorithms for Graphics and Image Processing. Computer Science Press, MD (1982)

14. Baggs, R.A., Tamir, D.E.: Image Registration Using Dynamic Space Warping. In: The International Conference on Artificial Intelligence and Pattern Recognition, Florida (2008)
15. Keogh, E., et al.: LB_Keogh Supports Exact Indexing of Shapes under Rotation Invariance with Arbitrary Representations and Distance Measures. In: International Conference on Very Large Data Bases (2006)
16. Candès, E., Wakin, M.: An introduction to Compressive Sampling. IEEE Signal Processing Magazine 25(2), 21–30 (2008)
17. Romberg, J.: Imaging Via Compressive Sampling. IEEE Signal Processing Magazine 25(2), 14–20 (2008)
18. Chan, W.L., Charan, K., Takhar, D., et al.: A Single-Pixel Terahertz Imaging System Based on Compressed Sensing. Applied Physics Letters 93, 121105 (2008)
19. Chan, W.L., Moravec, M.L., Baraniuk, R.G., Mittleman, D.M.: Terahertz Imaging with Compressed Sensing and Phase Retrieval. Optics Letters 33(9), 974–977 (2008)
20. Stern, A., Javidi, B.: Random Projections Imaging with Extended Space-Bandwidth Product. IEEE/OSA Journal of Display Technology 3(3), 315–320 (2007)
21. Stern, A.: Compressed Imaging System with Linear Sensors. Optics Letters 32(21), 3077–3079 (2007)
22. Rivenson, Y., Stern, A.: Compressed Imaging with Separable Sensing Operator. IEEE Signal Processing Letters 16(6), 449–452 (2009)
23. Rivenson, Y., Stern, A.: Practical Compressive Sensing of Large Images. In: International Conference on Digital Signal Processing, Greece (2009)
24. Gehm, M.E., John, R., Brady, D.E., Willett, R.M., Schulyz, T.J.: Single Shot Compressive Spectral Imaging Using a Dual Disperser Architecture. Optics Express 12(21), 14013–14027 (2007)
25. Fergus, R., Torralba, A., Freeman, W.T.: Random Lens Imaging, MIT Technical Report, MIT-CSAILTR-2006-058 (2006)
26. Mahalanobis, A.: Compressive and Computational Sensing. In: Seventh International Workshop on Information Optics, France (2008)
27. Brady, D.J., Choi, K., Marks, D.L., Horisaki, R., Lim, S.: Compressive Holography. Optics Express 17, 13040–13049 (2009)
28. Choi, K., Horisaki, R., Hahn, J., Lim, S., Marks, et al.: Compressive Holography of Diffuse Objects. Applied Optics 49, H1–H10 (2010)
29. Miller, R.A., Burr, C.U., Tai, Y., Psaltis, D.: A Magnetically Actuated MEMS Scanning Mirror. In: SPIE, vol. 2678, pp. 47–52 (1996)
30. Cattan, E., Haccart, T., Velu, G., Remiens, D., Bergaud, C., Nicu, L.: Piezoelectric Properties of PZT Films for Microcantilevers. Sensors and Actuators 74, 60–64 (1999)
31. Kahn, H., Juff, M.A., Heuer, J.H.: The TiNi Shape Memory Alloy and its Applications for MEMS. Journal of Micromechanical and Microengineering 8, 213–221 (1998)
32. Pizzi, M., Koniachkine, K., Bassino, E., Sinesi, S., Perlo, P.: Electrostatic Microshutter-Micromirror Array for Light Modulation Systems. In: Proceeding of the SPIE, vol. 3878, pp. 164--171 (1999)
33. Pizzi, M., Koniachkine, V., Nieri, M., Sinesi, S., Perlo, P.: Electrostatically Driven Film Light Modulators for Display Applications. Microsystems Technologies 10, 17–21 (2003)
34. Stockley, J., Sharp, G., Doroski, D., Johnson, K.: High-Speed Analog Achromatic Intensity Modulator. Optics Letters 19, 758 (1994)
35. Tamir, D.E., Park, C., Yoo, B.: The Validity of Pyramid K-means. In: SPIE Conference on Optics and Photonics / Optical Engineering and Applications, CA (2007)

36. Arkin, E.M., Chiang, Y.J., Held, M., Mitchell, J., Sacristan, V., Skiena, S.S., Yang, T.C.: On Minimum-Area Hulls. Algorithmica 21, 119–136 (1998)
37. Ziv, J., Lempel, A.: A Universal Algorithm for Sequential Data Compression. IEEE Transactions on Information Theory IT 23(3), 337–343 (1977)
38. Ziv, J., Lempel, A.: Compression of Individual Sequences via Variable-Rate Coding. IEEE Transactions on Information Theory IT 24(5), 530–536 (1978)

Optics Goes Where No Electronics Can Go: Zero-Energy-Dissipation Logic

H. John Caulfield

Fisk University,
1000 17th Ave., N.,
Nashville, TN 37208
hjc@fisk.edu

Abstract. Optical computing has a seemingly eternal problem. It always appears to be in competition with electronic computing. Moore's law and the advantages of digital over analog processing make pure electronics superior in almost every case. Optical computing uses come when the signal is already in the optical domain or when it is used to reduce the heat load in hybrid optical-electronic chips. I describe here work done with a number of bright opticists and logicians over the last four years that produces using optics logic that dissipates no energy and accommodates whatever bandwidth the input and output laser modulation affords. Moreover, we can show why electronics alone can never accomplish those important properties.

Keywords: Optical logic, zero energy, unlimited bandwidth.

1 Introduction

Optical computing has existed for many decades now and has been through many cycles of excitement and depression [1,2,3]. It seems now to be in its best condition ever, because there are niches it seems to fill (optical communication) and next generation chips that will contain optics and electronics. Though these matters are quite important, they will not be reviewed here. Rather, here we report on a totally new capability that optics alone can do. In this, there can never be competition. For the first time, optical computing has a field to itself. That field is sequential logic that does two things that seem at first to be beyond possibility for either electronics or optics:

- Perform logic with zero energy dissipation
- Operate at whatever speed can be modulated onto a laser beam The zero energy operation was shown to be thermodynamically allowable many years ago [4,5,6,7]. The reader should note, that the papers cited here are the ones we consider to be the foundation papers. The literature in the field is huge. A limited but useful bibliography of that field has been published online [8].

What Landauer noted [4] was that a binary logic gate takes in two bits but outputs only one. The erasure of one bit costs at least $kTln2$ in energy, where k

S. Dolev and M. Oltean (Eds.): OSC 2010, LNCS 6748, pp. 78–85, 2011.

is the Boltzmann constant and T is the ambient temperature. This is a very tiny amount of energy. But if we aim for high speed, say, 10^9 of those per second, which requires a huge amount of power. And current electronic computers use millions of times more energy than this minimum.

So far as I can discern, it was Bennett [5,6] who took the next critical step. Like most good ideas, it was very simple after he published it. If we do not erase that extra output bit, there is no price to pay for it. They tended to call that undesired bit a "garbage bit." But logic gates so constructed were reversible. Information was transformed but not lost. In a sense nothing was lost. If gates using that strange kind of gate could be made, then lossless logic might appear momentarily.

Fredkin and Toffoli each contributed computationally complete reversible logic gates [7]. This made the growing community of researchers in the field quite excited. Perhaps the end was near. Someone might invent a suitable zero energy or at least a very low energy logic gate soon. They were wrong.

Feynman [9] contributed his own gate and published it in an optics journal. That was really the start of the quest that started me down the road of optical reversible logic. With various colleagues, most notably Joseph Shamir [10], I set out to see if those gate could be made optically. The result was a kind of graceful failure. Later, I will describe precisely the mistake we made. Nevertheless, we accomplished several nice things, only one of which I mention in the Introduction, namely: Both Joe Shamir and I became very interested in this field.

In about 2003, I saw an opportunity to assemble a team of bright opticists and bright logicians to try again what we failed to do earlier. Our team (grouped by affiliation) was:

Joseph Shamir, The Technion (consultant)
Andrey Zavalin, Fisk University (opticist)
Lei Qian, Fisk University (logician)
Chandra Vikram, Fisk University (opticist)
John Caulfield, Fisk University (opticist)
Jim Hardy, Idaho State (logician)
Jonathan Westphal, Idaho State (logician)
Liz Golden, Idaho State (logician)
Steve Blair, Utah (integrated optics)

Like everyone before us, we struggled but began to see some progress a few years ago. But it was a shock to us when we realized in late 2007 that:

- We had solved the problem
- We knew why everyone else had failed.

Those are the stories I want to tell in this paper.

There were things we knew and things we thought we knew. What we actually knew is shown in Figure 1.

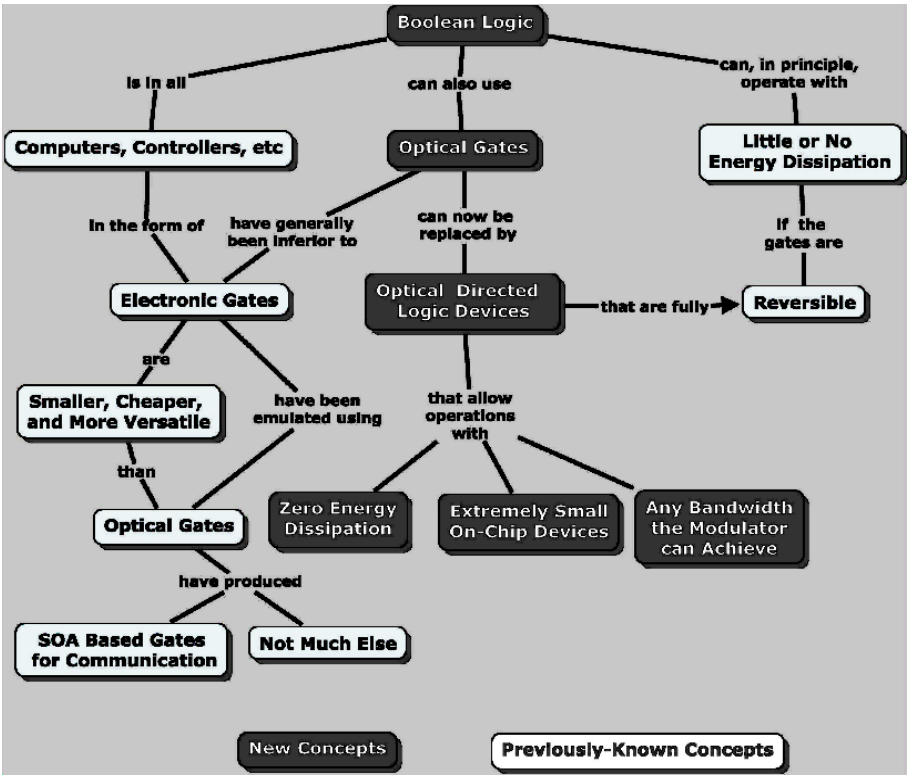


Fig. 1. Our initial insight was that the zero-energy-dissipation requires making the all of the operations linear. That is obvious, but a surprising number of people forget it, for reasons to be discussed below.

A linear operation can be made lossless in information content. A reversible system must conserve information. The arithmetic version of reversibility matter makes it easier to understand. Consider the relationship:
$2 + 3 = 5$
Initially, it seems that information has been added. We might not have known that the sum of 2 and 3 is 5. Nevertheless, that is implied by the data (2 and 3) and the instructions (add the numbers). So it really adds nothing that was not already implied. To the contrary, $2 + 3 = 5$ destroys information. Given the answer – 5 – there is no way to find what instructions and data led to that answer. Was it "add 2 and 3"? Perhaps it was add 4 and 1? Perhaps it was: divide 10 by 2. An information conservative operation is $2 + 3 = 3 + 2$. No information is lost, so it could be reversed if we wanted.

So we reasoned that, as there are no lossless gain mechanisms, by definition; each component must be both lossless and reversible. This suggested optics, because optics is inherently linear and reversible. To accomplish nonlinear optics we have to insert some appropriate material that interacts with the light. Of the past papers in optical logic only a very few have sought to do logic using the only

way it could conceivably be done: with only linear components. Those papers were those of my colleagues in this work and me [11,12]. And none of them actually solved the problem.

The originators of this line of thinking [4,5,6,7,8] had in mind only the energy needed to perform the logic operations. Neither they nor we were interested in the energy needed to enter the data not in the energy needed to read out the data.

Historically, we solved the problem twice. Out first solution was limited in that the lossless logic device we produced could be programmed electrically to perform any logical function without energy dissipation. For many purposes, this device (called a Generalized Optical Logic Element or GOLE) is not useful, as its input mechanism was electronic but its output was in the phase of light in the output relative to the phase in another beam. This gave it very limited use and limited cascadability [11,12]. It will not be discussed more extensively here, because it is a wonderful example of an intellectual error we and everyone else had made.

But also, it demonstrated that a passive switch was possible. And that turns out to be vital. Figure 1 shows why that relationship between interferometers and intensity switches. Interferometers can be adjusted so that the two outputs can be either 0 and 1 or 1 and zero, depending on the relative phase of the two input beams. That functions as a binary switch so long as the two beams are present. Electronics and other nonlinear means can latch the output states, while interferometry cannot. This passive, linear switch is the key to our inventions, see Figure 1.

Two things we thought we knew and did not. The most obvious thing about doing reversible logic is that it must be done by linear lossless gate designs such as those of Fredkin, Toffoli, Feynman, and so forth.

The second thing we and everyone else knew that was wrong concerned the relationship of speed to energy loss. Feynman [9] explained the problem with elegant simplicity. Fortunately, he was completely wrong.

The errors made were largely in the implicit assumptions that:

- The solution must involve logic gates that are themselves lossless,
- The gates must be electronic.

The solutions we found violated both of those "obvious" assumptions. We did not use gates of any previously known kind. And, we did not use electronics in computing the answer. Feynman's supposed inescapable speed problem is that the less energy per gate we use, the slower we go. That is based on the need to urge electronic signals along with an electric field. But that is totally irrelevant to optics. Light moves. If it does not move, it is not light. No urging is needed.

2 Correcting the Errors We All Made

A major step was taken by Hardy and Shamir [13]. Our earlier work [10] showed clearly why optical Fredkin gates could not produce zero energy logic. The gate

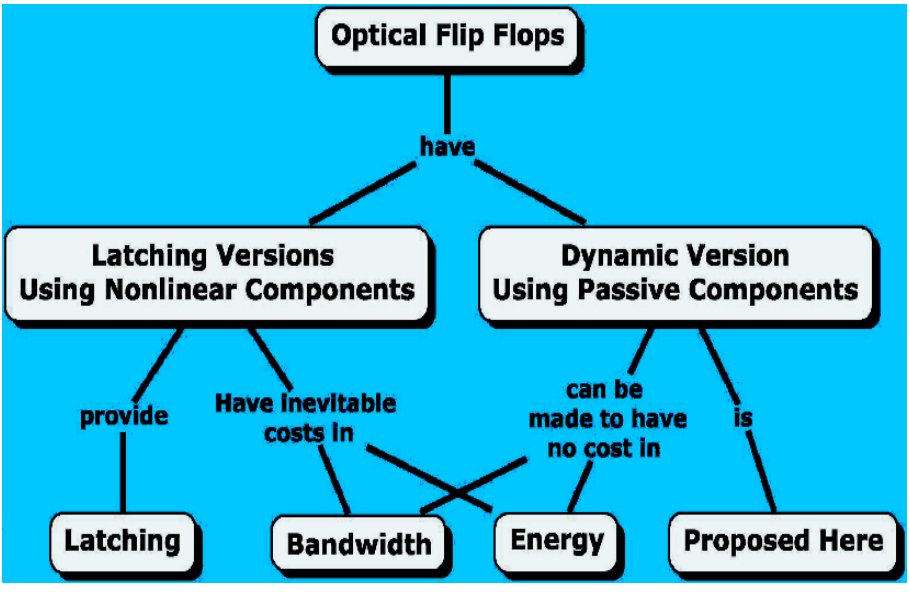


Fig. 2. Conventional logic uses intensity for logic, but that is inherently nonlinear. So it cannot lead to lossless operations. Our optical logic uses interferometry for phase (not intensity) so it can be switched linearly. This clears the path for lossless logic.

has three inputs and three outputs. But one of the inputs and one of the outputs was electronic. The other two inputs and two outputs were optical. But to synthesize even simple logical operations with a Fredkin gate, it must be possible for any output to connect to any input to the next device. With mixed optics and electronics, this cannot work. Hardy and Shamir [13] devised gates that did not require such mixing and thus can be made all optical in terms of the input signals. That matched what optics is good at with the need to evaluate arbitrary sequences of logic operations. What optics does is move (Maxwell's equations require that light travels at the fixed speed c, and Einstein showed that c was the universe's speed limit.). It is easy to direct where the light travels in waveguides, so the need is to figure out how to use the direction of light travel to compute the outcome of a sequence of logical operations. This is what they did in what they called Directed Logic. This is what they did. The results of some simple operations are shown in Figure 2. It gets the right answer, but not in the conventional way.

The earlier assumptions on speed would predict extremely slow operations of any solution to the zero-energy problem. In fact, these gates will work at any speed at which a laser can be modulated. The fastest modulation rates conceivable would be several wavelengths long. The only speed-limiting effect in our devices is clock skew. We need to make all optical path lengths equal; to within a fraction of several wavelengths. This is straightforward to do in optics. The waveguides cab be made that good, and electrooptic correction is quite feasible.

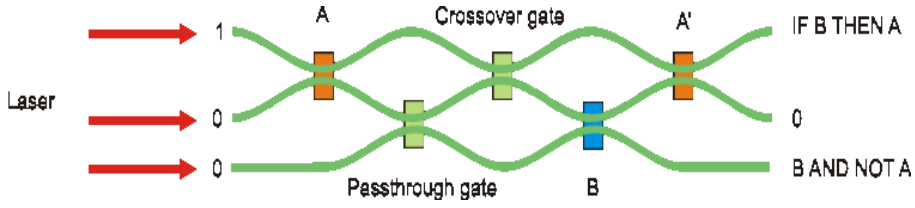


Fig. 3. The flow of light through waveguides and passive interferometric switches can implement any sequence of logic operations. We show one simple but representative operator.

The last step we have taken is to go back to the regular connection fabric originally suggested by Shamir et al. [10]. It looks nothing like any previous logic gate – electronic or optical. But it works.

3 Why We (and Others) Failed for So Many Decades

Quite often scientists are correct about saying what cannot be done, if their implicit rules apply. We explicitly violated the implicit assumptions of prior workers. Our earlier attempt to solve the problem explicitly violated the electronics assumption. But we retained the assumption that we must use some sort of constructively complete gates that are themselves lossless.

4 Remaining Problems

Two problems remain:

- First, we would like to see how small such systems can be made. At the moment, perhaps tens of microns is required. But, slow light can allow the interaction lengths to be made shorter. If we sacrifice a little on energy, we may be able to use plasmonics to make the devices smaller. If we sacrifice a little in bandwidth, we can use resonant rings that can now be made very small.
- Second, we can accomplish the zero-energy feat by avoiding measurements until the end. The system is analog and subject, therefore, to error accumulation. Even here, however, there may be a hope of significant BER (Bit Error Rate). There are two complementary outputs. Measuring both places the decision in the domain of hypothesis testing. This is terra incognito. As we were preparing this paper, we ran across a paper by Peres [14] who uses this same error reduction approach in the kind of optical processor he found so fascinating: those that harness entanglement. This is very encouraging to us, not only because of Peres's great record in his kind of quantum computing but also because his work can be a template for ours.

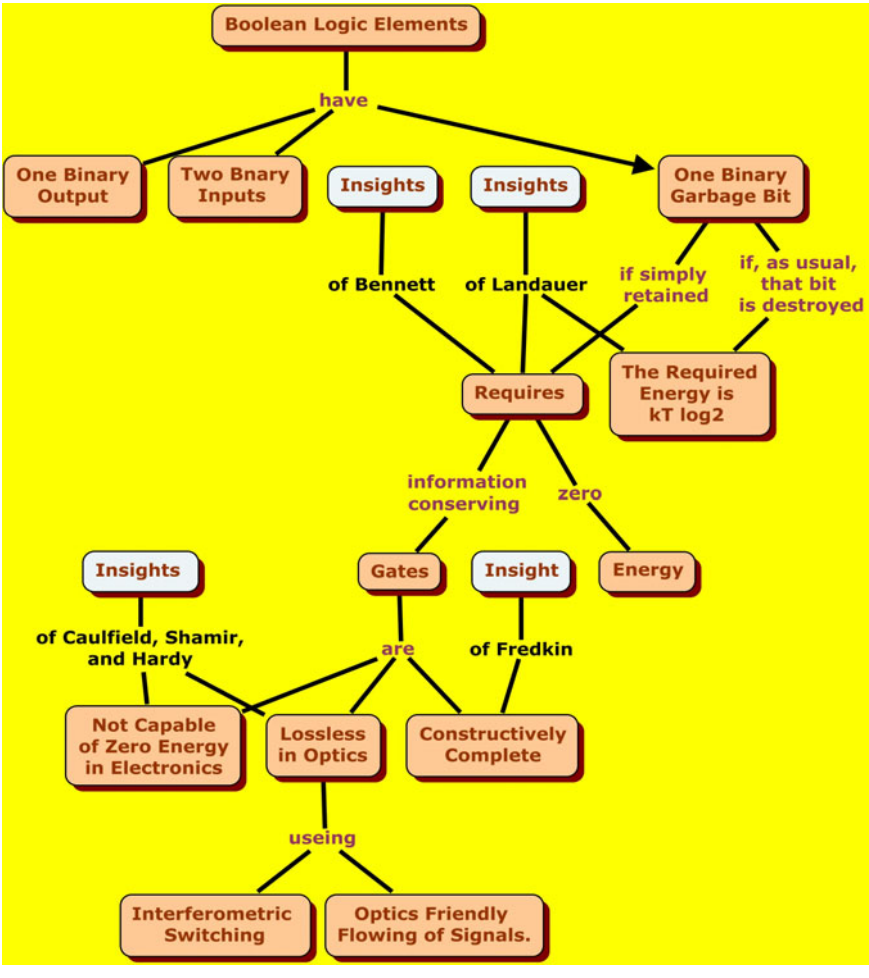


Fig. 4. This is a lossy compression of the history of the project discussed here. Everyone involved made significant contributions but only the lead players on any part of the path toward lossless logic are shown for compactness

5 Summary Diagram

Everything of technical importance in this paper is summarized in Figure 4.

6 Conclusions

The number of implicit assumptions and their persuasive obviousness seem to us what has caused so many brilliant scientists to fail in attempts to make zero-energy-dissipation logic and then be able to use it for high speed operations.

We too fell victim to these obvious things for 20 years. We think other "impossible" tasks should be reexamined from time to time to find if they contain implicit assumptions that can be avoided. This has worked for us in several fields and almost certainly others do this as well. The task we set out to solve was impossible, given the tacit, unnoticed assumptions we and others made. Removing inessential assumptions allowed a team of experts in different fields to solve the now-possible problem.

Acknowledgment. This work was sponsored by the United States Air Force Office of Scientific Research.

References

1. Caulfield, H.J., Vikram, C., Zavalin, A.: Optical Logic Redux, Optik. International Journal for Light and Electron Optics 117, 199–209 (2006)
2. Shamir, J.: Optics in computing, - 40 year later, Critical Technologies for the Future of Computing. In: Proc. SPIE, San Diego, July 30 - August 4, vol. 4109, pp. 4109–4115 (2000)
3. Shamir, J.: Optical computing remains in shadow, EETimes (1122) (July 17, 2000), http://www.techweb.com/se/directlink.cgi?EET20000717S0072
4. Landauer, R.: Irreversibility and heat generation in the computing process. IBM Journal of Research and Development 5, 183–191 (1961)
5. Bennett, C.H.: Logical reversibility of computation. IBM Journal of Research and Development 17, 525–532 (1973)
6. Bennett, C.H.: The Thermodynamics of Computation – A Review. International Journal of Theoretical Physics 21, 905–940 (1982)
7. Fredkin, E., Toffoli, T.: Conservative logic. International Journal of Theoretical Physics 21(3-4), 219–253 (1982)
8. Perkowski, M.: Bibliography of reversible and quantum logic and computing, http://web.cecs.pdx.edu/~mperkows/PQLG/biblio.html
9. Feynman, R.: Quantum mechanical computers. Optics News 11, 11–20 (1985)
10. Shamir, J., Caulfield, H.J., Miceli, W., Seymour, R.J.: Optical Computing and the Fredkin Gates. Appl. Opt. 25, 1604–1607 (1986)
11. Caulfield, H.J., Soref, R.A., Qian, L., Zavalin, A., Hardy, J.: Generalized Optical Logic Elements – GOLEs. Optics Communications 271, 365–376 (2007)
12. Caulfield, H.J., Soref, R.A., Vikram, C.S.: Universal reconfigurable optical logic with silicon-on-insulator resonant structures. Photonics and Nanostructures 5, 14–20 (2007)
13. Hardy, J., Shamir, J.: Optics inspired logic architecture. Optics Express 1(5), 150–165 (2007)
14. Peres, A.: Reversible logic and quantum computers. Physical Review A 32, 3266–3276 (1985)

Nanotechnology Based Optical Solution for NP-Hard Problems*

(Extended Abstract)

Eyal Cohen[1], Shlomi Dolev[1], Sergey Frenkel[2], Rami Puzis[3] and Michael Rosenblit[4]

[1] Department of Computer Science
Ben Gurion University of the Negev, Israel
[2] Institute of Informatics Problems
Russian Academy of Science, Russia
[3] Deutsche Telekom Laboratories
Ben Gurion University of the Negev, Israel
[4] Ilze Katz Institute for Nanoscale Science & Technology
Ben Gurion University of the Negev, Israel

Abstract. We present a design for a micro optical architecture for solving instances of NP-hard problems, using nano-technology. The architecture is using pre-processed masks to block some of the light propagating through them. We demonstrate how such a device could be used to solve instances of Hamiltonian-cycle and the Permanent problems.

1 Introduction

An optical computer is a device that uses light rather than electric current to perform computations. Optical computers have several advantages over their electronic counterparts. For example, light beams, unlike electric currents, can intersect without interference among the beams, even when they are confined to two dimensions.

Using fiber optics and laser on silicon within a processing unit instead of electronic connections can improve a computer's speed due to the elimination of resistance, crosstalks, soft errors and the like. When using light beams within a processing unit, there is a need to avoid optical to digital and digital to optical conversions. A straightforward approach for creating an optical processing unit is to implement optical logical gates (e.g. AND, OR, NOT) and design architectures that directly map the current VLSI designs to general-purpose all-optical processors [8,17].

The overhead involved in creating a general purpose processor would be too high in comparison to one designed to perform specific primitives. For example, special purpose, optical devices designed to solve a limited set of hard combinatorial problems

* Partially supported by Deutsche Telekom, the ICT Programme of the European Union under contract number FP7-215270 (FRONTS), Rita Altura Trust Chair in Computer Sciences, and the Lynne and William Frankel Center for Computer Sciences. Emails: {dolev,eyalco}@cs.bgu.ac.il, rmichael@bgu.ac.il, fsergei@mail.ru, puzis@bgu.ac.il. An extended version appears as TR of the Dept. of Computer Science, BGU.

S. Dolev and M. Oltean (Eds.): OSC 2010, LNCS 6748, pp. 86–99, 2011.

(e.g., NP- and #P complete or EXPTIME complete) can be utilized by the industry right away and also can later help in building general purpose optical computers.

Efficient solutions to hard combinatorial problems have been the goal of many researchers. Conventional methods for solving hard combinatorial problems include exhaustive searches, approximation algorithms, or heuristics. Currently, only exhaustive search algorithms guarantee finding the optimal solution [10]. However, for large problems, these methods result in exponential execution time.

Electronic computers are not structured like mechanical computers, such as the Babbage machine [13]. Similarly, optical computers architecture may be different from their electronic counterparts. Light has many features that can be exploited to ease on the computation. For example, some success in using many beams in free space for computing has been recently reported [15]. The design of [15] is based on parallel optical multiplication. The use of similar multiplication devices to solve NP-complete problems was suggested in [20,22], where the authors propose a device which is designed to solve bounded NP-complete problems.

Reif et al. [21] prove that the problem of ray tracing, in three-dimensional optical systems consisting of a finite set of mirrors with endpoints that are rational coordinates, is PSPACE-hard. In [7] we suggest using a mapping similar to the non-deterministic Turing machine by (amplifying and) splitting beams. The mapping can be viewed as a theoretical proof of the existence of a solution rather than an efficient solution. Knowing, that a solution exists for a (bounded) NP-complete problem instance, we seek the most efficient solution in terms of: the number of beams used; the number of optical elements (or location in space used to represent a computation state); the energy needed; the maximum number of beams that should be split from a single source beam (fan-out); and the number of locations a beam needs to visit (and possibly split) from its creation until its final detection on arrival.

Note that we are not solving NP-complete problems with polynomial (in the input length) time and space, but suggesting that an optical approach may be promising in solving larger instances of hard combinatorial problems. In fact, we use exponential space to solve exhaustively an instance in linear time. The design suggested in [7] was extended in [3] for all (six) basic NP-complete problems listed in [9]: Hamiltonian circuit, permanent, clique, vertex cover, partition [14], 3-SAT [2], and 3D-matching [12]. Note that polynomial reduction between NP-complete problems is not used there, since we are concerned with the inflation of instance size while the target is solving the largest possible instance of each of these problems.

Development of a prototype of miniature optical computer architecture capable of solving large instances of NP-complete combinatorial problems requires a multidiciplinary effort. Namely, algorithm theory, information theory, physical optics, and nanotechnology, should be the basis for achieving significant success. In this paper we present the first steps in this multi-disciplinary research. We examine a nano-scale architecture suitable for solving the Hamiltonian-cycle problem and for computing a binary permanent. We add an analysis of the fan-in, fan-out in an effort to find an efficient design. An efficient design should have a low depth which is defined by the maximal number of locations a beam traverses from the source to the detectors.

2 Optical Solutions for Hamiltonian-cycle and Permanent Problems

One of the most interesting problems solved using optical computing is the Hamiltonian-cycle problem. Different designs for an optical architecture, that solves the NP-complete Hamiltonian-cycle problem, have been suggested. The architecture we have suggested in [7] and the architectures suggested in [18,11] are based on splitting and delaying the light beams. The Hamiltonian-cycle problem can also be solved following either one of the two approaches of the traveling beams architecture we have suggested in [3]. The first approach in [3] can also be used to solve the permanent problem. The first approach involves mapping the graph nodes to physical locations in space and the propagation of beams according to the edges of the input graph instance. The second approach propagates beams along a computation tree such that the leaves represent all possible solutions and the delay in propagation from the root to each leaf corresponds to the "value" of the specific combination. We also present in [3] a very different architecture called the coordinated holes in a mask-made black box (see Figure 1). In this architecture a set of masks with "holes" is chosen from $n(n-1)$ pre-computed masks, according to the input instance. A solution to the problem instance exists only if the combined masks do not block all beams.

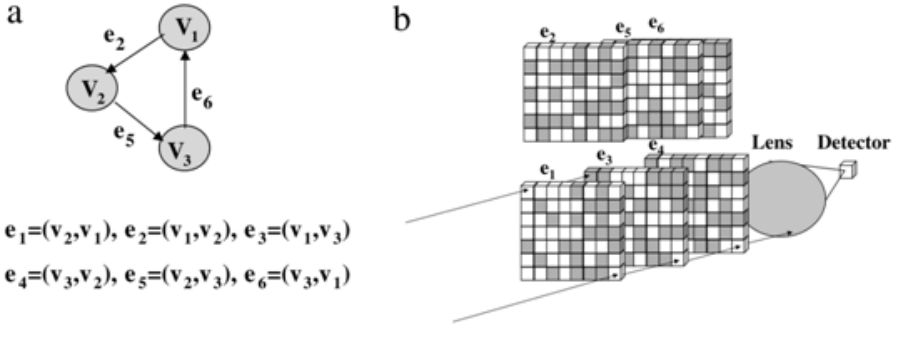


Fig. 1. Illustration of the holes in masks-made, black-box architecture [3]

Optical copying has been identified as an important computation powerful technique in e.g., [23]. In [6] we implement a specific set of primitives that allows us to have an efficient pre-processing stage based on mask creation by copying matrices. This allows defining a problem instance by choosing only a small number of predefined masks to obtain the result. In addition, in [6,1] we present a technique for solving problems that are hard-in-average, namely, the shortest vector problem and integer version of the permanent (see Figure 2).

2.1 Optical Considerations

In order to provide more effective and miniaturized implementation of optical algorithms mentioned above, we will study the possibility of using new techniques for light

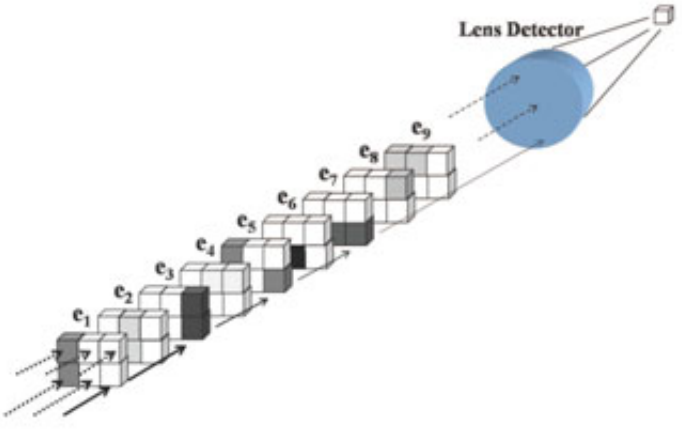


Fig. 2. Architecture for solving the (3×3 matrix) permanent [1]

beam masking. Recent nanotechnology developments have resulted in some interesting components based on meta-materials, non-liner media and complex miniature waveguides.

As the optical computing community seeks to develop all-optical integrated circuits, researchers have created a new process for making complex miniature waveguides that can steer optical signals in three dimensions through solid materials.

We focus on and present miniature optical mask method. Masks are used, arranged and designed such that the output signal holds a certain pattern which constitutes a computed data. We present simulations done on suggested optical mask setup.

Some other methods that may be used and are not discussed here are:

- Micro resonator array with complex waveguide structure. Resonators are used to enhance and amplify certain signals while eliminating other signals, usually discriminating signals by their frequency. A waveguide structure may be used to redirect signals to different location in space. A complex combination of the two is possible and allows us to get a desired output by using the right design.
- Micro arrays with active elements. When light encounters elements organized in certain patterns e.g. lattice, the light is influenced by the geometry of the lattice together with the effect of a single element. This may result in effects such as phase shifting, polarizing, switching, frequency shifting, pass compensating and loss compensating.
- Other non-linear elements.

3 Preprocessing Techniques

Our primary objective is to design optical computer architecture for solving arbitrary bounded size instances of various hard problems (e.g. NP- and #P-complete and EXPTIME). The new developed architecture will be based on programmable masks. The architecture will contain several mask layers, each of which can be turned on and off as part of the input definition. Each layer will be represented by a matrix with opaque,

transparent, or partially transparent cells. The output is determined by analyzing beams that have traversed all layers. While investigating the possible layouts, we will also seek to increase the variety of problems, that can be solved using this architecture, as well as the maximal size of the problem instances. We will update the optical computing-based algorithms suggested previously in order to optimize their effectiveness in such architecture.

3.1 Hamiltonian-Cycle

The Hamiltonian-cycle problem tries to find a closed path that goes through all vertices of a given graph, while visiting each vertex exactly once. We solve the directional hamiltonian cycle problem where edges of the graph may be directional.

The Hamiltonian-cycle problem is NP-complete problem thus there is no known polinomial time algorithm to solve it. We use exhaustive search, which imply that the number of possible cycles grows factorialy with the number of vertices of the graph.

Binary Matrix Representation. A full graph with n vertices could be represented by a binary matrix. The matrix has $n(n-1)$ and $(n-1)!$ rows, where each column represents an edge in the full graph, and each row represents a possible path. Every matrix element identified by a possible path and an edge. Examples of $n = 3$ and $n = 4$ matrices are M_3 and M_4 respectively,

$$M_3 = \begin{pmatrix} 0\,1\,1\,0\,0\,1 \\ 1\,0\,0\,1\,1\,0 \end{pmatrix}$$

$$M_4 = \begin{pmatrix} 1\,0\,0\,0\,1\,0\,0\,0\,1\,1\,0\,0 \\ 1\,0\,0\,0\,0\,1\,1\,0\,0\,0\,0\,1 \\ 0\,1\,0\,0\,0\,1\,0\,1\,0\,1\,0\,0 \\ 0\,1\,0\,1\,0\,0\,0\,0\,1\,0\,1\,0 \\ 0\,0\,1\,1\,0\,0\,0\,1\,0\,0\,0\,1 \\ 0\,0\,1\,0\,1\,0\,1\,0\,0\,0\,1\,0 \end{pmatrix}.$$

The system should receive a vector as an input which represents the existence of edges in the graph. If an edge exist on the input graph, the corresponding vector element has the value 0 and if it does not exist its value is 1. The edges are organized in vectors in the appropriate matrices. The organization of vectors that correspond to matrices M_3 and M_4 are represented by vectors V_3 and V_4 respectively.

$$V_3 = \begin{pmatrix} e_{12} \\ e_{13} \\ e_{21} \\ e_{23} \\ e_{31} \\ e_{32} \end{pmatrix}$$

$$V_4 = \begin{pmatrix} e_{12} \\ e_{13} \\ e_{14} \\ e_{21} \\ e_{23} \\ e_{24} \\ e_{31} \\ e_{32} \\ e_{34} \\ e_{41} \\ e_{42} \\ e_{43} \end{pmatrix}.$$

We can multiply the vector by the matrix. If the result vector contains a 0 (a zero) the Hamiltonian cycle is possible.

Creating binary masks. We now describe how to create the binary matrix. The algorithm presented in [3] used is an extending algorithm. That means that the algorithm uses M_n to create M_{n+1}. It extends $(n-1)! \times [n(n-1)]$ matrix to a $n! \times [n(n+1)]$ matrix.

The algorithm uses an optical copying method to produce the masks. We use optical lithography to copy entire sections of the original matrix to the new matrix. Each iteration the algorithm copies an entire column of the original matrix. Thus $(n-1)!$ matrix elements are copied at one. When this is done n times, we get a column with $n!$ matrix elements. This eliminates the need of factorial in the time complexity, even though the output is a factorial sized matrix.

3.2 The Permanent Problem

The Permanent problem is defined by

$$Perm(A) = \sum_{\sigma} \prod_{i=1}^{n} A_{i\sigma(i)},$$

where A is an $n \times n$ matrix, and the summation is over all permutations σ on n elements. We can try to solve the binary matrix permanent problem. A 2×2 binary matrix could be rearranged to be represented by a vector by reorganizing the elements and taking their logical inverse,

$$A_2 = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$$

$$V_2 = \begin{pmatrix} \neg a_{11} \\ \neg a_{12} \\ \neg a_{21} \\ \neg a_{22} \end{pmatrix}.$$

Multiplying the vector by an appropriate matrix could give us the multiplications we need to sum upon by using De Morgan's laws,

$$M_2 = \begin{pmatrix} 1\,0\,0\,1 \\ 0\,1\,1\,0 \end{pmatrix}$$

$$M_2 \cdot V_2 = \begin{pmatrix} \neg a_{11} + \neg a_{22} \\ \neg a_{12} + \neg a_{21} \end{pmatrix} = \begin{pmatrix} \neg(a_{11} \cdot a_{22}) \\ \neg(a_{12} \cdot a_{21}) \end{pmatrix}.$$

We can see that this problem became very similar to the Hamiltonian-cycle problem as it is presented by a vector multiplication by a matrix. This means that in order to calculate these multiplications, we can use the same architecture as the Hamiltonian-cycle. The sum of the multiplications could be done by collecting all beams by lens for example, and using a detector that measures the intensity. This way we can determine the sum of the beams that managed to pass through the masks.

Next we show how our previous proposed optical setups [20,3] will be miniaturized and optimized for maximum possible problem size.

4 Nano Based Optical Computer Design

Design and implementation of nano-scale optical computing devices will serve as a proof of concept and may lead to a dramatic increase in computing capabilities.

To implement the architecture, we implement the masks using lithography methods[16,26].The size of the mask will be in the range of $20 \times 20cm^2$ to $10 \times 10cm^2$ to solve an instace of the Hamiltonian-cycle when the number of nodes in the graph is $n = 15$. The masks will be produced with high precision and designed for wavelengths that will allow us to increase a density of the optical information pixels. The material that will be chosen will be transparent in the short wavelength range.

The resulting mask will be a monolithic design, meaning that light will propagate through a material with a high refracrion index. Light will dissipate elsewhere where the refraction index is low. This is similar to light propagating through fiber optics.

We will build a layer selection mechanism, based on MOEMS that will make it possible to activate any subset of the layers. We will enclose the results in a miniature package. Fulfillment of the above objectives and the use of several packages will further advance the development of a new generation of optical computers which are considerably smaller in size than the current generation and which offer a more convenient input definition and choice of computation tasks.

Note that we are not solving NP-complete problems with polynomial (in the input length) time and space, but we suggest that an optical approach may be promising in solving larger instances of hard combinatorial problems. We start with a general design that serves as an existence proof and as a base for finding a more efficient designs for specific tasks.

4.1 Optical Computation Architectures

We consider and test the methods of creating the masks. The first method considers a straight forward approach where masks block light exactly as the algorithm suggests.

An optical algorithms makes use of (exponentially large) masks to solve a number of NP-hard problems. High-resolution optical films are employed in the input and output planes of the optical system in order to synthesize a large binary matrix [3]. The influence of background noise, low light intensity, as well unwanted duplications was examined. We note that mask creation by an iterative optical synthesis process is an essential part of the optical computer design. The size of the mask, the manufacturing precision, and the possibility to increase the density of the optical information pixels are very important parameters for practical use. To solve these problems we suggest exploiting modern nanotechnologies implemented to create metalized photo-patterns. This includes the following principal directions:

(a) To develop a method of topologically transforming a computer generated masks to standard vector graphic format compatible with Laser writer and e-line. The working field will be divided into several sub-fields.

(b) To create a set of patterns using topological files of sub-fields and alignment marks to combine these sub-field patterns.

(c) Assembling of patterns into a working optical mask will be performed by projective lithography using AER photo multiplication with a decreasing factor of 10.

After the fabrication process we will get a set of nano-scale structures which will represent optical algorithems through masks implimenting in set of samples each of which maps the graph nodes to the phisical locations. In the final stage we will add a layer selection mechanism and enclose the masks in a miniature package.

Another possible architecture is to use patterns with wavelength and sub-wavelength grating structures, using different materials including meta-materials.

4.2 Selection Mechanism

There are several possibilities to select mask setup corresponds to problem under investigation. A selection mechanism can be based on MOEMS (MEMS), that will make it possible to activate any subset of the masks.

A problem that may rise when using MOEMS is that we can not eject an entire mask out of the setup. We suggest enlarging each mask by adding transparent rows between each row of the original masks. We also need to allow light to propagate in every second row. Thus by moving the mask in a distance of just one row we control the mask as if it was transparent or not (See Figure 3). Using MOEMS in moving a distance of $\sim 1\mu m$ should be much simpler than a mechanism that completly removes masks out of the system.

Another possibility for the selection meachanism is to use materials and structures that allow tunability in light transparence. We are considering both methods and will enclose the setup in a miniature package.

5 Nano Based Masks Performance Simulation

When light propagates through the masks setup, eventually some of the light may pass through all the masks in some of the pixels. We call each of the paths a channel. We simulate this system of masks simulating monolithic nano-structures. Propagating light

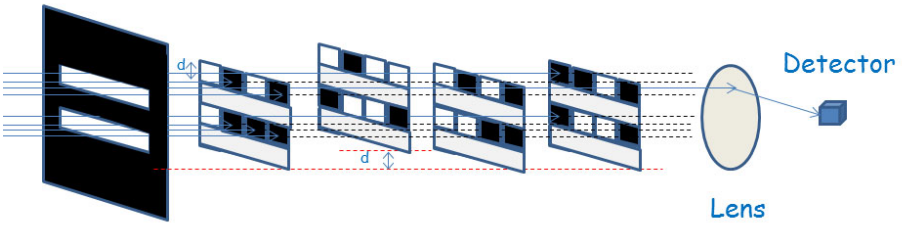


Fig. 3. Proposed mask setup. Each second row may transmit light. Each mask may or may not participate in blocking the light by moving it a distance d which is the height of a row of the masks. In this example the second mask (from the left) is not participating and it is moved a distance d. The masks are seperated for the sake of illustration. In our implementation plan masks will be closer together.

in nano scale monolithic structure behaves like a light inside a waveguide. The entire setup is attaced to a light source, ensuring that every channel will get the same light level. There is a need for a careful design of the masks in order to avoid unwanted mode coupling, multimode-interferences, PDL, etc. That may result in errors in calculations when light from one channel arrives to another. We refer to this problem as channel crosstalk.

We simulated the architecture using the BeamProp (RSoft) and OptiFDTD (Optiwave) software.

We examined several structures, materials and wavelengths for our design to allow optimization of optical mask size. Design, simulation and tolerance analysis were done for monolithic structures based on silicon, taking into account fabrication ability, and possible photonics complication (mode coupling, input mode mismatch, polarization dependence, etc.). On the next stage we will implement fabrication nanotechnologies to create a set of monolithic patterns based on silicon wafers.

Simulations were done for silicon - silicon nitride structures for $n = 3, 4 \ldots$ masks. Simulations for silicon - silica should be the next step.

5.1 Simulation Model Setup

A preliminary simulation was done to choose an optimal set of variables such as the witdh of the pixel, the difference in the width of two neighbouring pixel, the distance between the channels, and the wavelength. The simulation was done for a 3×3 setup, where only the central pixel should allow light to propagate. We can change the width of the neighboring pixel and compare the amount of light emitted through the central pixel to the light emitted through other pixels. We would like to maximize the ratio of the emission of the central pixel over the other pixels. The values of the actual simulation were selected as the optimal values by different aspects of the photonics complications.

The setup we chose is $n = 4$ Hamiltonian cycle problem (Figure 4). In all the masks four pixels should allow light to propagate. The desired result depends on the mask setup configuration and show how many pixels can emit light.

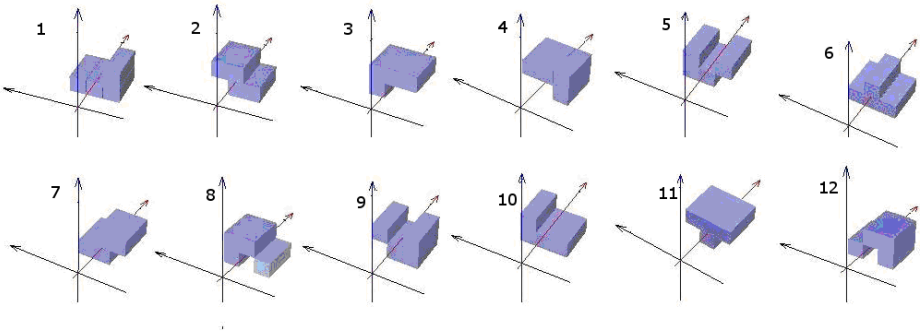
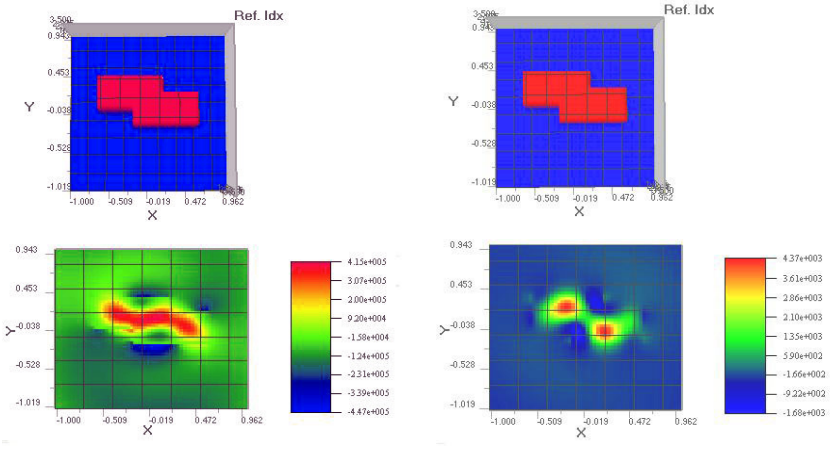


Fig. 4. Example masks for a Hamiltonian cycle with $n = 4$

All simulations of this example were done for silicon-air masks with a length of $dz = 100\mu m$, and a $z_0 = 10\mu m$ waveguide that directs the light into the setup. Thus using three masks the output of the setup would be in $z = 310\mu m$. The free space wavelength used was $\lambda = 1.3\mu m$. The refraction index $n = 3.5$ used for inside the material is suitable for silicon. The distance between the center of neighboring channel is $0.6\mu m$.

Figure 5 shows a single mask #2 (Figure 4) with different polarization of the input light. The two output patterns differ and would suggest that in this specific mask pattern, a Transverse Electric (TE) polarization is prefrable as its transmission is more similar to the shape of the mask.



(a) TM Polarizetion (b) TE Polarizetion

Fig. 5. Transmitance results of one mask (drawn above) for (a) TM and (b) TE polarizations

Next we have to check if we avoid crosstalk, that is, making sure that light that propagates in one channel continues to do so, and do not stimulate propagation in another channel. Several mask setups were simulated to check if it satisfies those demands.

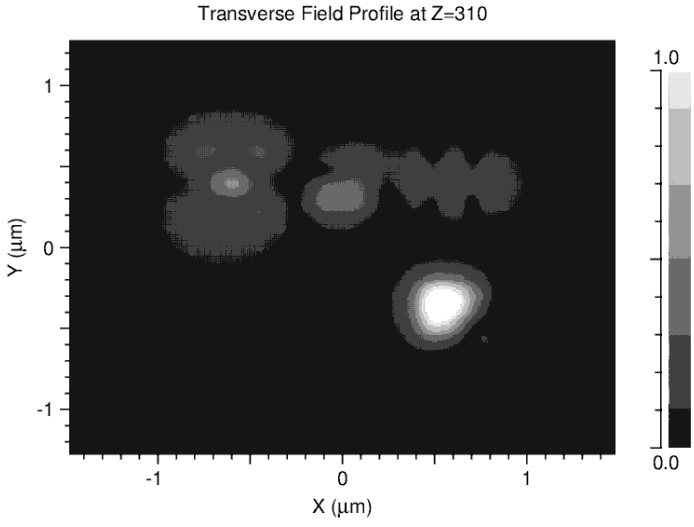


Fig. 6. Output of a 3 masks setup. Masks $1, 2, 4$ from Figure 4

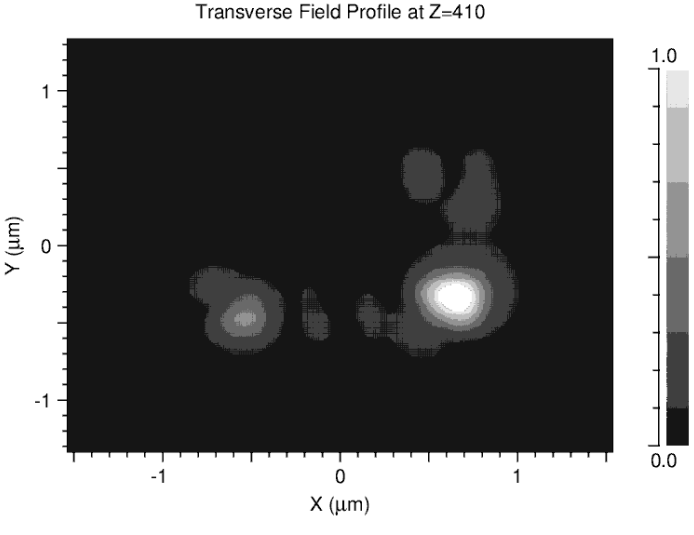


Fig. 7. Output of a 4 masks setup. Masks $1, 2, 4, 10$ from Figure 4

We would like to check what happens in the system when only one pixel should allow propagation.

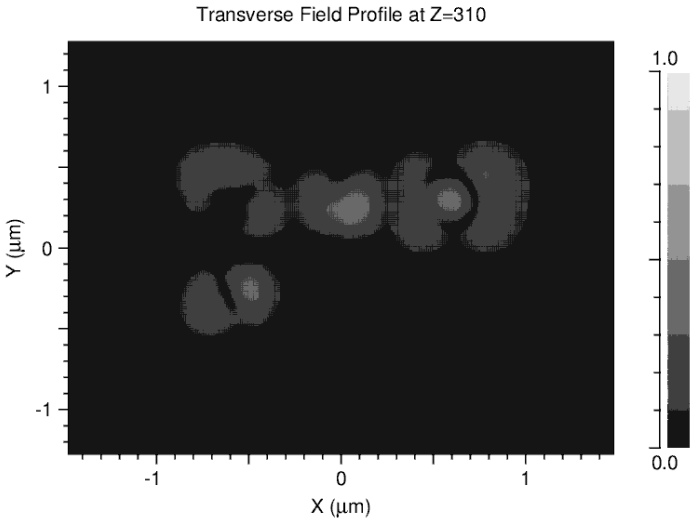


Fig. 8. Output of a 3 masks setup. Masks $1, 2, 3$ from Figure 4

Figure 6 presents the output of a setup of three masks. The masks were $1, 2, 4$ (Figure 4). Using the setup in Figure 6 allows light to propagate only through the lower right pixel. The light on the other pixels should dissipate or scatter.

In Figure 7, another mask was added such that the masks used are $1, 2, 4, 10$ (Figure 4). This setup is checking if adding more masks keeps the light propagating through the lower right pixel, and that it does not stimulate propagation on other pixels.

Another important setup simulation was done using masks that should block all the light. This time we used masks $1, 2, 3$ and the output is displayed in Figure 8. The same results should be achieved using all masks since this setup blocks all the pixels and more masks will only lower the transmittance.

6 Concluding Remarks

The existence of recent industrial attempts to produce optical processing devices (e.g., [15]) as well as the limited implementations in our laboratory (e.g., [20]) encourage us to believe that our new designs will be used in practice for solving combinatorial tasks, at least when there are real-time constraints.

We will produce a design that solves the Hamiltonian-cycle problem. The device will use the holes in masks-made architecture produced and pre-processed by projective lithography mask-copying. The selection mechanism will use MOEMS to move and align the masks.

Simulations of a monolithic design were done to prove that a practical design is feasible.

References

1. Anter, A., Dolev, S.: Optical solution for hard on average #P-complete instances. Natural Computing (2010)
2. Cook, S.A.: The complexity of theorem-proving procedures. In: Proc. of the 3rd Ann. ACM Symp. On Theory of Computing, pp. 151–158 (1971)
3. Dolev, S., Fitoussi, H.: Masking traveling beams: Optical solutions for NP-complete problems, trading space for time. Theoretical Comp. Science 411, 837 (2010)
4. Dolev, S., Fitoussi, H.: Primitive Operations for Graph-Optical Processor. In: 6th Haifa Workshop on Interdisciplinary Applications of Graph Theory, Combinatorics, and Algorithms (May 2006)
5. Dolev, S., Fitoussi, H.: The Traveling Beams: Optical Solutions for Bounded NP-Complete Problems, Technical report #07-04, Ben Gurion University of the Negev (January 2007)
6. Dolev, S., Korach, E., Uzan, G.: A Method for Encryption and Decryption of Messages, PCT Patent Application WO 2006/001006 (January 2006)
7. Dolev, S., Yuval, N.: Optical implementation of bounded non-deterministic Turing machines, US Patent 7,130,093 B2, January 2005, Filed (May 2004)
8. Feitelson, G.: Optical Computing: A Survey for Computer Scientists. MIT Press, Cambridge (1988)
9. Garey, M.R., Johnson, D.S.: Computers and Intractability, a guide to the theory of NP completeness. W. H. Freeman and Company, San Francisco (1979)
10. Gutfreund, D., Shaltiel, R., Ta-Shma, A.: If NP Languages are Hard on the Worst-Case, then it is easy to find their Hard Instances. Journal of Computational Complexity (2007)
11. Haist, T., Osten, W.: An Optical Solution For The Traveling Salesman Problem. Opt. Express 15, 10473–10482 (2007)
12. Hopcroft, J.E., Karp, R.M.: An algorithm for maximum matching in bipartite graphs. SIAM J. Computing 2, 225–231 (1973)
13. Hyman, A.: Charles Babbage: Pioneer of the Computer. Princeton University Press, Princeton (1982)
14. Karp, R.M.: Reducibilty among combinatorial problems. Complexity of Computer Computations, 85–103 (1972)
15. Lenslet LTD, `http://www.hpcwire.com/hpcwire/hpcwireWWW/03/1017/106185.html`
16. Mann, H.J., Ulrich, W., Seitz, G.: 8-Mirror microlithography projection objective, US Patent 2004/0012866 A1, January 2004, Filed (April 2003)
17. McAulay, A.D.: Optical computer architectures. John Wiley, Chichester (1991)
18. Oltean, M.: A Light-Based Device for Solving the Hamiltonian Path Problem. In: Calude, C.S., Dinneen, M.J., Păun, G., Rozenberg, G., Stepney, S. (eds.) UC 2006. LNCS, vol. 4135, pp. 217–227. Springer, Heidelberg (2006)
19. Oltean, M., Muntean, O.: Solving the subset-sum problem with a light-based device. In: Natural Computing, Springer, Heidelberg (2007)
20. Shaked, N.T., Messika, S., Dolev, S., Rosen, J.: Optical solution for Bounded NP-Complete Problems. Journal of App. Optics 46, 711 (2007)
21. Reif, J.H., Tygar, D., Yoshida, A.: The Computability and Complexity of Optical Beam Tracing. In: 31st Annual IEEE Symposium on Foundations of Computer Science, pp. 106–114 (1990); Also The Computability and Complexity of Ray Tracing. Discrete and Computational Geometry 11, 265-287 (1994)
22. Tamir, D.E., Shaked, N.T., Wilson, P.J., Dolev, S.: Electro-Optical DSP of Tera Operations per Second and Beyond (Extended Abstract). In: Dolev, S., Haist, T., Oltean, M. (eds.) OSC 2008. LNCS, vol. 5172, pp. 56–69. Springer, Heidelberg (2008)

23. van Emde Boas, P.: Machine Models and Simulation. In: Volume, A. (ed.) Handbook of Theoretical Computer Science, Volume A: Algorithms and Complexity (A) pp. 1–66 (1990)
24. Woods, D.: Optical Computing and Computational Complexity. In: Calude, C.S., Dinneen, M.J., Păun, G., Rozenberg, G., Stepney, S. (eds.) UC 2006. LNCS, vol. 4135, pp. 27–40. Springer, Heidelberg (2006)
25. Woods, D., Gibson, J.P.: Lower Bounds on the Computational Power of an Optical Model of Computation. In: Calude, C.S., Dinneen, M.J., Păun, G., Jesús Pérez-Jímenez, M., Rozenberg, G. (eds.) UC 2005. LNCS, vol. 3699, pp. 237–250. Springer, Heidelberg (2005); Journal version, Natural Computing 79(1), 95–108 (2008)
26. Xiajun, W., Zhao, X., Bermak, A., Boussaind, F.: An AER based CMOS Polarization Image Sensor with Photo-aligned Micropolarizer Array. In: Calude, C.S., Dinneen, M.J., Păun, G., Jesús Pérez-Jímenez, M., Rozenberg, G. (eds.) UC 2005. LNCS, vol. 3699, pp. 95–108. Springer, Heidelberg (2005); Journal version. Natural Computing 7(1), 95–108 (2008)

Holographic Computation of Balanced Succinct Permanent Instances*

(Extended Abstract)

Shlomi Dolev[1], Nova Fandina[1], and Joseph Rosen[2]

[1] Department of Computer Science
Ben Gurion University of the Negev, Israel
{dolev,fandina}@cs.bgu.ac.il
[2] Department of Electrical and Computer Engineering
Ben Gurion University of the Negev, Israel
rosen@ee.bgu.ac.il

Abstract. Galperin and Wigderson proposed a succinct representation for graphs, that uses number of bits that is logarithmic in the number of nodes. They proved complexity results for various decision problems on graph properties, when the graph is given in a succinct representation. Later, Papadimitriou and Yannakakis showed, that under the same succinct encoding method, certain class of decision problems on graph properties becomes exponentially hard. In this paper we consider the complexity of the Permanent problem when the graph/matrix is given in a restricted succinct representation. We present an optical architecture that is based on the holographic concept for solving balanced succinct permanent problem. Holography enables to have exponential copying (roughly, $n \times n$ in each iteration) rather than constant copying (e.g., doubling in each iteration).

1 Introduction

The paradigm change in current computing from a single core to multi-core is a dramatic change, a change from sequential computing to parallel computing. Optical computing has benefits in parallel computations. Previous recent works suggested optical solutions for *NP*-Complete problems. In this work we present optical solution for a restricted variant of *NEXP* time hard task.

Optical holography is a technique for recording/reading data by using physical properties of coherent light, interference and diffraction. To record a simple *hologram* the laser beam is split into two separate beams of light. One beam (an object beam) illuminates the object. As a result some light reflects from the object and falls on the recording medium. Simultaneously the second beam (a reference beam) is directed to the same recording medium. The interaction of the two beams creates an interference pattern, called a hologram of an object. When the original reference beam illuminates the hologram, diffraction occurs and the object beam is reconstructed. The light observed by eyes defines the original object in three dimensions.

* Partially supported by the Lynne and William Frankel Center for Computer Science, Ben-Gurion University of the Negev, Israel. ICT Programme of the European Union under contract number FP7-21570 (FRONTS), and Rita Altura Trust Chair in Computer Sciences.

S. Dolev and M. Oltean (Eds.): OSC 2010, LNCS 6748, pp. 100–112, 2011.

There are various applications of holography in different fields. For example in security, medicine and art. We are interested in *holographic processing and holographic data storage*, as techniques for processing and maintaining data.

Holographic data storage approach encodes the data in holograms that are stored in the volume of a photosensitive medium, rather than (the typical) two dimensional storage devices. Thus allowing to store a huge amount of data that is proportional to the volume of the storage device. Today's technology enables to write approximately a terabit of data into holographic crystal of one cubic centimeter. Future technology may significantly increase the amount of information [5]. In addition to the high density, holographic data storage provides an ability for fast parallel reading/writing by using a single flash of laser light. Modern holographic memory devices allow data transfer rate as high as a billion bits per second.

All these advantages have inspired us to start a farther investigation of holographic data storage taking it in the direction of computing rather than only storage.

The current paper introduces a new approach that may be used in employing optics for solving exponential hard computational problem.

Definition 1. *Let A be an $n \times n$ matrix, such that:*

$$\forall\, a_{ij} \in A\,, \quad a_{ij} \in Z\,, \quad 0 \leq a_{ij} \leq p$$

for some natural number p. A permanent of matrix A is defined as:

$$perm(A) = \sum_{\sigma} \prod_{i=1}^{n} A_{i\sigma(i)}$$

Where the summation is over all the permutations σ of n elements.

In [7] Valiant showed that the permanent problem of integer matrix is *#P-Complete*. In [4] Lipton proved that the Permanent is *random-self-reducible* implying that the permanent is hard on the worst case as it is hard on the average case.

In order to find a problem that is *NEXP time hard on average*, we refer to the notation of *succinct representation* of graph, that was proposed by Galperin and Wigderson [3]. Papadimitriou and Yannakakis showed that certain *NP-Complete* problems on graph with succinct inputs are *NEXP time hard* [6].

We consider the permanent problem with succinctly represented inputs, as follows:

Definition 2. *Succinct Permanent problem defined as:*

input: An $O(\log^k n)$ sized boolean circuit C representing an $n \times n$ integer matrix A (with bounded entries) where k is some constant integer.

output: the permanent of matrix A.

Boolean circuits with only one output gate represent a binary matrix. To represent a general integer matrix a circuit with $O(k \times \log n)$ output gates can be used. We can represent such a circuit with $O(k \times \log n)$ number of circuits with one output gate. Each output gate corresponds to one bit in the binary representation of the integer number.

Therefore, this scheme of representation enables to encode integer matrices with both positive and negative values up to $O(n^k)$.

Notice, that the length of representation of the value of permanent of succinctly represented matrix can be exponential in the size of the input. Therefore, we suggest a decision version of the permanent with succinct inputs.

Definition 3. *Succinct Zero-Permanent problem defined as:*

input: An $O(\log^k n)$ sized boolean circuit C representing an $n \times n$ integer matrix A (with bounded entries) where k is some constant integer.

output: Permanent(A) == 0.

The aim of this paper is to establish the computational complexity of Succinct Zero-Permanent problem and introduce a new efficient optical device for solving it. In the sequel we prove that the Succinct Zero-Permanent problem is NEXP time hard and present a new efficient optical device for solving instances of the Succinct Zero-Permanent that are represented by balanced circuits, or random combination of such circuits.

The rest of the paper is organized as follows: in Section 2 the complexity of Succinct Permanent problem is considered. Section 3 describes the input for the optical device. In Sections 4 and 5 we describe the integrated part of the optical architecture and its holographic implementation. The description of the complete optical device for solving succinct permanent problem is provided in Section 6. We complete the paper with discussions and certain directions for future research.

2 Complexity of the Succinct Permanent Problem

Various optical solvers for *NP* Complete problems were proposed. They are beneficial due to the assumption that $P \neq NP$. We suggest considering computational problems that are provable harder than the problems in *P*. Namely, *NEXP* time hard problems.

Definition 4. *Let ϕ be a boolean formula. Then, $\#\phi$ denote the number of satisfying assignments of ϕ. Let C_ϕ be a succinct circuit representation of formula ϕ. Then, $\#C_\phi$ denote the number of satisfying assignments of ϕ.*

Next we prove that the Succinct Zero-Permanent is NEXP time hard in the worst case.

Theorem 1. *Succinct Zero-Permanent is NEXP time hard.*

Proof. By results in [6] Succ-3SAT decision problem is *NEXP* time hard. We reduce this problem to Succinct Zero-Permanent in the following way. In [7] Valiant presented a polynomial time reduction from #3 SAT to the Permanent problem of integer matrix. Given an instance ϕ of the 3SAT, the reduction constructs a directed, weighted graph G, with weights $-1, 0, 1, 2, 3$, such that

$$Permanent(G) = 4^{t(\phi)} \times s(\phi)$$

where $t(\phi)$ is twice the number of occurrences of literals in ϕ, minus the number of clauses in ϕ, and $s(\phi)$ is the number of satisfying assignments of ϕ.

This construction is based on the composition of graph structures in highly regular fashion. An output graph G is obtained by combination a polynomial number of *track*, *interchange* and *junction* structures. The order and type of structures are predefined by boolean formula and can be efficiently obtained from the formula. Namely, given two $c|\log x|$-bit integers, the indices of nodes of graph G, it can be determined in polynomial time (of the length of the integers) whether there is an edge between these nodes. The algorithm reads at most polylogarithmic number bits of ϕ. Call this algorithm A.

Let f denote the reduction transformation of Valiant. Given C_ϕ - an O $(\log^k n)$-sized boolean circuit representation of ϕ, combine an algorithm A with C_ϕ to obtain a polylogarithmic description of the graph $G = f(\phi)$. Finally, using a relation between $Permanent(G)$ and the number of satisfying assignments of ϕ, we obtain: if $Permanent(G) == 0$, then $\#\phi = 0$, and therefore there is no satisfying assignments for ϕ; if $Permanent(G) > 0$, then $\#\phi > 0$ implying the existence of satisfying assignments for ϕ. Note, that $4^{t(\phi)}$ is a positive integer number. □

Therefore, the Succinct Zero-Permanent of integer matrix (with both positive and negative values) problem is a good candidate for being solved with optical computing techniques. But does the hardness on the worst case suffice? We are interesting in problem that is hard on most of the inputs or on a big fraction of the inputs. We conjecture that the Succinct Zero-Permanent problem is hard on average as it is hard on the worst case. That implies that it is exponentially hard to solve this problem in most of the inputs.

3 Circuits as Inputs

We start with a description of the format of the input data supported by the Succinct Permanent solver device.

Encoding scheme for input. As was mentioned earlier, we solve a computational problem assuming succinctly encoded inputs. Various techniques for a short description of the data were proposed. We use the Small Circuit Representation (SCR) scheme, proposed by Galperin and Wigderson [3].

The input is a boolean $O(\log^k n)$ sized circuit, for some k, representing a particular matrix of size $n \times n$. The exact form of such SCR circuits is not a theoretical concern. In other words, the internal configuration, such as, the types of the used logic gates, fan-in/fan-out of this gates, is not important for achieving the complexity results. It remains the same, as long as the chosen model of circuits is complete, and as long as $O(\log^k n)$ space is used. We call the model of circuits with n inputs *complete*, if and only if, for each boolean function with n variables there is a boolean circuit from the model that computes this function.

We will consider circuits that contain only AND, OR and NOT gates. In particular, we choose to deal with boolean formulas, i.e, boolean circuits with $fan - out = 1$ for all gates. It is a well known fact, that such a model is complete [8]. Without loss of generality we can assume that NOT operations are applied only to the input variables of the circuit. If negations appear higher up in the circuit we can move them down to the

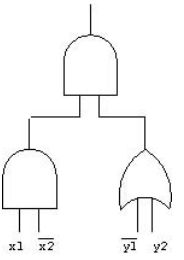


Fig. 1. Small Boolean circuit that represents a 4×4 matrix

0	0	0	0
0	0	0	0
1	1	0	1
0	0	0	0

Fig. 2. Matrix that represented by circuit in Fig. 1

inputs, using DeMorgan's low. By using the associative and the commutative properties of the AND and OR operations we assume that all gates in the circuit has $fan-in = 2$. Thus, we consider boolean circuits that have a topology of a binary tree.

Furthermore, in this work, we will assume a topology of a complete binary tree without duplication of the inputs. We call such a circuits a *balanced* boolean circuits. To obtain a larger set of boolean circuits, we combine a random set of balanced circuits with OR or AND operations. We believe that such a random set of instances is hard on average. Fig. 1 and Fig. 2 depict a boolean circuit and the 4×4 matrix it represents.

The optical architecture we build consists of two main parts: a holographic preprocessing unit and an optical Permanent solver. The preprocessing unit is an optical device that generates all matrices that can be defined by balanced small circuits using a bounded space to store this data. Thus, this device defines the size of the inputs of the Succinct Permanent problem that can be solved by the architecture. For each integer n the Preprocessing unit generates all matrices of size $n \times n$, that have a balanced Small Circuit Representation in $\log(\log n)$ number of iteration steps. This unit runs only once to build such matrices. The Succinct Permanent solver acts as follows: given a balanced boolean circuit representing a matrix, the solver finds an appropriate matrix stored in the medium, and outputs it to the Permanent solver to obtain a solution.

4 Preprocessing Unit

In this section we provide a detailed description of the method used by the Preprocessing unit to generate the matrices. In the preprocessing phase we generate and store all matrices of size $n \times n$, that has a balanced Small Circuit Representation. In this section we describe an algorithm for this task.

The idea is to produce a general matrix that contains all matrices of size $n \times n$ that can be encoded with $O(\log^k n)$ sized balanced boolean circuit. We accomplish this task by using an iterative algorithm. It starts with a general matrix corresponding to boolean circuits with one level of gates. By iterative copying phases, the algorithm extends the general matrix of i leveled circuits to a general matrix corresponding to $i+1$ leveled circuits.

Definition 5. *Let BC_i be the set of all balanced boolean circuits that contains i (full) levels of gates.*

Note, that circuits in BC_i have $2 \times 2^{(i-1)}$ input variables, and therefore they represent a $2^{(2^{(i-1)})} \times 2^{(2^{(i-1)})}$ matrices.

Definition 6. *Let G_i be an $2^{2^i} \times 2^{(2^{(i+1)}-1)}$ binary matrix. We call it a General Matrix for BC_i, if and only if, for each column J_c of G_i there exists a boolean circuit $C \in BC_i$ such that*

$$M(C) = J_c \qquad and,$$

for each $C' \in BC_i$ there exists a column J'_c in G_i such that

$$M(C') = J'_c$$

where $M(C)$ is a matrix encoded by the circuit C.

Note that each column of G_i represents a matrix that can be encoded by circuit with i levels of gates. There are $2^{(2^{(i+1)}-1)}$ such a matrices, each of size $2^{(2^{(i-1)})} \times 2^{(2^{(i-1)})}$. Next we describe the initialization stage and the inductive step.

Initialization stage. Method starts with building an initial General Matrix G_1. Each column in G_1 is a matrix that can be encoded by a boolean circuit from B_1.

Induction stage. Algorithm 1, that is described in Fig. 4 extends G_i to G_{i+1}. Algorithm 1 defines a skeleton for G_{i+1} in terms of *blocks* of fixed size (lines 2 to 4). Each cell of G_i turns to be a block of the size of G_i. The order of blocks is exactly the order of cells in G_i. These blocks are filled with data according to G_i (lines 5 to 11). The data contained in G_i is copied into all blocks that correspond to cells with value 1. All the rest of the blocks are filled with the value 0. All blocks together compose the part of G_{i+1}, that we call ${H^1}_{i+1}$. See Fig. 5 for a schematic description of the obtained matrix

M_1	M_2	M_3	M_4	M_5	M_6	M_7	M_8
0	0	0	1	0	1	1	1
0	0	1	0	1	0	1	1
0	1	0	0	1	1	0	1
1	0	0	0	1	1	1	0

Fig. 3. The General Matrix for BC_1

1: $\diamond$ **Build** ${H^1}_{i+1}$**:**
2: define a new binary matrix of size $2^{(2^{(i+1)})} \times 2^{(2^{(i+2)}-2)}$
3: divide this matrix into $2^{2^{(i+1)}} \times 2^{(2^{(i+1)}-1)}$ sized *blocks*. {As a result we get a structure composed of 2^{2^i} rows of blocks, each row contains of $2^{(2^{(i+1)}-1)}$ blocks.}
4: enumerate blocks in fashion of matrix indexing. {For $1 \leq \hat{i} \leq 2^{2^i}, 1 \leq \hat{j} \leq 2^{2^i}$ block $B_{\hat{i},\hat{j}}$ is in $\hat{i}$'s row and $\hat{j}$'s column of blocks.}
5: **for** $i = 1$ to 2^{2^i} **do**
6: **for** $j = 1$ to $2^{(2^{(i+1)}-1)}$ **do**
7: **if** $(G_i[i,j] == 1)$ **then**
8: $B_{\hat{i},\hat{j}} \leftarrow G_i$
9: **end if**
10: **end for**
11: **end for**
12: $\diamond$ **Build** ${H^2}_{i+1}$ **:**
13: ${H^2}_{i+1} \leftarrow$ NOT $({H^1}_{i+1})$
14: ${H^2}_{i+1} \leftarrow$ REFLECT $({H^2}_{i+1})$
15: $\diamond$ $G_{i+1} \leftarrow ({H^1}_{i+1}) \circ {H^2}_{i+1}$

Fig. 4. Algoritm 1. Create General Matrix.

0	0	0	Copy of G_1	0	Copy of G_1	Copy of G_1	Copy of G_1
0	0	Copy of G_1	0	Copy of G_1	0	Copy of G_1	Copy of G_1
0	Copy of G_1	0	0	Copy of G_1	Copy of G_1	0	Copy of G_1
Copy of G_1	0	0	0	Copy of G_1	Copy of G_1	Copy of G_1	0

Fig. 5. ${H^1}_2$ part of the General Matrix for G_2

${H^1}_{i+1}$ following the first part of an iteration. The second part, ${H^2}_{i+1}$, is obtained by negation and reflection applied to ${H^1}_{i+1}$ (lines 12 to 14 of the code). Finally,

$$G_{i+1} = {H^1}_{i+1} \circ {H^2}_{i+1}$$

where the $\circ$ operation is a concatenation.
The correctness of the Algorithm 1 is proved in the next theorem.

Theorem 2. *Given G_i Algorithm 1 returns G_{i+1}.*

Proof. For each cell in G_{i+1} we define a double index

$$\langle\langle \hat{i},\ k\rangle, \langle \hat{j},\ l\rangle\rangle.$$

The first pair indicates the row's coordinates, and the second corresponds to the coordinates of the column. To get the second index in each pair, we enumerate rows and columns in each block.

To prove correctness, we prove the following two lemmas.

Lemma 1. *For each $C \in BC_{i+1}$ there is column J of G_{i+1}, s.t. M(C)=J.*

Proof. Let C be some circuit from BC_{i+1}.
First case: an output gate of C is an AND gate. Let L, R be the circuits, s.t. $C =$ (L AND R). Therefore $L, R \in BC_i$.
Hence, by the induction assumption there exist J_L, J_R columns of G_i s.t.

$$M(L) = J_L \qquad M(R) = J_R$$

Let j_l, j_r be the indices of these columns in G_i. Consider a column M in G_{i+1} with index $\langle \hat{j}_l,\ j_r \rangle$. We show that

$$M(C) = M$$

and this will complete the proof.
We need to show that for all values of boolean variables $x_1, x_2, ..., x_{2^i} \quad y_1, y_2, ..., y_{2^i}$ it holds that

$$(1)\ \ M(C)[x_1\ x_2\ \ldots x_{2^i}\ ,\ \ y_1\ y_2\ \ldots\ y_{2^i}] \ =\ M[x_1\ x_2\ \ldots\ x_{2^i},\ \ y_1\ y_2\ \ldots\ y_{2^i}]\ \ (2)$$

Note, that by definition

$$(1) \ \ = \ \ C[x_1 \ldots x_{2^i}\ ,\ \ y_1 \ldots y_{2^i}]$$

where $C[x_1\ x_2 \ldots\ x_n]$ is an output value of C on inputs $x_1, ..., x_n$.

$$(2) \ \ = \ \ G_{i+1}[\ \langle \widehat{x_1 \ldots x_{2^i}}\ ,\ \ y_1 \ldots y_{2^i} \rangle\ ,\ \langle \hat{j}_l,\ j_r \rangle\ ]$$

Recall that $C =$ (L AND R), therefore,

$$C[x_1 \ldots x_{2^i}\ ,\ \ y_1 \ldots y_{2^i}] = L[x_1 \ldots x_{2^i}]\ AND\ R[y_1 \ldots y_{2^i}]$$

.
If $L[x_1 \ldots x_{2^i}] = 0$ then $G_i[x_1 \ldots x_{2^i}\ , j_l] = 0$ and $C[x_1 \ldots x_{2^i}\ ,\ \ y_1 \ldots y_{2^i}] = 0$. Hence, by the construction, block $B_{\widehat{x_1 \ldots x_{2^i}}\ ,\ \hat{j}_l}$ in G_{i+1} is filled with 0. In particular $G_{i+1}[\ \langle \widehat{x_1 \ldots x_{2^i}}\ ,\ \ y_1 \ldots y_{2^i} \rangle\ ,\ \langle \hat{j}_l,\ j_r \rangle\ ] = 0$.
If $L[x_1 \ldots x_{2^i}] = 1$ then $G_i[x_1 \ldots x_{2^i}\ , j_l] = 1$ and $C[x_1 \ldots x_{2^i}\ ,\ \ y_1 \ldots y_{2^i}] = R[y_1 \ldots y_{2^i}]$. Hence, by construction, G_i is copied to block $B_{\widehat{x_1 \ldots x_{2^i}}\ ,\ \hat{j}_l}$ in G_{i+1}. Therefore, column with index j_r in this block is a copy of the column with index j_r of G_i. It follows that $G_{i+1}[\ \langle \widehat{x_1 \ldots x_{2^i}}\ ,\ \ y_1 \ldots y_{2^i} \rangle\ ,\ \langle \hat{j}_l,\ j_r \rangle\ ] = R[y_1 \ldots y_{2^i}]$.
This completes the first case of the Lemma 1. Consider the second case.
Second case: an output gate of C is an OR gate.
We make a reduction to the previous case. Define a circuit $\hat{C}$ = $NOT(C)$. By DeMorgan low an output of $\hat{C}$ is an AND gate. Hence, there is column M in G_{i+1}, s.t. the

matrix encoded by $\hat{C}$ is M. Therefore, C represents a matrix that is $NOT(M)$. By construction, there is some column $\hat{M}$ in G_{i+1} s.t. $\hat{M} = NOT(M)$. □

Lemma 2. *For each column J of G_{i+1} there is $C \in BC_{i+1}$, s.t. M(C)=J.*

Proof. Let J be some column of G_{i+1}.
First case: column J is in the $(H^1{}_{i+1})$.
Let $\langle \hat{j}_l,\ j_r \rangle$ be the index of the J in G_{i+1}. By induction assumption, for the columns C_{j_l} and C_{j_r} in G_i there are circuits L and R in B_i s.t.

$$M(L) = C_{j_l} \qquad M(R) = C_{j_r}$$

Define circuit

$$B = (L\ AND\ R)$$

Note, that $C \in B_{i+1}$ as required.
Use the same arguments as in the previous Lemma to prove that $M(B) = J$.
Second case: column J is in the $(H^2{}_{i+1})$.
Hence, by construction, there is column $\hat{J}$ in $(H^1{}_{i+1})$ s.t. $J = NOT(\hat{J})$. Therefore, there is a circuit C that represents $\hat{J}$. Applying a logical NOT to the output of C and DeMorgan low results in a circuit $\hat{C}$ that represents J.

□

Lemma 1 proves the if condition, namely, that if C is a circuit form BC_{i+1} then there is a column in G_{i+1} that is represented by C, and Lemma 2 proves the only if condition, hence the correctness of the theorem.

□

5 Preprocessing by Holography

We propose an optical architecture using holographic approach to implement the above algorithm. To the best of our knowledge, this is the first technique that can perform exponential copying, namely producing n copies, where n is the size of the input, rather than producing a constant number of copies in each iteration. The optical device works in phases. In phase i the input is G_i recorded on a photosensitive film and the output is G_{i+1} recorded on a photosensitive film, which is the input to the next phase. For the initialization phase we prepare a film with G_1. For the induction phase we introduce the system described in Fig. 6.

Implementation of a phase consists of two stages: hologram recording and reconstruction.

Recording stage. Laser beam illuminates an input matrix placed on an input plane. Then the beam is split into two separated beams. One of the beams is directed to a $4f$ correlator that is capable of performing a duplication of the input matrix. Another beam propagates through a magnification system designed to enlarge and duplicate an input matrix. Output beams from both systems are directed to a photograph plate where their interference pattern is recorded.

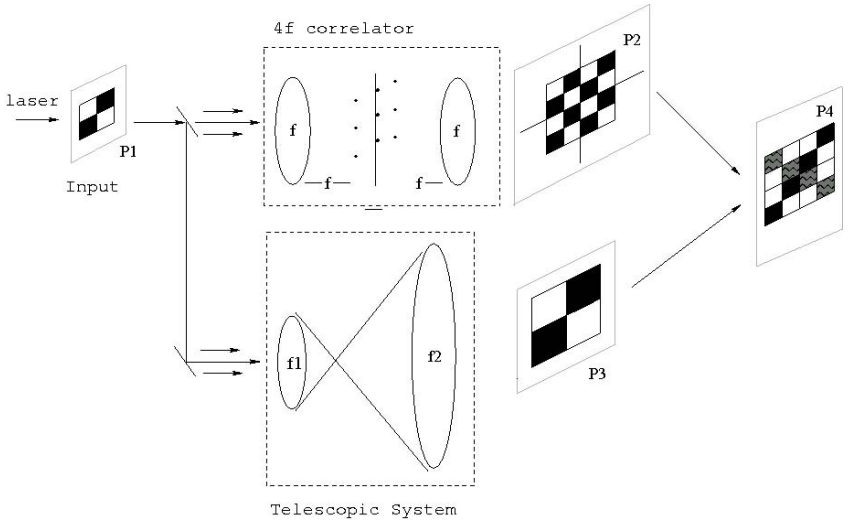


Fig. 6. Hologram Recording

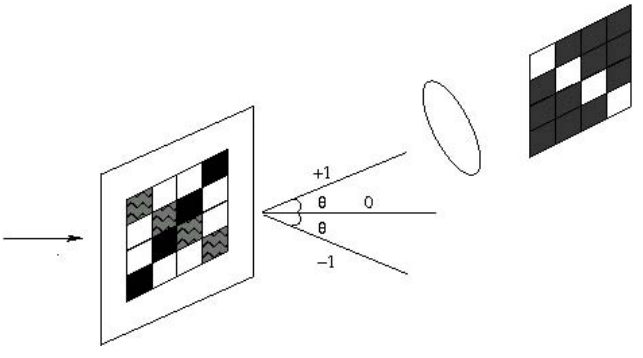


Fig. 7. Hologram Reconstruction

As shown in Fig. 6 an input binary matrix is represented on a film and placed on the input plane P1. A white rectangle on the slide represents a logical $'1'$ in the matrix, and a black rectangle on the slide represents a logical $'0'$.

An output from the $4f$ system is presented on plane P2. In each phase we calibrate the distance between holes in a pinholes array to yield the right gap between the duplications, according to the size of the input matrix on the i-th phase.

Enlarged image of the input matrix is on plane P3. A scale rate is determined by focal lengths of two lenses in the following form:

$$M = \frac{f_2}{f_1}$$

Plan P4 represents the image recorded on the film. Interference fringes are created only at positions that correspond to logical $'1'$ at both beams. We develop this film and use it in the reconstruction stage.

Reconstruction stage. To get a final representation of the input matrix to the next phase we perform a reconstruction of the hologram as depicted in Fig. 7.

The hologram is illuminated by laser. The beams with direction angle θ are transmitted by lens toward a photosensitive plate. The resulting image is recorded on the film and is used as input matrix for the next phase.

Note, that we described a system for building part of G_{i+1}. In particular, the result of the described phase is ${H^1}_{i+1}$. We add a pinhole array device to the telescopic system to enable both enlargement and duplication of the input matrix. Thus, we will get two identical resulting films. Developing one of them in a positive form and another in a negative form to obtain ${H^1}_{i+1}$ and ${H^2}_{i+1}$.

6 An Optical Architecture for Succinct Permanent Problem

Now we present the whole architecture for solving the balanced succinct permanent problem. Given an input in the form of a circuit we need to output its related matrix in a fast manner using optics. Then use the matrix as input to an existing optical architecture for solving a permanent problem.

Several optical devices for solving a permanent problem were proposed. In [2] Dolev and Fitoussi suggest architecture for computing permanent of binary matrices in polynomial time. In this device the answer is obtained by detecting an intensity of the light, that propagates trough built set of masks. Thus, to solve Zero-Succinct Permanent problem we only should detect if there some light that propagates through the masks. We assume the existence of such a device and use an optical architecture for solving the permanent problem of integer valued matrix, that was proposed in [1] by Anter and Dolev. This device solves instances of integer matrices with positive values. We should modify this architecture to handle matrices with negative values as well. We leave this task to our future work. We mention that our recent results conclude that a decision version of Succinct Permanent problem, applying modulo given prime operation is NEXP time hard. Thus, we can solve it using optical device in [1] without any modifications. In fact, for input boolean circuit, we should just detect if the result is zero modulo a prime (randomly chosen from a small set of primes).

We represent a balanced boolean circuit C with $2 \times \log n$ inputs in heap form. Inputs with NOT operation are represented by the label $'N'$. All other inputs are represented by label $'G'$. AND, OR operations are represented by labels $'A'$, $'O'$ respectively. Following we present a simple recursive algorithm that returns an index of the column of the General matrix $G_{\log(2\times\log n)}$ representing a matrix of input circuit C.

The algorithm for computing the index in the matrix, called Algorithm 2, appears in Fig. 8. *Left(Right)Child* denotes a heap rooted at left (right) child of H. Algorithm 2 performs 2 recursive calls (lines 3 to 4). According to the type of the output gate of the input circuit, and the results of the recursive calls, algorithm computes an exact index of the column (lines 5 to 9). *sizeof* refers to the number of the columns of the matrix. It can be computed directly from n. Algorithm 2 finishes the recursive calls when H_1 and H_2 are heaps with only one gate, and thus, it can return an answer immediately. The correctness of the algorithm follows from the correctness of Algorithm 1.

Using Algorithm 2 and acousto-optic modulators the architecture outputs a matrix encoded by balanced boolean circuit. Then, an optical permanent solver provides a

Require: heap H representing balanced boolean circuit C with $2 \times \log n$ inputs
Ensure: index of the column of General Matrix that corresponds to $M(C)$
1: $H_1 \leftarrow LeftChild\,(H)$
2: $H_2 \leftarrow RightChild\,(H)$
3: $i \leftarrow MatrixIndex\,(H_1)$
4: $j \leftarrow MatrixIndex\,(H_2)$
5: **if** $H[1] ==$ 'A' **then**
6: **return** $(i-1) \times sizeof\left(G_{\log(\log n)}\right) + j$
7: **end if**
8: **if** $H[1] ==$ 'O' **then**
9: **return** $(i-1) \times sizeof\left(G_{\log(\log n)}\right) + j + sizeof\left(G_{\log(\log n)}\right)^2$
10: **end if**

Fig. 8. Algoritm 2. Computing an index of General Matrix.

solution. Several acousto-optic modulators can select randomly in parallel several balanced boolean circuits. By applying optical AND and OR operations to chosen circuits we obtain a combined circuit that encodes a more general instance. In this way we construct a bigger set of instances. We conjecture that such a random set encode hard instances.

7 Discussion

We presented a solution for the balanced Zero-Succinct Permanent that is a restrictive case of the Zero-Succinct Permanent problem. To solve instances that corresponds to integer matrices (not only binary matrices) we represent an integer matrix by composing a set of $O(k \times \log n)$ balanced boolean circuits, obtaining a short boolean circuit with $O(k \times \log n)$ outputs. Outputs of such a circuit represent in a natural way an integer value in binary basis.

Given the proven hardness on average of the integer Permanent problem, we conjecture that integer Zero-Succinct Permanent problem is hard on the worst case as it is on the average.

The preprocessing stage architecture uses a photosensitive film for data storage. We mention, that holographic data storage may be used instead, for obtaining a faster performance and a reduction in the physical size of the architecture.

References

1. Anter, A., Dolev, S.: Optical solution for hard on average #p-complete instances (using exponential space for solving instances of the permanent). Natural Computing 9, 891–902 (2010)
2. Dolev, S., Fitoussi, H.: Masking traveling beams: Optical solutions for np-complete problems, trading space for time. Theor. Comput. Sci. 411(6), 837–853 (2010)
3. Galperin, H., Wigderson, A.: Succinct representations of graphs. Inf. Control 56, 183–198 (1984)

4. Lipton, R.: New directions in testing. In: Distributed Computing and Cryptography. DIMACS Series on Discrete Mathematicsand Theoretical Computer Science, vol. 2, pp. 191–202 (1991)
5. Moon, C.R., Mattos, L.S., Foster, B.K., Zeltzer, G., Ko, W., Manoharan, H.C.: Quantum Phase Extraction in Isospectral Electronic Nanostructures. Science 319(5864), 782–787 (2008)
6. Papadimitriou, C.H., Yannakakis, M.: A note on succinct representations of graphs. Inf. Control 71, 181–185 (1986)
7. Valiant, L.G.: The complexity of computing the permanent. Theoretical Computer Science 8(2), 189–201 (1979)
8. Wegener, I.: The complexity of Boolean functions. Wiley-Feubner Series in Computer Science. B. G. Teubner and John Wiley & Sons, Chichester (1987)

Being Analog

Sunny Bains

Electrical and Electronic Engineering Department,
Imperial College London, United Kingdom
s.bains@imperial.ac.uk

Abstract. The conventional wisdom has been that artificial intelligence is all about algorithms, and therefore restricted by the Turing model of computation. In fact embodied machines (those that interact with the outside world via sensors and actuators) cannot reasonably be modelled as being 'digital' in the conventional sense, or even as Turing machines. Far from being a disadvantage, by understanding and exploiting analog nature of these machines we have the opportunity to increase power efficiency, improve learning and discrimination, and provide better adaptability to changing circumstances.

1 Big Artificial Intelligence

Since long before Roger Penrose wrote his misleading book, *The Emperor's New Mind* [6], people have been confused about what they mean by artificial intelligence (AI). Sometimes they mean entirely *virtual* intelligence: that is, intelligence that only operates inside the closed-off virtual world of a machine (like artificial life). Sometime they mean intelligent systems that take information from the outside world and pass commands out to it, but entirely mediated by fixed analog-to-digital (a/d) converters, feature extractors and so forth that decide what input the 'intelligence' should operate on before it gets to see it.

But when I use the term AI, I mean what probably we all think of as the eventual goal of artificial intelligence: a thinking, acting, sentient android or robot that can adapt to new tasks and learn—as we do—through experience. To differentiate, I use the term Big AI. Of course, it would be pointless if such machines were identical to us. They would be faster or stronger or cleverer than we are. However, they would use many of the same mechanisms for intelligent behavior as we do, because these mechanisms have been tested by evolution and found worthy.

I would argue that interest in Big AI requires the abandonment of the traditional practice of engineering, while in another sense it takes the discipline to a new level. Traditional engineering involves a specification, which might be something like: "build me a machine that can do *this task*, and that satisfies the following constraints." The constraints could be power efficiency, size, safety, compatibility with other systems, whatever. The task in question could be a single task or it could be a *finite* set of tasks, but whatever it is, it is specific. Thus, without such a specification, we could not set about building a system in the old way.

S. Dolev and M. Oltean (Eds.): OSC 2010, LNCS 6748, pp. 113–118, 2011.

So how would we specify a human, or its machine equivalent, an android? It's true, what a human can do is constrained. But is it finite? Could we write down every little task we might want a machine to be able to perform and then program it to do that in every conceivable set of circumstances? Of course not: we wouldn't try.

This is why machine learning—both for the most sophisticated and trivial of applications—is such a huge topic of interest. We understand that machines are going to have to learn what to do as well as how to do it. The building of systems that can't do what's necessary, but which can learn, is creating a new level of engineering: a kind of meta-engineering if you like.

2 The Analog Shell

But this shifting of the problem from *knowing* what you need to *learning* it does not just apply to the intelligence side, *it applies to the sensors and actuators too.* If you don't yet know all of the environments a machine will have to work in during its lifetime, or all the tasks it will have to perform, how can you accurately decide how much you need from its sensors? And how will you decide which systems in the brain or nervous system should receive the sensor data and which do not need it? And how will you even decide what constitutes a sensor and what does not?

The fact is that physical objects 'process' physical information whether they want to or not. They are inherently coupled to all of the physical forces around them. Windows, for instance—by which I mean the glass things that let you look out of buildings, not the operating system—vibrate in the presence of noise. This is why you can shine a laser onto them and, based on the changing reflection, use them as a kind of microphone. Windows are not sensors by the term's conventional meaning, but they *do sense.* Neither are they actuators in the conventional sense, but they *do actuate.*

This is not just an abstract argument. Parts of the body that would not be accepted as 'sensors' by any conventional engineer are, in fact, used in human cognition. The most dramatic demonstration of this that I've come across is the percussionist Evelyn Glennie who, though born hearing, became profoundly deaf. According to her own website [4], "Evelyn spent a lot of time when she was young (with the help of Ron Forbes her percussion teacher at school) refining her ability to detect vibrations. She would stand with her hands against the classroom wall while Ron played notes on the timpani (timpani produce a lot of vibrations). Eventually Evelyn managed to distinguish the rough pitch of notes by associating where on her body she felt the sound with the sense of perfect pitch she had before losing her hearing. The low sounds she feels mainly in her legs and feet and high sounds might be particular places on her face, neck and chest."

So our brains are able to leverage incidental information (information that most engineers would describe as noise) and use it to solve problems. Why would we want to prevent the machines we create from being able to do this as well? Because as soon as you start to over-engineer—that is, be too specific about

what you want a particular part of your machine to achieve and how you expect it to perform—you risk preventing that leg or ear or muscle from being used to help perform other tasks that you haven't thought of yet.

From a practical point of view, what I am arguing is that if you have a nice analog sensor (as most are), why are you ignoring part of what it's trying to communicate by forcing it to 'speak' through the filter an analog-to-digital converter? Not only does this force the signal into permitted digital states, but it can mask the underlying signal shape and important temporal dynamics.

Neuromorphic engineering—the goal of which is to create mainly (though not exclusively) analog circuits and systems that behave with similar dynamics to those found in the brains, senses, and nervous systems of real animals—is full of examples where processing the analog signal *in the analog domain* can give you a huge advantages in power and speed. For a full description of how these advantages are possible, how the approach relates to Big AI, many examples of successful neuromorphic engineering projects, and a description of some of the alternative computing technologies that can be thought of as being compatible with the approach (including optical) please see [2] and Chapter 2 of [3] and references therein.

A colleague, Chris Toumazou, describes this problem of getting information from the analog real world to the virtual world of computers as *the analog shell*. What's interesting about this metaphor is that it implies that you need to think about where you put barrier between the analog and digital domains: too early and you either throw away information or leave yourself with lots of (expensive) digital signal processing to do.

However, I prefer to think of it not a shell, but maybe a set of shells (like Russian dolls). At each level of processing, the information is more processed, more abstract, less analog, more digital, until eventually you can safely do a straight a/d conversion (or even just treat the analog information as digital) without losing anything.

The advantage of this is that you don't have to approach the problem of Big AI in either an entirely analog or digital way: you can be pragmatic about it. It may make sense for those working on high-level reasoning or mathematics to work using more conventional computing (digital/Turing) models. Likewise, however, you would not want to have to use entirely digital technology for the huge mass of signals being received by an android body: analog solutions will probably be more appropriate for the initial signal processing. The difficult engineering problem can then be defined as making sure that the information moves smoothly and appropriately from one regime to the other.

3 Physical Beings, Physical Computation

However, I would argue that analog computation is just a subset of the bigger challenge to traditional approaches to AI. Certainly, the misunderstanding of continuous approaches is a major problem for many in this field: hence the name of the paper. But the deeper problem is that computer scientists forget

that not every system has to turn a task into manipulation of 'information' in order to perform useful processing.

There are numerous examples of what I would call 'physical computing': using the physics of a device to straightforwardly process a signal. Familiar examples would include mechanical scales, sundials, even lenses that can take an image and Fourier Transform it *at the speed of light.* Such processes are inherently efficient. Which is the quickest way to make a robot limb safer: by calculating the forces impinging on it, working out from the context what is going on, and giving a command to withdraw the limb; or by putting springs in the limb, so making it compliant? What is the fastest way to find matching pictorial elements on images, by cutting them into strings of bits and using complex algorithms to reconstruct their geometry and make nearest-neighbor comparisons? Or to optically filter the Fourier Transform of one image with another in the time it takes to refresh a display (optical correlation)?

It is true that physical solutions are rarely as widely applicable as their information-based counterparts, just as custom chips are usually more efficient but less adaptable than general-purpose processors. But with a problem as difficult as trying to fit the intelligence, acuity, and dexterity of a person into a machine, surely it is necessary to take advantage of every efficiency available?

And, even if this were not the case, even if we saw no advantage whatsoever in using efficient analog, optical, mechanical or chemical processing, even then, there is a problem. Because androids and robots do *not* live in a virtual world and therefore *cannot* be understood purely as Turing machines. *By definition,* any robot that acts in the world is acting in an analog way (if analog has any meaning at all, those of you want to argue about discrete space-time models can see the arguments in my thesis [3]). Wheels spin in through an uncountably infinite number of states, limbs move through an uncountably infinite number of points.

For computer scientists, this can be a terrifying idea. Essentially, what I am claiming is that people are not computers. I'm not the first, of course: Turing said it himself and there is a whole field of people who are interested in the idea of applying dynamical systems theories to both 'real' and artificial intelligence.

But again, this should not be seen as a problem for AI researchers, but rather an opportunity. Embodied artificial intelligences *need not be constrained* by the computational limits of Turing machines. Researcher Hava Siegelmann has shown that recurrent analog neural networks (which seem to be the kind of networks that our brains use) can perform functions that would normally be considered super-Turing [7].

4 The Drawbacks

Where all the evidence points to the idea that physical computation can improve computational power, energy efficiency and speed, why has the switch to analog and physical approaches been, relatively speaking, so slow? There are many reasons. It is partly because the theoretical framework isn't there yet. A goal of

my research [1] was to start to build a bridge between the Turing and physical worlds, but it will take concentrated practical and theoretical work to develop a set of robust guidelines that help engineers to understand which problems are best tackled in an analog way, in a digital way, or in a more direct physical way.

Another problem is that, to be blunt, we've all got a bit lazy. Doing research in analog or dynamical systems is complex (which I mean in both the everyday and mathematical senses of the word) and mathematically challenging. Programming is much easier (and doesn't require you to send things to be fabricated). There is an increasing shortage of analog electronics engineers, which means that—even when digital solutions are less efficient, more expensive, etc.—they are sometimes easier to implement.

Finally, a major problem—at least in my experience—has come from ignorance within the computer-science community (and electrical engineering too). People think that the assumptions and approximations made in the Turing machine model and communications theory can be applied to all of physics. This is simply not true.

In a way, this is the opposite of the mistake that Penrose makes: he assumes that computers are not physical, and therefore entirely constrained by Turing. What I am saying is that this constraint on computers may be true, but only applies to machines that operate in an entirely virtual world.

Yet another part of the problem may be the fact that we all understand the advantages of digital for communication and media. I'm not arguing that digital isn't (generally) a better technology to use in these fields. But the problem of trying to communicate and receive a specific message is completely different than the problem of an intelligent agent has when trying to suck as much information out of the environment as its sensors will allow in order to perform an *as yet unknown task.*

To say it another way, communication theory and the Turing model both have, as inherent assumptions, that there is a *known* acceptable level of approximation that can be tolerated. This assumption does not hold in Big AI. For some task a single bit response from a sensor may be sufficient. For other tasks many bits may be needed, with sophisticated averaging over time and space. The intelligent system must learn what it needs for what tasks, and what information it can afford to throw away. If it doesn't, it either risks not having the information it needs when necessary, or wasting a lot of time and energy processing signals at resolutions that are simply too high.

5 Moving Forward

There are many avenues that can progress the physical computation agenda in AI, and people are working on a number of them. As well as the other disciplines I've described (neuromorphic, optical, mechanical) there are relatively new approaches to computing, like probabilistic electronics that (see e.g. [5]) aim to adjust the power devoted to a task to the need for precision in that task.

From my point of view, however, that harmonizing framework, the pulling together of one discipline from many disparate fields, that is missing. Perhaps it is too early for that.

While we are waiting, however, the best thing that we can all do, as computer scientists, as roboticists, as electronic engineers, is to keep an open mind. We should be constantly questioning whether the theories, methods, and approaches we are used to applying are really suitable for the job in hand. Do the assumptions that the theories were based on hold for this task? We should keep looking to biology for alternative solutions. How does the brain/spinal cord/retina solve this problem? And we should keep asking ourselves whether we are trying to be too controlling as engineers. Are we hardwiring behaviors rather than making it possible for the machine to learn to do what we need it to?

The technical challenges are not insubstantial. But before we can start to tackle them, we must open our eyes to the fact that Big AI really is not the same as the other engineering fields we are used to.

Further Reading

For brevity, references have been kept to a minimum in this paper. For further discussion of the practical problems of, and technologies that can be used in, Big AI, please see [2] and Chapter 2 of [3]. For further discussion of the theoretical issues surrounding Big AI, see Chapter 3 of [3]. Subsequent thesis chapters describe a simple model that incorporates the Turing machine into physical computation, and discuss how interpretations of physics may affect the theoretical implications of this model. The thesis and all papers are freely available at: http://www.sunnybains.com.

References

[1] Bains, S.: Intelligence as Physical Computation. AISB Journal (1), 225–240 (2003)
[2] Bains, S.: Extending neuromorphic engineering beyond electronics. In: Proc. Brain Inspired Cognitive Systems, Stirling (2004)
[3] Bains, S.: Physical computation and embodied artificial intelligence, PhD Thesis, The Open University (2005)
[4] Glennie, E.: Evelyn's Hearing (1996), http://www.evelyn.co.uk/hearing.htm (accessed 2004)
[5] Palem, K.V.: Energy aware computing through probabilistic switching: a study of limits. IEEE Trans. Computing, 1123–1137 (2005)
[6] Penrose, R.: The Large, the Small, and the Human Mind. Cambridge University Press, Cambridge (1997)
[7] Siegelmann, H.T.: Neural Networks and Analog Computation: Beyond the Turing Limit. Birkhäuser, Boston (1999)

Exceeding the Diffraction and the Geometric Limits of Imaging Systems: A Review

Zeev Zalevsky

School of Engineering, Bar-Ilan University, 52900 Ramat-Gan, Israel
zalevsz@eng.biu.ac.il

Abstract. In this review paper we explain the diffraction and the geometric related limits of an imaging system and discuss several practical techniques to overcome those limitations.

Keywords: Super resolution; Diffraction limit; Geometric limit.

1 Introduction

What actually resolution is? It is the finest spatial feature that an imaging system can resolve. Resolution of optical systems is restricted by diffraction (Lord Rayleigh, Abbe) [1], by the geometry of the detector [2] and by the noise equivalence of its pixels [2]. The diffraction limitation of resolution is proportional to the product between the optical wavelength and the F number of the imaging lens (the ratio between the focal length and the diameter of the lens) and it relates to the angular span of rays that are diffracted from a small spatial feature in the object while still being collected by the imaging lens [3]. The geometric resolution is limited by the spatial density and the area of each one of the pixels in the detector [2]. The more spatially dense the pixels are the higher is the Nyquist sampling frequency is. The smaller each pixel is the closer it is to an ideal sampling with an array of Dirac's delta functions. Noise equivalent resolution is originated by the internal noises existing within each pixel of the detector (electronic noises, shot noises etc) and basically what it means is that after passing the previous two limitations the signal is still being sampled with sufficient number of bits while its signal to noise ratio is sufficiently large to be detected.

In this paper we present a short review of the space bandwidth adaptation process (section 2), how to overcome the resolution limitation due to diffraction (section 3), how to overcome the resolution limitation due to the geometry of the detector (section 4) and how to use light not only to see better but also to sense vibrations (section 5). The paper is concluded in section 6.

2 SW Adaptation Process

If a spatial detail is not resolved, is it hopeless? The answer is NO if a priori information on the object is available. There are various types of a priori information:

S. Dolev and M. Oltean (Eds.): OSC 2010, LNCS 6748, pp. 119–130, 2011.

we may know that the imaged object is one dimensional, it has polarization restricted information, the signal is temporally restricted, wavelength restricted, the object has a unique shape etc. Having a priori knowledge of the signal may lead to super resolution using an SW (space-bandwidth) adaptation process [4]. The SW was recently defined as a chart rather as a number (as it was before that definition) allowing in a better way to see graphically the spatial-spectral distribution of degrees of freedom that need to be detected by the imaging sensor. This chart is computed using a phase space distribution such as the Wigner transform that is being properly threshold yielding a binary space-frequency distribution indicating where there are (when it has value of one) or where there are no (where the SW has value of zero) degrees of freedom:

$$SW(x,\nu)=\begin{cases}1 & for \quad all \quad \langle W(x,\nu)\rangle > W_{tresh} \\ 0 & otherwise\end{cases} \tag{1}$$

$W(x,\nu)$ is the Wigner distribution. W_{Tresh} is the threshold according to which the SW is computed and which determines the spectral-spatial regions containing too low energy than to be considered as a degree of freedom.

In the adaptation process [5,6] we adapt the SW of the signal to the acceptance SW of the system. Obviously the basic condition for succeeding of doing so is not only having the a priori knowledge of how one can do the adaptation but also that the area of the SW of the signal (its number of degrees of freedom) is equal or smaller than the one of the system (the overall number of degrees of freedom that the system can receive over all of its domains such as time, polarization, wavelength, space etc).

3 Diffractive Type of Super Resolution

The basic configuration for performing the diffractive type of super resolution is related to time multiplexing [7,8]. In this basic configuration a high resolution periodic spatial encoding structure is physically attached to the inspected object. Similar decoding structure (scaled according to the magnification of the imaging system) is attached to the detection array. Both structures are moving along the transversal direction. The ratio between their velocities is according to the magnification factor of the imaging system. When the encoding structure is static the spatial spectrum of the inspected object is replicated in the aperture plane. Thus all the high frequency spectral slots, that before were cut out by the aperture of the imaging lens, are now folded into the aperture. However, the folded spectral slots are mixed together and one cannot separate between them all. Now when the encoding grating is shifted then due to a generated Doppler like effect each spectral replication receives a different Doppler shift (because each replication is a different diffraction order of the encoding grating and thus the direction of diffraction of the various orders has different component along the direction of movement (the higher the diffraction order is, the higher the Doppler shift is). Therefore, the mixed spectral slots can be separated. The decoding grating does exactly this. It performs additional spectral replication with proper opposite Doppler shift added to each replication. After performing time averaging in the detector all the non desired replicated terms are averaged to zero (the terms in

which the Doppler correction made by the decoding grating did not cancel the Doppler shift generated by the encoding grating) and the rest are reinforced and construct a synthetic aperture transmitting high spatial frequencies. In order to obtain proper extension of the synthetic aperture the basic period of the encoding grating should generate spectral replications having separation equal to the size of the imaging aperture. In that case the resolution improvement factor equals to the number of diffraction orders [9,10].

In Fig. 1 we present an example of a time multiplexing super resolving experiment [10] where the upper row presents the high resolution image of three different objects. The middle row shows how they are seen in a low resolution imaging system and the lower row is the experimentally obtained reconstruction using time multiplexing. One may see the significant improvement in resolution and the resemblance between the upper and the reconstructed lower row.

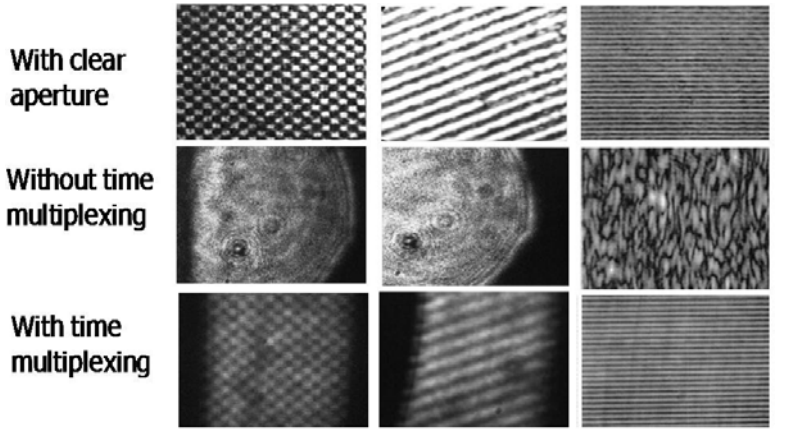


Fig. 1. Example of experimentally obtained time multiplexing based super resolution

An extension of this technique is related to replacing the encoding grating by a projected structure which can even be not periodic but random as obtained with a diffuser that when being illuminated projects random speckle patterns on top of the inspected object [11,12]. Another extension is replacing the decoding structure (that matches the encoding one) with a digital decoding [13]. In this process each image in the set is digitally multiplied by the decoding pattern before performing the summation (time integration). The next extension includes passive time multiplexing approach for detecting moving objects. There instead of projecting the pattern one uses the background of a moving object as the encoding/decoding structure [14]. However, as in the previous modifications for the basic technique one had to know the encoding pattern in high resolution in order to properly perform the decoding. Here as well the process requires knowing the background at high resolution and using it to decode the set of low resolution images. In addition this modification of the basic operation principle allows super resolving only the outline or the contour of the moving object.

In order to avoid knowing the encoding/decoding pattern completely one may use rain drops which are falling and thus have relative movement in respect to the inspected target [15]. Although each drop is blurred due to the low resolution imager, the blurred spots are sparse and thus they are sufficiently separable which allows one to digitally allocate the center of each moving spot and to construct the high resolution decoding pattern including a set of small spots (that correspond to the centers of the moving blurred spots being the rain drops).

The very same concept can be applied for near field microscopy to allow seeing sub wavelength features without the need to apply powerful laser illumination. In this approach the inspected sample is inserted into a fluid with metallic nano particles. Similar to the rain droplets approach, the nano particles are used as the encoding pattern while the required decoding function is extracted from the set of low resolution patterns having the moving nano particles [16]. In Fig. 2 we present one image from the set of captured images where one may see the resolution target on top of which we have the floating gold nano particles (seen as black spots).

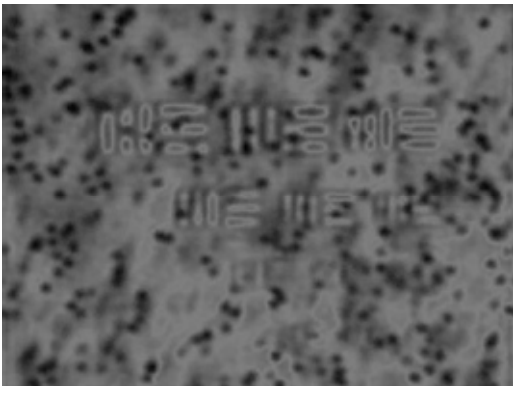
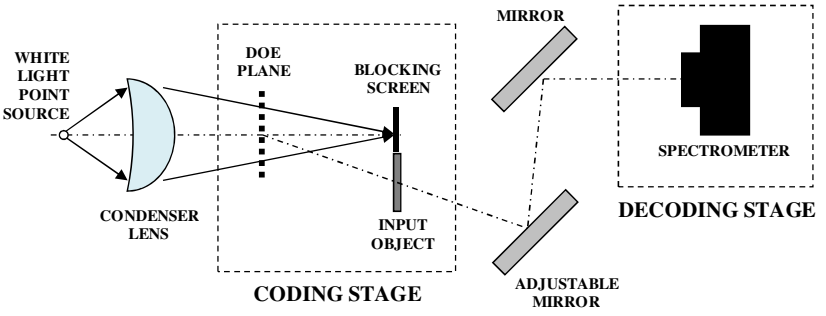


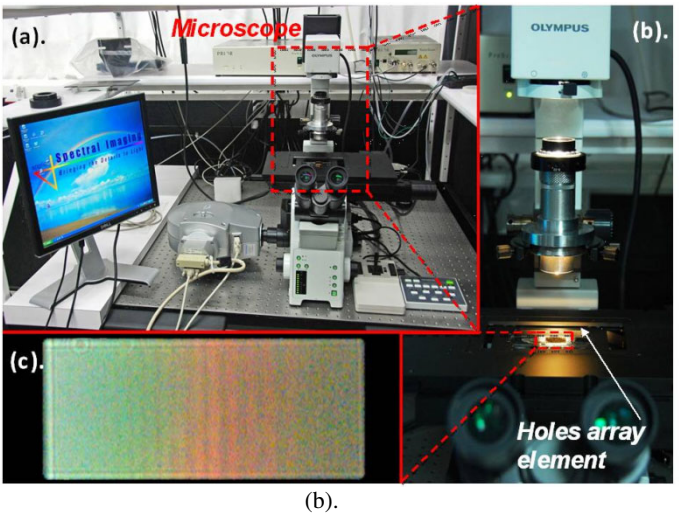
Fig. 2. Gold nano particles based time multiplexing super resolving approach for near field microscopy applications

Another extension to the time multiplexing approach can be coherence (lateral or axial) [17,18] or polarization encoding [19] while the coherence state and the polarization state of the synthesized illumination are time varying to properly encode, in a different and a decodable way, each spatial degree of freedom of the input object.

Another interesting approach deals with wavelength multiplexing [20-22]. The basic concept includes illumination of a diffractive optical element (DOE) that directs different wavelengths to different spatial positions. The input object is placed at the position where the wavelengths' spreading is obtained. The various wavelengths perform mapping of the spatial information such that when the various spatial degrees of freedom are summed all together (e.g. when the 2D image is transmitted through a single mode fiber), the mixed information can still be separated using similar DOE which will redirect again different wavelength each having its own intensity (that corresponds to the transmission of the input object in the spatial location that was illuminated by that specific wavelength) to the proper spatial position. As second step



(a).



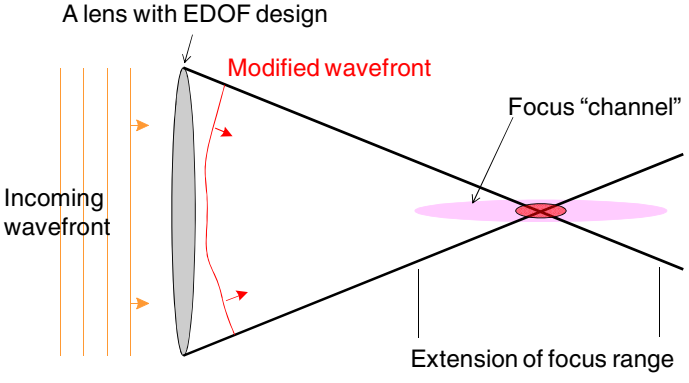
(b).

Fig. 3. Wavelength multiplexing super resolving approaches. (a). Wavelength multiplexing based microscope without objective lens. (b). Wavelength multiplexing based microscope with non periodic holes array. In part (c) of the figure one may see the spread of color generated by the nano holes array.

instead of using the decoding DOE which is identical to the encoding one, we may use a spectrometer which does similar procedure of extracting the intensity of all the wavelengths [23]. Since each wavelength corresponds to a different spatial position one can extract the 2D high resolution image. This technique was applied for microscopy application. It schematic sketch is seen in Fig. 3(a). In that configuration a microscope without an objective lens was realized.

One interesting derivation includes replacing the encoding DOE with non periodic metallic nano holes array [24]. This array has holes at different periods that are designed in such a way such that from every hole a slightly different wavelength is radiating. The inspected sample is placed in a proximity to the encoding holes array. The decoding as before can be done with a spectrometer. The advantage of using the nano holes array is that first although the holes are small (being sub wavelength holes) the energetic efficiency defined as the ratio between the illuminated radiation and the

radiation that goes through the holes is close to 100%. Second, this encoding element significantly reduces the volume of the proposed super resolving system where the distance between the encoding element and the inspected sample can be a few microns. Third, this element allows in a very simple way to obtain 2D, and not only 1D, wavelengths encoding and consequently 2D super resolved reconstruction. The experimental setup with the experimental results showing the spatial spread of wavelengths can be seen in Fig. 3(b).



(a).

(b).

Fig. 4. Extended depth of focus. (a). The schematic description of the operation principle of interference based all-optical EDOF approach. (b). Experimental results for the interference based EDOF approach: Comparison between results obtained without the EDOF element (left) and with the element (right).

Extended depth of focus (EDOF) is an important derivation of the field of super resolution while it is basically being an axial super resolving approach which is directly related to the obtainable lateral resolution limit. There are many approaches attempting to extend the focal depth either by providing refractive solutions [25-27] or diffractive based techniques [28]. Most of the existing approaches are also not all optical approaches and thus they require digital post processing to extract the high resolution information.

One new all-optical direction for extending the depth of focus is based upon interference [29,30]. The imaging lens basically sums all the rays passing through it in its focal spot. By controlling the relative phase at which all the rays are summed together one may realize a constructive interference in a focus "channel" as marked in Fig. 4(a) and a disruptive interference around this channel yielding an extension of the focal range. The control of the phases of the various rays passing through the aperture of the lens is done by engraving spatially gradually changing profile on the surface of the lens. Since the required range for the phase addition is $0\text{-}2\pi$ the engraving is very thin (less than a micron). The spatial change is so gradual such that the diffracted energy directed outside the zero diffraction order is negligible (unlike in diffractive solutions). An experimental example of the proposed approach may be seen in Fig. 4(b) where one may clearly see the increase in the focal range. The next stage, since we are dealing with an all-optical EDOF solution, was to try this concept for ophthalmic applications where the proposed technology was embedded in spectacles, contact lenses and intra ocular lenses [31-33]. The first two ophthalmic devices were tested in clinical trials while the last one was tested in an optical bench following the model of the human eye. The proposed concept was demonstrated to show significant improvement for subjects having presbyopia (the lack of accommodation capability of the human eye) and astigmatism as well as for myopia and hyperopia. Maximal improvement of close to 3.00 Diopters in the depth of focus was demonstrated over the optical bench as well as in the clinical trials.

4 Geometrical Type of Super Resolution

As previously mentioned, the geometric type of resolution reduction is obtained due to spatial under sampling which can be solved by a procedure called micro-scanning [34,35] in which a plurality of images are captured while having relative sub pixel shifts between each other. A more fundamental limitation is related to the fact that the sampling is not ideal since each pixel performs spatially localized averaging rather than functioning as a Dirac delta (as required by an ideal sampling).

To overcome the second limitation two basic concepts may be used. At first a periodic mask can be positioned in the intermediate image plane or to be attached to the detector or to be projected over the inspected object [36,37]. The period of the encoding pattern is the pixels pitch and it basically reshapes the spatial responsivity of each pixel such that its point spread function (PSF) can be deconvolved (i.e. to be inverse filtered in the Fourier domain). An example of applying such an approach to microscopy may be seen in Fig. 5. The upper and lower rows are two different objects. In the left column one may see the high resolution reference objects capture with a detector having small pixels, in the middle column the low resolution readout

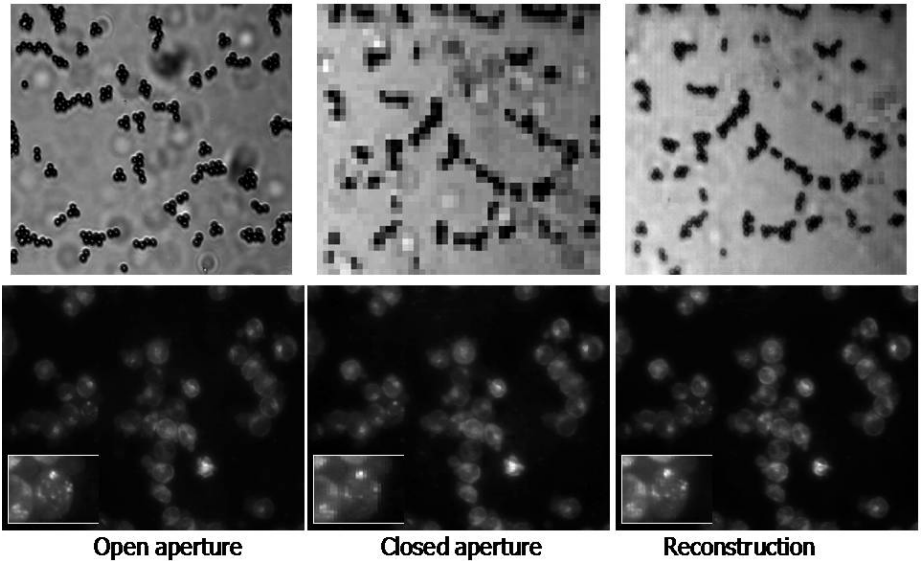


Fig. 5. Experimental results with projected periodic structure to obtain geometric super resolution

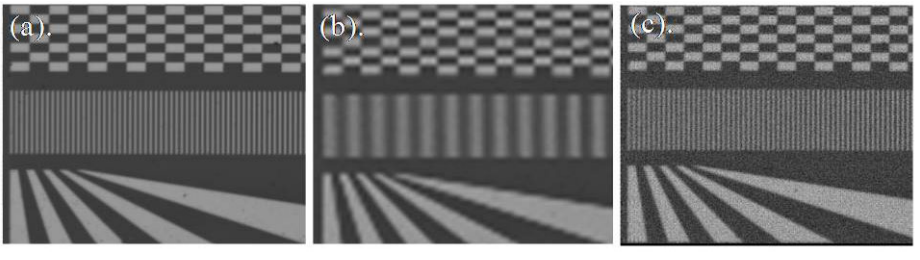


Fig. 6. Results for geometric super resolution obtained with non periodic encoding mask positioned in the intermediate image plane

obtained with a detector with large pixels and in the right column the high resolution reconstruction (obtained from a detector with the large pixels while applying the above mentioned projection approach) which resembles the high resolution reference.

A different concept can include treating the entire image as one by positioning of a random non periodic binary mask (once again in the intermediate image plane) [38]. This mask adds equations since in a priori known spatial positions the transmission should be zero due to the added mask. By having additional equations and by expressing the loss of resolution in a linear algebra representation, in which the readout vector (represents the low resolution captured image) equals to a matrix (representing the resolution reduced imaging system) multiplied by the high resolution vector (the original high resolution information), a reconstruction may be obtained. The reconstruction is obtained by inverting the matrix and obtaining the original high resolution image (expressed as a vector). Due to the additional equations

added by the mask, the matrix becomes an invertible matrix. An example of such an approach is seen in Fig. 6 where in Fig. 6(a) we present the high resolution reference, in Fig. 6(b) the captured low resolution image and in Fig. 6(c) the high resolution reconstruction.

5 Sensing with Light

Light can also be used not only to see better but also to sense other important information. For instance one application is usage of tracking of the trajectories of secondary speckle patterns, which are generated when an object is being illuminated, in order to compute the 3D movement profile of the object [39]. As an experimental

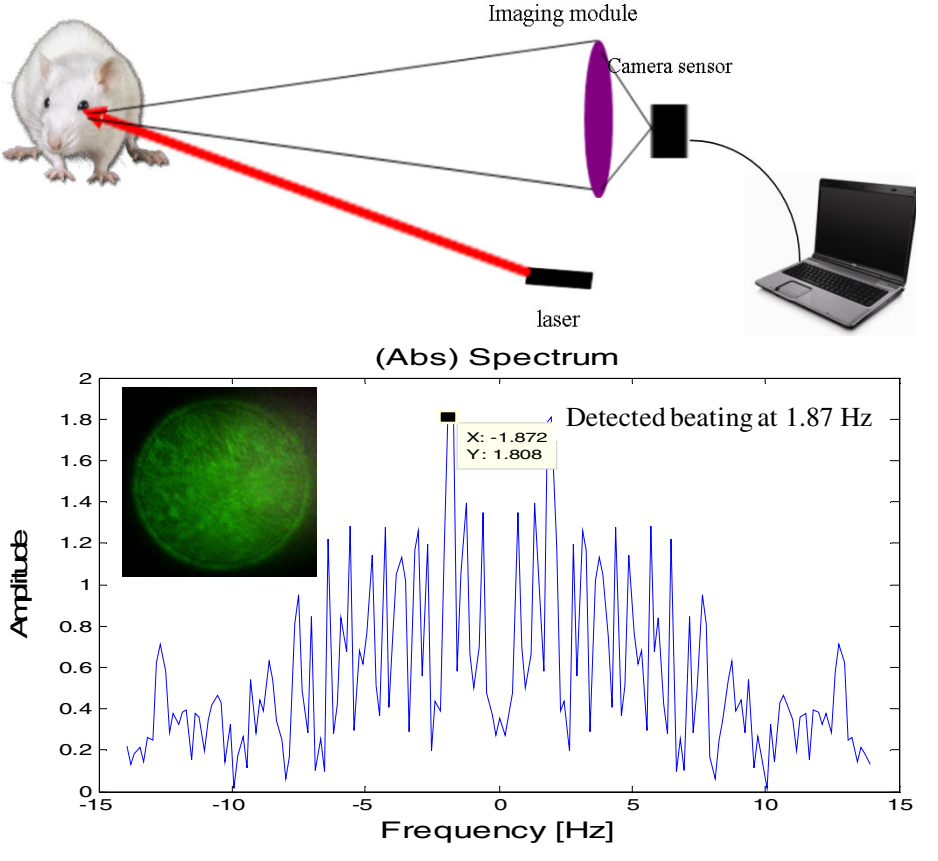


Fig. 7. Usage of reflected secondary speckle patterns to track 3D movement with nano metric accuracy. In this experiment we track the breathing of a mouse by measuring the reflections from its cornea.

example, in Fig. 7 we illuminate the cornea of a mouse and perform image processing that tracks the movement of the reflected secondary speckle patterns. From the performed tracking we extract a beating rate of 1.87Hz which corresponds to the breathing of the mouse.

In the upper part of Fig. 7 we can see the schematic sketch of the experimental setup and in the lower part the spectrum of the movement of the reflected speckle patterns. In the left upper corner we see one of the speckle images captured as part of the set of images that were analyzed to extract the 1.87Hz beating.

6 Conclusions

In this paper we have presented a short overview of various applicable techniques used to overcome the diffraction and the geometric related resolution limitations. We also showed that light can be used not only in order to see better but also in order to monitor breathing. Further applications are constantly being developed in this vivid field of research.

References

1. Abbe, E.: "Beitrage zur theorie des mikroskops und der mikroskopischen wahrnehmung. Arch. Mikrosk. Anat. 9, 413–468 (1873)
2. Zalevsky, Z., Mendlovic, D.: Optical Super Resolution. Springer, New-York (2002)
3. Zalevsky, Z., Mendlovic, D., Lohmann, A.W.: Progress in Optics. In: Wolf, E. (ed.) Optical System with Improved Resolving Power, vol. XL, Elsevier, Amsterdam (1999)
4. Lohmann, A.W., Dorsch, R.G., Mendlovic, D., Zalevsky, Z., Ferreira, C.: About the space bandwidth product of optical signal and systems. J. Opt. Soc. Am. A 13, 470–473 (1996)
5. Mendlovic, D., Lohmann, A.W.: Space-bandwidth product adaptation and its applications to superresolution: fundamentals. J. Opt. Soc. Am. A 14, 558–562 (1997)
6. Mendlovic, D., Lohmann, A.W., Zalevsky, Z.: Space-badnwidth– product adaptation and its application for superresolution - examples. J. Opt. Soc. Am. A 14, 563–567 (1997)
7. Francon, M.: Amelioration de resolution d'optique. Nuovo. Cimento. Suppl. 9, 283–290 (1952)
8. Lukosz, W.: Optical systems with resolving powers exceeding the classical limits. J. Opt. Soc. Am. 56, 1463–1472 (1967)
9. Mendlovic, D., Lohmann, A.W., Konforti, N., Kiryuschev, I., Zalevsky, Z.: One dimensional superresolution optical system for temporally restricted objects. Appl. Opt. 36, 2353–2359 (1997)
10. Mendlovic, D., Kiryuschev, I., Zalevsky, Z., Lohmann, A.W., Farkas, D.: Two dimensional superresolution optical system for temporally restricted objects. Appl. Opt. 36, 6687–6691 (1997)
11. Garcia, J., Zalevsky, Z., Fixler, D.: Synthetic aperture superresolution by speckle pattern projection. Opt. Exp. 13, 6073–6078 (2005)
12. Fixler, D., Garcia, J., Zalevsky, Z., Weiss, A., Deutsch, M.: Speckle Random Coding for 2-D Super Resolving Fluorescent Microscopic Imaging. Micron. 38, 121–128 (2007)
13. Shemer, A., Mendlovic, D., Zalevsky, Z., Garcia, J., Martinez, P.G.: Super resolving optical system with time multiplexing and computer decoding. Appl. Opt. 38, 7245–7251 (1999)

14. Zalevsky, Z., Garcia, J., Ferreira, C.: Super Resolved Imaging of Remote Moving Targets. Opt. Let. 31, 586–588 (2006)
15. Zalevsky, Z., Saat, E., Orbach, S., Mico, V., Garcia, J.: Exceeding the resolving imaging power using environmental conditions. Appl. Opt. 47, A1–A6 (2008)
16. Gur, A., Fixler, D., Micó, V., Garcia, J., Zalevsky, Z.: Linear optics based nanoscopy. Opt. Exp. 18, 22222–22231 (2010)
17. Zalevsky, Z., Garcia, J., Garcia-Martinez, P., Ferreira, C.: Spatial information transmission using orthogonal mutual coherence coding. Opt. Let. 20, 2837–2839 (2005)
18. Mico, V., García, J., Ferreira, C., Sylman, D., Zalevsky, Z.: Spatial Information Transmission Using Axial Temporal Coherence Coding. Opt. Lett. 32, 736–738 (2007)
19. Sylman, D., Zalevsky, Z., Mico, V., Garcia, J.: Super-Resolved or Field of View Enlarged Imaging based upon Spatial Depolarization of Light. Opt. Commun. 283, 1715–1719 (2010)
20. Kartashev, A.I.: Optical systems with enhanced resolving power. Opt. Spectrosc. 9, 204–206 (1960)
21. Armitage, J.D., Lohmann, A.W., Parish, D.P.: "Superresolution image forming systems for objects with restricted lambda dependence. Jpn. J. Appl. Phys. 4, 273–275 (1965)
22. Alexandrov, S.A., Sampson, D.D.: "Spatial information transmission beyond a system's diffraction limit using optical spectral encoding of the spatial frequency. J. Opt. A: Pure Appl. Opt. 10, 25304 (2008)
23. Schwarz, A., Weiss, A., Fixler, D., Zalevsky, Z., Micó, V., García, J.: One-Dimensional Wavelength Multiplexed Microscope without Objective Lens. Opt. Commun. 282, 2780–2786 (2009)
24. Gur, A., Aharoni, R., Zalevsky, Z., Garini, Y., Mico, V., Garcia, J.: Lensless Superresolved Microscopy based on Sub-Wavelength Non-Periodic Holes Array Plate, High and Super Resolution Imaging (SHRI) Conference in Lipica, Slovenia (September 2009)
25. Sauceda, A., Ojeda-Castaneda, J.: High focal depth with fractional-power wavefronts. Opt. Lett. 29, 560–562 (2004)
26. Chi, W., George, N.: Electronic imaging using a logarithmic asphere. Opt. Lett. 26, 875–877 (2001)
27. Dowski, E.R., Cathey, W.T.: Extended depth of field through wave-front coding. Appl. Opt. 34, 1859–1866 (1995)
28. Ben-Eliezer, E., Zalevsky, Z., Marom, E., Konforti, N.: All-optical extended depth of field imaging system. J. Opt. A: Pure Appl. Opt. 5, S164–S169 (2003)
29. Zalevsky, Z., Shemer, A., Zlotnik, A., Ben-Eliezer, E., Marom, E.: All-optical axial super resolving imaging using low-frequency binary-phase mask. Opt. Express 14, 2631–2643 (2006)
30. Raveh, I., Zalevsky, Z.: All-optical axially multi-regional super resolved imaging. Opt. Express 15, 17912–17921 (2007)
31. Zalevsky, Z., Ben Yaish, S., Yehezkel, O., Belkin, M.: Thin spectacles for myopia, presbyopia and astigmatism insensitive vision. Opt. Express 15, 10790–10803 (2007)
32. Zlotnik, A., Ben Yaish, S., Yehezkel, O., Lahav-Yacouel, K., Belkin, M., Zalevsky, Z.: Extended Depth of Focus Contact Lenses for Presbyopia. Opt. Lett. 34, 2219–2221 (2009)
33. Ben Yaish, S., Zlotnik, A., Raveh, I., Yehezkel, O., Belkin, M., Zalevsky, Z.: Intra-Ocular Omni-Focal Lens with Increased Tolerance to Decentration and Astigmatism. Journal of Refractive Surgery 26(1) (January 2010)
34. Fortin, J., Chevrette, P., Plante, R.: Evaluation of the microscanning process. In: Proc. SPIE, vol. 2269, pp. 271–279 (1994)

35. Borman, S., Stevenson, R.: Super-resolution from image sequences - A review. In: Proc. Midwest Symposium on Circuits and Systems, pp. 374–378 (1998)
36. Zalevsky, Z., Mendlovic, D., Marom, E.: Special sensor masking for exceeding system geometrical resolving power. Opt. Eng. 39, 1936–1942 (2000)
37. Fixler, D., Garcia, J., Zalevsky, Z., Weiss, A., Deutsch, M.: Pattern projection for subpixel resolved imaging in microscopy. Micron. 38, 115–120 (2007)
38. Borkowski, A., Zalevsky, Z., Javidi, B.: Geometrical super resolved imaging using non periodic spatial masking. J. Opt. Soc. Am. A 26, 589–601 (2009)
39. Zalevsky, Z., Beiderman, Y., Margalit, I., Gingold, S., Teicher, M., Mico, V., Garcia, J.: Simultaneous remote extraction of multiple speech sources and heart beats from secondary speckles pattern. Opt. Express. 17, 21566–21580 (2009)

Author Index