ICQNM 2015

The Ninth International Conference on Quantum, Nano/Bio, and Micro Technologies

ISBN: 978-1-61208-431-2

August 23 - 28, 2015

Venice, Italy

ICQNM 2015 Editors

Vladimir Privman, Clarkson University - Potsdam, USA

Victor Ovchinnikov, Aalto University, Finland

ICQNM 2015

Foreword

The Ninth International Conference on Quantum, Nano and Micro Technologies (ICQNM 2015), held between August 23-28, 2015 in Venice, Italy, continued a series of events covering particularly promising theories and technologies. The conference covered fundamentals on designing, implementing, testing, validating and maintaining various kinds of materials, systems, techniques and mechanisms related to quantum-, nano- and micro-technologies.

Quantum technologies and nano technologies have a great potential to transform communications telecommunications infrastructure and communication protocols, and computers and networking devices. Nanotechnologies and micro-technologies already made their mark on smart materials, nano-medicine, nano-devices, molecular manufacturing, biotechnology, metrology, airspace.

The advancements in material science and computer science have allowed the building, launching and deploying of space exploration systems that continually do more and more as they become smaller and lighter. As an example, carbon nano-tubes have been created that are 250 times stronger than steel, 10 times lighter, and transparent. Similar advances are occurring in glass, plastics and concrete. Spacecraft are being launched, with hulls that are composed of carbon fibers, a light weight high strength material.

Electronic devices, medicine, environment, metrology, aerospace programs, clothes and materials, telecommunications, cryptography, semiconductors, manufacturing, and other domains are impacted by the progress on the areas mentioned above. Particularly, micro imaging, nano-medicine: (drug delivery; nano-particles i.e. viruses; proteins.), bio-nanostructures: (nano-tubes, nano-particles), microsystems, micro fluidics: (including nano-fluidics, modeling; fabrication and application), micro instrumentation / implantable microdevices (miniaturized bio-electronic systems, etc.) and micro sensors benefits from the progress on quantum, nano and micro technologies.

Developing nanoscale-manufactured robots presents fabrication and control challenges. The evolution of mechatronics system and robotic system requires advanced functions for control. Special methods and technologies have been developed to design, analyze, build, controls, and apply micro/nano-robotic systems for biotechnology, medical, information technology, materials, etc. A particular application of nano-robots would be in carrying out projects in hostile environments, utilizing local materials and local energy. Ultra-miniature robotic systems and nano-mechanical devices will be the biomolecular electro-mechanical hardware of future manufacturing and biomedical industry.

Nowadays, there are tremendous attempts to develop new bio-molecular machines, components that can be assembled in nano-devices. Bio-robotics entities are able to manipulate the nano-world components, convey information from the nano/nano to the nano/macro world and navigate at the nano-environment level. Additionally, they are able to self replicate, leading to the bio-robot factory. Protein-based nano-motors and nano-robots, as well as biomolecular components interfaces.

Quantum cryptography uses the uncertainty principle of quantum physics to provide a safe but public means for transmitting vital, secret information. A quantum public key distribution system depends on the uncertainty principle to ensure secrecy. Special protocols correlations and composability algorithms ensure similar functionality as in non-quantum systems. The security related tracks cover a series of events focusing on quantum security aspects. On the quantum protocol side, automated proofs of security and probabilistic model-checking methods have been suggested. Research teams focus on quantum key distribution and aspects related to key composability and correlations. Limitations are mainly related to physical devices and polarization control.

We take here the opportunity to warmly thank all the members of the ICQNM 2015 Technical Program Committee, as well as the numerous reviewers. The creation of such a high quality conference program would not have been possible without their involvement. We also kindly thank all the authors who dedicated much of their time and efforts to contribute to ICQNM 2015. We truly believe that, thanks to all these efforts, the final conference program consisted of top quality contributions.

Also, this event could not have been a reality without the support of many individuals, organizations, and sponsors. We are grateful to the members of the ICQNM 2015 organizing committee for their help in handling the logistics and for their work to make this professional meeting a success.

We hope that ICQNM 2015 was a successful international forum for the exchange of ideas and results between academia and industry and for the promotion of progress in quantum, nano and micro technologies.

We are convinced that the participants found the event useful and communications very open. We hope Venice provided a pleasant environment during the conference and everyone saved some time for exploring this beautiful city.

ICQNM 2015 Chairs:

ICQNM Advisory Chairs
Vladimir Privman, Clarkson University - Potsdam, USA
Christian Kollmitzer, AIT Austrian Institute of Technology GmbH, Austria
Victor Ovchinnikov, Aalto University, Finland

ICQNM 2015 Research/Industry Chairs
Marco Genovese, Italian Metrological Institute (INRIM) -Torino, Italy
Keiji Matsumoto, National Institute of Informatics, Japan

ICQNM 2015 Special Area Chairs
QSEC
Masahito Hayashi, Nagoya University, Japan
Fluidics
Alireza Azarbadegan, University College London (UCL), UK
Quantum algorithms and quantum complexity
Francois Le Gall, The University of Tokyo, Japan

Quatum control

Daoyi Dong, University of New South Wales, Australia

ICQNM 2015

Committee

ICQNM Advisory Chairs

Vladimir Privman, Clarkson University - Potsdam, USA
Christian Kollmitzer, AIT Austrian Institute of Technology GmbH, Austria
Victor Ovchinnikov, Aalto University, Finland

ICQNM 2015 Research/Industry Chairs

Marco Genovese, Italian Metrological Institute (INRIM) -Torino, Italy
Keiji Matsumoto, National Institute of Informatics, Japan

ICQNM 2015 Special Area Chairs

QSEC
Masahito Hayashi, Nagoya University, Japan

Fluidics
Alireza Azarbadegan, University College London (UCL), UK

Quantum algorithms and quantum complexity
Francois Le Gall, The University of Tokyo, Japan

Quantum control
Daoyi Dong, University of New South Wales, Australia

ICQNM 2015 Technical Program Committee

Andrew Adamatzky, University of the West of England - Bristol, U.K.
Gerardo Adesso, University of Nottingham, UK
Irina Buyanova, Linkoping University, Sweden
Weimin M. Chen, Linköping University, Sweden
Taksu Cheon, Kochi University of Technology - Tosa Yamada, Japan
Sang H. Choi, NASA Langley Research Center, USA
Mihaela Corneanu, Banat's University of Agricultural Sciences and Veterinary Medicine, Romania
Sorin Cotofana, TU Delft, The Netherlands
Brahim Dennai, ENERGARID Laboratory, France
Sao-Ming Fei, Capital Normal University - Beijing, China
Akihiko Fujiwara, Japan Synchrotron Radiation Research Institute - Hyogo, Japan
Juan Carlos García-Escartín, Universidad de Valladolid, Spain
Yuval Gefen, The Weizmann Institute of Science, Israel
Marco Genovese, Italian Metrological Institute (INRIM) -Torino, Italy
Bonnie Gray, Simon Fraser University, Canada
Masahito Hayashi, Nagoya University, Japan

Copyright Information

Table of Contents

Naturally Inspired SERS Substrate Properties of Silver Nanoparticles Deposited on TiO_2-Coated Insect Wings

Ichiro Tanahashi and Yoshiyuki Harada
Nanomaterials and Microdevices Research Center
Osaka Institute of Technology
Osaka, Japan
e-mail: ichiro.tanahashi@oit.ac.jp, yoshiyuki.harada@oit.ac.jp

Abstract— **Ag nanoparticles were photocatalytically deposited on TiO_2-coated insect wings (Ag/TiO_2-coated insect wings). Wings of dragonfly, cicada and butterfly with different micro/nanostructured surfaces were investigated as substrates based on natural materials. We have investigated the surface enhanced Raman scattering (SERS) substrate properties of the six types of Ag/TiO_2-coated insect wings. The size and shape of the deposited Ag nanoparticles were influenced by the surface structures of the insect wings. The SERS signal intensities of Rhodamine 6G adsorbed on the Ag/TiO_2-coated insect wings with spherical-shaped Ag nanoparticles of 100-250 nm diameter and flake-shaped Ag nanoparticles were more than one order of magnitude larger than those on the Ag thin film grown by sputtering on glass slides.**

Keywords-metal nanoparticle; silver; SRES; insect; wing.

I. INTRODUCTION

In recent years, surface-enhanced Raman scattering (SERS) has attracted much attention because of its application in the detection of trace amount of chemical and biological agents. As the SERS-active materials, roughened metal films and metal nanoparticles have been investigated [1][2]. In the case of metal nanoparticles, the local electromagnetic field enhancement induced by the excitation of the localized surface plasmon resonance (LSPR) is the major mechanism for SERS. It is known that the size and shape of the metal nanoparticles affect their LSPR properties. In particular, silver (Ag) nanoparticles with clear LSPR absorption are the most suitable candidate for SERS-active materials. However, the appropriate size and shape of the Ag nanoparticles for SERS-active materials are still not clear.

The size and shape of the deposited metal nanoparticles on the substrates are also largely affected by the surface structures of the base substrates. The artificial base substrates such as glasses, ceramics and plastics are widely used. On the other hand, some biological surfaces of insect wings are structured at the micro/nanometer scale. Thus, scientists have been trying to develop the new types of SERS-active substrates using natural biological materials such as cicada and butterfly wings [3][4]. In our previous studies [5], we have reported the preparation and SERS properties of the Ag nanoparticles on TiO_2-coated cicada wings with nanopillar array structures. However, it is not clear what types of micro/nanostructured surfaces of insect wings are more suitable for the naturally inspired SERS-active substrates. In this paper, we have reported the characterization, optical and SERS properties of the six types of Ag/TiO_2-coated insect wings.

II. EXPERIMENTAL DETAILS

We used three species of dragonflies, two species of butterflies and one species of cicada. All the investigated insects were captured in the Kansai region of Japan. Wing samples were collected from *Sympetrum darwinianum* (Summer Darter, small to medium-sized skimmer dragonfly with clear and transparent wings): denoted as dragonfly (a), *Mnais costalis* (small-sized one with orange color wings): dragonfly (b), *Mnais pruinosa* (small-sized one with clear and transparent wings): dragonfly (c), *Panantica sita* (Chestnut Tiger, the Danaid group butterfly, black and bluish-white subhyaline markings on the forewings): butterfly (d), *Parnassius citrinarius* (the Parnassiinae one, white and subhyaline markings on the forewings): butterfly (e) and *Cryptotympana facialis* (a black cicada with clear and transparent wings): cicada (f). The male dorsal forewings of the samples were used in the experiments.

The detailed preparation processes of the Ag/TiO_2-coated insect wings have been described previously [5]. By using TiO_2 (Ishihara Sangyo Kaisha, ST-K211) as a photocatalyst, Ag metal was produced by photoreduction of Ag^+ ions.

Photographs of the insect specimens and prepared samples were taken using a digital camera. Scanning electron microscopy (SEM) analysis, X-ray diffraction (XRD) and UV-Vis absorption spectra measurements were carried out to characterize the bare insect wings and Ag/TiO_2-coated insect wings.

SERS measurements were performed by irradiating the sample with 50 mW of 514.5 nm line of an Ar^+ laser in back scattering geometry at room temperature. The laser beam was focused on a spot with a diameter of 2 μm and the data acquisition time was 1 s. The intensities of Raman spectra of

10^{-3} mol L^{-1} Rhodamine 6G (R6G, 2 μL) adsorbed on various samples were compared. As a reference, R6G (2 μL) adsorbed on a mirror-like smooth surface Ag thin film prepared by an RF magnetron sputtering system was used.

III. Results and Discussion

In the SEM micrographs, the wings of the dragonfly (a), (b) and (c) showed sub-micrometer-scale relief structures. The pitch of the concavo-convex pattern of the dragonfly (a), (b) and (c) increased in that order. The wing membrane (bluish-white subhyaline marking part) of the butterfly (d) had a wrinkle-like irregular shape surface. A small number of scales were observed on the bluish-white subhyaline markings. The wing scales of the butterfly (e) possessed uniform ordered ridge structure. The wings of the cicada (f) had a dense nanopillar array structure [5].

In all the investigated six types of Ag/TiO_2-coated insect wings, the color of the wings was changed to metallic gray after the UV irradiation indicating the formation of Ag metal on the insect wings. In the case of the Ag/TiO_2-coated insect wings of the dragonfly (a) shown in Fig. 1, (c) and cicada (f), densely stacked Ag nanoparticles with 100-250 nm in diameter were seen. The mean diameter of Ag nanoparticles of the Ag/TiO_2-coated insect wing of the dragonfly (c) was smaller than those of the dragonfly (a) and cicada (f). In the case of the Ag/TiO_2-coated insect wings of the butterfly (d) (membrane) and dragonfly (b) shown in Fig. 1, flake-shaped Ag nanoparticles with various sizes were seen.

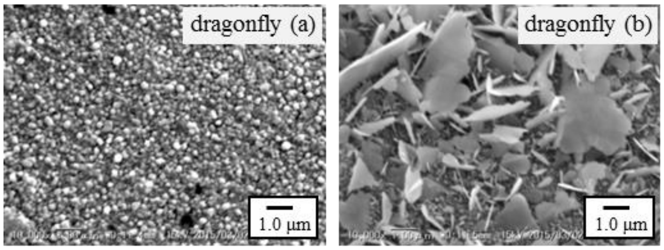


Figure 1. SEM micrographs of Ag/TiO_2-coated insect wings of the dragonfly (a) and (b).

On the other hand, the uniform ordered ridge structures of the Ag/TiO_2-coated insect wings of the butterfly (e) (scales) were covered with the Ag films and small number of nanoparticles. Among the investigated six types of insect wings, the size and shape of the Ag nanoparticles deposited on the TiO_2-coated insect wings were largely influenced by the surface structures such as sub-micrometer-scale relief and nanopillar array ones.

In the XRD patterns of the six types of Ag/TiO_2-coated insect wings, it was seen that the peak at $2\theta=38.1^{\circ}$ which was assigned to the (111) reflection line of cubic Ag. From the results of the XRD, the TiO_2-coated insect wings were successfully covered by the metallic Ag.

In the optical absorption spectra of the Ag/TiO_2-coated insect wings of the dragonfly (a), (c) and cicada (f), the broad absorption band peaking at about 430 nm was observed. The band was due to the LSPR absorption of the Ag nanoparticles. These results were in accordance with the SEM observations. Thus, the Ag nanoparticles could be made on the TiO_2-coated sub-micrometer-scale relief and nanopillar array structured wings by the photoreduction.

SERS spectra of R6G adsorbed on the six types of Ag/TiO_2-coated insect wings and the Ag thin film sputtered on a glass slide were measured. In the SERS measurements, R6G (10^{-3} mol L^{-1}) as standard remarks was dripped and dried on the surface of the Ag/TiO_2-coated insect dorsal forewings and the Ag thin film. In all the SERS spectra, the distinct peaks and a broad band from 600 to 1800 cm^{-1} were seen. The observed Raman peaks were characteristic of R6G [6]. The peak intensities at ca. 1649 cm^{-1}（the C-C stretching mode) of the six types of Ag/TiO_2-coated insect wings were compared with that of the Ag thin film. The intensities of the Ag/TiO_2-coated insect wings of dragonfly (c), butterfly (d) and (e) were about 1-4 times larger than that of the Ag thin film. On the other hand, the intensities of the Ag/TiO_2-coated insect wings of dragonfly (a), (b) and cicada (f) were respectively about 15, 10 and 10 times larger than that of the Ag thin film.

The Ag/TiO_2-coated insect wings with large-area, low-cost and high-SERS performance seem to be a promising candidate for the naturally inspired SERS-active substrates.

References

[1] Y. Wu, B. Zhao, W. Xu, B. Li, Y. M. Jung, and Y. Ozaki, “Near-Infrared Surface-Enhanced Raman Scattering Study of Ultrathin Films of Azobenzene-Containing Long-Chain Fatty Acids on a Silver Surface Prepared by Silver Mirror and Nitric Acid Etched Silver Foil Methods,” Langmuir, 1999, **15**, pp. 4625-4629.

[2] W. Xie, P. Qui, and C. Mao, “Bio-Imaging, Detection and Analysis by Using Nanostructures as SERS Substrates,” J. Mater. Chem., 2011, **21**, pp. 5190-5202.

[3] P. R. Stoddart, P. J. Cadusch, T. M. Boyce, R. M. Erasmus, and J. D. Comins, “Optical Properties of Chitin: Surface-Enhanced Raman Scattering Substrates Based on Antireflection Structures on Cicada Wings,” Nanotechnology, 2006, **17**, pp. 680-686.

[4] Y. Tan, et al., “Morphological Effects on Surface-Enhanced Raman Scattering from Silver Butterfly Wing Scales Synthesized via Photoreduction,” Langmuir, 2011, **27**, pp. 11742-11746.

[5] I. Tanahashi and Y. Harada, “Naturally Inspired SERS Substrates Fabricated by Photocatalytically Depositing Silver Nanoparticles on Cicada Wings,” Nanoscale Res. Lett., 2014, **9**, pp. 1/298-5/298.

[6] Y. Lu, G. L. Liu, and L. P. Lee, “High-Density Silver Nanoparticle Film with Temperature-Controllable Interparticle Spacing for a Tunable Surface Enhanced Raman Scattering Substrate,” Nano. Lett., 2005, **5**, pp. 5-9.

Competitive Mechanisms of Resistive Switching in Nanooxide Based Memory Cells

Dmitriy V. Stremous[1], Aleksandr L. Danilyuk[2], Denis A. Podryabinkin[3], Victor E. Borisenko[4]

Belarusian State University of Informatics and Radioelectronics

P. Browka 6, 220013, Minsk, Belarus

e-mails: stremous@list.ru[1], danilyuk@nano-center.org[2], arm@tut.by[3], borisenko@bsuir.by[4]

***Abstract* — Atomic migration and electronic switching of bi-stable centers in conducting filaments formed in nanooxide based resistive random access memory (RRAM) cells are modeled and analyzed as competitive mechanisms determining their operation frequency. They are demonstrated to be mediated by the filament growth dynamics. Atomic migration is shown to be responsible for a slow change of the filament resistivity with typical switching times in the millisecond range. Fast switching with the shortest nanosecond delay can be achieved using bi-stable electronic centers in the filaments. Possible configurations of such centers are discussed.**

Keywords - resistive switching; nanooxide; memory cell.

I. INTRODUCTION

Currently, metal nanooxide-based resistive switching memory is being studied extensively as one of the most competitive candidates for non-volatile memory applications because of its simple structure, rapid switching and excellent scalability. The mechanisms of resistive switching phenomena in oxides can be quite diverse. One possible mechanism is electronic switching, originating from a trap/de-trap process through defects in the oxides [1]. Another possible mechanism is ionic switching, which is usually attributed to the formation/rupture of conductive filaments (CFs), which may consist of oxygen vacancies or metal precipitates [2][3]. It is illustrated in Figure 1.

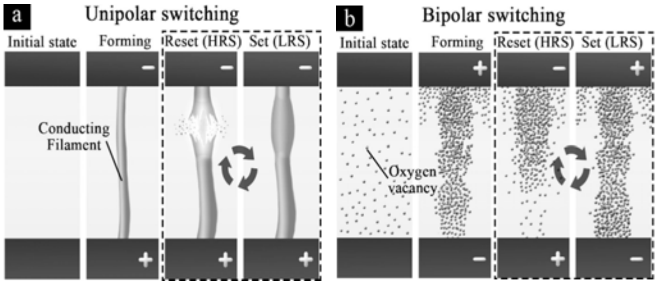


Figure 1. Dynamic growth and formation/rupture processes of filament-type resistive switching mechanism in unipolar (a) and bipolar switching (b) modes [3].

The process of CFs formation and their transition from an initial high resistance state (HRS) to a low resistance state (LRS) are interpreted as a dielectric soft breakdown associated with the migration of oxygen ions toward the anode, and the formation of CFs in the bulk oxide connecting both electrodes. In the unipolar reset process where the reset occurs at the same polarity as the set, joule-heating-assisted diffusion of oxygen ions from the anode occurs and the surrounding oxides rupture the CFs by recombining with oxygen vacancies or by the re-oxidation of metal precipitates. Cooling down and formation of a functional region in the CFs occur during the set process (Figure 1). In the bipolar reset process that occurs at the polarity opposite to the set process, oxygen ions drifting in the applied electric field rupture the CFs.

Combined diffusion and drift ion transport may be considered to be a driving mechanism of the slow switching [4]. Fast electronic switching is provided by the synchronized process of the capture and release of mobile charge carriers on the bi-stable trapping states in the regions of the formation/rupture of CFs. These states are caused by the presence of oxygen vacancies. Both atomic and electronic mechanisms are modeled and discussed in this paper.

The dynamic model of CFs formation and the switching model of bi-stable trapping states in the oxide are described in Section II. Section III presents results of the computer simulations based on the above models for HfO_2-based metal/insulator/metal nanostructures with the emphasis to their application for resistive random access memory (RRAM) cells.

II. MODELS

The dynamics of CFs formation are considered to be determined by the motion of their constituent ions in the electric field. The growth rate of the filament is defined by the drift speed of the moving ions. Supposing hope jumping of the ions over the energy barrier U it is [5]:

$$dl/dt = 2dv\exp\left(-qU/kT\right)\sinh\left(qVd/2kT(h-l)\right), \quad (1)$$

where l is the filament length, d is the hopping site distance, v is the characteristic ion hop attempt frequency, k is the Boltzmann's constant, T is the absolute temperature, q is the electron charge, h is the overall thickness of the oxide, V is the applied voltage.

In the model considering electronic states, switching of bi-stable traps in the oxide is supposed to be responsible for the interplay between its HRS and LRS. Such switching is induced by a thermal noise. We simulate it using the following equation [6]:

$$dy/dt = ay - by^3 + A\cos(\Omega t + \varphi) + \sqrt{2D}\xi(t), \quad (2)$$

where *y* is the generalized coordinate of the traps characterizing the type of its potential energy, *a* and *b* are the potential parameters, and *A*, Ω, and φ are the amplitude, frequency and the phase of the center of oscillations, respectively, *D* is the noise intensity, and the function $\xi(t)$ denotes a zero-mean Gaussian white noise.

A sub-circuit is used to calculate the switching threshold voltage and resistance of the resistive memory cell and its dependence on the sweep rate and operating temperature as it is described in [7].

III. Results and discussion

The length of CFs in HfO_2 was calculated to grow exponentially with time like it is shown in Figure 2a. The external bias determines the characteristic time of the growth. A significant increase of the growth rate is reached at the external bias of 2–2.5 V.

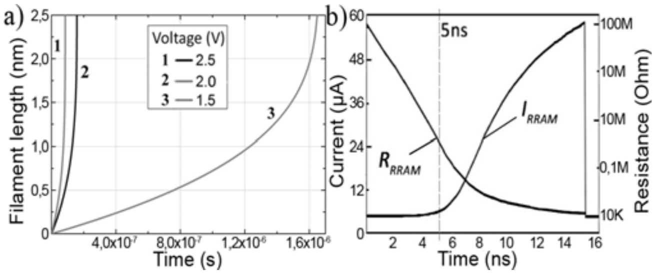


Figure 2. Dynamics of conductive filament growth in 5 nm HfO_2 under various applied voltages (a). RRAM dynamics during pulse programming (5 V) at 300 K (b).

SPICE Modeling of the dynamics of the CFs growth in HfO_2 during programming voltage pulse of 5 V that depends on the operating temperature shows that the RRAM cell does not switch on immediately. It takes approximately 4–6 ns. The switching event is determined by the sharp increase in the conductivity marked by the dotted line in Figure 2b. Switch-ON time depends exponentially of the external bias. This result shows an importance of the dynamic effects involving a range of transient processes.

As for atomic mechanisms of ON/OFF switching of the FCs formed, oxygen and/or metal atom diffusion for the distances comparable to the FCs diameters (few nanometers) was calculated to proceed within milliseconds even at the melting point of HfO_2.

Modeling of electronic switching of bi-stable traps was carried out for HfO_2 with the following parameters: the thermal ionization energy of the traps of 0.5 eV, the oscillation frequency of 10–12 GHz, the trap concentration of 10^{19} cm^{-3}, and the noise intensity of 0.08–0.15. The bi-stable potential of the traps under an influence of a weak periodic modulation was confirmed to make a transition from one state to another only under the effect of the noise.

Figure 3a shows that the impact of noise leads to the switching of the traps in HfO_2 from one metastable state to another within few nanoseconds. An increasing intensity of the noise stimulates variations in the output signal $y(t)$, causing formation of metastable states. The switching time decreases with an increase of the amplitude *A* and the impact frequency of the noise.

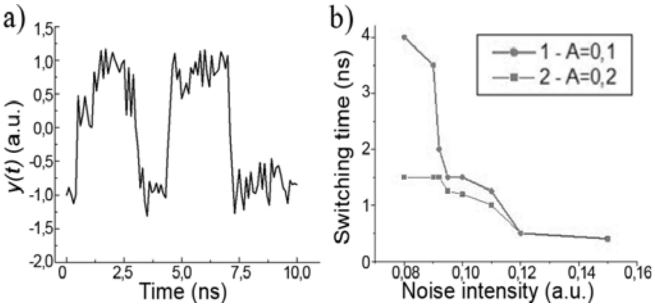


Figure 3. Switching dynamics (a) and the switching time *vs* noise intensity at different amplitudes noise amplitudes *A* (b).

Numerical estimations of the electron switching time of the trapping states showed it to be of about few nanoseconds. It can decrease below 1 ns at relatively high amplitudes and intensities of the noise (Figure 3b).

IV. Conclusion

The performed modeling shows that comprehensive understanding of the mechanisms of resistance switching in RRAM cells should include consideration of ion migration, the growth of conductive filaments and their rupture, as well as the capture and release of electrons including metastable trapping states. Experimental identification and numerical characterization of these states are important to increase operating frequencies of RRAMs.

Acknowledgment

The authors thank Belarusian Fund for Fundamental Researches for the financial support.

References

[1] M. A. Danilyuk, *et al.*, “Multiphonon ionization of traps formed in hafnium oxide by electrical stress,” Phys. Stat. Sol. A, vol 210, no 2, Feb. 2013, pp. 361–366, doi:1002/pssa.201228083.

[2] X. Wu, *et al.*, “Intrinsic nanofilamentation in resistive switching,” J. Appl. Phys. vol. 113, no. 11, Nov. 2013, pp. 114503 (6 pages), doi:dx.doi.org/10.1063/1.4794519.

[3] A. Mehonic, *et al.*, “Electrically tailored resistance switching in silicon oxide,” Nanotechnology, vol. 23, no. 45, Nov. 2012, pp. 455201 (9 pages), doi:10.1088/0957-4484/23/45/455201.

[4] F. Pan, S. Gao, C. Chen, C. Song, and F. Zeng, “Recent progress in resistive random access memories: Materials, switching mechanisms, and performance,” Mater. Sci. & Eng., vol. 83, Sep. 2014, pp. 1–59, doi:10.1016/j.mser.2014.06.002.

[5] A. Sawa, “Resistive switching in transition metal oxides,” Materials Today, vol. 11, no. 6, June 2008, pp. 28–36, doi:10.1016/S1369-7021(08)70119–6.

[6] L. Gammaitoni, P. Hänggi, P. Jung, and F. Marchesoni, “Stochastic resonance,” Rev. Mod. Phys., vol. 70, no. 1, Jan. 1998, pp. 223–287, doi:10.1103/RevModPhys.70.223.

[7] P. Sheridan, *et al.*, “Device and SPICE modeling of RRAM devices,” Nanoscale, vol. 3, no. 9, Aug. 2011, pp. 3833–3840, doi:10.1039/C1NR1.

Strong Visible Light Emission from Silicon Nanocrystals Embedded into a Silicon Carbide Film

Chul Huh, Tae-Youb Kim, Chang-Geun Ahn, and Bong-Kyu Kim
IT Convergence Technology Research Laboratory, Electronics and Telecommunications Research Institute (ETRI), Daejeon 305-700, Republic of Korea
E-mail: chuh@etri.re.kr

***Abstract*—We report the strong visible light emission from silicon (Si) nanocrystals (NCs) embedded in a Si carbide (SiC) film. Compared to Si NC light-emitting diode (LED) by employing the Si nitride (SiN_x) film as a surrounding matrix, the turn-on voltage of the Si NC LED with the SiC film was significantly decreased by 4 V. This was attributed to a smaller barrier height for injecting the electrons into the Si NCs due to a smaller band gap of SiC film than a SiN_x film. The electroluminescence spectra increase with the forward voltage, indicating that the electrons are efficiently injected into the Si NCs in the SiC film. The power efficiency of the Si NC LED with the SiC film was 1.56 times larger than that of the Si NC LED with the SiN_x film. The Si NCs in a SiC film show unique advantages, and are a promising candidate for application in optical devices.**

Keywords-silicon nanocrystals, Silicon carbide, Light-emitting diode, Electroluminescence

I. Introduction

Recently, since the optical band gap of Si NCs can be easily tuned by changing the size of NCs due to a quantum confinement effect, Si NCs are of particular interest as a light-emitting diode (LED) covering whole visible wavelength range [1][2]. Si-rich oxide (SRO) film has been generally used as the surrounding matrix to synthesize the Si NCs [3][4]. The SRO film, however, has disadvantages in the formation of Si NCs as the surrounding matrix due to the trapping of the electrons in localized levels in the band gap of Si NC, a relatively high annealing temperature (> 1000 °C), and the high operating voltage (> a few tens of V) caused by a large band gap of SRO film (> 9 eV). In our previous result [5], well-organized Si NCs in the silicon nitride (SiN_x) film grown by a plasma enhanced chemical vapor deposition (PECVD) at a low temperature (250 °C) showed a clear quantum confinement effect depending on the size of Si NC, resulting that the band gap of Si NCs could be tuned from the near infrared (1.38 eV) to the ultraviolet (3.02 eV) range. In addition, we fabricated the mesa-type Si NC LED by applying an amorphous silicon carbide (SiC) film as an electron injection layer [6][7]. Because SRO and SiN_x films, however, have a relatively high band gap (approximately 9 eV for SRO and 5.3 eV for SiN_x, respectively), the tunneling probability between Si NCs decreases due to a high barrier height, resulting that the high operating voltage is required to inject the current into Si NCs. In the previous result [7], we have *in-situ* grown the well-organized Si NCs in a SiC matrix by using a PECVD. It was found that Si NCs in a SiC film showed a quantum confinement effect depending on the size of the Si NCs. In this work, we report the strong visible light emission from the LEDs by employing the Si NCs in a SiC film. The turn-on voltage of the Si NC LED was approximately 5 V and also decreased by 4 V compared to that of the Si NC LED by using Si NCs in SiN_x film. In addition, the wall-plug efficiency (WPE) was increased by 56 % compared to the Si NC LED with a SiN_x film.

II. Experimental

Si NCs in a SiC film with a thickness of 50 nm were *in-situ* grown at 250 ºC by conventional plasma enhanced chemical vapor deposition (PECVD). Ar-diluted 10 % SiH_4 and CH_4 were used as the reactant gases. The plasma power, chamber pressure, and substrate temperature for the growth were fixed at 5 W, 500 mTorr, and 250 °C, respectively. The flow rates of SiH_4 and CH_4 gases were 60 and 10 sccm, respectively. An amorphous SiC layer (~ 10 nm) doped with phosphorous, which was used to inject the electrons into the Si NCs, was deposited onto Si NC layer at 300 ºC by using a PECVD. ITO layer (100 nm) used as a transparent current spreading layer was deposited onto n-SiC layer at 150 ºC by using a pulsed laser deposition method. The size of Si NC LED fabricated was 300 um×300 um. The Si NC LED with Si NCs in a SiN_x matrix was also fabricated to compare the performance. The structure of Si NC LED investigated here was ITO (100 nm)/n-SiC (10 nm)/Si NCs in SiC (50 nm)/p^+-Si, as shown in Figure 1(a).

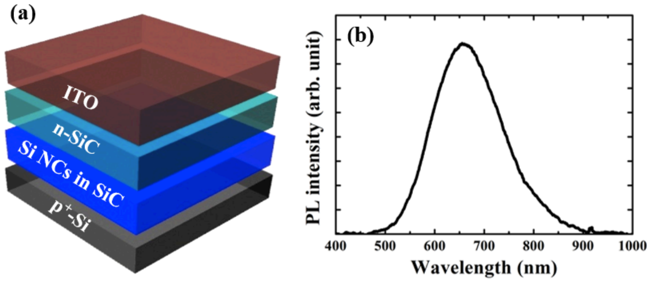


Figure 1. (a) A schematic diagram of the LED structure. (b) A RT PL spectrum taken from the Si NCs in a SiC film.

ISBN: 978-1-61208-431-2

III. RESULTS AND DISCUSSION

A RT photoluminescence (PL) spectrum taken from the Si NCs in a SiC film is shown in Figure 1(b), which is centered at ~ 650 nm. The average size of Si NCs into the SiC film was ~ 9 nm, which was confirmed by a HRTEM [8].

The *I-V* characteristics of Si NC LEDs with SiC and SiN_x films measured at RT are shown in Figure 2(a), respectively. The turn-on voltage of Si NC LED with the SiC matrix was decreased by around 4 V compared to that of Si NC LED with the SiN_x matrix.

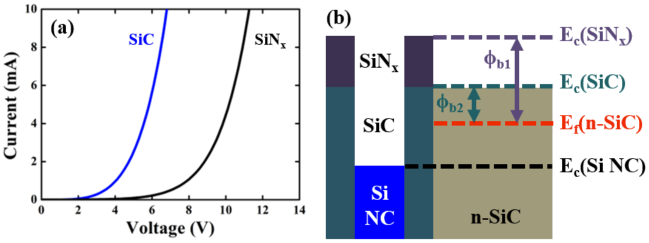


Figure 2. (a) *I-V* characteristics of Si NC LEDs with SiC and SiN_x films measured at RT, respectively. (b) A schematic band gap diagram of the Si NC LED structure with SiN_x and SiC films. ϕ_{b1} and ϕ_{b2} are the barrier height for the SiN_x film and the SiC film, respectively.

The schematic band gap diagram of the Si NC LED structure with SiN_x and SiC films is shown in Figure 2(b). As can be clearly seen in Figure 2(b), the barrier height (ϕ_{b2}) for the SiC film is lower than that (ϕ_{b1}) for the SiN_x film. The barrier height is very crucial to inject the electrons into the Si NCs from the transparent current spreading layer. The smaller the barrier height, the better the electron injection into the Si NCs. As the barrier height decreases, the electrons injected into the Si NCs for an external applied voltage to the LED can be increased. The electron transport between the Si NCs can be, therefore, significantly increased, resulting that the electrical performance of Si NC LED can be greatly improved. Therefore, lowering the turn-on voltage of Si NC LED with a SiC film was attributed to a lower barrier height for injecting the electrons into the Si NCs from the transparent current spreading layer caused by a lower band gap of SiC film (~ 2.5 eV) compared to SiN_x film (~ 5.3 eV).

The electroluminescence (EL) spectra of the Si NC LED as a function of forward voltage measured at RT are shown in Figure 3(a). We found that both PL (shown in Figure 1(b)) and EL showed a similar peak centered at 650 nm. This indicates that the PL and EL processes could be related to the same origin. Even though the center of EL peak was around 650 nm, overall EL peak showed the broad spectrum in the range of 500 ~ 850 nm. EL intensity was increased with the the forward voltage, as shown in Figure 3(a).

The light output power of Si NC LED with a SiC film as a function of the forward voltage is shown in Figure 3(b). With increasing the forward voltage, the light output power was linearly increased, indicating that the light emission was increased as more electrons and holes were injected into the Si NCs into a SiC film. The WPE, which means the power efficiency (output power/input power), is very important in LED applications. Based on the *I-V* data and light output power, the WPE of Si NC LED with a SiC film at 20 mA (~ 8 V) was estimated to be 1.94×10^{-9} %. The WPE was increased by 56 % compared to the Si NC LED with a SiN_x film. Inset shows the optical microscope image of light emission from the Si NC LED measured at a forward voltage of 12 V. As shown in the inset of Figure 3(b), the uniformity of light emission was quite good.

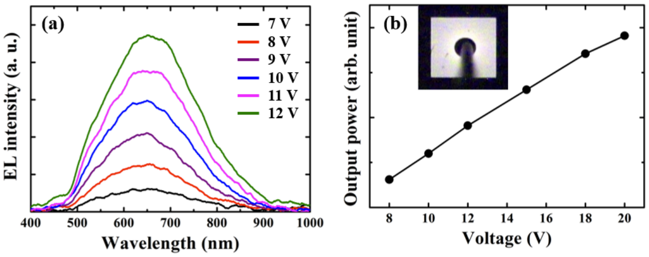


Figure 3. (a) EL spectra of the Si NC LED as a function of forward voltage measured at RT. (b) Light output power of Si NC LED with a SiC film as a function of the forward voltage. Inset shows the optical microscope image of light emission from the Si NC LED with a SiC film measured at a forward voltage of 12 V.

IV. CONCLUSION

A strong visible electroluminescence from Si NCs embedded into a SiC film was demonstrated. Compared to the Si NC LED by employing the SiN_x film, the electrical characteristics of the Si NC LED with the SiC film were significantly improved. This was originated from a smaller barrier height for injecting the electrons into the Si NCs due to a smaller band gap of the SiC film than the SiN_x film. Moreover, WPE of the Si NC LED with the SiC film was enhanced by 56 % compared to that of the Si NC LED with the SiN_x film. The light emission originated from the Si NCs in the SiC film was quite uniform.

ACKNOWLEDGMENT

This work was partly supported by the Internal Research Program (14RC1200) and the Pionier Research Program (14RC2300) funded by the ETRI.

REFERENCES

[1] L. Pavesi, L. Dal Negro, C. Mazzoleni, G. Franzò, and F. Priolo, Nature 408, 2000, pp. 440-444.

[2] R. Huang et al., Appl. Phys. Lett. 92, 2008, pp. 181106-1-181106-3.

[3] M. L. Brongersma, A. Polman, K. S. Min, E. Boer, T. Tambo, and H. A. Atwater, Appl. Phys. Lett. 72, 1998, pp. 2577-2579.

[4] N. Lalic and J. Linnros, J. Lumin. 80, 1999, pp. 263-267.

[5] T.-Y. Kim et al., Appl. Phys. Lett. 85, 2004, pp. 5355-5357.

[6] C. Huh et al., Adv. Mater. 22, 2010, pp. 5058-5062.

[7] C. Huh et al., Appl. Phys. Lett. 100, 2012, pp. 181108-1-181108-5.

[8] T.-Y. Kim, C. Huh, N.-M. Park, C.-J. Choi, and M. Suemitsu, Nanoscale Res. Lett. 7, 2012, pp. 634-1-634-5.

Nanoporous Silicon as an Electrode Material for Li-ion Batteries

Andrew A. Leshok[1], Dmitriy A. Sasinovich[2], Sergey K. Lazarouk[3], Victor E. Borisenko[4]

Department of Micro- and Nanoelectronics

Belarusian State University of Informatics and Radioelectronics

P. Browka 6, Minsk, Belarus

e-mails: andrew@nano.bsuir.edu.by[1], ovich@tut.by[2], serg@nano.bsuir.edu.by[3], borisenko@bsuir.by[4]

***Abstract* — We have developed a novel technique providing fabrication of nanoporous silicon films on nonsilicon substrates. Magnetron sputtering of an Al+Si composite target with subsequent selective etching off the aluminum from the deposited film were used to produce nanoporous silicon films on stainless steel substrates. These films were found to have a highly nanoporous skeleton formed by connected 20-200 nm silicon grains. They demonstrated an efficient Li accumulation/release during charging/discharging cycles combined with a high mechanical durability.**

Keywords - nanoporous silicon; Li-ion battery.

I. Introduction

There has been an increasing research activity in development of high-capacity, reliable and durable energy sources, in particular Li-ion batteries, for compact electronic devices, electric vehicles and various energy systems. Major efforts are focused on finding new and improving existing materials for electrodes of the batteries. At that point silicon (Si) is of a considerable interest because it is characterized by a much higher charge/discharge capacity as compared to the commonly used carbon materials [1]-[3]. Meanwhile, a principal problem of Si anodes in the batteries is their mechanical durability during volume expansion accompanying the lithiation process resulting in cracking of the electrodes. Nanostructuring of the electrodes could overcome this problem.

In this paper we present experimental results on fabrication and study of nanostructured Si anodes for Li-ion batteries demonstrating appropriate accumulation properties and durability. Section II describes in detail preparation of the samples and the methods used for their analysis. The results and their discussion are presented in Section III.

II. Samples and Experimental Technique

In the experiments performed, stainless steel foils were used as substrates. Composite Al+Si films were deposited by magnetron sputtering of Al target with Si insets covering 40 % of the erosion area (see [4], [5] for more details). The thickness of the deposited films was about 0.7 µm. Then the samples were immersed in H_3PO_4 water solution at 35 °C for 5 min in order to etch off Al selectively and to get nanostructured Si skeleton revealed. Subsequently, the samples with the flat surface area of 1.5 cm^2 were used as anodes in prototype Li-ion cells where cathode and polymer separator were taken from commercial Li-ion batteries. 1 M $LiClO_4$ solution in propylene carbonate was used as an electrolyte in the prototype cells.

Charging of the cells was performed at two current densities of 6.67 mA/cm^2 and 33.33 mA/cm^2 using Autolab potentiostat/galvanostat PGSTAT100N. Before and after charge/discharge cycling the anodes were analyzed with scanning electron microscopy (SEM).

III. Results and Discussion

The nanostructured Si film after selective removal of Al is shown in Figure 1. Highly porous skeleton of Si is visible to be formed by connected 20-200 nm grains. Al is completely etched off while the Si crystalline grains left look to be covered with amorphous Si like it was previously observed in [6].

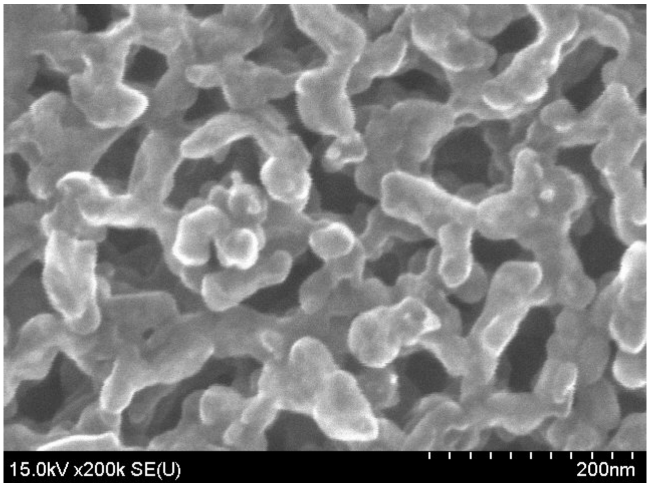


Figure 1. Surface SEM image of Al+Si film after selective etching off Al.

A voltage-time characteristic of the prototype Li-ion cell at charge/discharge cycling is shown in Figure 2. In general, it confirms conventional operation of Li-ion cell after the first charge cycle when the initial lithiation of the anode takes place.

ISBN: 978-1-61208-431-2

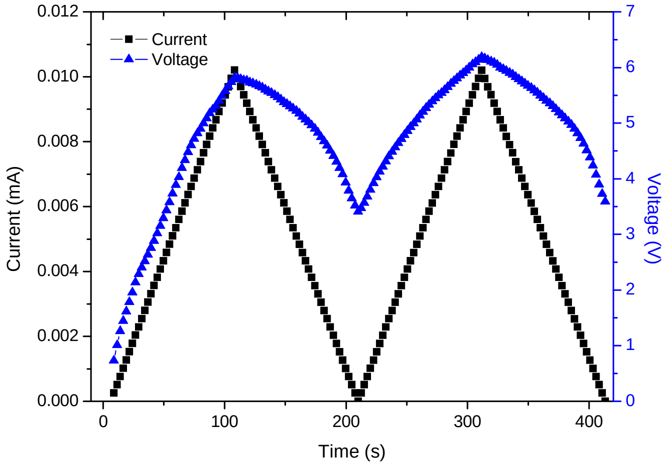


Figure 2. Typical voltage-time characteristic of the prototype Li-ion cell at cyclic lithiation. The straight lines show the test current variation for reference.

Self-discharge dynamics of the prototype Li-ion cell is shown in Figure 3. Despite the losses (of about 35 % within an hour) associated with the imperfection of the experimental system and the presence of leaks, the fabricated prototype Li-ion cell has demonstrated considerable potential for further improvement.

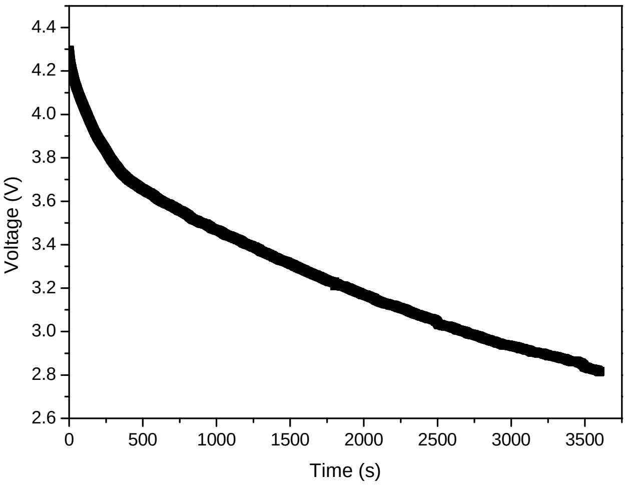


Figure 3. Self-discharge dynamics of the prototype Li-ion cell.

The nanostructured Si film morphology after 20 charge/discharge cycles is shown in Figure 4. The test sample has retained the original coral-like structure, but microcracks have appeared therein. The number of the microcracks correlates with the current density used for charging the cells. Nevertheless, microcracks have not led to the fragmentation of the whole structure. The microcracks appearance is associated with the quite hard test conditions owing to high current density through the cell.

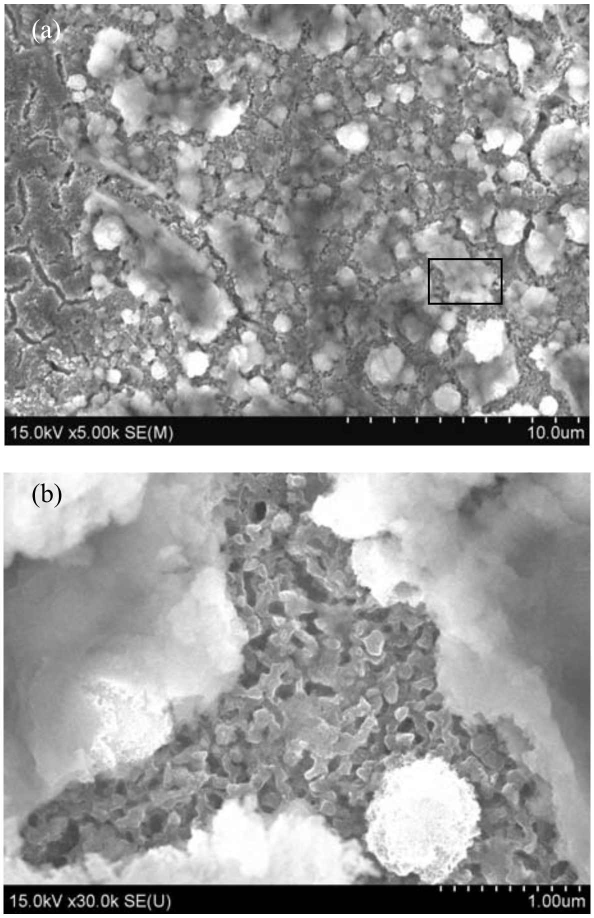


Figure 4. Surface SEM images of nanostructured Si film after 20 charge/discharge cycles: a) general view b) fragment of the structure.

Rather thick cracked layer can be also seen on the upper surface of the anode. Apparently, it is a solid-electrolyte interphase (SEI), which is formed on the anode surface as a product of the electrolyte reduction [7]. Formation of thick SEI layers on Si nanostructures was observed in [3], [8]. They were noted to play a stabilizing role providing an increase of mechanical durability of anode materials.

It is obvious that the coral-like structure is more mechanically strong in comparison to needle or tubular ones because its grains are interconnected to each other over the whole film thickness.

IV. Conclusion

The coral-like nanostructured Si films fabricated by magnetron sputtering of the composite Al+Si target and selective etching off Al have clearly demonstrated an ability to accumulate Li during charging and to maintain a certain level of voltage on a discharge time. Herewith, they have excellent mechanical durability even at high charging current densities which is their undoubted advantage. An extended study of the role of porosity and crystal structure of the grains has to optimize their performance in Li-ion batteries.

Acknowledgment

The work is financially supported by National Research Program “NANOTECH” (Project 2.4.18). The authors thank “Belmicroanaliz” for SEM analysis of the samples.

References

[1] U. Kasavajjula, C. Wang, and A. J. Appleby, “Nano- and bulk-silicon-based insertion anodes for lithium-ion secondary cells,” J. Power Sources, vol. 163, no. 1, Jan. 2007, pp. 1003-1039, doi:10.1016/j.jpowsour.2006.09.084.

[2] Candace K. Chan, *et al*., “High-performance lithium battery anodes using silicon nanowires,” Nature Nanotechnol., vol. 3, no. 1, Jan. 2008, pp. 31-35, doi:10.1038/nnano.2007.411.

[3] G. V. Li, *et al.*, “Structural transformation of macroporous silicon anodes as a result of cyclic lithiation processes,” Semiconductors, vol. 47, Sep. 2013, pp. 1275-1281, doi: 10.1134/S1063782613090133.

[4] A. A. Leshok, P. S. Katsuba, and V. B. Vysotskii, “Formation of nanostructured silicon by magnetron sputtering of an Al+Si composite target” Proc. Of International Conference Nanomeeting-2013, May 2013, pp. 362-365, ISBN: 978-981-4460-17-0.

[5] P. Katsuba, P. Jaguiro, S. Lazarouk, and A. Smirnov, “Stable electroluminescence of nanostructured silicon embedded into anodic alumina,” Physica E, vol. 41, no. 5, May 2009, pp. 931-934, doi:10.1016/j.physe.2008.08.051.

[6] V. A. Terekhov, *et al.*, “Specific features of the electronic and atomic structures of silicon single crystals in the aluminum matrix,” Phys. Sol. State, vol. 56, no. 12, Dec. 2014, pp. 2543-2547, doi:10.1134/S1063783414120336.

[7] P. Balbuena, and Y. Wang, Lithium-Ion Batteries Solid – Electrolyte Interphase. London: Imperial College Press, ch. 1, p.2, 2004.

[8] Emmanuel Ossei-Wusu, Ala Cojocaru, Hauke Hartz, Jürgen Carstensen, and Helmut Föll, “Silicon nanowires made via macropore etching for superior Li ion batteries,” Phys. Stat. Sol. A, vol. 208, no. 6, June 2011, pp. 1417-1421, DOI: 10.1002/pssa.201000031.

Challenges in Modeling Delayed Erosion due to Degradation of Novel Polyanhydride Biomaterials

Vladimir Privman and Sergii Domanskyi
Department of Physics, Clarkson University
Potsdam, NY 13699, USA
e-mail: privman@clarkson.edu

Katie L. Poetz and Devon A. Shipp
Department of Chemistry and Biomolecular Science,
Clarkson University
Potsdam, NY 13699, USA

***Abstract* — The long induction period of degradation proceeding without a noticeable mass loss was found in the recent experiments on highly cross-linked photopolymerized polyanhydrides. In order to model the observed phenomenon, we describe the kinetics of several processes such as the intake of water followed by the hydrolytic degradation of the cross-linked polyanhydride matrix. We survey the model and experimental data fitting, which suggest that the long induction interval is caused by the nonlinear dependence of the degradation rates on the local water concentration in the material, suggesting a breakdown of the standard rate-equation approach.**

Keywords — Biomaterials; cross-linked polyanhydrides; erosion; hydrolysis; rate equations.

I. INTRODUCTION

A better understanding of the mechanisms and processes of degradation and erosion of various polymeric biomaterials [1]-[16] can be achieved by means of modeling. Here we review recent work on numerical modelling of cross-linked polyanhydrides [1], a specific biodegradable polymer material that has beneficial chemical and physical properties utilized in applications, because of its undergoing degradation and erosion in aqueous environments [2]-[5]. In a recently developed modeling approach [1], we aimed at explaining an important property found in the experiments [17][18]: a long induction interval of water intake prior to the noticeable erosion (mass loss) in highly cross-linked polyanhydrides produced via thiol-ene photopolymerization. This property might allow additional control in medical applications, such as drug release [19][20]. Utilizing a new theoretical model, we conclude that the observed long induction interval can be explained by a kinetic effect of the breakdown of the rate-equation approach, typically used for reaction-diffusion systems.

Specifically for medical applications [5][21]-[24], predictable degradation of biodegradable polymer materials is crucial for drug delivery capsules. Furthermore, biocompatibility is important for orthopedic applications. Some polyanhydrides actually have compressive strengths similar to the human cortical bone [24][25]. In addition to predictable degradation and biocompatibility, polymer implants can help avoid multiple surgeries and incorporate healing drugs [22] for long-term delivery during the implant degradation and erosion. Other benefits include making stress shielding unnecessary and absence of corrosion processes [22]. Furthermore, such materials could facilitate in tissue engineering [5] and development of new bio-adhesives [23].

In the experiments [17] that preceded the modeling work reported here, the authors investigated several aspects of the degradation kinetics of polyanhydrides synthetized by thiol-ene photopolymerization. They experimentally measured and reported several quantities including the time-dependent mass of the surviving eroding cross-linked polyanhydride sample, release rate of a chosen drug-mimicking substance (a hydrophilic dye) in an aqueous environment, and rate of anhydride bond breakage by hydrolysis. Other quantities that might have affected the process kinetics included the pH-dependent degradation product solubility, and the average pK_a of that product. The hydrolysis-involving experiment leading to the polymer matrix swelling and connectivity reduction was carried out at 37° C in phosphate buffered saline (PBS) solution at pH 7.4. A substantial induction period of water intake, about 10 h, before the onset of the noticeable erosion was observed in these experiments, unlike the earlier degradation experiments [4][9][13][22][26]-[29] done for linear polymers only, where typically no indications of such a property were reported, for which phenomenological modeling was reported [12].

In the experiments of interest [17], as the sample takes up the water and the anhydride bonds are subjected to hydrolysis, the connectivity of the polymer network is reduced. Consequently, low molecular weight degradation products and possibly small pieces are released into the solution. We focused on explaining the cause of the experimentally observed [17] induction interval preceding substantial mass loss of the degrading sample [1]. The

ISBN: 978-1-61208-431-2

aforementioned effect is illustrated on Figure 1, top panel, where the experimental data [17] and the model fit curves [1] are combined. The details of the system and of the modeling approach can be found in [1]; they are surveyed here.

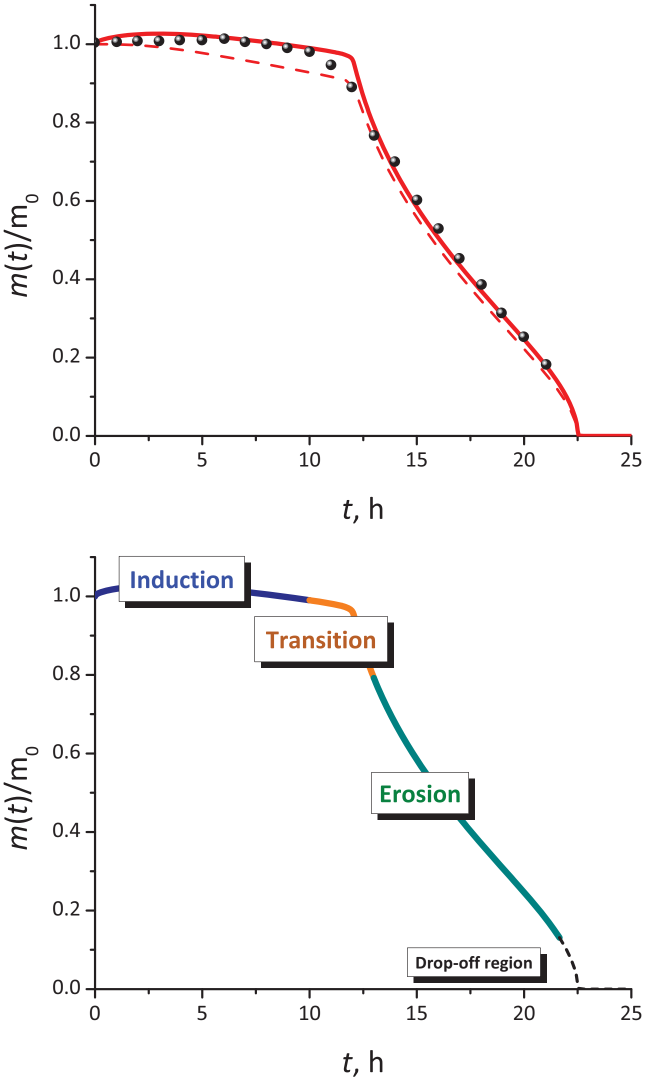


Fig. 1. The relative mass of the remaining undissolved sample as a function of time. Top panel: Black spheres represent the experimental data points [17], measured in hourly intervals. The solid line is the model fit [1] for the sample's relative mass, whereas the dashed line is the polymer-fraction mass in the remaining sample. Bottom panel: Stages of the model-predicted erosion [1].

For the recently developed and studied highly cross-linked amorphous polyanhydrides [17][18][31][32], the existing modeling work [6]-[16][30] on bulk and surface eroding polymers is not applicable as it does not provide an explanation of the long induction interval, comparable to the overall erosion timescale. Figure 1 illustrates the primary modeling results of our new modeling approach [1], reviewed here. The objectives of the modeling were to explore the kinetic mechanisms of the relevant processes and influence of various system parameters. We concluded that the long induction interval of water intake without measurable erosion can be attributed to an effect of the rate-equation description breakdown due to altered water reactivity.

We note that the polymer materials examined in [17] are amorphous. Most of the previous modeling and experimental work was done on semi-crystalline polyanhydrides. For the latter, the presence of two phases adds complexity to the erosion, as the amorphous and crystalline phases can exhibit different degradation behaviors [8].

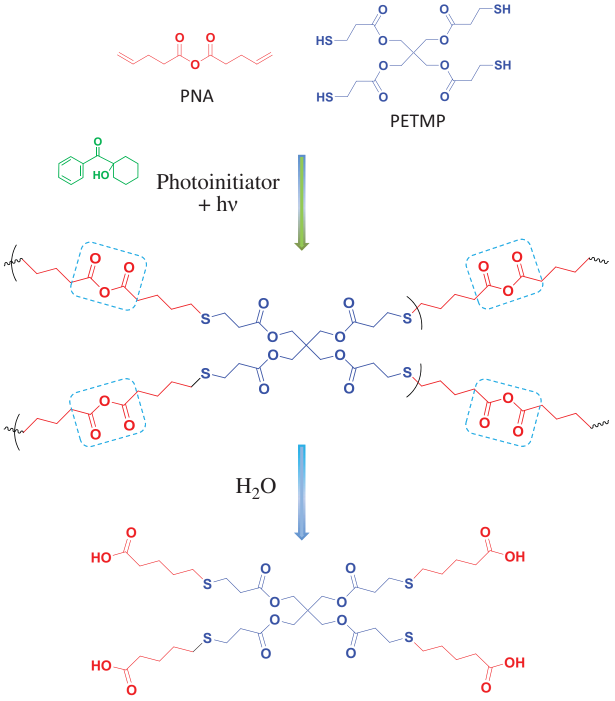


Fig. 2. The scheme shows PNA and PETMP molecules (top section) used in synthesizing polyanhydrides. In the cross-linked structure (middle section) the dashed-line boxes mark the anhydride functional groups, which degrade by hydrolysis, resulting in the polymer breakup into small units (bottom section).

For linear non-cross-linked polymers used in previous works, polymer molecular weight was a determining factor in the erosion process. Polyanhydrides explored here have highly cross-linked structure in addition to being amorphous. These polyanhydrides were synthesized [17] through the thiol-ene photopolymerization of pentaerythritol tetrakis(3-mercaptopropionate) (PETMP) and 4-pentenoic anhydride (PNA); see Figure 2. This type of polymerization, termed a radical-mediated thiol-ene reaction, is often referred to as "click" chemistry, meaning it is highly efficient and easy to perform [33]-[35]. For the system considered here, reaction of the PETMP (1 mole equivalent) and the PNA (2 mole equivalent), yields the highly cross-linked network structure shown in Figure 2. The PNA monomer contains the anhydride functionality that readily undergoes hydrolysis, thus these highly cross-linked polymers degrade in aqueous environments. The polymers were synthesized by combining 0.27 ml PETMP, 0.26 ml PNA, and 0.6 mg photoinitiator (1-hydroxycyclohexyl phenyl ketone) in a silicone mold (10 × 10 × 2 mm) and irradiating with UV light for 15 minutes.

The samples were subsequently degraded in 100 mL of phosphate buffered saline (PBS) solution at 37° C and pH 7.4. Mass of the sample was measured every hour after removing it from the solution and wiping off the excess of the solution and loose (not cross-linked) layer. After each measurement, the PBS solution was renewed to maintain pH 7.4. In the rest of this survey, in Section II we describe the modeling approach. Section III offers a concluding discussion.

II. The modeling approach

The degradation and erosion processes are controlled by many parameters. The breakup of the polyanhydride bonds by hydrolysis is typically described using rate equations that introduce rate constants. For the process of erosion we considered diffusion of water into the structure and diffusion of the detached small degradation products within and out of the polymer matrix. Kinetic parameters might be pH-dependent, specifically, the rate of hydrolysis [9][11][26][28][36][37] and solubility of the degradation products [10][13][27].

The buffer diffusion into the sample may also need to be included into the model. We introduced additional parameters for the receding sample boundary, at which variation of buffer, degradation products and water concentration occurs as the sample degrades. The swelling of the outer layer of the polymer sample [17] was also considered [1], yielding more parameters for modeling.

The details of the model are described in [1]. Here, we survey the findings. Consider the standard rate equations, used in earlier modeling [6][7][12][30][38] for the degradation of the polymer matrix, here in a shape of a narrow slab as in the experiment [17], with *x* measured from the middle (and *t* representing time),

$$\begin{aligned}\frac{\partial u_4(x,t)}{\partial t} &= -k_4 u_4 u_{\mathrm{w}};\\ \frac{\partial u_i(x,t)}{\partial t} &= k_{i+1}u_{i+1}u_{\mathrm{w}} - k_i u_i u_{\mathrm{w}},\ i=1{,}2{,}3;\\ \frac{\partial u_0(x,t)}{\partial t} &= k_1 u_1 u_{\mathrm{w}} + D_0\frac{\partial^2 u_0}{\partial x^2};\\ \frac{\partial u_{\mathrm{w}}(x,t)}{\partial t} &= D_{\mathrm{w}}\frac{\partial^2 u_{\mathrm{w}}}{\partial x^2}.\end{aligned} \tag{1}$$

Here we have several adjustable parameters, including the rate constants $k_{i=1,2,3,4}$, molar concentrations $u_{i=0,1,2,3,4}$ of the 4-, 3-, 2-, 1-(cross-)linked (within the network) units, as well as 0-linked disconnected small diffusing units. These 0-linked units are assumed to diffuse with the average diffusion constant D_0. The water (with molar concentration u_{w}) diffuses into the matrix with diffusion constant D_{w}. The reaction terms $k_i u_i u_{\mathrm{w}}, i = 1{,}2{,}3{,}4$, in the rate equations quantify the hydrolytic break up of the network.

The sample boundary was defined at $x = X_{\mathrm{B}}(t)$, such that the total amount of the cross-linked material at this *x* value dropped to some reference fraction, *g* (the parameter introduced in [1]), of the initial concentration of the fully cross-linked network at time *t* = 0,

$$\textstyle\sum_{i=1}^{4} u_i(X_{\mathrm{B}}(t),t) = g u_4(0). \tag{2}$$

Several modifications of the above description are required for $x > X_{\mathrm{B}}(t)$ where some quantities undergo qualitative changes in properties at the boundary.

The set of equations (1) involves 7 parameters, and only some of them can be found in the literature at least approximately, e.g., D_{w} in such a polymeric material environment [6][39]-[42]. Other parameters can be only estimated, rather than obtained precisely from data fitting because of the noise in the experimental data, or possibly the model not fitting the considered data. Thus, we are left with several rate constants, etc., including some of the parameters defined at the boundary and outside the sample (described in [1]), and even more quantities that can be fitted in relation to the observed swelling, which we also modeled, see [1] where all the relevant parameter values are given. The available experimental data has only two well-defined time scales as the key measurable properties to fit (see Figure 1). One is the induction time, the other is the erosion time scale measured by the rate of erosion in the approximately linear decay regime.

These two time scales could not be reproduced, as discussed in [1], even when involving numerous adjustable parameters. The standard rate-equation model that included (1), with various boundary considerations and with the inclusion of swelling into the model, etc., was insufficient to fit the experiments data.

Figure 3, top panel, demonstrates how for different values of the rate constants, $k_{i=1,2,3,4}$, no substantial delay time (the induction region) can be produced, cf. Figure 1. By changing some of the parameters (rate constants), the time scale of the fast erosion region can be adjusted, but the duration of the delay remains small, see Figure 3, bottom panel. Furthermore, there is a “bottleneck” effect, also exemplified in Figure 3, in that only a single rate constant is the rate-limiting one. This generally leaves much less freedom in the parameter determination by data fitting than one would expect from the large number of the introduced model parameters, which is actually typical for models of such chemical kinetics.

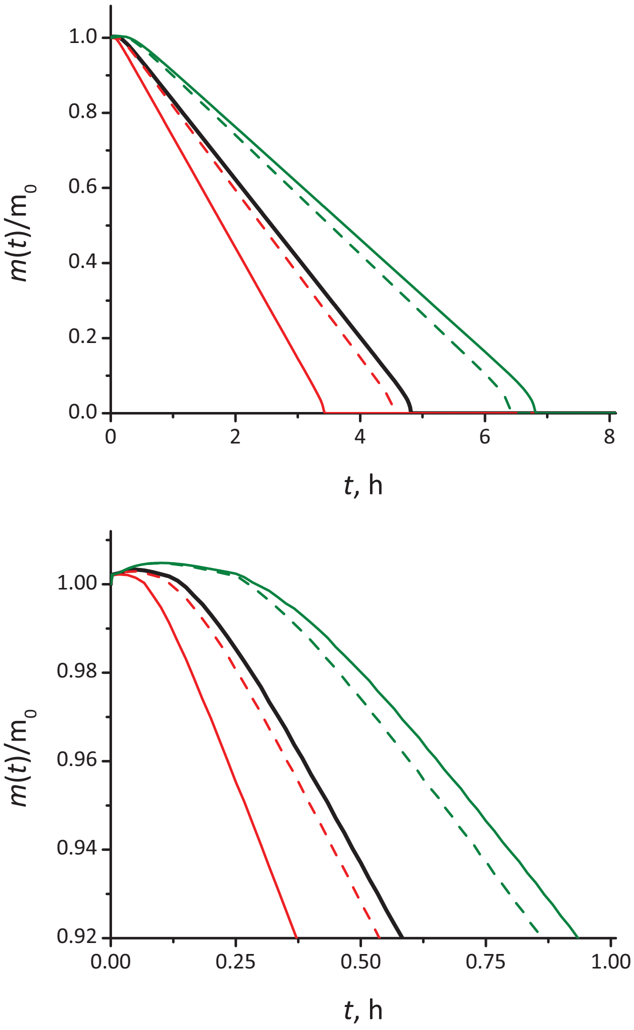


Fig. 3. Top panel: relative mass of the cross-linked polyanhydride sample as a function of time in the standard-rate-equation model. The solid red line was produced with all the rate constants, k_i, doubled as compared to the black line. The green line was obtained by instead halving all the rate constants together. The bottleneck effect is illustrated as follows: The dashed red line represents the case when only k_2 is doubled. The dashed green line is for the case when $k_{1,2}$ are halved. Bottom panel: the initial fast variation time scale is magnified. The parameter values are given in [1].

The overall finding is that the standard rate equation model is not suitable to explain the experimentally observed large induction period. We mention again that several other experiments, which involved less cross-linked networks did not report the long induction time preceding erosion [4][9][13][22][26][27][29][37], however in some of them the delay stage can be observed [7][9][17][18][43]-[45]. For different model modifications, the rate constants in (1) were found [1] to be the basic rate-determining parameters. These equations were common to the different model variations considered. Therefore, the validity of the rate equations (1) should be questioned, and we found that the concentration-of-water dependence is the reason. The model, even in its simplified variant provides a good fit of the data if each of the reaction terms in (1) is modified according to

$$k_i u_i u_{\mathrm{w}} \Rightarrow k_i u_i u_{\mathrm{w}} f(u_{\mathrm{w}}),\ i = 1,2,3,4, \tag{3}$$

with a single-parameter, u_∞, function

$$f(u_{\mathrm{w}}) = \frac{u_\infty}{u_\infty + u_{\mathrm{w}}}, \tag{4}$$

that phenomenologically describes the deviation of the original u_{w}-dependence of the reaction rates from linear, here saturating as a finite value, u_∞, for large $u_{\mathrm{w}}(\gg u_\infty)$.

Only such a model modification yielded the curves in Figure 1, where the new parameter, u_∞, controls the induction time. We also noted that the shape and sharpness of the transition region from induction to fast erosion, see Figure 1, bottom panel, are mostly controlled by the assumed initial degree of cross-linking near the sample boundary and the diffusion constant, D_0, for detached 0-connected units. However, the given data are too noisy for the precise fitting of this diffusion constant. Additionally, the model-predicted sharp drop-off region at the end of the erosion time, see Figure 1, bottom panel, is difficult to measure experimentally as the sample integrity is mostly compromised when its remaining mass is below ~20%.

III. Conclusion

We found evidence that the induction delay time correlates with the water reactivity in hydrolysis in the considered system. One possible explanation of this property is that the dense amorphous network prevents fast enough local water equilibration by diffusion as the water concentration increases. The local reaction rates are then effectively slowed down. However, assuming a standard water diffusion mechanism in the amorphous polymer matrix, statistical-mechanics considerations [46][47] make this explanation questionable.

An alternative explanation could be that a dense cross-linked polymer network prevents resupply of buffer by diffusion from outside the sample, and the reaction rates are lowered due to local pH changes. Another possible explanation of the observed decrease in water reactivity is due to a completely different effect: During the erosion and continuous water intake, some of it will be entrapped in small crevices and only part of that water will be surface-reacting with the surrounding network, causing further anhydride bonds breakup. Such explanation for surface-only reactivity has been noted in a different context [10][13].

These possibilities cannot be sorted out by modeling alone, and they pose a challenge for experimental studies to yield microscopic, rather than the presently available primarily macroscopic degradation/erosion data, providing information on local cross-linked polyanhydride material properties.

ACKNOWLEDGEMENT

Funding by the NSF (grant CBET-1403208) is gratefully acknowledged.

REFERENCES

[1] S. Domanskyi, K. L. Poetz, D. A. Shipp, and V. Privman, “Reaction-diffusion degradation model for delayed erosion of cross-linked polyanhydride biomaterials,” Phys. Chem. Chem. Phys. 2015, vol. 17, pp. 13215-13222.

[2] S. Liu, R. Maheshwari, and K. L. Kiick, “Polymer-Based Therapeutics,” Macromol. 2009, vol. 42, pp. 3–13.

[3] R. Langer, “Drug Delivery and Targeting,” Nature 1998, vol. 392, pp. 5–10.

[4] A. Göpferich et al., “Drug delivery from bioerodible polymers: systemic and intravenous administration,” ChemInform. 1995, vol. 26, pp. 242–277.

[5] R. Langer and J. Vacanti, “Tissue Engineering,” Science 1993, vol. 260, pp. 920–926.

[6] S. N. Rothstein, W. J. Federspiel, and S. R. Little, “A unified mathematical model for the prediction of controlled release from surface and bulk eroding polymer matrices,” Biomater. 2009, vol. 30, pp. 1657–1664.

[7] Y. Chen, S. Zhou, and Q. Li, “Mathematical Modeling of Degradation for Bulk-Erosive Polymers: Applications in Tissue Engineering Scaffolds and Drug Delivery Systems,” Acta Biomater. 2011, vol. 7, pp. 1140–1149.

[8] A. Göpferich and J. Tessmar, “Polyanhydride degradation and erosion,” Adv. Drug Deliv. Rev. 2002, vol. 54, pp. 911–931.

[9] F. Burkersroda, L. Schedl, and A. Göpferich, “Why Degradable Polymers Undergo Surface Erosion or Bulk Erosion,” Biomater. 2002, vol. 23, pp. 4221–4231.

[10] A. Göpferich and R. Langer, “Modeling monomer release from bioerodible polymers,” J. Controlled Release 1995, vol. 33, pp. 55–69.

[11] A. Göpferich, “Mechanisms of polymer degradation and erosion,” Biomater. 1996, vol. 17, pp. 103–114.

[12] J. Siepmann and A. Göpferich, “Mathematical modeling of bioerodible, polymeric drug delivery systems,” Adv. Drug Deliv. Rev. 2001, vol. 48, pp. 229–247.

[13] M. J. Kipper and B. Narasimhan, “Molecular Description of Erosion Phenomena in Biodegradable Polymers,” Macromol. 2005, vol. 38, pp. 1989–1999.

[14] K. Zygourakis, “Discrete simulations and bioerodible controlled release systems,” Polym. Prepr. ACS 1989, vol. 30, pp. 456–457.

[15] K. Zygourakis and P. A. Markenscoff, “Computer-aided design of bioerodible devices with optimal release characteristics: a cellular automata approach,” Biomater. 1996, vol. 17, pp. 125–135.

[16] K. Zygourakis, “Development and temporal evolution of erosion fronts in bioerodible controlled release devices," Chem. Eng. Sci. 1990, vol. 45, pp. 2359–2366.

[17] K. L. Poetz et al., “Photopolymerized Cross-Linked Thiol-Ene Polyanhydrides: Erosion, Release and Toxicity Studies," Biomacromol. 2014, vol. 15, pp. 2573–2582.

[18] D. A. Shipp, C. W. McQuinn, B. G. Rutherglen, and R. A. McBath, “Elastomeric and degradable polyanhydride network polymers by step-growth thiol-ene photopolymerization,” Chem. Commun. 2009, vol. 2009, pp. 6415–6417.

[19] K. E. Uhrich, S. M. Cannizzaro, R. S. Langer, and K. M. Shakesheff, “Polymeric systems for controlled drug release,” Chem. Rev. 1999, vol. 99, pp. 3181–3198.

[20] A. G. Thombre, “Theoretical Aspects of Polymer Biodegradation: Mathematical Modeling of Drug Release and Acid-Catalyzed Poly(Ortho-Ester) Biodegradation,” Biodegradable Polymer Plastics 1992, vol. 109, pp. 214–225.

[21] A. Domb, J. P. Jain, and N. Kumar, ch. 3. “Polyanhydrides,” In “Handbook of Biodegradable Polymers: Synthesis, Characterization and Applications,” A. Lendlein and A. Sisson (Eds.), Wiley-VCH: Weinheim, Germany, 2011, pp. 45–75.

[22] D. S. Muggli, A. K. Burkoth, and K. S. Anseth, “Crosslinked polyanhydrides for use in orthopedic applications: Degradation behavior and mechanics,” J. Biomed. Mater. Res. 1999, vol. 46, pp. 271–278.

[23] R. S. Langer and D. L. Wise, “Medical Applications of Controlled Release, Vol. 1: Classes of Systems,” CRC Press: Boca Raton, 1984.

[24] K. E. Uhrich, S. E. M. Ibim, C. T. Laurencin, and R. Langer, “Degradation of Poly(Anhydride-Co-Imides): Novel Polymers for Orthopedic Applications,” MRS Proc. 1995, vol. 394, pp. 41–48.

[25] K. E. Uhrich et al., “Synthesis and Characterization of Degradable Poly(Anhydride-Co-Imides),” Macromol. 1995, vol. 28, pp. 2184–2193.

[26] K. W. Leong, B. C. Brott, and R. Langer, “Bioerodible Polyanhydrides as Drug-Carrier Matrices. I: Characterization, Degradation, and Release Characteristics,” J. Biomed. Mater. Res. 1985, vol. 19, 941–955.

[27] A. Göpferich and R. Langer, “The Influence of Microstructure and Monomer Properties on the Erosion Mechanism of a Class of Polyanhydrides,” J. Polym. Sci. A 1993, vol. 31, pp. 2445–2458.

[28] E. Park, M. Maniar, and J. Shah, “Effects of Model Compounds with Varying Physicochemical Properties on Erosion of Polyanhydride Devices,” J. Controlled Release 1996, vol. 40, pp. 111–121.

[29] A. Dong, J. Zhang, K. Jiang, and L. Deng, “Characterization and in Vitro Degradation of Poly(Octadecanoic Anhydride),” J. Mater. Sci. Mater. Med. 2008, 19, pp. 39–46.

[30] C. K. Sackett and B. Narasimhan, “Mathematical modeling of polymer erosion: Consequences for drug delivery,” Int. J. Pharm. 2011, vol. 418, pp. 104–114.

[31] B. G. Rutherglen, R. A. McBath, Y. L. Huang, and D. A. Shipp, “Polyanhydride Networks from Thiol-Ene Polymerizations,” Macromol. 2010, vol. 43, pp. 10297–10303.

[32] B. D. Fairbanks, T. F. Scott, C. J. Kloxin, K. S. Anseth, and C. N. Bowman, “Thiol-Yne Photopolymerizations: Novel Mechanism, Kinetics, and Step-Growth Formation of Highly Cross-Linked Networks,” Macromol. 2009, vol. 42, pp. 211–217.

[33] C. E. Hoyle, T. Y. Lee, and T. Roper, “Thiol–enes: Chemistry of the past with promise for the future,” J. Polym. Sci. A, 2004, vol. 42, pp. 5301–5338.

[34] C. E. Hoyle and C. N. Bowman, “Thiol–Ene Click Chemistry,” Angew. Chem., Int. Ed., 2010, vol. 49, pp. 1540–1573.

[35] C. E. Hoyle, A. B. Lowe, and C. N. Bowman, “Thiol-click chemistry: a multifaceted toolbox for small molecule and polymer synthesis,” Chem. Soc. Rev., 2010, vol. 39, pp. 1355–1387.

[36] S. H. Hilal, “Estimation of Hydrolysis Rate Constants of Carboxylic Acid Ester and Phosphate Ester Compounds in Aqueous Systems from Molecular Structure by SPARC,” U.S. Environmental Protection Agency: Washington, DC, 2006.

[37] E. Park, M. Maniar, and J. Shah, “Water Uptake in to Polyanhydride Devices: Kinetics of Uptake and Effects of Model Compounds Incorporated, and Device Geometry on Water Uptake,” J. Controlled Release 1996, vol. 40, pp. 55–65.

[38] J. Heller and R. W. Baker, In “Controlled Release of Bioactive Materials,” R. W. Baker (Ed.), Academic Press: NY, 1980.

[39] R. Langer and N. Peppas, “Chemical and physical structure of polymers as carriers for controlled release of bioactive agents: a review,” J. Macromol. Sci. C 1983, vol. 23, pp. 61–126.

[40] P. Neogi, “Diffusion in Polymers,” Marcel Dekker: NY, 1996.

[41] R. Parthasarathy, A. Misra, J. Park, Q. Ye, and P. Spencer, “Diffusion Coefficients of Water and Leachables in Methacrylate-Based Crosslinked Polymers Using Absorption Experiments,” J. Mater. Sci. Mater. Med. 2012, 23, pp. 1157–1172.

[42] A. G. Thombre and K. J. Himmelstein, “A Simultaneous Transport–Reaction Model for Controlled Drug Delivery from Catalyzed Bioerodible Polymer Matrices,” AIChE J. 1985, vol. 31, pp. 759–766.

[43] A. D'Emanuele, J. Hill, J. Tamada, A. Domb, and R. Langer, “Molecular Weight Changes in Polymer Erosion,” Pharmaceutical Research 1992, vol. 9, pp. 1279–1283.

[44] S. S. Shah, Y. Cha, and C. G. Pitt, “Poly(glycolic acid-co-DL-lactic acid): diffusion or degradation controlled drug delivery?” J. Controlled Release 1992, vol. 18, pp. 261–270.

[45] M. L. Johnson and K. E. Uhrich, “Concurrent release of admixed antimicrobials and salicylic acid from salicylate-based poly(anhydride-esters),” J. Biomed. Mater. Res. A 2008, vol. 91, pp. 671–678.

[46] V. Privman and M. D. Grynberg, “Fast-Diffusion Mean-Field Theory for *k*-Body Reactions in One Dimension,” J. Phys. A 1992, vol. 25, pp. 6567–6575.

[47] V. Privman and M. Barma, “Random Sequential Adsorption on a Line: Mean-Field Theory of Diffusional Relaxation,” J. Chem. Phys. 1992, vol. 97, pp. 6714–6719.

Reflection from Irregular Array of Silver Nanoparticles on Multilayer Substrate

Victor Ovchinnikov
Department of Aalto Nanofab
School of Electrical Engineering, Aalto University
Espoo, Finland
e-mail: Victor.Ovchinnikov@aalto.fi

Abstract—**Reflection from silver irregular arrays of nanostructures on quartz and oxidized silicon substrates is studied. It is shown that localized plasmon resonances in reflectance spectra cannot be easily identified by their peaks like it is done in case of extinction spectra. To clarify positions of resonances optical properties of samples are analyzed in relation to their design and morphology. Extinction and reflection from as prepared, plasma etched and SiO_2 covered samples are compared. It is concluded that coupling between nanoparticles, phase shift of scattered light and reflection from film interfaces lead to additional features in reflectance spectra in comparison with extinction ones. Recommendations for identification of plasmon resonances in reflectance spectra are proposed.**

Keywords-Ag nanoparticle; surface plasmon resonance; reflectance; extinction; irregular array.

I. INTRODUCTION

Plasmonic nanostructures are widely used in sensors, metamaterials, solar cells, photonics and spectroscopy [1]-[5]. Effective application of these structures is based on localized surface plasmon resonance (LSPR) demonstrated in ultraviolet, visible and infrared. The wavelength of LSPR depends on material and geometry of nanostructures, their interaction with each other and electromagnetic properties of environment, including substrate and capping layers. Despite on near field nature, LSPR can be observed in far field optical measurements due to variation in optical properties of the studied structures. Extinction is the most popular method of LSPR registration due to its simple implementation and straightforward interpretation, i.e., maximum and width of extinction peak correspond LSPR wavelength and damping, respectively. However, extinction can be measured only for non-opaque structures, e.g., for nanostructures on transparent substrates or for plasmonic colloidal particles. Furthermore, extinction spectra are not effective for overlapped peaks, when spectral deconvolution is not obvious.

LSPR on opaque substrates can be visualized by different kinds of reflection and scattering measurements. However, peak and trough of specular reflectance do not correspond to LSPRs and spectrum analysis becomes problematic. Scattering measurements require special arrangement of light illumination (dark field) to separate low scattering signal from strong reflection background. It limits range of measured samples by plasmonic nanostructures on substrate surface and, for example, plasmonic nanoparticles inside of dielectric matrix cannot be analyzed. In case of correlated scattering centers, i.e., when array of coupled nanostructures is analyzed, correspondence between scattered peaks and LSPRs is broken and LSPRs should be observed in specular reflectance. Additionally, scattering results are obtained in arbitrary units and cannot be used for comparison of different experiments. This happens, due to the problems with measurement of scattered reference spectrum for calibration procedure. In contrast, reflectance reference spectrum can be easily obtained for any material. Combination of total reflectance and diffuse reflectance is especially useful for analyzing plasmon structures, because the last one provides measurement of scattering in absolute values.

In this paper, we propose to use total reflectance for identification of LSPR on opaque and transparent substrates. It is demonstrated that wavelength position of LSPR correlates with peak and trough of reflectance in a clear way. Furthermore, overlapped peaks manifest themselves separately in reflectance spectra and can be easily distinguished.

This paper is organized in a following way. In the subsequent Section II, the details of sample preparation and the measurement procedures are presented. In Section III, the results of the work are demonstrated by scanning electron microscope (SEM) images as well as reflection and extinction spectra of the fabricated samples. The effect of the substrate and dielectric layers on reflectance of silver nanoparticles is discussed in Section III as well. In Section IV, the conclusions are drawn.

II. METHOD

Quartz or crystalline Si wafers (4” in diameter, 0.5-mm-thick) were used as substrates. The Al_2O_3 layer was grown on the substrate by atomic layer deposition (ALD), and SiO_2 layer was created by thermal oxidation of the Si wafer. Silver layers, with a thickness of 15 nm, were deposited by electron-beam evaporation with the deposition rate of 0.5 nm/s. Nanoparticle arrays were fabricated by ion beam mixing (IBM) or annealing of silver films. In case of IBM, Ag films were irradiated by 400 keV Ar ions at normal incidence and at low (1×10^{16} Ar/cm^{-2}) or high (2×10^{16} Ar/cm^{-2}) ion fluence to produce the nanoparticles as reported elsewhere [6]. One

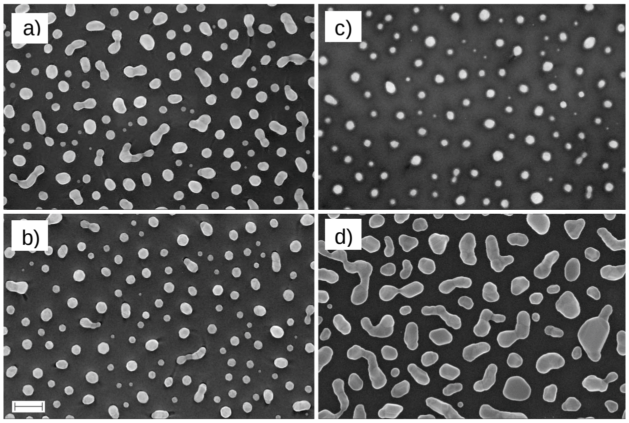


Figure 1. Plan SEM images of low (a) and high (b) dose Ar IBM samples, Xe IBM sample (c) and annealed sample (d). Scale bar is 200 nm.

sample was processed by IBM with Xe ions at dose 6×10^{15} Ar/cm^{-2}. In the case of annealing, silver films were heated at 350 ºC during 10 minutes. Annealing was done in diffusion furnace in nitrogen ambient. Further details about samples and processing can be found elsewhere [7][8]. To cover the nanoparticles with a SiO_2 layer we used a plasma enhanced chemical vapor deposition (PECVD) technique. Metal deposition and ion irradiation were perfomed at room temperature. ALD and PECVD processes were run at low temperatures 200 ºC and 170 ºC, respectively to avoid Ag oxidation. To examine the nanoparticle formation in the structures created, the images of the samples were taken with a Zeiss Supra 40 field emission scanning electron microscope. Three such images, depicting effect of IBM dose and mixing ions are shown in Figure 1(a)-(c). One more image of the sample prepared by annealing of silver film is presented in Figure 1(d). The details of nanopartile size distribution and sample surface morphology can be found elsewhere [6][9].

The optical extinction and reflection spectra were measured with a PerkinElmer Lambda 950 UV-VIS spectrometer in the range from 250 to 850 nm. Reflectance spectra at the angle of light incidence 8º were obtained by using an integrating-sphere detector incorporated in the spectrometer. Either total reflectance or diffuse reflectance only can be measured by placing spectralon plate at the specular reflectance angle or removing it, respectively.

III. Analysis of Reflection and Extinction Spectra

In this section, we study spectra of silver nanostructures on different substrates. In subsections III-A and III-B, visible parts of spectra are discussed, while the subsection III-B is devoted to UV features of the spectra.

A. Ag Nanoparticles on a Quartz Substrate

Silver nanostructures on a weekly reflecting substrate without any additional layers between nanostructures and the substrate are studied in this subsection. It simplifies spectrum analysis due to excluding from consideration interference effects. In Figure 2 (a)(b) extinction, reflection and scattering

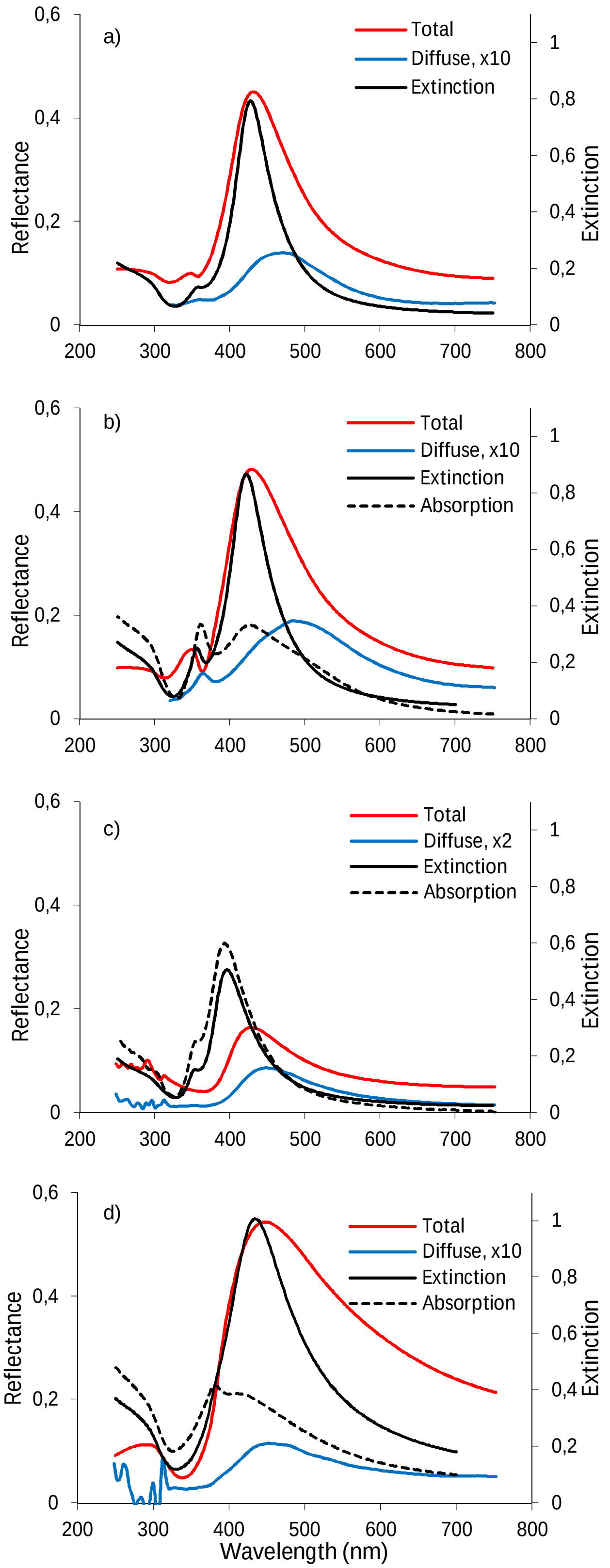


Figure 2. Spectra of low (a) and high (b) dose Ar mixed samples, RIE treated sample (c) and SiO_2 capped sample (d).

of silver nanoislands on quartz substrate are demonstrated for high and low dose of IBM, respectively. The corresponding SEM images of the samples are given in Figure 1. In the spectra, there are distinctly visible two areas: right one (wavelength more than 400 nm) with broad and intense peak in visible (VIS) range and left one (wavelength is less than 400 nm) with weak peak in ultraviolet (UV) range. Further, we call these parts as VIS and UV, respectively. The high amplitude peak is usually attributed to dipolar LSPR, whereas the low amplitude one is connected with quadrupolar LSPR [3] [9][10]. Theoretically, LSPR exhibits itself at the same wavelength in extinction and scattering [2]. However, it is valid only for isolated nanoparticles without size variation. In Figure 2 (a)(b) we observe difference in peak positions for extinction and diffuse reflection, while coinciding for extinction and total reflection peaks. Standard explanation of the observed difference is the size variation of plasmon nanoparticles. Scattering cross-section is higher for larger nanoparticles possessing lower frequency LSPR, while extinction cross-section is higher for smaller nanoparticles having LSPR at higher frequency. As a result, extinction and scattering peaks are separated. The same argument is used for explaining an increased full width at a half maximum (FWHM) of peaks in comparison with calculated ones [2]. Peak asymmetry is usually explained by shape deviation of nanoparticles from sphere to ellipsoid. It results in splitting of one LSPR in two separate resonances (redshifted and blueshifted), which can lead to observable shape of dipolar peak.

UV resonance manifestation is usually attributed to valley near 360 nm in total reflection as well as to peak at 350 nm in extinction [6][9][11] and is ascribed to quadrupolar resonance. There is also a peak at 330 nm in total reflection clearly visible in low dose sample (Figure 2 (b)). As a whole, UV features are more intense in low dose sample than in high dose one, but dipolar peak intensity is practically the same in both samples.

In Figure 2 scattering spectra are also shown. In comparison with total reflectance, for which wavelength of maximum is independent on IBM dose, difference in position of diffuse reflection peaks is 20 nm for the low and high dose samples. Intensity of scattering is 20 times less than intensity of total reflection and we can equate it with specular reflection. The samples scatter most of radiation in direction of specular reflection. It is only possible, if all radiating points, i.e., silver nanoparticles work in phase and have similar radiation patterns. If we consider our samples as diffracting gratings, then specular reflection is possible at the zero-order grating condition on the period Λ, which is expressed as [12]

$$\Lambda < \frac{\lambda}{n \sin \theta + n}, \quad (1)$$

where λ is wavelength, θ is incident angle and n is refractive index of ambient. For θ = 8º and n = 1 the inequality (1) is simplified to $\Lambda < \lambda$. This condition is fulfilled for all wavelengths in our experiments and provides specular reflection of the arrays despite of scattering of any separate nanoparticle. Phase shift $\Delta\varphi$ appearing between incident and emitted radiation can be calculated as [1]

$$\Delta\varphi = arctan \frac{2\beta\omega}{\omega_0^2 - \omega^2}, \quad (2)$$

where β is a damping constant, ω_0 is the plasmon resonance frequency and ω is the frequency of incident wave. According to (2), $\Delta\varphi$ is changed from 0º at low frequency to 180º at high frequency and is equal 90º at the resonance wavelength. In irregular array, we can consider the most probable phase shift $\Delta\varphi_A$ of the whole array and variable phase shift $\Delta\varphi_P$ of an individual particle. If $\Delta\varphi_A$ and $\Delta\varphi_P$ are different for the same wavelength, then the particle is atypical and contributes to scattering, in the opposite case the particle takes part in specular reflection. The largest difference $\Delta\varphi_A$ - $\Delta\varphi_P$ happens at resonance wavelength of the atypical nanoparticle and leads to highest scattering intensity of the nanoparticle. Therefore, the peak of diffuse reflection indicates wavelength of LSPR for ensemble of atypical particles, which shape and size are far away from the most probable ones. This peak is redshifted relatively specular LSPR due to larger size and asymmetrical shape of atypical particles. Therefore, diffuse reflection is more sensitive to shape and size fluctuations of nanoparticles than specular reflection. In Figure 2 (a)(b) the low dose sample has higher intensity and redshift of scattering than the high dose sample, because with increasing of IBM dose particle shape variation diminishes and particle size distribution converges to average size. As a result of analysis of Figure 2 (a)(b), we can conclude that extinction and diffuse reflection both demonstrate positions of LSPRs. However, these positions are attributed to LSPRs of different nanoparticles, they do not coincide and the difference exceeds 50 nm.

The low dose sample was additionally treated by reactive ion etching (RIE) as reported elsewhere [5][13][14]. As a result, oxide between Ag nanoislands was removed and SiO_2 pillars with a height of 50 nm were fabricated. Ag nanoparticles were left at the top of pillars. The purpose of experiment was to change the dielectric environment and to reduce possible coupling between nanoparticles. The obtained spectra of pillar sample are shown in Figure 2 (c). Additionally, in Figure 2 (b)(c) is also shown absorption $A = 1 - T - R$, where T and R are transmittance and reflectance, respectively. The dipolar extinction, absorption and scattering peaks were blueshifted on 30 - 50 nm, due to replacing SiO_2 (ε = 2.5) between nanoparticles by air (ε = 1). However, the peak of total reflection was left at the same position. Intensity of extinction and total reflection decreased 1.5 and 3 times, respectively, but intensity of absorption and scattering increased 2 and 2.5 times, respectively. The UV peak of absorption and extinction (peak and valley in reflection) was replaced by shoulder in absorption (extinction) and broad depression in reflection. Based on absorption results in VIS part (Figure 2 (b)(c)) we conclude that above mentioned deviation between absorption LSPR and reflection LSPR is attributed to different sensitivity of both methods to coupling of nanoparticles. We suppose that short wavelength dipolar LSPR (around 420 nm) visible in absorption corresponds to

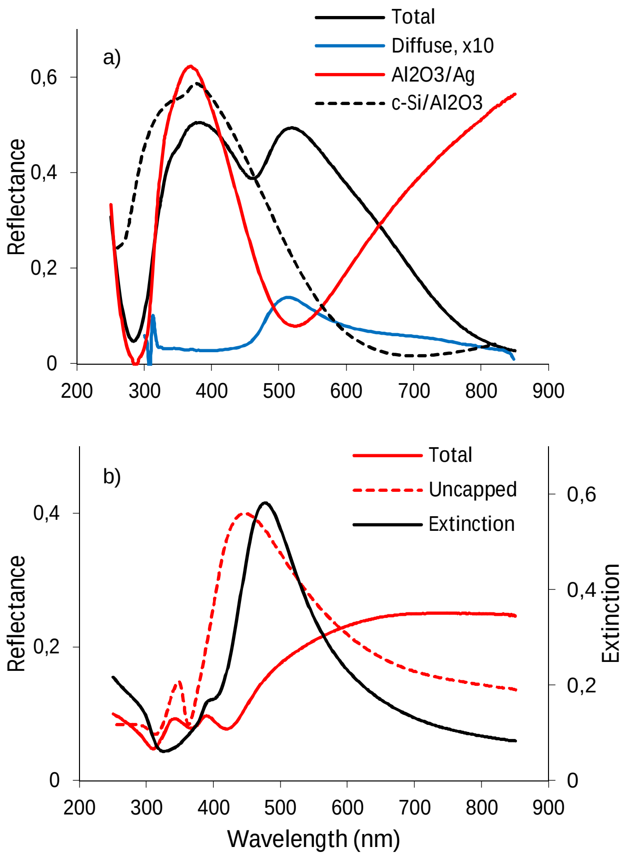


Figure 3. Spectra of Xe mixed Ag nanoparticles on 100 nm Al_2O_3 /*c*-Si (a) and of high dose Ar mixed nanoparticles on quartz, capped by 73 nm SiO_2/15 nm Ag (b).

resonance of isolated, not coupled dipoles (nanoparticles). The long wavelength LSPR (around 500 nm) visible in total reflection corresponds to coupled dipoles, where electromagnetic field is concentrated between dipoles in SiO_2. Removing of SiO_2 causes significant increase of absorption and decrease of reflection, because coupled particles transform in isolated ones. However, large nanoparticles at a small distance from each other remain coupled after RIE. They contribute in total reflection, but at longer wavelengths due to size dependence of LSPR. As a result, this redshift compensates the blueshift due to decrease of ε and peak wavelength of total reflection is not changed.

One more variation of dielectric environment was done by covering (capping) of Ag nanoparticles by oxide layer. It was realized by IBM of silver layer covered by 12 nm of SiO_2, which resulted in Ag nanoparticles embedded inside of SiO_2 matrix. Figure 2 (d) shows spectra of Ag nanoparticles capped by SiO_2. Intensity of VIS extinction is a little higher than in uncapped sample, due to higher amount of silver (capping layer prevents Ag sputtering during IBM). Intensity of total reflection increases due to stronger coupling between nanoparticles in the medium with higher ε. It is known that radiation of dipole pair (dimer) is more intensive than simple sum of isolated dipole radiations [4]. The more dimers appear after capping of nanoparticles, the higher is radiation intensity. The same effect with opposite sign decreases the total reflection of RIE processed samples (Figure 2 (c)). Coupling also makes VIS peaks much wider, due to increased extinction and reflection at long wavelengths. Scattering in capped sample is weak due to total internal reflection in SiO_2 layer. Absorption band near 400 nm is quite broad, because it is attributed to combine effect of redshifted quadrupolar resonance (390 nm in SiO_2) and dipolar LSPR.

B. Ag Nanoparticles on a Multilayer Substrate

In the subsection III-A, we demonstrated that substrate facilitates coupling between nanoparticles. It, in turn, increases intensity of radiation and results in splitting of one-peak LSPR in blueshifted (transversal) LSPR and redshifted (longitudinal) LSPR. Furthermore, multilayer substrate provides additional reflection interfaces and corresponding interference patterns in reflection. We start analysis of multilayer substrates from demonstration of phase shifting during LSPR. Figure 3 (a) shows spectra of Xe mixed Ag nanoparticles on 100 nm thick Al_2O_3 layer above *c*-Si substrate. Here is also given reflectance spectrum of 15 nm thick Ag film above Al_2O_3 before IBM. This spectrum demonstrates transparency of 15 nm thick silver film and high quality interference in Al_2O_3 optical cavity with Si and Ag mirrors. The minimum at 290 nm and maximum at 380 nm are close to theoretical interference extrema, calculated with bulk silver optical constants. After formation of nanoislands, the minimum at 530 nm is replaced by new one at 460 nm, which is quite close to wavelength of scattering peak at 510 nm, i.e., LSPR of Ag nanoparticles. When in the optical cavity one of the mirrors is replaced with plasmonic structure, the phase shift balance for extrema is [10]

$$2\Delta\varphi_{prop} + \Delta\varphi_{refl} + \Delta\varphi_{pl} = N\pi, \tag{3}$$

where $\Delta\varphi_{prop}$ is the phase shift due to propagation of the wave through cavity, $\Delta\varphi_{refl}$ is the phase shift upon reflection at the cavity mirror, $\Delta\varphi_{pl}$ is the phase shift on plasmon structure, N is integer. Wavelength of local minimum in Figure 3 (a) is 460 nm, what corresponds $2\Delta\varphi_{prop} = 1.47\pi$ ($n_{Al2O3} = 1.69$). Additional phase shift $\Delta\varphi_{pl} = 0.53\ \pi$ provides condition for destructive interference ($\Delta\varphi_{refl} = \pi$) at 460 nm according to (3). This interference happens between light waves emitted by upper and lower lobes of dipole radiation pattern. These lobes are coherent and in phase immediately after scattering. Then, the lower wave radiated in the substrate acquires additional phase shift during propagation and reflection and interferes with upper wave. As a result, presence of optical cavity splits LSPR in two local extrema visible in total reflection. In contrast to extinction, where LSPR exhibits only one peak, the reflection spectrum demonstrates peak and valley at frequencies below and higher than resonance one. For example, in Figure 3 (a) LSPR is redshifted on 0.03π from the local minimum, because $\Delta\varphi_{pl} = 0.53\ \pi$ and $\Delta\varphi = 0.5\ \pi$ at the resonance.

To validate this suggestion for quadrupolar LSPR, the high dose Ar mixed Ag nanoparicles on quartz substrate were covered by 73 nm thick SiO_2 layer and 15 nm thick silver.

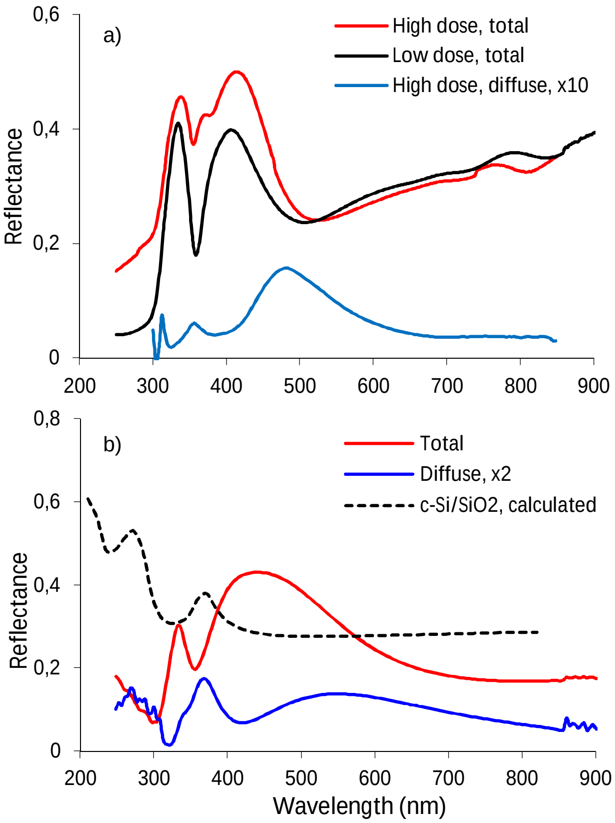


Figure 4. Spectra of Ar mixed Ag nanoparticles on 20 nm SiO_2 /*c*-Si (a) and of annealed Ag nanoparticles on 36 nm SiO_2 /*c*-Si (b).

Total reflection for uncapped nanopaticles, as well as extinction and total reflection after deposition of SiO_2/Ag are shown in Figure 3 (b). Without capping oxide the obtained reflectance spectrum is similar to spectrum in Figure 2 (b). Capping of Ag nanoparticles by SiO_2 and silver results in interference picture and reduced intensity of reflection in UV part. The minimum at 430 nm is connected with destructive interference in the capping SiO_2 layer, i.e., $2\Delta\varphi_{\text{prop}} = \pi$. According to extinction spectra (Figure 3 (b)), the quadrupolar LSPR is redshifted in SiO_2 and takes place at 390 nm. Phase shift of this LSPR $\Delta\varphi_{\text{pl}}$ transfers uprising trend in reflection after minimum at 430 nm in decreasing trend between 395 nm and 370 nm. At wavelength of local minimum 370 nm $2\Delta\varphi_{\text{prop}}$ =1.15 π, $\Delta\varphi_{\text{refl}} = \pi$ at the uppermost Ag film. It means that additional phase shift 0.85 π due to LSPR will bring reflection to destructive interference. The quadrupolar LSPR is located at phase distance 0.85 π -0.5 π = 0.35 π from the minimum at 370 nm, i.e., between 370 and 395 nm.

Figure 4 (a) demonstrates total and diffuse reflection for Ar mixed Ag nanoparticles prepared on Si substrate covered by 20 nm of SiO_2. The fabrication procedure is the same as for samples on quartz substrate in Figure 2 (a)(b). The obtained spectra look similar to spectra in Figure 2 (b) in UV part. They contents the same peak 330 nm and valley 360 nm. Nevertheless, VIS parts of the spectra in Figure 2 (b) and Figure 4 (a) are different. There are two reasons for this: splitting of LSPR due to dipole coupling and reflection from SiO_2/Si interface. According to position of the scattering peak LSPR of isolated nanoparticles happens at wavelength shorter than 480 nm. It provides rise of reflection around 400 nm. The increase of total reflection for wavelengths more than 510 nm is explained by hybridized longitudinal LSPR. The valley at 510 nm is overlapping point of coupled and isolated resonances. Phase shift increasing of isolated LSPR takes place between 410 nm and 510 nm. At the wavelength of 410 nm, the $2\Delta\varphi_{\text{prop}} = 0.29\pi$ ($n_{\text{SiO2}} = 1.47$) and $\Delta\varphi_{\text{refl}} = \pi$. Additional phase shift $\Delta\varphi_{\text{pl}} = 0.71\ \pi$ leads to peak at 410 nm. Again, LSPR in reflection can be found between local extrema, at the phase distance 0.21 π from the peak. The reflection spectrum of high dose sample demonstrates one more quadrupolar resonance at 390 nm (compare with extinction in Figure 3 (b)). It is attributed to Ag nanoparticles submerged in SiO_2 during IBM [9].

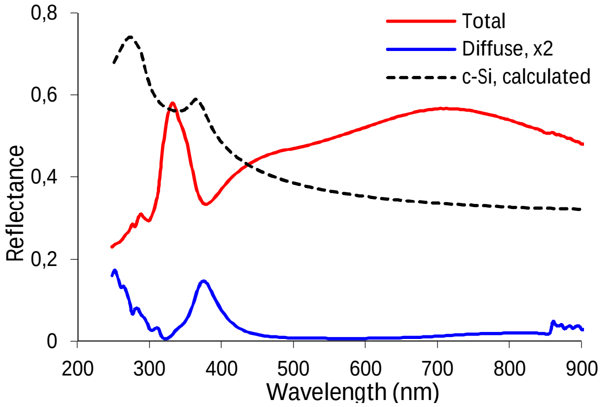


Figure 5. Spectra of annealed Ag nanoparticles on bare *c*-Si (a).

Spectra of Ag nanoparticles prepared by annealing on oxidized Si substrate (36 nm thick SiO_2) are given in Figure 4 (b). They are similar to spectra in Figure 4 (a) in UV part. The VIS part of Figure 4 (b) has a broader dipolar peak and redshifted valley as compared with similar features in Figure 4 (a). Dipole coupling happens mainly through Si substrate, due to higher ε of silicon. This means that thicker oxide decreases coupling and LSPR splitting. It can be observed as less separation of LSPR wavelengths for isolated and coupled nanoparticles. As a result, two peaks are transformed in one broad peak. The trough is redshifted, because the longitudinal LSPR is blueshifted.

Figure 5 demonstrates reflection and scattering of silver nanoparticles on bare silicon substrate. This sample was prepared by annealing. It has the same spectrum features as IBM samples on quartz (Figure 2 (b)), only UV minimum of reflection is redshifted to 380 nm. Due to higher ε dipolar coupling is very strong and broad range of split LSPRs from 450 nm to 900 nm is observed.

C. UV features of the spectra

All studied samples, exclude the 100 nm Al_2O_3/Ag and capped ones, demonstrate 330 nm peak and 360 nm trough in the UV part of reflection spectra. Moreover, in the RIE processed sample, both features were observed in as prepared nanoparticle array and disappeared after RIE (Figure 2 (c)(d)).

Therefore, existing of reflective surface with phase shift close to π below Ag nanoparticles is essential for obtained results. Wavelength of UV peak (330 nm) does not depend on nanostructure shape and substrate ε. The peak can be attributed to variation of reflection phase at Ag/air and Ag/substrate interfaces according with wavelength and coincides with maximum of silver refractive index n_{Ag}. Replacing air by SiO_2 (capping) or reflective surface by pillars changes reflection conditions and the mentioned UV features disappear.

Position of UV peak in absorption (360 nm) strictly coincides with the peak of diffuse reflection (Figure 2 (b), Figure 4 and Figure 5). It means that this feature is a quadrupole LSPR [10]. However, its position weakly depends on dielectric environment and for Si substrate with high ε valley is moved only to 375 nm (Figure 5). At the same time, quadrupolar LSPR of submerged in oxide Ag particles is shifted to 390 nm (Figure 4 (a)). Additionally, extinction coefficient of silver has minimum and reflection has maximum at the same wavelength 360 nm (see calculated spectra at Figure 4 (b) and Figure 5), which in turn, facilitates destructive interference. We believe that all these factors contribute in stable position of 360 nm valley.

IV. CONCLUSIONS

We have demonstrated that peaks and valleys in reflectance spectrum of nanoparticles on multilayer substrates do not correspond directly to plasmon resonances. However, it is possible to identify position of LSPR with accuracy of FWHM of resonance band. This position is situated between two local extrema of reflectance spectrum, corresponding phase shift variation during LSPR. The spectrum features in UV range may be attributed either to quadrupolar resonance or to variation of reflection phase. In the first case, the valley position depends on dielectric environment and geometry of plasmonic structures. In the second case, the peak is fixed at 330 nm and observed only in nanostructures having silver/air interface and formed on smooth reflecting surface. The obtained results can be used in analysis and design of plasmonic nanostructures on opaque substrates.

ACKNOWLEDGMENT

This research was undertaken at the Micronova Nanofabrication Centre, supported by Aalto University.

REFERENCES

[1] R. Ameling et al., “Cavity-enhanced localized plasmon resonance sensing”, Appl. Phys. Lett., v.97, 2010, pp. 253116-1 – 253116-3.

[2] E. C. Le Ru and P. G. Etchegoin, “Principles of Surface-Enhanced Raman Spectroscopy and Related Plasmonic Effects”, Elsevier, 2008, 688 p.

[3] S.Pillai, K. R. Catchpole, T. Trupke, and M.A. Green, “Surface plasmon enhanced silicon solar cells”, J. Appl. Phys., v.101, 2007, pp. 093105-1 - 093105-8.

[4] P. Biagioni, J.-S. Huang, and B. Hecht, “Nanoantennas for visible and infrared radiation”, Rep. Prog. Phys., v.75, 2012, pp. 024402-1 - 024402-40.

[5] 5. A. Shevchenko, V. Ovchinnikov, and A. Shevchenko “Large-area nanostructured substrates for surface enhanced Raman spectroscopy”, Appl. Phys. Lett., v. 100 (17), 2012, pp. 171913-1 - 171913-4.

[6] V. Ovchinnikov and A. Priimagi, “Anisotropic Plasmon Resonance of Surface Metallic Nanostructures Prepared by Ion Beam Mixing”, Proceedings of the First International Conference on Quantum, Nano and Micro Technologies (ICQNM'07), 2-6 January 2007, Guadeloupe, French Caribbean, IEEE XPlore Digital Library, ISBN: 978-1-4244-3131-1, icqnm, 2007, pp. 3-8.

[7] V. Ovchinnikov, “Analysis of Furnace Operational Parameters for Controllable Annealing of Thin Films”, Proceedings of the Eighth International Conference on Quantum, Nano/Bio, and Micro Technologies (ICQNM 2014), November 16 - 20, 2014, Lisbon, Portugal, ThinkMind Digital Library (ISBN: 978-1-61208-380-3), 2014, pp. 32-37.

[8] V. Ovchinnikov, “Effect of Thermal Radiation during Annealing on Self-organization of Thin Silver Films”, Proceedings of the Seventh International Conference on Quantum, Nano and Micro Technologies (ICQNM 2013), August 25-31, 2013, Barcelona, Spain, ThinkMind Digital Library (ISBN: 978-1-61208-303-2), 2013, pp. 1-6.

[9] V. Ovchinnikov, “Formation and Characterization of Surface Metal Nanostructures with Tunable Optical Properties”, Microelectronics Journal, v.39(3-4), 2008, pp. 664-668.

[10] E. Thouti, N. Chander, V. Dutta, and V. K. Komarala, “Optical properties of Ag nanoparticle layers deposited on silicon substrates”, J. Opt., v.15, 2013, 035005-1 - 035005-7.

[11] V. Ovchinnikov and A. Shevchenko, “Surface Plasmon Resonances in Diffusive Reflection Spectra of Multilayered Silver Nanocomposite Films”, Proceedings of the Second International Conference on Quantum, Nano and Micro Technologies (ICQNM 2008), 10-15 February 2008, Sainte Luce, Martinique, IEEE Computer Society Digital Library, ISBN: 978-1-4244-4228-7, icqnm, 2008, pp. 40-44.

[12] Tian Sang et al., “Systematic study of the mirror effect in a poly-Si subwavelength periodic membrane”. J. Opt. Soc. Am. A, 26 (3), 2009, pp. 559-565.

[13] V Ovchinnikov, A Malinin, S Novikov, and C Tuovinen, “Silicon nanopillars formed by reactive ion etching using a self-organized gold mask”, Physica Scripta (T79), 1999, pp. 263-265.

[14] A. Shevchenko and V. Ovchinnikov, “Magnetic Excitations in Silver Nanocrescents at Visible and Ultraviolet Frequencies”, Plasmonics, 4(2), 2009, pp. 121-126.

Clamped-Clamped Microbeam Resonators of Enhanced Higher Order-Modes Response and Wide Bandwidth

Nizar R. Jaber, Abdallah Ramini, Mohammad I. Younis
Physical Science and Engineering Division
King Abdullah University of Science and Technology (KAUST)
Thuwal 23955-6900, Saudi Arabia
e-mail: Nizar.jaber@kaust.edu.sa, Abdallah.ramini@kaust.edu.sa, Mohammad.younis@kaust.edu.sa

***Abstract*—In this study, we present an experimental investigation of electrically actuated clamped-clamped microbeam resonators. The objective is to excite the higher order modes of the microbeams using partial electrodes with shapes that induce strong excitation of the mode of interest. The devices are fabricated using polyimide as a structural layer coated with Nickel from top and Chrome and Gold layers from bottom. Using a high frequency laser Doppler vibrometer, the first three resonance frequencies are revealed via white noise signals. Then, we studied the nonlinear dynamics of the microbeams near these resonance frequencies by applying forward frequency sweep with different electro dynamical loading conditions. The reported results prove the ability to excite higher order modes effectively using partial electrodes. Using a half electrode, the second mode is excited with high amplitude compared with almost zero response using the full electrode. Also, we present an experimental study of an electrically actuated clamped-clamped microbeam under a two-source harmonic excitation. The first frequency is swept around the first and second mode of vibration where the second one is fixed. The excitation of additive and subtractive type resonances is highlighted. In addition, we show that by properly tuning the frequency and the amplitude of the excitation force, the frequency bandwidth of the resonator is increased. Such micro-resonator is shown to be promising in gas and mass detection applications.**

Keywords-Electrostatic; multifrequency; resonator; higher order modes.

I. INTRODUCTION

Micro-electromechanical systems (MEMS) are devices and technologies that have been evolved from the microelectronics industry. Researchers have investigated different ways to create MEMS devices using microelectronic fabrication method. MEMS devices have attractive features, such as the smaller size, the ability to work in harsh environment, and power efficiency [1][2]. In particular, MEMS resonators composed mainly of microbeams are the main building block of many MEMS sensors and actuators that are used in variety of applications,such as toxic gas sensors [3], mass and biological sensors [4-7], temperature sensors [8], force and acceleration sensors [9], and earthquake detectors [10]. MEMS resonators are excited using different types of forces, such as piezoelectric [11][12], thermal [13], and electrostatic [10][14].

Electrostatic and electrodynamic excitations are the most commonly used method because of simplicity and availability [14][15]. The nonlinear dynamics of electrostatically actuated resonator is studied intensively and thoroughly in [14- 20].

MEMS resonators excited near their higher order modes have been proposed for mass and gas detection. Exciting the resonators near their higher order modes improve the sensitivity and the quality factor of the mass sensor. In [12], the sensitivity S_n and the quality factor Q_n of a resonant cantilever is defined as

$$S_n = \frac{\omega'_n - \omega_n}{m} \approx \frac{-\omega_n}{2m_{eff}} \qquad (1)$$

$$Q_n = 3\pi b h \omega_n (256\mu)^{-1} \qquad (2)$$

where m is the cantilever mass, m_{eff} is the n^{th} mode effective mass of the cantilever, ω_n is the resonance frequency of the cantilever, ω'_n is the final resonance frequency after detecting a mass, b is the beam width, h is the thickness and μ is the air viscosity. As noticed from (1) and (2) the sensitivity and the quality factor are directly proportional to the excited mode number. High quality factor implies a sharper and stable resonance peak. This can be achieved through increasing ω_n and decreasing m_{eff} ; both are achieved through high-order mode excitations.

There is an increasing demand to develop resonant sensors with large frequency band and low power consumptions [3][4][21]. An efficient approach to improve the vibration of resonator and increase the frequency band is to use parametric excitation [16], secondary resonance [21], slightly buckled resonators [22] and multi frequency excitation [23]. A device with tunable resonant frequency for energy harvesting application is designed and tested in [24]. They increased the resonant frequency band up to ±20% of the original resonant frequency using a permanent magnet. The effect of the double potential well systems on the resonant frequency band and their application in energy harvesting application is reviewed in [25]. Also, Cho et al. [26] they designed and characterized a carbon nanotube based nano-resonator for mass detection application, they proved that the resonator bandwidth is directly proportional to the forcing amplitude.

Recent research highlighted the interesting dynamics of mixed frequency excitation and their applications as sensors and actuators. Mixed frequency excitation of a micro mirror is studied intensively by Ilyas et al. [23]. They proposed mixed frequency excitation as way to improve bandwidth in resonators. Parametrically and harmonically excited a micro-ring gyroscope at two different frequencies is investigated in [27]. Using this method, they increased the signal to noise ratio and improved the performance of the gyroscope. In [28], they fabricated and characterized a device that harvests electromagnetic energy at the three different resonance frequencies of vibration. Moreover, the method of multi frequency excitation is implemented in [29] to perform mechanical logic operation where each frequency carries a different bit of information. In [30], the frequency mixing is exploited on the atomic force microscope (AFM) resonator to generate a high resolution imaging and extract the surface properties.

Our goal is to excite the higher modes by changing the lower electrode configuration such that it, to some extent, resembles the excited mode shape of the clamped-clamped micro-beam. As we can see in Figure 1, we use full electrode to excite first mode, half electrode to excite the second mode, and two-third electrodes spaced out along the beam length to excite the third mode.

Motivated by the interesting dynamic behavior of clamped-clamped microstructures excited by a multi-frequency electrical source and their wide range of applications, this paper also investigates the dynamics of the resonator shown in Figure 1 experimentally.

The paper is organized as follows. In Section 2, we present the fabrication process of the microbeams. The characterization setup and methodology used in extracting the data is presented in Section 3. In Section 4, we discuss and report the results to single frequency excitation near the first and second mode of vibration. The response to a multi frequency source is presented in Section 5. Summary of the results and the potential applications is discussed in Section 6.

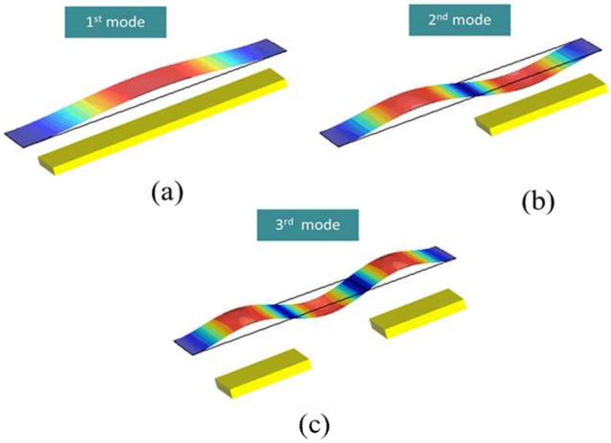


Figure 1. Clamped-clamped mode shapes with different lower electrode configuration. (a) Full electrode. (b) Half electrode. (c) Two-third electrode.

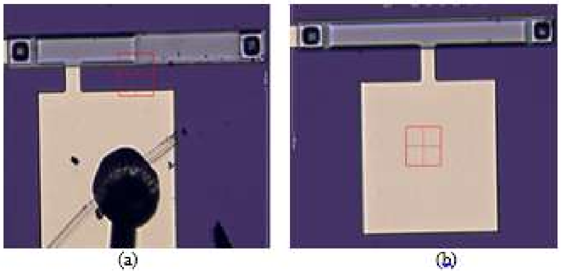
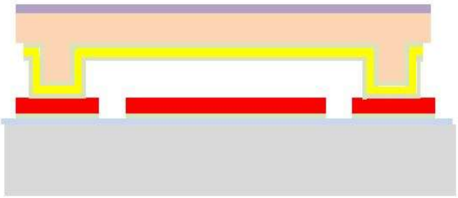


Figure 2. A top view picture of the fabricated microbeam and the actuation pad. (a) half lower electrode configuration. (b) full lower electrode configuration.

II. FABRICATION

The clamped-clamped microstructure resonators are fabricated using the in-house process developed in [31][32]. The process consists of six physical layers shown in Figure 2 and set of rules that defines the allowed configuration and minimum feature size. The microstructure consists of a $6\mu m$ polyimide structural layer coated with $500nm$ nickel layer from top and $50nm$ chrome $250nm$ gold and $50nm$ chrome from bottom. The chrome layer enhances the adhesion properties between gold and another materials The Nickel layer protects the microbeam during the polyimide etch. The lower electrode is placed directly underneath the microstructure and composed of gold and chrome layers. It spans half of the microstructure length and provides the actuation force to the resonator. The two electrodes are separated with a $2\mu m$ air gap. When the two electrodes are connected to an external excitation voltage, the resonator vibrates in the out-of-plane direction. Figure 3 shows a picture illustrating the various layers of the fabricated resonator. A $500\ \mu m\ SiO_2$ is thermally grown to provide insulation and enhance the adhesion properties between the lower electrode and the wafer.



Color	Material	Thickness	Length	Width	Modulus of Elasticity	Density
⬜	Silicon Wafer	$5\ mm$				
🟦	SiO_2	$500\ \mu m$				
🟩	Cr	$50\ nm$	$400\ \mu m$	$40\ \mu m$	$279\ GPa$	$7190\ Kg/m^3$
🟥	Au	$500\ nm$			$79\ GPa$	$19300\ Kg/m^3$
🟨	Au	$250\ nm$	$400\ \mu m$	$40\ \mu m$	$79\ GPa$	$19300\ Kg/m^3$
🟧	Pi	$6\ \mu m$	$400\ \mu m$	$50\ \mu m$	$8.5\ GPa$	$1400\ Kg/m^3$
🟪	Ni	$500\ nm$	$400\ \mu m$	$50\ \mu m$	$200\ GPa$	$8908\ Kg/m^3$

Figure 3. Cross sectional view of the fabricated microbeam.

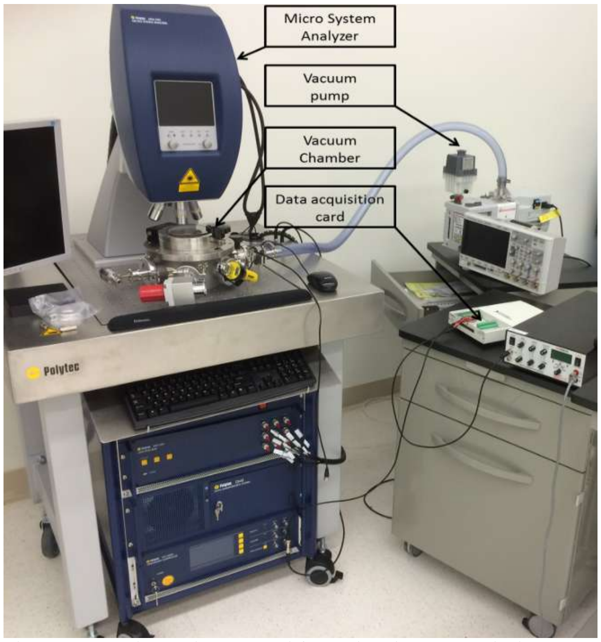


Figure 4. Experimental setup used for testing the MEMS device.

III. CHARECTRIZATION

In this section, we describe the experimental set up used for testing the device and measuring the initial profile, gap thickness and the out-of-plane vibration. The experimental setup consists of a micro system analyzer (MSA) under which the microstructure is placed to measure the vibration, data acquisition cards and amplifier to provide actuation signals of wide range of frequencies and amplitudes, a vacuum chamber equipped with ports to pass the actuation signal and measure the pressure. Also, the chamber is connected to a vacuum pump that reduces the pressure as low as $4mTorr$. The microbeam movement is measured using the laser-Doppler vibrometer. The laser beam was focused onto the microstructure using the microscope and the x-y positioning stage. The measurement is based on interferometry, where the laser beam is split into two beams: one focused on the moving structure and the other one focused on a reference target. The difference between the two beams in phase and distance traveled will be translated into displacement or velocity of the microstructure as a function of the frequency and time. The setup is shown in Figure 4.

A. *Topography Characterization*

The initial profile of the microstructure is revealed using an optical profilometer. After defining the vertical scanning range and exposure time, a 3-D map of the microbeam is generated, as depicted in Figure 5. The combined thickness of the microstructure and air gap is around $9\mu m$, which is slightly smaller from the design nominal value of $9.35\mu m$. Also, the microbeam total length is $400\mu m$ and the profile is fully straight without any curvature or curling.

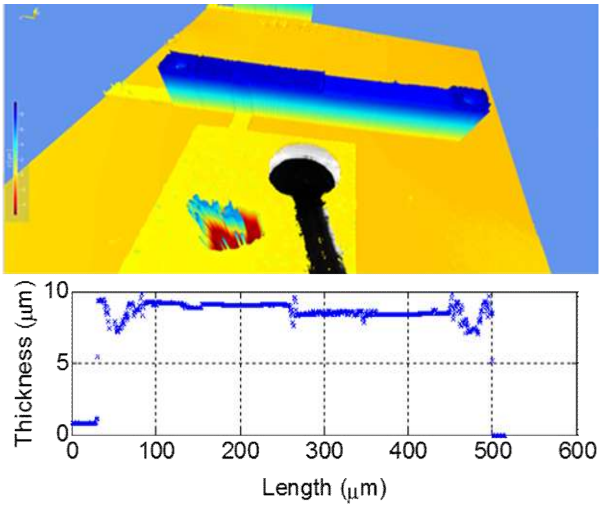


Figure 5. A 3-D map of the microbeam with half lower electrode configuration profile as seen from top.

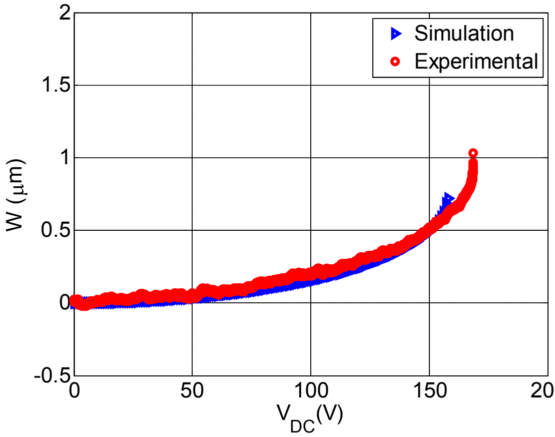


Figure 6. Static profile of the microbeam with half electrode actuation.

B. *Static Characterization*

To characterize the device we initially biased the microbeam by a slow DC ramp voltage - generated using the data acquisition card - and measured the static deflection. The experimental result is reported in Figure 6. The deflection increases until it exhibits pull in at $168V$ for a clamped beam with half electrode actuation.

C. *Natural frequencies*

We experimentally measured the first three natural frequencies by connecting the wire-bonded chip to the MSA function generator and applying a white noise signal. The MSA measures the microstructure vibration at the laser point position. Also, it has the capability to reveal the mode shape of the vibration.

The microstructure is excited with a white noise of $V_{DC} = 30V$ and $V_{AC} = 50V$. The vibration at different points along the beam length is scanned to extract the vibration mode shapes and resonance frequencies. The acquired frequency response curve is shown in Figure 7 and it reveals the values of the first three natural frequencies $\omega_1 = 160\ kHz,$ $\omega_2 =\ 402\ kHz,$ and $\omega_3 = 738\ kHz.$ The acquired vibration mode shapes of the test are reported in Figure 8. We notice at ω_1, as shown in Figure 8(a) all points are vibrating whereas at ω_2 (see Figure 8(b)), the mid points are nodal points. Also, at ω_3 (see Figure 8(c)), there are two nodal points. These results match the clamped-clamped structure first, second and third vibration mode shapes.

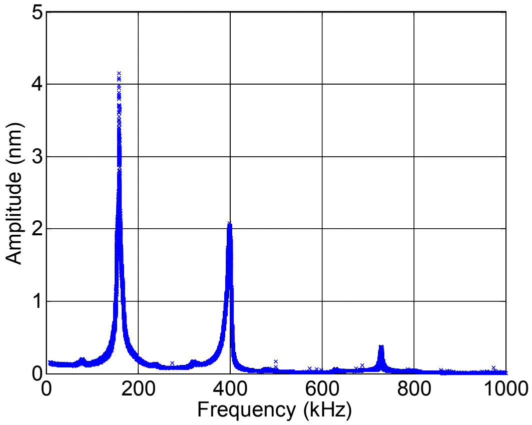
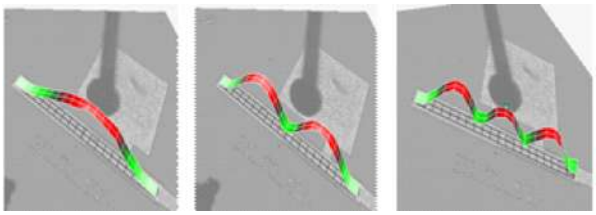


Figure 7. Frequency response curve of the microbeam with half electrode actuation to white noise actuation signal. $V_{DC} = 30V$ and $V_{AC} = 50V$ at $4 mTorr$ chamber pressure.



(a) (b) (c)

Figure 8. The vibrational mode shapes of the microbeam with half lower electrode configuration at (a) $\omega_1 = 160kHz$, (b) $\omega_2 = 402kHz$ and $\omega_3 = 738kHz$.

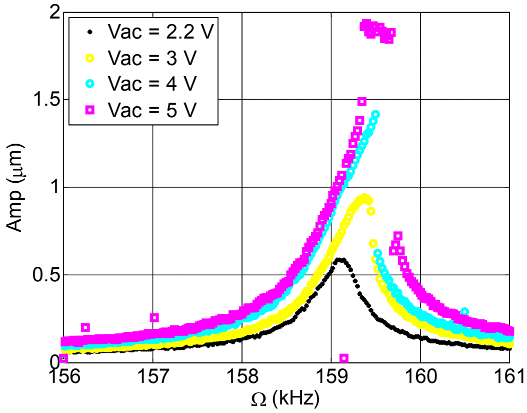


Figure 9. Frequency response curve for the microbeam near the first resonance with full lower electrode configuration. $V_{DC} = 5V$.

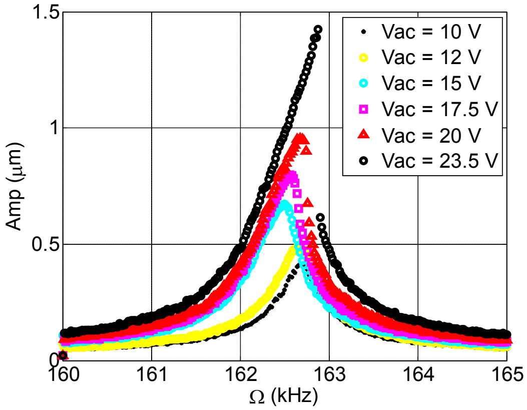


Figure 10. Frequency response curve for the microbeam near the first resonance with half lower electrode configuration. $V_{DC} = 5V$.

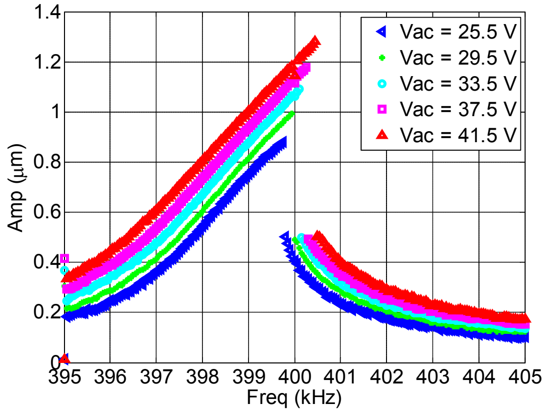


Figure 11. Frequency response curve for microbeam near the second resonance with half electrode actuation. $V_{DC} = 30V$.

IV. HIGHER ORDER MODES RESULTS

We experimentally investigate the nonlinear response of the microbeams near the first two resonance frequencies via frequency sweep tests. The micro-beams are excited using the data acquisition card and the vibration is detected using the MSA. The excitation signal is composed of an AC signal V_{AC} superimposed to a DC signal V_{DC}. The frequency response curve is generated by taking the steady state amplitude of the motion and focusing the laser at the mid-point of the microbeam for the first measurement and at quarter of the beam length for the second mode

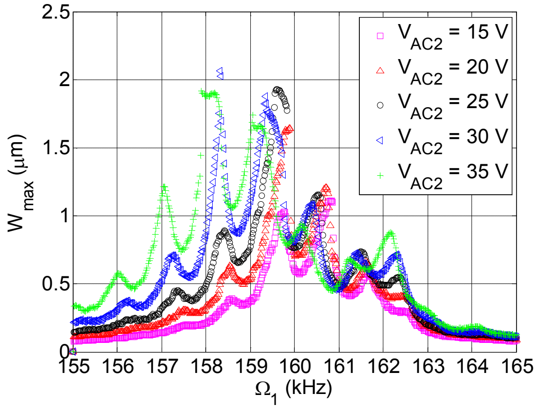


Figure 12. Frequency response curve for $V_{DC} = 15V$, $V_{AC1} = 5V$ and $\Omega_2 = 1kHz$ near the first resonance.

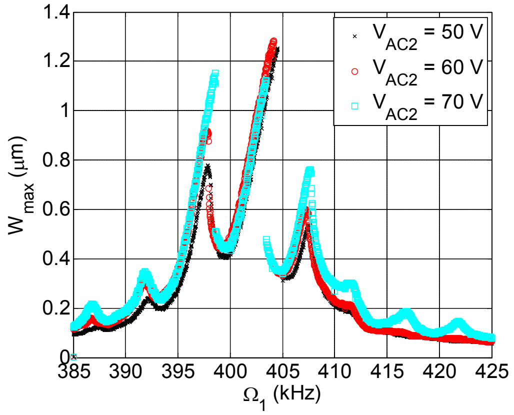


Figure 13. Frequency response curve for $V_{DC} = 15V$, $V_{AC1} = 20V$ and $\Omega_2 = 5kHz$ near the second resonance.

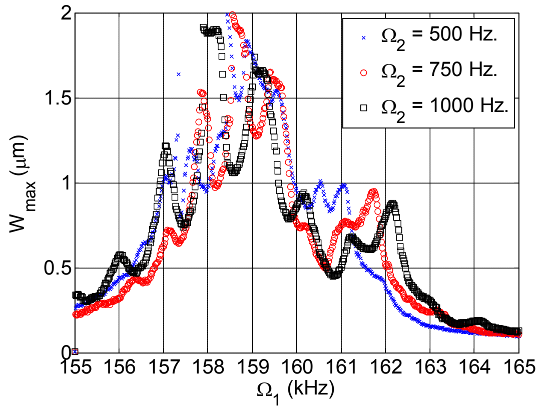


Figure 14. Frequency response curve for different values of Ω_2 at $V_{DC} = 15V$, $V_{AC1} = 5V$ and $V_{AC2} = 35V$ near the first resonance.

measurement. In the following subsection, the frequency response curves are reported for the different microbeams with full and half electrode actuation at different electrodynamical loadings and at $4mTorr$ chamber pressure.

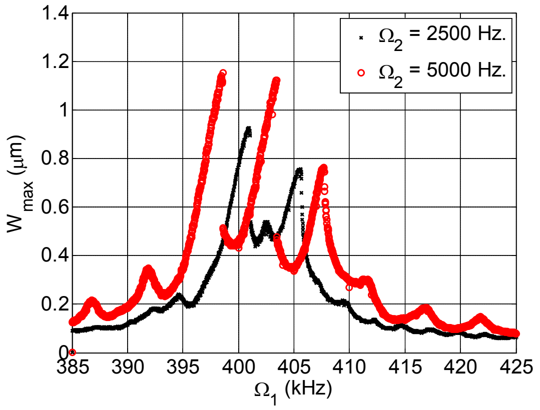


Figure 15. Frequency response curve for different values of Ω_2 at $V_{DC} = 15V$, $V_{AC1} = 20V$ and $V_{AC2} = 70V$ near the second resonance.

A. First Mode

Figures 9 and 10 show the variation of the frequency response curve as the AC voltage is increased for different lower electrode configuration near the first resonance. As expected, using the full electrode configuration we achieved higher amplitude near the first mode compared with the half electrode configurations under the same conditions of electro dynamical loading and vacuum chamber pressure. Also, a hardening effect is reported due to the cubic nonlinearities from mid-plane stretching.

B. Second Mode

Figure 11 shows the variation of the frequency response curve as the AC voltage is increased near the second mode resonance for the half electrode configuration. Using a half electrode, the second mode is excited with high amplitude compared with no response using the full electrode configuration. In addition, a hardening effect is reported due to the cubic nonlinearities.

V. MULTIFREQUENCY EXCITATION RESULTS

We experimentally investigated the nonlinear response of the microbeams to a multi-frequency excitation near the first and second resonance frequencies. The microbeams are excited using the data acquisition card and the vibration is detected using the MSA. The excitation signal composed of two AC signals V_{AC1} and V_{AC2} of different frequencies Ω_1 and Ω_2 superimposed to a DC signal V_{DC}.

The generated frequency response curves near the first resonance are presented in Figure 12. Each curve shows the frequency response for different values of V_{AC2}. The results are obtained by sweeping the frequency Ω_1 of the first source V_{AC1} around the first mode and fixing the second source V_{AC2} frequency Ω_2 at $1\,kHz$. The swept source voltage V_{AC1} and the DC voltage are fixed at 5 V and 15 V, respectively. Figure13 shows the result of sweeping the first source frequency V_{AC1} around the second mode while fixing the second source frequency V_{AC2} at $5kHz$. The swept source voltage V_{AC1} and the DC voltage are fixed at 20 V and 15 V, respectively. The chamber pressure is fixed at $4mTorr$. The curves highlight the effect of V_{AC2} on the combination resonances where new resonance peaks appear at frequencies of additive type at $(\Omega_1+\Omega_2)$, $(\Omega_1+2\Omega_2)$, $(\Omega_1+3\Omega_2)$ and subtractive type at $(\Omega_1-\Omega_2)$, $(\Omega_1-2\Omega_2)$, $(\Omega_1-3\Omega_2)$. These resonances arise due to the cubic nonlinearity coming from the midplane stretching effect and the electrostatic force. Also, as V_{AC2} increases the response curves tilt towards the lower frequency values (softening). Figure 14 and 15 show the result for different values of Ω_2 under the same electrodynimacal loading condition near the first and second resonance frequencies, respectively. As Ω_2 decreases, a continuous band of high amplitude is formed.

VI. Conclusions

In this paper, we investigated the dynamics of clamped-clamped microbeam with full and half lower electrode configuration. These microbeams are electrically actuated by an AC source with variable frequency superimposed to a DC voltage. We proved the ability to excite the second mode resonance by using a half electrode configuration. Also, a hardening behavior is reported due to the cubic nonlinearities among all the excited modes due to the dominating effect of mid-plane stretching. In addition, we investigated the dynamics of clamped-clamped microbeams electrically actuated by two harmonic AC sources with different frequencies superimposed to a DC voltage. Moreover, the ability to excite the combination resonance of additive and subtractive type is proved.

This capability of exciting the higher order modes can have a promising application in MEMS-based mass, gas and humidity sensors. In addition, the ability to broaden the bandwidth of the resonator is shown by reducing the frequency of the fixed source. These capabilities of generating multiple peaks and a high response band with ability to control its amplitude and location can have a promising application in increasing the resonator band width, mechanical logic circuits, energy harvesting and mass sensing.

References

[1] B. Kim, et al., "Using MEMS to build the device and the package," in Solid-State Sensors, Actuators and Microsystems Conference, 2007. TRANSDUCERS 2007. International, 2007, pp. 331-334.

[2] B. Kim, et al., "CMOS Compatible Wafer-Scale Encapsulation MEMS Resonators," in ASME 2007 InterPACK Conference collocated with the ASME/JSME 2007 Thermal Engineering Heat Transfer Summer Conference, 2007, pp. 499-504.

[3] S. Subhashini and A. Vimala Juliet, "Toxic gas sensor using resonant frequency variation in micro-cantilever," in Sustainable Utilization and Development in Engineering and Technology (STUDENT), 2012 IEEE Conference on, 2012, pp. 87-91.

[4] S. Dohn, R. Sandberg, W. Svendsen, and A. Boisen, "Enhanced functionality of cantilever based mass sensors using higher modes," Applied Physics Letters, vol. 86, pp. 233501, 2005.

[5] E. Gil-Santos, et al., "Nanomechanical mass sensing and stiffness spectrometry based on two-dimensional vibrations of resonant nanowires," Nature nanotechnology, vol. 5, pp. 641-645, 2010.

[6] M. Hanay, et al., "Single-protein nanomechanical mass spectrometry in real time," Nature nanotechnology, vol. 7, pp. 602-608, 2012.

[7] T. P. Burg, et al., "Vacuum-packaged suspended microchannel resonant mass sensor for biomolecular detection," Microelectromechanical Systems, Journal of, vol. 15, pp. 1466-1476, 2006.

[8] W.-T. Hsu, J. R. Clark, and C. T.-C. Nguyen, "A resonant temperature sensor based on electrical spring softening," in Tech. Dig., 11th Int. Conf. on Solid-State Sensors Actuators (Transducers' 01), Munich, Germany, 2001, pp. 1484-1487.

[9] H. C. Kim, S. Seok, I. Kim, S.-D. Choi, and K. Chun, "Inertial-grade out-of-plane and in-plane differential resonant silicon accelerometers (DRXLs)," in Solid-State Sensors, Actuators and Microsystems, 2005. Digest of Technical Papers. TRANSDUCERS'05. The 13th International Conference on, 2005, pp. 172-175.

[10] A. Ramini, K. Masri, and M. I. Younis, "Electrostatically actuated resonant switches for earthquake detection," in Mechatronics and its Applications (ISMA), 2013 9th International Symposium on, 2013, pp. 1-7.

[11] B. Piekarski, et al., "Fabrication of suspended piezoelectric microresonators," Integrated Ferroelectrics, vol. 24, pp. 147-154, 1999.

[12] D. Jin, et al., "High-mode resonant piezoresistive cantilever sensors for tens-femtogram resoluble mass sensing in air," Journal of Micromechanics and Microengineering, vol. 16, pp. 1017, 2006.

[13] G. Rinaldi, M. Packirisamy, and I. Stiharu, "Quantitative boundary support characterization for cantilever MEMS," Sensors, vol. 7, pp. 2062-2079, 2007.

[14] M. I. Younis, MEMS Linear and Nonlinear Statics and Dynamics: Mems Linear and Nonlinear Statics and Dynamics vol. 20: Springer Science & Business Media, 2011.

[15] J. F. Rhoads, S. W. Shaw, and K. L. Turner, "Nonlinear dynamics and its applications in micro-and nanoresonators," Journal of Dynamic Systems, Measurement, and Control, vol. 132, pp. 034001, 2010.

[16] R. Kalyanaraman, G. Rinaldi, M. Packirisamy, and R. Bhat, "Equivalent area nonlinear static and dynamic analysis of electrostatically actuated microstructures," Microsystem technologies, vol. 19, pp. 61-70, 2013.

[17] F. M. Alsaleem, M. I. Younis, and H. M. Ouakad, "On the nonlinear resonances and dynamic pull-in of electrostatically actuated resonators," Journal of Micromechanics and Microengineering, vol. 19, p. 045013, 2009.

[18] J. F. Rhoads, V. Kumar, S. W. Shaw, and K. L. Turner, "The non-linear dynamics of electromagnetically actuated microbeam resonators with purely parametric excitations," International Journal of Non-Linear Mechanics, vol. 55, pp. 79-89, 2013.

[19] M. Younis and A. Nayfeh, "A study of the nonlinear response of a resonant microbeam to an electric actuation," Nonlinear Dynamics, vol. 31, pp. 91-117, 2003.

[20] M. Younis, "Multi-mode excitation of a clamped–clamped microbeam resonator," Nonlinear Dynamics, vol. 80, pp. 1531-1541, 2015/05/01 2015.

[21] A. H. Nayfeh and M. I. Younis, "Dynamics of MEMS resonators under superharmonic and subharmonic excitations," Journal of Micromechanics and Microengineering, vol. 15, p. 1840, 2005.

[22] M. Bagheri, M. Poot, M. Li, W. P. Pernice, and H. X. Tang, "Dynamic manipulation of nanomechanical resonators in the high-amplitude regime and non-volatile mechanical memory operation," Nature nanotechnology, vol. 6, pp. 726-732, 2011.

[23] S. Ilyas, A. Ramini, A. Arevalo, and M. I. Younis, "An Experimental and Theoretical Investigation of a Micromirror Under Mixed-Frequency Excitation." (DOI:10.1109/JMEMS.2014.2386285).

[24] V. R. Challa, M. Prasad, Y. Shi, and F. T. Fisher, "A vibration energy harvesting device with bidirectional resonance frequency tunability," Smart Materials and Structures, vol. 17, pp. 015035, 2008.

[25] R. Harne and K. Wang, "A review of the recent research on vibration energy harvesting via bistable systems," Smart Materials and Structures, vol. 22, pp. 023001, 2013.

[26] H. Cho, M.-F. Yu, A. F. Vakakis, L. A. Bergman, and D. M. McFarland, "Tunable, broadband nonlinear nanomechanical resonator," Nano letters, vol. 10, pp. 1793-1798, 2010.

[27] B. Gallacher, J. Burdess, and K. Harish, "A control scheme for a MEMS electrostatic resonant gyroscope excited using combined parametric excitation and harmonic forcing," Journal of Micromechanics and Microengineering, vol. 16, pp. 320, 2006.

[28] H. Liu, Y. Qian, and C. Lee, "A multi-frequency vibration-based MEMS electromagnetic energy harvesting device," Sensors and Actuators A: Physical, vol. 204, pp. 37-43, 2013.

[29] I. Mahboob, E. Flurin, K. Nishiguchi, A. Fujiwara, and H. Yamaguchi, "Interconnect-free parallel logic circuits in a single mechanical resonator," Nature communications, vol. 2, pp. 198, 2011.

[30] D. Forchheimer, D. Platz, E. A. Tholén, and D. B. Haviland, "Model-based extraction of material properties in multifrequency atomic force microscopy," Physical Review B, vol. 85, p. 195449, 2012.

[31] L. Marnat, A. A. Carreno, D. Conchouso, M. G. Martinez, I. G. Foulds, and A. Shamim, "New Movable Plate for Efficient Millimeter Wave Vertical on-Chip Antenna," IEEE Transactions on Antennas and Propagation, vol. 61, pp. 1608-1615, 2013.

[32] A. Alfadhel, A. A. Arevalo Carreno, I. G. Foulds, and J. Kosel, "Three-Axis Magnetic Field Induction Sensor Realized on Buckled Cantilever Plate," Magnetics, IEEE Transactions on, vol. 49, pp. 4144-4147, 2013.

On the Robustness of Quantum Key Distribution with Classical Alice (photons-based protocol)

Michel Boyer

Département IRO
Université de Montréal, Canada
Email: boyer@iro.umontreal.ca

Tal Mor

Computer Science Department
Technion, Israel
Email: talmo@cs.technion.ac.il

***Abstract*—Quantum Key Distribution (QKD) with classical Bob, has been suggested and proven robust. Following this work, QKD with classical Alice was also suggested and proven robust. The above protocols are ideal in the sense that they make use of qubits. However, in the past, well-known QKD protocols that were proven robust and even proven unconditionally secure, when qubits are used, were found to be totally insecure when photons are used. This is due to sensitivity to photon losses (e.g., Bennett's two-state protocol) or sensitivity to losses combined with multi-photon states (e.g., the photon-number-splitting attack on the weak-pulse Bennett-Brassard protocol, BB84). Here, we prove that QKD with classical Alice is still robust when photon losses and even multi-photon states are taken into account. Our method can pave the road to robustness and security analysis of various other two-way QKD protocols.**

***Index Terms*—Cryptography; Quantum Mechanics.**

I. Introduction

A two-way Quantum Key Distribution (QKD) protocol in which one of the parties (Bob) uses only classical operations was recently introduced [1,2]. A very interesting extension in which the originator always sends the same state $|+\rangle = (|\mathtt{0}\rangle + |\mathtt{1}\rangle)/\sqrt{2}$ (where $\mathtt{0}$ and $\mathtt{1}$ are used to denote bits, to avoid confusion with the integers 0 and 1 used as occupancy numbers) while in [1] all four states, $|\mathtt{0}\rangle$, $|\mathtt{1}\rangle$, $|+\rangle$ and $|-\rangle = (|\mathtt{0}\rangle - |\mathtt{1}\rangle)/\sqrt{2}$, are sent, is suggested by Zou *et al.* [3]. In both "semi-quantum" key distribution (SQKD) protocols the qubits go from the originator Alice to (classical) Bob and back to Alice. Bob randomly either reflects a received qubit without touching its state (those are deemed CTRL bits), or measures it in the standard (classical) basis and sends back his result as $|\mathtt{0}\rangle$ or $|\mathtt{1}\rangle$ (those are deemed SIFT bits). These two operations, "doing nothing" or measure-resend in the standard (computation) basis, are called classical [1,2] for obvious reasons; in principle, semi-quantum protocols might be simpler to implement than fully quantum ones.

Following [4], we prefer to call the originator in [3] Bob (and not Alice), and to call the classical party Alice: usually in quantum cryptography, Alice is the sender of some non-trivial data, e.g., she is the one choosing the quantum states. The originator in [3] does not have that special role, as the state $|+\rangle$ is always sent (and we could even ask Eve to generate it). The classical person is then the one actually choosing a basis and knowing which of the three state ($|\mathtt{0}\rangle$, $|\mathtt{1}\rangle$, or $|+\rangle$) is sent back to the originator, thus it is natural to name that classical person Alice. We call the originator Bob, and we call the SQKD protocol of Zou et al "QKD with classical Alice". Note that QKD with classical Alice was also suggested, independently of [3], by Lu and Cai [5]. As proven in [4], QKD with Classical Alice (the protocol suggested in [3]) is completely robust against eavesdropping.

Here, we use Fock-space representation to extend the QKD with classical Alice protocol to the important case in which Alice and Bob use photons and not merely ideal qubits. We first extend the proof of robustness to include photon loss, and subsequently, also multi-photon states. To the best of our knowledge, such a general analysis has not yet been provided for any of the (many) two-way QKD protocols, including ping-pong protocols and (experimental and commerical) plug and play protocols, hence our approach and method may have major influence on future robustness and security analysis of QKD. A related (photonic) security analysis was recently done for a different two-way QKD protocol, see Section 4.2 in [6].

Such extensions from qubits to photons are far from trivial; on the contrary, often, robustness is actually lost when trying to deal with photons rather than qubits.

As a first example, in the two-state scheme (known as the Bennett'92 — B92 scheme), when qubits are assumed to be carried by photons, photon losses cause a severe problem: if Eve can replace a lossy channel by a lossless one, she might be able to get full information without causing errors at all, using an "un-ambiguous state discrimination" attack. As a second example, in the four-state scheme (known as the Bennett-Brassard'84 — BB84 scheme), when qubits are assumed to be carried by photons, photon losses combined with multi-photon pulses cause a severe problem: if Eve can replace a lossy channel by a lossless one, and can measure photon numbers (via a non-demolition measurement), she might be able to get full information without causing errors at all [7], using a "photon number splitting" attack.

II. The Fock space notations

The Fock space notations that serve as an extension of a qubit are as follows: in the standard (z) basis, the Fock basis vector $|0,1\rangle$ stands for a single photon in a qubit-state $|0\rangle$ and the Fock basis vector $|1,0\rangle$ stands for a single photon in a qubit-state $|1\rangle$. Naturally, the Hadamard (x) basis qubit-states are given by the superposition of those Fock states so that $[|0,1\rangle \pm |1,0\rangle]/\sqrt{2}$ stand for a single photon in a qubit-state $|\pm\rangle = (|0\rangle \pm |1\rangle)/\sqrt{2}$. The general state of this photonic qubit can then be written as $\alpha|0,1\rangle + \beta|1,0\rangle$, with $|\alpha|^2 + |\beta|^2 = 1$.

This photonic qubit lies in a much larger space called Fock space. The first natural extension is $|0,0\rangle$ that describes the lack of photons (the vacuum state), a case of great practical importance, as it enables dealing properly with photon loss. The next extension of a very high practical importance is that $|2,0\rangle$ describes two (indistinguishable) photons in the same qubit-state $|1\rangle$, $|0,2\rangle$ describes two (indistinguishable) photons in the same qubit-state $|0\rangle$, and $|1,1\rangle$ describes two (in this case, distinguishable) photons, one in the qubit-state $|0\rangle$, and one in the qubit-state $|1\rangle$. This case (a six dimensional space, describing two or less photons) was found very important in the photon number splitting attack [7], as prior to that analysis, experimentalists assumed that the only impact of high loss rate is on the bit-rate and not on security.

In general, if a single photon can be found in two orthogonal states (these are called "modes" when discussing photons), then $|n_1, n_0\rangle$ represents n_1 (respectively n_0) indistinguishable photons in a qubit-state $|1\rangle$ (resp. $|0\rangle$). The numbers n_0 and n_1 are then called the occupation numbers of the two modes. From now on, the notations $|0\rangle \equiv |0,1\rangle$, $|1\rangle \equiv |1,0\rangle$, $|+\rangle = (|0,1\rangle + |1,0\rangle)/\sqrt{2}$ and $|-\rangle = (|0,1\rangle - |1,0\rangle)/\sqrt{2}$ will be used interchangeably. Similarly, since the single photon can also be found in $|0,1\rangle_x \equiv |+\rangle$ and $|1,0\rangle_x \equiv |-\rangle$ (namely, the x basis), then $|n_-, n_+\rangle$ represents n_- (resp. n_+) indistinguishable photons in qubit-state $|-\rangle$ (resp. $|+\rangle$).

More generally, one may consider more than two modes. For instance, the four modes $|n_{1b}, n_{1a}, n_{0b}, n_{0a}\rangle$ are the generalization of qu-quadrit (say a photon in one of two arms a or b, and one of two orthogonal polarizations, denoted 0 or 1).

III. The Classical Alice protocol, Dealing with Losses

The originator Bob sends Alice qubits in the state $|+\rangle$ and keeps in a quantum memory all qubits he received back from her. If Bob does not hold a memory to keep the qubits, he measures them upon reception at random in the standard (z) or the Hadamard (x) basis. Only CTRL bits measured in the x basis, and SIFT bits measured in the z basis, are used. That does not modify the conceptual proof (but in a security proof it would mean that they need to send more qubits to start with). When N qubits have been sent and received, (classical) Alice announces publicly which qubits she reflected (without disturbing them); the originator Bob then checks that he received $|+\rangle$ and not $|-\rangle$ on those positions (CTRL). For the (SIFT) qubits measured by Alice in the standard (classical) $\{|0\rangle; |1\rangle\}$ basis, a sample is chosen to be checked for errors (TEST). The remaining SIFT bits serve for obtaining a final, secure key, via error correction and privacy amplification, as in any conventional QKD protocol.

A. Defining the (limited) "photonic QKD with classical Alice" protocol

The qubits are embedded in the 3-dimensional, 2-mode Fock space containing the qubit states $|1,0\rangle$ and $|0,1\rangle$ and the vacuum state $|0,0\rangle$. The Hilbert space describing Alice+Bob states is (for now) the subspace

$$\mathscr{H}_{AB} = \mathrm{Span}\left(|0,1\rangle, |1,0\rangle, |0,0\rangle\right) \subseteq \mathscr{F} \qquad (1)$$

of the more general 2-mode Fock space ($\mathscr{F}$).

In this photonic protocol, Bob is always sending the $|+\rangle$ state. Losses or vacuum states are modeled by the state $|0,0\rangle$, and thus, we must define Alice's and Bob's operations when such states occur. Losses normally come from the interaction with the environment; as usual, the (worst case) analysis gives Eve total control on the environment. Classical Alice can either SIFT or CTRL [3,4]. In the SIFT mode, Alice's "measurement" is described (WLG) with the adjunction of a probe, extending $\mathscr{H}_{AB}$ to $\mathscr{H}_A \otimes \mathscr{H}_{AB}$, a unitary transformation and a measurement of her probe in the standard basis. Such a description is meant to match the general framework of measurements in quantum information, and may not correspond to the actual physical measurement performed by Alice. Using the Fock-space notations, it is assumed that Alice adds a two-mode probe in a state $|0,0\rangle_A$ to get the state $|0,0\rangle_A |+\rangle_{AB}$. Alice then performs one of the following two operations (with $|n_1,n_0\rangle_{AB}$ in the z, i.e., the standard basis):

$$U_{\text{CTRL}} |0,0\rangle_A |n_1,n_0\rangle_{AB} = |0,0\rangle_A |n_1,n_0\rangle_{AB} \qquad (2)$$

$$U_{\text{SIFT}} |0,0\rangle_A |n_1,n_0\rangle_{AB} = |n_1,n_0\rangle_A |n_1,n_0\rangle_{AB} \qquad (3)$$

then she measures her probe in the standard classical basis and sends Alice+Bob's state to Bob; in the case described by (2) (CTRL) she needs not measure, still the probe and its measurement are added there only to make the description uniform; Bob's original state ($|+\rangle_{AB}$) is reflected back to him, undisturbed. In the case described by (3) (SIFT), Alice gets the outcome $n_1 n_0$, and the state $|n_1,n_0\rangle_{AB}$ is sent to Bob. Note that, in order to analyze the enlarged space of the protocol, we *had to* add the definition of Alice's operation on the added state, $|0,0\rangle_{AB}$. Our choice of $U_{\text{SIFT}} |0,0\rangle_A |0,0\rangle_{AB} = |0,0\rangle_A |0,0\rangle_{AB}$ is the most natural way of extending Alice's SIFT operation, and it thus becomes part of our definition of the protocol "Photonic-QKD with classical Alice".

Naturally, when Bob measures in the classical (z) basis, he also measures the same three states as Alice, $|n_1,n_0\rangle$ with $n_0 + n_1 \leq 1$. However, the space $\mathscr{H}_{AB}$ (1) is also spanned by the orthonormal basis $\{ |+\rangle, |-\rangle, |0,0\rangle \}$, thus Bob (who is not limited to being classical) can perform a measurement in this generalized x basis of the qutrit.

B. Eve's attack on the (photonic) classical Alice protocol

Eve performs her attack in both directions; from Bob to Alice, Eve applies U; from Alice to Bob, Eve applies V. We may assume, WLG, that Eve is using a fixed probe space $\mathscr{H}_E$ for her attacks in both directions. The attack from Bob to Alice produces a state of the form $|E_{01}\rangle |0,1\rangle_{AB} + |E_{10}\rangle |1,0\rangle_{AB} + |E_{00}\rangle |0,0\rangle_{AB}$ (namely $\sum_{n_1,n_0 \ | \ n_0+n_1 \leq 1} |E_{n_1 n_0}\rangle |n_1,n_0\rangle_{AB} \in \mathscr{H}_E \otimes \mathscr{H}_{AB}$), where the $|E_{ij}\rangle$ are non normalized (and potentially non-orthogonal) vectors in $\mathscr{H}_E$. With Alice's probe attached we obtain

$$|\Psi\rangle = \sum_{n_1,n_0 \ | \ n_0+n_1 \leq 1} |E_{n_1 n_0}\rangle |0,0\rangle_A |n_1,n_0\rangle_{AB} \ , \qquad (4)$$

in $\mathscr{H}_E \otimes \mathscr{H}_A \otimes \mathscr{H}_{AB}$. In particular, if Eve does nothing then $|E_{10}\rangle = |E_{01}\rangle = |E_{00}\rangle \equiv |E\rangle$ and the state in Alice+Eve's hands, prior to Alice's operation, is $|E\rangle |0,0\rangle_A |+\rangle_{AB}$.

Going back to the general case, if Alice applies U_{CTRL}, then the state in Eve+Alice hands (after Alice's CTRL action) is still $|\Psi\rangle$. However, if Alice applies U_{SIFT}, the resulting global state in Eve+Alice's hands is

$$\sum_{n_1,n_0 \ | \ n_0+n_1 \leq 1} |E_{n_1 n_0}\rangle |n_1,n_0\rangle_A |n_1,n_0\rangle_{AB} \qquad (5)$$

and after Alice has measured her probe, she gets some output ($\{00,01,10\}$), and some (non normalized) residual state that she sends back to Bob.

Once Alice has performed her measurements and sent $|i,j\rangle_{AB}$ back to Bob via Eve, the resulting global state (fully in Eve's hands) is given by Table I, where the $|\psi_{ij}\rangle$ are not normalized, and where the $|E_{ij}\rangle$ were chosen by Eve. Eve now applies a unitary V on $\mathscr{H}_E \otimes \mathscr{H}_{AB}$ and then sends Bob his part of the resulting state.

TABLE I
GLOBAL STATE AFTER ALICE'S MEASUREMENT

Measurement	*State (non normalized)*
00	$|\psi_{00}\rangle = |E_{00}\rangle |0,0\rangle_{AB}$
01	$|\psi_{01}\rangle = |E_{01}\rangle |0,1\rangle_{AB}$
10	$|\psi_{10}\rangle = |E_{10}\rangle |1,0\rangle_{AB}$
CTRL	$|\psi\rangle = |\psi_{00}\rangle + |\psi_{01}\rangle + |\psi_{10}\rangle$

C. A proof of robustness

For Eve to stay undetectable, if Alice measured $|0,0\rangle$ (namely, the outcome 00) in the SIFT mode, then Bob should have a probability zero of measuring 01 or 10, thus, a probability zero of receiving the states $|0,1\rangle$ or $|1,0\rangle$. Similarly if Alice measured 10 (01), then Bob should have a probability zero of measuring 01 (10); he could however get a loss, 00. The resulting (non normalized) Eve+Bob residual states thus take the form $|\psi'_{00}\rangle = V |\psi_{00}\rangle = |H_{00}\rangle |0,0\rangle_{AB}$ when a loss arrives, and otherwise,

$$|\psi'_{01}\rangle = V |\psi_{01}\rangle = |F_{01}\rangle |0,1\rangle_{AB} + |H_{01}\rangle |0,0\rangle_{AB} \qquad (6)$$

$$|\psi'_{10}\rangle = V |\psi_{10}\rangle = |F_{10}\rangle |1,0\rangle_{AB} + |H_{10}\rangle |0,0\rangle_{AB} \ . \qquad (7)$$

Finally, V being linear, the (normalized) residual state if Alice applied CTRL is $|\psi'\rangle \equiv V |\psi\rangle = |\psi'_{00}\rangle + |\psi'_{01}\rangle + |\psi'_{10}\rangle$.

In order to check CTRL bits, Bob measures $|\psi'\rangle$ in the x basis and checks if he gets a photon in the illicit state

$|-\rangle$. To avoid that, Eve must make sure that the overlap between Eve-Bob's state $|\psi'\rangle$ and Bob's state $|-\rangle$ is zero. This results with another limitation on Eve's attack: the norm of ${}_{AB}\langle -\big|\big(|F_{01}\rangle|0,1\rangle_{AB}\big) + {}_{AB}\langle -\big|\big(|F_{10}\rangle|1,0\rangle_{AB}\big)$ must be 0; namely, $|F_{01}\rangle\langle - \mid 0,1\rangle + |F_{10}\rangle\langle - \mid 1,0\rangle = \big(|F_{01}\rangle - |F_{10}\rangle\big)/\sqrt{2} = 0$, i.e., $|F_{01}\rangle = |F_{10}\rangle = |F\rangle$ for some (non normalized) state $|F\rangle \in \mathscr{H}_E$. The final global states (6) and (7) if Alice measured 01 and 10 are thus (respectively)

$$|F\rangle|0,1\rangle_{AB} + |H_{01}\rangle|0,0\rangle_{AB} \tag{8}$$

$$|F\rangle|1,0\rangle_{AB} + |H_{10}\rangle|0,0\rangle_{AB} \ , \tag{9}$$

and if Bob does not get a loss, Eve's final state is $|F\rangle$ whether Bob measures $|0,1\rangle$, i.e., the bit 0, or $|1,0\rangle$, i.e., the bit 1. Eve's final probe is, thus, independent of all of Alice's and Bob's measurements, and is unentangled with their state.

Eve can thus get no information on the bits Alice and Bob agree upon without being detectable. That reasoning can be done inductively bitwise to get robustness with N qubits.

IV. The Classical Alice Protocol, Dealing with Losses and Multi-Photon Pulses

In practice, there are not just losses: when qubits are encoded using photon pulses, there may be more than one photon per pulse, giving the eavesdropper more tools to get information on the SIFT bits. We now allow the Hilbert space to contain all photonic states of the above-mentioned two modes. Namely, we consider all states $|n_1, n_0\rangle$ with $n_0 + n_1 \geq 0$. As before, we *must* specify Alice's and Bob's operations on those states.

A. Defining the (full) "photonic QKD with classical Alice" protocol

If Alice and Bob can distinguish one from more than one photon, extending the results of the earlier section is rather trivial; in brief, Eve becomes limited to the same space as in the previous section, or else she will be noticed.

The interesting extension is when Alice and Bob are limited, and cannot tell a single photon pulse from a multi-photon pulse. It is conventional to say that they have "detectors" and not "counters". This, of course, is in contrast to Eve who has counters, and who can do whatever physics allows.

We now assume a specific realization of the Fock states, to make the limitation on the measurements more clear. We assume that the two classical states, $|0\rangle$ and $|1\rangle$, describe two pulses on the same arm, such that the photon can either be in one pulse, in the other, or in a superposition, such as the (non-classical) state $|+\rangle$. Measurements are applied onto the two modes separately, using two detectors, thus a state $|1,1\rangle$, as well as any state $|n_1, n_0\rangle$ with both $n_1 \geq 1$ and $n_1 \geq 1$, can be identified as an error. That will be enough to guarantee robustness.

As before, we assume that Alice's CTRL operation is given by (2), yet now, with n_0 and n_1 being any non-negative integers. Let $\hat{n}_1 = 1$ if $n_1 \geq 1$, else $\hat{n}_1 = 0$; similarly, $\hat{n}_0 = 1$ if $n_0 \geq 1$, else $\hat{n}_0 = 0$. To model properly the use of a detector that clicks when noticing one or more photons, it is assumed that in the SIFT mode Alice still attaches a probe in the $|0,0\rangle_A$ state. Now she applies the following transform, U_{SIFT}, on $\mathscr{H}_A \otimes \mathscr{H}_{AB}$ where $\mathscr{H}_A = \text{Span}\big(|0,0\rangle_A, |0,1\rangle_A, |1,0\rangle_A, |1,1\rangle_A\big)$ and $\mathscr{H}_{AB}$ is $\mathscr{F}$, Alice+Bob's 2-mode photonic space:

$$U_{\text{SIFT}}|0,0\rangle_A|n_1,n_0\rangle_{AB} = |\hat{n}_1,\hat{n}_0\rangle_A|n_1,n_0\rangle_{AB} \ . \tag{10}$$

Alice then measures her probe in the $|0,0\rangle_A$, $|0,1\rangle_A$, $|1,0\rangle_A$ and $|1,1\rangle_A$ basis; she cannot distinguish $|n_1,0\rangle$ with $n_1 \geq 2$ from $|1,0\rangle$, yet she can distinguish $|1,1\rangle$ from $|1,0\rangle$. When $n_1 \geq 1$ or $n_0 \geq 1$ she sees $\hat{n}_1 = 1$ or $\hat{n}_0 = 1$ (respectively); if both $n_1 \geq 1$ and $n_0 \geq 1$ then she measures her probe in a state $|1,1\rangle_A$; this is telling her that the state she received is illicit.

We need to *carefully* define Alice's operation on the states she receives, as the robustness analysis depends on the residual state after Alice's "measurement", which Alice sends back to Bob; we now consider two legitimate options for defining that state. In one, which we could call "the conventional measure-resend approach", we assume that depending on which detector clicks, the state $|0,1\rangle$ or the state $|1,0\rangle$ (or the state $|0,0\rangle$ if no detector clicked) is then sent back to Bob. However, now Eve could prepare the state $\big(|0,2\rangle + |2,0\rangle\big)/\sqrt{2}$ and send it to Alice; in CTRL mode the same state will return to Eve, while in SIFT mode only a single photon (or none) will be given back to Eve. Thus, Eve (who can measure the number of photons) will easily decode Alice's operation, and will be able to measure (and resend) in case of SIFT, or send the state $\big(|0,1\rangle + |1,0\rangle\big)/\sqrt{2}$ back to Bob in case of CTRL.

We thus stick here to a different way of defining the residual state after Alice's action: we simply assume that the state $|n_1, n_0\rangle$ is sent back to Bob in both (10) and (2). Incidently, that attack above is an example of a simple tagging attack. In a separate work (in preparation) we present a modified photonic classical Alice protocol that prevents many other tagging attacks, including the one suggested in [8] as an attack against QKD with classical Bob ([1]); see also [9].

B. Eve's attack on the (photonic) classical Alice protocol

Eve performs her attack in both directions using a fixed probe space $\mathscr{H}_E$; from Bob to Alice, Eve applies U; from Alice to Bob, Eve applies V. The attack from Bob to Alice produces a state of the form $\sum |E_{n_1 n_0}\rangle |n_1, n_0\rangle_{AB} \in \mathscr{H}_E \otimes \mathscr{H}_{AB}$ where $\mathscr{H}_{AB} = \mathscr{F}$. With Alice's probe attached we obtain

$$|\Psi\rangle = \sum |E_{n_1 n_0}\rangle |0,0\rangle_A |n_1, n_0\rangle \ , \qquad (11)$$

in $\mathscr{H}_E \otimes \mathscr{H}_A \otimes \mathscr{H}_{AB}$. In particular, if Eve does nothing then $|E_{n_1 n_0}\rangle \equiv |E\rangle$ independently of n_1 and n_0, and the state in Alice+Eve's hands, prior to Alice's operation, is $|E\rangle |0,0\rangle_A |+\rangle_{AB}$.

Going back to the general case, if Alice applies U_{CTRL}, then the state in Eve+Alice hands (after Alice's CTRL action) is still $|\Psi\rangle$. However, if Alice applies U_{SIFT}, the resulting global state in Eve+Alice's hands is

$$\sum |E_{n_1 n_0}\rangle |\hat{n}_1, \hat{n}_0\rangle_A |n_1, n_0\rangle_{AB} \ ; \qquad (12)$$

after Alice has measured her probe she gets some output ($\{00, 01, 10, 11\}$), and some complicated (non normalized) residual state (sent then back to Bob) that we soon analyze.

Eve now attacks that residual state on the way back from Alice to Bob using the unitary V acting on both her probe and the state sent by Alice to Bob (see below). Eve then sends Bob his part of the resulting state.

C. A proof of robustness

Alice's measuring abilities put a constraint on the state $|\Psi\rangle$ for Eve not to be detectable: Alice's probability of measuring $|1,1\rangle_A$ according to that model must be zero, or else Eve can be noticed. It is thus required that $|E_{n_1 n_0}\rangle = 0$ for $n_1 \times n_0 \neq 0$. Therefore, Eve+Alice's state when Alice applies U_{SIFT} must take the form

$$\sum_{n_0 \geq 1} |E_{0 n_0}\rangle |0,1\rangle_A |0, n_0\rangle + \sum_{n_1 \geq 1} |E_{n_1 0}\rangle |1,0\rangle_A |n_1, 0\rangle \qquad (13)$$

$$+ |E_{00}\rangle |00\rangle_A |0,0\rangle \ . \qquad (14)$$

Once Alice has performed her measurements and sent $|i,j\rangle_{AB}$ back to Bob via Eve, the resulting global state (fully in Eve's hands) is given by Table II where $|\psi_{ij}\rangle$ are not normalized, and where the $|E_{ij}\rangle$ were chosen by Eve. Eve now applies a unitary V on $\mathscr{H}_E \otimes \mathscr{H}_{AB}$ and then sends Bob his part of the resulting state.

Recall that Eve attacks now using the unitary V acting on the residual state in $\mathscr{H}_E \otimes \mathscr{H}_{AB}$, and then she sends Bob his part of the resulting state. Bob's measuring abilities put more constraints on the state $|\psi\rangle$ for Eve not to be detectable. In case the SIFT bit is used for TEST, Bob's probability of measuring 11 must be zero, no matter what Alice measured. Furthermore, for Eve to stay undetectable, if Alice measured $|0,0\rangle$ (namely, the outcome 00) in the SIFT mode, then Bob should have a probability zero of measuring 01 or 10, thus, a probability zero of receiving the states $|1,0\rangle$ or $|0,1\rangle$. Similarly if Alice measured 10 (01), then Bob should have a probability zero of measuring 01 (10); he could however get a loss, 00. The resulting (non normalized) Eve+Bob residual states thus take the form $|\psi'_{00}\rangle = V|\psi_{00}\rangle = |H_{00}\rangle |0,0\rangle_{AB}$ when a loss arrives, and

$$\begin{aligned} |\psi'_{01}\rangle = V|\psi_{01}\rangle &= \sum_{n_0 \geq 1} |F_{0 n_0}\rangle |0, n_0\rangle_{AB} + |H_{01}\rangle |0,0\rangle_{AB} \\ |\psi'_{10}\rangle = V|\psi_{10}\rangle &= \sum_{n_1 \geq 1} |F_{n_1 0}\rangle |n_1, 0\rangle_{AB} + |H_{10}\rangle |0,0\rangle_{AB} \end{aligned} \qquad (15)$$

otherwise; V being linear, the (normalized) residual state if Alice applied CTRL is $|\psi'\rangle \equiv V|\psi\rangle = |\psi'_{00}\rangle + |\psi'_{01}\rangle + |\psi'_{10}\rangle$.

In order to check CTRL bits, Bob measures $|\psi'\rangle$ in the x basis and checks if he gets at least one photon in any illicit state, such as $|-\rangle$; more precisely, he measures $|\psi'\rangle$ in the Fock basis $|n_-, n_+\rangle_x$ corresponding to the x basis of single photon states, and aborts if he gets $n_- > 0$ (if the detector for $|-\rangle$ photons clicks). To avoid that, Eve must make sure that the overlap between Eve-Bob's state $|\psi'\rangle$ and each state of the form $|n_-, n_+\rangle_x$ with $n_- > 0$ is zero. This results with another limitation on Eve's attack.

TABLE II
THE STATE IN EVE'S HANDS AFTER ALICE'S MEASUREMENT WHEN LOSSES AND MULTI-PHOTON PULSES ARE ALLOWED

Measurement	*Residual state (in Eve's hands)*
00	$\|\psi_{00}\rangle = \|E_{00}\rangle \|0,0\rangle_{AB}$
01	$\|\psi_{01}\rangle = \sum_{n_0 \geq 1} \|E_{0 n_0}\rangle \|0, n_0\rangle_{AB}$
10	$\|\psi_{10}\rangle = \sum_{n_1 \geq 1} \|E_{n_1 0}\rangle \|n_1, 0\rangle_{AB}$
CTRL	$\|\psi\rangle = \|\psi_{00}\rangle + \|\psi_{01}\rangle + \|\psi_{10}\rangle$

Lemma 1. *If Bob has a zero probability of measuring any state $|n_-, n_+\rangle_x$ with $n_- > 0$, then $|F_{01}\rangle = |F_{10}\rangle$, and $|F_{0n}\rangle = |F_{n0}\rangle = 0$ for $n > 1$.*

Proof (sketch).
Let $|\psi'\rangle = \sum_{n_0 \geq 1} |F_{0 n_0}\rangle |0, n_0\rangle + \sum_{n_1 \geq 1} |F_{n_1 0}\rangle |n_1, 0\rangle +$

$|H\rangle\,|0,0\rangle$ be the Eve+Bob residual state. Let $|e^{(n)}\rangle = \big(|0,n\rangle + |n,0\rangle\big)/\sqrt{2}$ and $|o^{(n)}\rangle = \big(|0,n\rangle - |n,0\rangle\big)/\sqrt{2}$; it then holds that $|e^{(n)}\rangle$ (resp. $|o^{(n)}\rangle$) is a superposition with non zero amplitudes of the states $|n_-, n_+\rangle_x$ with n_- even (resp. n_- odd) and that moreover $|F_{0n}\rangle\,|0,n\rangle + |F_{n0}\rangle\,|n,0\rangle$ is equal to

$$\left[\frac{|F_{0n}\rangle + |F_{n0}\rangle}{\sqrt{2}}\right]|e^{(n)}\rangle + \left[\frac{|F_{0n}\rangle - |F_{n0}\rangle}{\sqrt{2}}\right]|o^{(n)}\rangle \qquad (16)$$

For $n = 1$ the probability of measuring $|1,0\rangle_x$ must be 0. Since $\langle e^{(1)} \mid 1,0\rangle_x = 0$ (because 1 is odd) and $\langle o^{(1)} \mid 1,0\rangle_x \neq 0$, the probability of measuring $|1,0\rangle_x$ is zero iff $\big[|F_{01}\rangle - |F_{10}\rangle\big]/\sqrt{2} = 0$ i.e. $|F_{01}\rangle = |F_{10}\rangle$. For $n > 1$, the probabilities of measuring both $|n,0\rangle_x$ and $|n-1,1\rangle_x$ must be 0, implying that $|F_{0n}\rangle + |F_{n0}\rangle = 0$ and $|F_{0n}\rangle - |F_{n0}\rangle = 0$ and thus $|F_{n0}\rangle = |F_{0n}\rangle = 0$. More details are left for the full journal paper (see a preliminary version in [10, Appendix C], where is also provided the expansion of the x-basis Fock states $|n_-, n_+\rangle_x$ using the z-basis Fock states $|n_{\mathtt{1}}, n_{\mathtt{o}}\rangle$). □

Letting $|F\rangle = |F_{01}\rangle = |F_{10}\rangle$, Eve+Bob's final residual states given by (15), if Alice measured 01 and 10, are reduced to, strikingly, exactly the same states given (for the simpler case) by (8) and (15) (respectively). As before, if Bob measures in the z basis and gets a SIFT bit, Eve's final state $|F\rangle$ is the same whether Bob measured 0 or 1 and she thus can get no information on either Alice's measurement or Bob's result: the protocol is completely robust.

V. Conclusions

From the above analysis, we conclude that Bob must in the end, on CTRL bits, get either a loss or exactly the state $|+\rangle$, which he thinks he sent. This does not mean that Eve's attack is trivial (namely, she must send $|+\rangle$ to Alice, and do nothing on the way back). As the simplest non-trivial attack, Eve could prepare the state $|E\rangle[|0,2\rangle + |2,0\rangle]/\sqrt{2}$, and apply the transformation $V[|E\rangle\,|0,2\rangle] = |E\rangle\,|0,1\rangle; V[|E\rangle\,|2,0\rangle] = |E\rangle\,|1,0\rangle$ on the way back, without being noticed, but also, without gaining any information, as we proved here. It is important to combine our result with the use of decoy states, which is now the common practice in QKD. We believe that our result holds also if our analysis is applied to the recent practical implementation using a laser pulse train [11], but checking this is left for future work.

References

[1] M. Boyer, D. Kenigsberg, and T. Mor, "Quantum key distribution with classical Bob," *Phys. Rev. Lett.*, vol. 99, no. 14, p. 140501, 2007.

[2] ——, "Quantum key distribution with classical Bob," in *ICQNM 2007*, vol. 0. Los Alamitos, CA, USA: IEEE Computer Society, Jan. 2007, p. 10, URL: http://doi.ieeecomputersociety.org/10.1109/ICQNM.2007.18 [accessed: 2015:03:04].

[3] X. Zou, D. Qiu, L. Li, L. Wu, and L. Li, "Semiquantum-key distribution using less than four quantum states," *Phys. Rev. A*, vol. 79, no. 5, p. 052312, 2009.

[4] M. Boyer and T. Mor, "Comment on 'Semiquantum-key distribution using less than four quantum states'," arXiv, 2010, URL: http://arxiv.org/abs/1010.2221 [accessed: 2015:03:04].

[5] H. Lu and Q.-Y. Cai, "Quantum key distribution with classical Alice," *International Journal of Quantum Information (IJQI)*, vol. 6, no. 6, pp. 1195–1202, 2008.

[6] M. Lucamarini and S. Mancini, "Quantum key distribution using a two-way quantum channel," *Theoretical Computer Science*, vol. 560, Part 1, no. 0, pp. 46 – 61, 2014.

[7] G. Brassard, N. Lütkenhaus, T. Mor, and B. C. Sanders, "Limitations on practical quantum cryptography," *Phys. Rev. Lett.*, vol. 85, no. 6, pp. 1330–1333, 2000.

[8] Y.-g. Tan, H. Lu, and Q.-y. Cai, "Comment on 'Quantum key distribution with classical Bob'," *Phys. Rev. Lett.*, vol. 102, no. 9, p. 098901, 2009.

[9] M. Boyer, D. Kenigsberg, and T. Mor, "Boyer, Kenigsberg, and Mor reply," *Phys. Rev. Lett.*, vol. 102, no. 9, p. 098902, 2009.

[10] M. Boyer and T. Mor, "On the robustness of (photonic) quantum key distribution with classical Alice," arXiv, 2010, URL: http://arxiv.org/abs/1012.2418/ [accessed: 2015:03:04].

[11] T. Sasaki, Y. Yamamoto, and M. Koashi, "Practical quantum key distribution protocol without monitoring signal disturbance," *Nature*, vol. 509, no. 7501, pp. 475–478, 05 2014, ISBN: 0028-0836.

A Protocol for Synchronizing Quantum-Derived Keys in IPsec and its Implementation

Stefan Marksteiner

JOANNEUM RESEARCH GmbH
DIGITAL - Institute for Information
and Communication Technologies
Graz, Austria
Email: stefan.marksteiner@joanneum.at

Oliver Maurhart

AIT Austrian Institute of Technology GmbH
Digtal Safety & Security Department
Klagenfurt, Austria
Email: oliver.maurhart@ait.ac.at

***Abstract*—This paper presents a practical solution to the problem of limited bandwidth in Quantum Key Distribution (QKD)-secured communication through using rapidly rekeyed Internet Protocol security (IPsec) links. QKD is a cutting-edge security technology that provides mathematically proven security by using quantum physical effects and information theoretical axioms to generate a guaranteed non-disclosed stream of encryption keys. Although it has been a field of theoretical research for some time, it has only been producing market-ready solutions for a short period of time. The downside of this technology is that its key generation rate is only around 12,500 key bits per second. As this rate limits the data throughput to the same rate, it is substandard for normal modern communications, especially for securely interconnecting networks. IPsec, on the other hand, is a well-known security protocol that uses classical encryption and is capable of exactly creating site-to-site virtual private networks. This paper presents a solution which combines the performance advantages of IPsec with QKD. The combination sacrifices only a small portion of QKD security by using the generated keys a limited number of times instead of just once. As a part of this, the solution answers the question of how many data bits per key bit make sensible upper and lower boundaries to yield high performance while maintaining high security. While previous approaches complement the Internet Key Exchange protocol (IKE), this approach simplifies the implementation with a new key synchronization concept. Furthermore, it provides a Linux-based module for the AIT QKD software using the Netlink XFRM Application Programmers Interface to feed the quantum key to the IPsec cipher. This enables wire-speed, QKD-secured communication links for business applications.**

Keywords–Quantum Key Distribution; QKD; IPsec; Cryptography; Security; Networks.

I. INTRODUCTION AND MOTIVATION

Quantum cryptography, in this particular case quantum key distribution (QKD), has the purpose to ensure the confidentiality of a communication channel between two parties. The main difference to classical cryptography is that it does not rely on assumptions about the security of the mathematical problem it is based on, nor the computing power of a hypothetical attacker. Instead, QKD presents a secure method of exchanging keys by connecting the two communicating parties with a quantum channel and thereby supplying them with guaranteed secret and true random key material [1, p.743]. When the key is applied with a Vernam cipher (also called one time pad - OTP) on a data channel on any public network, this method provides the channel with information-theoretically (in other words mathematically proven) security [2, p.583]. The downside of this combination is the limitation to approximately twelve kilobits, shown in a practical QKD setup, due to physical and technical factors, since with OTP one key bit is consumed by one data bit [3, S.9]. This data rate does not meet the requirements of modern communications. Another practical approach came to the same conclusion and therefore uses the Advanced Encryption Standard (AES) instead of OTP [4, p.6]. To address this problem, this paper presents an approach to combine QKD with IPsec, a widespread security protocol suite that provides integrity, authenticity and confidentiality for data connections [5, p.4], by using QKD to provide IPsec with the cryptographic keys necessary for its operation. To save valuable key material, this solution uses it for more than one data packet in IPsec, thus increasing the effective data rate, which is thereby not limited to the key rate anymore. Furthermore, using this approach, the presented solution benefits from the flexibility of IPsec in terms of cryptographic algorithms and cipher modes. In contrast to most of the previous approaches, that supplemented the Internet Key Exchange (IKE) protocol or combine in some way quantum-derived and classical keys, this paper refrains from using IKE (for a key exchange is rather the objective of QKD, as described later) in favor of a specialized, lightweight key synchronization protocol, working with a master/slave architecture. The goal of this protocol is to achieve very high changing rates of purely quantum-derived keys on the communicating peers while maintaining the keys synchronous in a very resilient manner, which means to deal with suboptimal networking conditions including packet losses and late or supplicate packets. In order to fulfill this objective, the following questions need to be clarified:

- What is the minimum acceptable frequency of changing the IPsec key that will ensure sufficient security?
- What is the maximum acceptable frequency of changing the IPsec key to save QKD key material?
- Is the native Linux kernel implementation suitable for this task?
- How can key synchronicity between the communication peers be assured at key periods of 50 milliseconds and less?

As a proof of concept, this paper further presents a software solution, called QKDIPsec, implementing this approach in

C++. This software is intended to be used as an IPsec module for the multi platform hardware-independent AIT QKD software, which provides already a market-ready solution for OTP-based QKD. The module achieves over forty key changes per second for the IPsec subsystem within the Linux kernel. At present time, the software uses a static key ring buffer for testing purposes instead of actual QKD keys, for the integration of QKDIPsec into the AIT QKD software is yet to be implemented (although most of the necessary interfaces are already present). The ultimate goal is to deliver a fully operational IPsec module for the AIT QKD software.

Section II of this paper describes considerations regarding necessary and sensible key change rates, while Section III contains the architecture of the presented solution and the subsequent Section IV its implementation. Section V, eventually, contains testing results and the conclusions drawn.

II. Key Change Rate Considerations

The strength of every cryptographic system relies on the key length, the secrecy of the key and the effectiveness of the used algorithms [6, p.5]. As this solution relies on QKD, which generates a secret and true random key [7], this means that more effective algorithms and more key material are able to provide more cryptographic security. In this particular case, the used algorithms are already prescribed by the IPsec standard [8], therefore the security is mainly determined by the used key lengths, more precisely by the relation between the amount of key material and the amount of data, which should be as much in favor of the key material as possible - given the low key rate compared to the data rate, naturally the opposite is the case in practice. This Section aims on giving feasible upper and lower boundaries of key change rates (or key periods P_k, respectively) and, thus, how much QKD key material should be used in order to save precious quantum key material while maintaining a very high level of security. The two main factors determining the key period in practice are the used algorithms (via their respective key lengths - the longer the key, the more key bits are used in one key period) and the capabilities of QKD in generating keys. Current working QKD solutions (such as the one used by the AIT) provide a quantum key rate Q of up to 12,500 key bits per second at close distances, 3,300 key bits at around 25 kilometers and 550 key bits at around 50 kilometers distance [3, p.9].

In order to fully utilize the possibly QKD key rate and given the currently shortest recommended key length, which is 128 bits, a IPsec solution using quantum-derived keys should thus be able to perform around 100 key changes per second ($\frac{12{,}500}{128} \approx 97,65$), 50 for every communication direction (for IPsec connection channels are in principle unidirectional and therefore independent from each other even if they belong to the same bidirectional conversation). This corresponds to a key period P_k of around 20 ms, as it is a function of the Quantum key rate Q and the algorithm's key length k. The period for a bidirectional IPsec link is $P_K = (\frac{Q}{2k})^{-1}$. At longer key lengths, this period becomes longer, for a single change cycle uses more key material and, thus, less key changes are necessary to utilize the full incoming key stream, therefore this period $P_{k_{min}} = 20ms$ presents a feasible lower boundary for the key period. As stated above, the security of this system depends also on the data rate. Given a widespread data rate of 100 megabits per second, a key period of 20 ms and 128 key bits means a ratio of 8000 data bits per key bit (or short dpk, for the reader's convenience).

A landmark in this *security ratio* is 1 dpk. This rate would provide unconditional security when applied with OTP. For the cipher and hash suites included in the IPsec protocol stack, there is no security proof and therefore they are not unconditionally secure. However, applying an IPsec cipher (for instance AES) with an appropriately fast key change and restricted data rate to achieve 1 dpk is the closest match inside standard IPsec, especially when the block size equals the key size.

To define an upper boundary (and therefore a minimum standard for the high security application of the presented solution), a very unfavorable relation between data and key bits through a high-speed connection of 10 gigabits of data is assumed. A recent attack on AES-192/256 uses $2^{69.2}$ computations with 2^{32} *chosen plaintext* [9, p.1]. Because of the AES block size of 128 bits, this corresponds to $2^{32} * 2^{7} = 2^{39}$ data bits. Although this attack is currently not feasible in practice, as it works only for seven out of 12/14 rounds and also has unfeasible requirements to data storage on processing power for a cryptanalytic machine, it serves as a theoretical fundament for this upper boundary. A bandwidth of 10 gigabits per second equals approximately 9.3 gibitbs per second. This is by the factor of 64 (2^{6}) smaller than the amount of data for the attack mentioned above, which means that it requires 64 seconds to gather the necessary amount of data to (though only theoretically) conduct the attack. In conclusion (with AES-192/256), the key should be changed at least every minute ($P_{k_{max}} = 60s$), while the maximum allowed key period according to the IPsec standard lies at eight hours or 28,800 seconds [10].

For cryptographic algorithms operating with lower cipher block sizes (ω), the *birthday bound* ($2^{\frac{\omega}{2}}$) is relevant. The birthday bound describes the number of brute force attempts to enforce a collision with a probability of 50 percent, such that different clear text messages render to the same cypher text. With a block size of 64 (*birthday bound* = 2^{32}), the example speed of 10 gigabit per second above would lower the secure key period to under half a second. Because of this factor, using 64-bit ciphers is generally discouraged for the use with modern data rates[11, pp.1-3] (although the present rapid rekeying approach is able to cope with this problem). Regarding key lengths, 128 bits are recommended beyond 2031 [6, p.67] while key sizes of 256 bits provide *good protection* even against the use of Grover's algorithm in hypothetical quantum computers for this period [12, p.32].

III. Rapid Rekeying Protocol

This Section describes the *rapid rekeying protocol*, the purpose of which is to provide to IPsec peers with QKD-derived key material and keep these keys synchronous under the low-key-period conditions (down to $P_{k_{min}} = 20ms$) stated in Section II.

This protocol pursues the approach that with QKD, there is no need for a classical key exchange (for instance with IKE). Relevant connection parameters (like peer addresses) are available a priori (before the establishment of the connection) in point-to-point connections, whereas keying material is provided by QKD, mostly obsoleting IKE. Furthermore, IPsec only dictates an automatic key exchange, not specifically

IKE [5, p.48] and a protocol that only synchronizes QKD-derived keys (instead of exchanging keys) is therefore deemed sufficient, yet compliant to the IPsec standard. Consequently, it is an outspoken objective to create a slender and simple key synchronization protocol to increase performance and reduce possible sources of error. Another objective for key synchronization is robustness in terms of resilience against suboptimal network environment conditions. The protocol described in this paper uses two channels for encrypted communication: an Authenticaton Header (AH)-authenticated control channel (amongst other tasks, signaling for key changes) and an Encapsulating Security Payload (ESP)-encrypted data channel to transmit the protected data (see Figure 1). The reason for the use of AH on the control channel is that it only contains non-secret information, while its authenticity is crucial for the security and stability of the protocol. The necessary *security policies (SPs)* for the IPsec channels remain constant during the connection. There are four necessary SPs, one data and one control SP for each direction. The complete software solution will, delivered by the AIT QKD software, contain additionally the quantum channel for key exchange and a *Q3P* channel, whereby the latter is another protocol that provides OTP-encrypted QKD point-to-point links. These two additional channels are outside of this paper's scope.

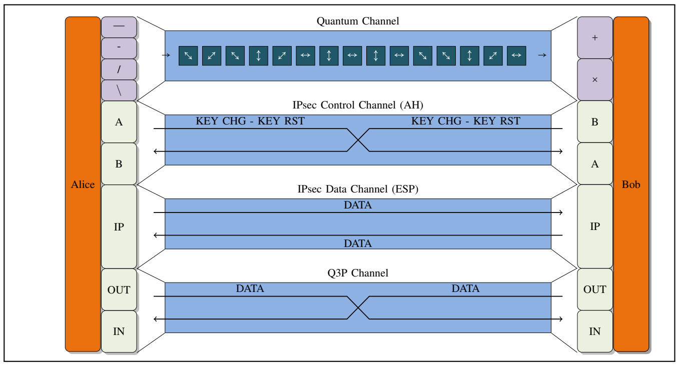


Figure 1. Rapid Rekeying Channel Architecture

The protocol itself follows, taking account of the unidirectional architecture of IPsec, a *master/slave* paradigm. Every peer assumes the master role for the connection in which the peer represents the sending part. When a key change is due (for instance because of the expiration of the key period), the master sends an according message (key change request) to the slave and the latter changes the key (as does the master). To compensate lost key change signals, every key change message contains the *security parameters index (SPI)* for the next-to-use key. The SPI is simply calculable for the peers through a salted hash whereby the salt and a initial seed value are QKD-derived and each SPI is a hash of its predecessor plus salt, which makes it non-obvious to third parties. This level of security is sufficient, for the SPI is a public value, included non-encrypted in every corresponding IPsec packet, making it a subject rather to non-predictability than to secrecy. Also, using only a seed and salt from QKD, the hashing method safes quantum keying material. As all necessary IPsec parameters are available beforehand, as well as the keys (through QKD), IPsec *security associations (SAs)* may be pre-calculated and established in advance (which are identified by unique SPIs). Permanently changing attributes during a conversation are only the SPI and the key, while all other parameters of an SA (for instance peer addresses, services, protocols) remain constant. The master calculates these two in advance and queues them for future use. Only one SA is actually installed (aplied to the kernel IPsec subsystem), for only one (per default, at least in Linux, the most recent) may be used to encrypt data. The slave, on the other hand, operates differently. For it identifies the right key to use based on the SPI, it may very well have multiple matching SAs installed. This makes key queuing expendable on the receiver side, while the SPI queuing is used as an indexer for lost key change message detection. For reasons of data packets arriving out of synchronization, SAs are not only installed beforehand, but also left in the system for some time on the receiver side, allowing it to process packets encrypted with both an older or newer key than the current one.

On every key change event, the master applies a new SA to the system (using the next following SPI/key from the queues), prepares a new SPI/key pair (SPI generation as mentioned above and acquirement of a new key from the QKD system) and deletes the deprecated data from both its queues and the IPsec subsystem. The slave also acquires a new SPI/key pair (the same the sender acquires) but installs it directly as an SA and only stores the SPI for indexing. It subsequently deletes the oldest SA from the system and SPI from the queue if the number of installed SAs exceeds a configured limit. To sum it up, on every key change event, the two peers conduct the following steps:

- the master acquires a new key and SPI and ads it to its queues
- it sends a key change request to the slave
- it fetches the oldest pair from the queue an installs it as a new SA, *replacing* the current one
- it deletes the deprecated pair from its queue
- the slave receives the key change request and also acquires a new SPI/key pair (the same as the master)
- it installs the pair as a new SA and the SPI into the indexing queue
- it deletes the oldest SA from the system and oldest SPI from the queue
- it sends a key change acknowledgement

This procedure keeps both of the installed SA types up to date. For instance, 50 installed SAs for the slave resulting in 25 queued SPI/key pairs on the master, for the latter does not need to store backward SAs. At the beginning, on every key change, SPI/key pair is acquired, while the already applied remain. When the (configurable) working threshold is met, additionally the oldest SA or SPI/key pair is deleted, keeping the queue sizes and number of installed SAs constant.

Figure 2 illustrates this process for a sender (*Alice*) and a receiver (*Bob*), where the arrows show the changes in case of an induced key change. Naturally, as with SPs, there are four SA types on a peer: one for data and control channels, each for sending (master) and receiving (slave). Each SA corresponds to an SPI and key queue on the master's side and one SPI queue on the slave's side, respectively.

As the data stream is independent from control signaling, this calculation in advance prevents the destabilization of the

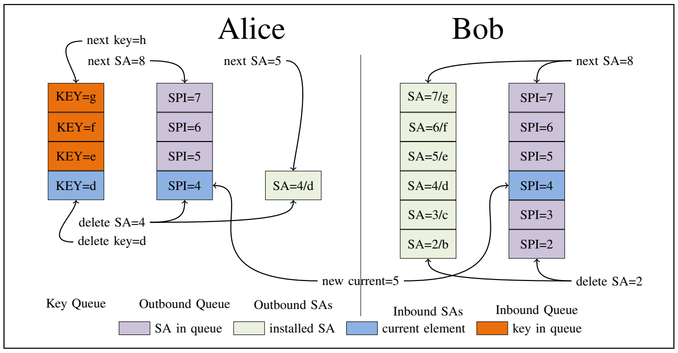


Figure 2. Key Change Process

key synchronization in case of lost and too early or too late arriving key change messages. The buffer of previously created SAs compensates desynchronization. For every receiver is able to calculate the according SPIs beforehand, it may, by comparing a received SPI with an expected, detect and correct the discrepancy by calculating the following SAs. Through this compensation process, there is neither need to interfere with the data communication nor to even inform the sender of lost key change messages; the sender may unperturbedly continue with data and control communications. This mechanisms make constant acknowledgements expendable and contribute thereby to a better protocol performance through omission of the round trip times for the majority of the necessary control messages. Because of this, acknowledgement messages (key change acknowledge) are still sent, but serve merely as a keepalive mechanism instead of true acknowledgements. In rare occasions, a key change message might be actually received, but the slave might not be able to apply the key for some reason (for instance issues regarding the QKD system or the Kernel). In this case, it reports the failure to the master with an appropriate message (key change fail). In case too many control packets go missing (what the receiver is able to detect by SPI comparisons and the sender by the absence of keepalive packets) or the key application fails, every peer is able to initiate a reset procedure (master or slave reset). The actual threshold of allowed and compensated missing messages is a matter of configuration and corresponds to the queue sizes for the SAs and therefore the ability of the system to compensate these losses. The master does not need to report key change fails, for it is in control of the synchronization process and might just initiate a reset if it is unable to apply its key. An additional occasion for a reset is the beginning of a conversation. At that point, the master starts the key synchronization process with an initial reset. A reset consists of clearing and refilling all of the queues and installed SAs. For the same reason as for the data channel, the authentication key for the control channel changes periodically. Due to the relatively low transmission rates on the control channel the key period is much longer (the software's default is 3 seconds) than on the data channel. As, therefore, control channel key changes are comparatively rare and reset procedures should only occur in extreme situations, both types implement a three way handshake. This is, on the one hand, because of the low impact on the overall performance due to the rare occurences, on the other hand due to higher impact of faulty packets. The control channel, however, implements the same SA buffering method as the data channel (only with AH SAs).

IV. Implementation

The presented solution, called *QKDIPsec*, consists of three parts (see also Figure 3):

- key acquisition;
- key application;
- key synchronization;

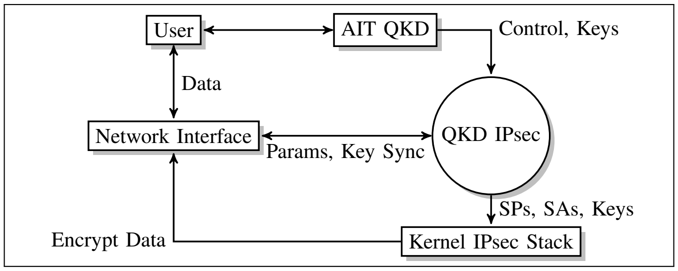


Figure 3. QKDIPsec Systems Context

Each of this tasks has a corresponding submodule inside QKDIPsec, while the overall control lies within the responsibility of the *ConnectionManager* class, which provides the main outside interface and instantiates the classes of said submodules using corresponding configuration. Also, all of these classes have corresponding configuration classes using a factory method pattern [13, p.134] and according configuration classes, decoupling program data and logic. The first task (key acquisition) is the objective of an interface to the AIT QKD software, the *KeyManager*, which provides the quantum key material. In this proof of concept, this class generates dummy key from a ring buffer, while it already has the according interfaces for the QKD software to serve as a class to acquire quantum key material and provide it in a an appropriate way to QKDIPsec. By now, only one function implementation is missing on the QKD software side to fully integrate QKDIPsec into the QKD software.

The second part (*KernelIPsecManager*) enters the acquired key directly into the Linux kernel, which encrypts the data sent to and decrypts the data received from a peer. Responsible for this part are a number of C++ classes, which control the SP and SA databases (SPD and SAD) within the Kernel's IPsec subsystem via the Linux *Netlink* protocol. Therefore, this solution uses the derived class *NetlinkIPsecManager*, but leaves the option to use other methods for kernel access as well. The reason for using Netlink to communicate with the kernel is that it was found the most intuitive of the available methods and that it is also able to handle not only the IPsec subsystem but a broad span of network functions in Linux. Furthermore, using a direct kernel API, as opposed to other IPsec implementations, omits middleware, both enhancing performance as well as eliminating potential source of error. Also using Netlink functions, this part governs the tunnel interfaces and routing table entries necessary for the communication via the classes *KernelNetworkManager* and *NetlinkNetworkManager* as well.

Netlink is a socket-oriented protocol and allows therefore the use of well-known functions from network programming.

The difference to the latter is that instead of network peers, communication runs within the system as *inter-process communication (IPC)*, through which also the kernel (via process ID zero) is addressable. Due to its network-oriented nature, a packet structure is used instead of function calls via parameters. This means that commands to the kernel (for instance to add a new SA) needs to be memory-aligned in the according packet structure and subsequently send to the kernel via a Netlink socket. A downside of Netlink during implementation was the complicated nature and weak documentation of its IPsec manipulation part (*NETLINK_XFRM*). While the Netlink protocol itself is present in every message in the form of its uniform header, the *NETLIN_XFRM* parts use a different structure plus individual extra payload attributes for every type of message (add and delete messages for both SAs and SPs), making the according class hierarchy rather inflated.

The key synchronization, eventually, is the main task of the *Rapid Rekeying Protocol*. As this is the very core of the solution, its implementation resides directly inside the connection manager. While it uses the classes mentioned above to acquire and apply the QKD keys in the manner discussed in Section III, it handles the key synchronization using sender and receiver threads (representing the master and slave parts, respectively), as well as a class for key synchronization messages. Within this class, also the described lost message compensation and reset, as well as initialization and clean-up procedures are implemented. The reset procedure may also include some re-initialization process for the QKD system, triggered via the *KeyManager*. This class also sets the clocking for the key changes, which is dynamically adjustable during runtime.

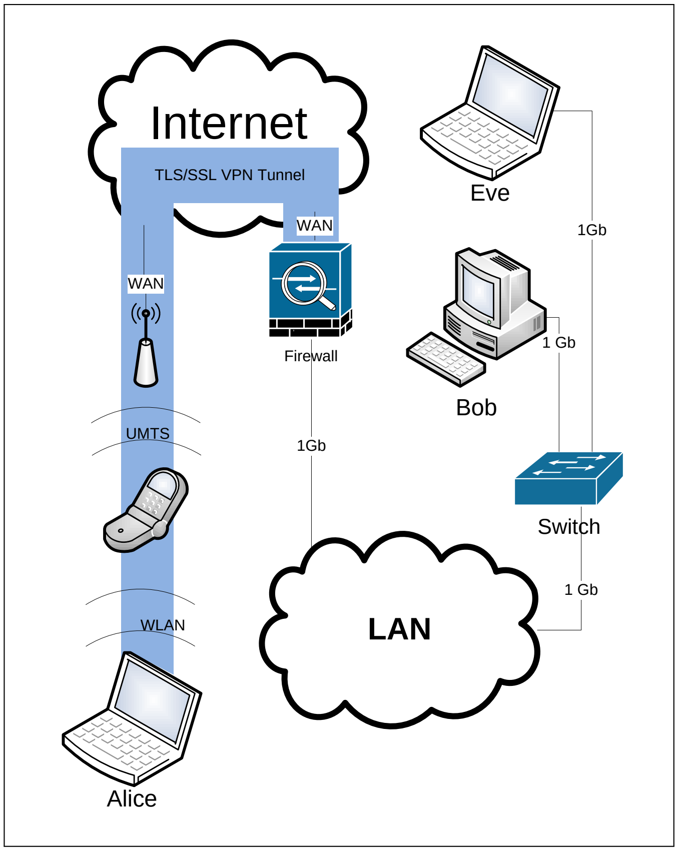


Figure 4. WAN Test Setup

V. Conclusion

The protocol design of the described solution aims on the one hand on speed and flexibility and on the other hand on fault tolerance, hence the architecture is as simple and lightweight as possible (including abandoning the IKE protocol). Due to this, very high IPsec key change rates can be achieved, even under harsh conditions. The solution was implemented in software using C++ and tested on two to five year-old Linux computers (Alice and Bob), both in a gigabit Local Area Network (LAN) and a UMTS-Wide Area Network (WAN) environment (the latter further aggravated by combining it with WLAN and an additional TLS-based VPN tunnel - see Figure 4) by means of data transfer time measurement and ping tests, as well as validation of the actual key changes by a Wireshark network sniffer (Eve).

Table I shows the results in seconds (four trials each, separated by slashes) of data transmission and in percent on ping tests within the mentioned LAN and WAN environments with various configurations: unencrypted, standard IPsec and QKDIPsec with different encryption algorithms, the latter also with different key periods. In these tests, both data transfer and ping were initiated by one peer (*Alice*). While the ping test was continuous, the data transfer consisted each of one data transfer from *Alice* to *Bob* and vice versa. The test file used on the LAN was a video file of 69.533.696 bytes size, while the WAN file was also a video, but only 1.813.904 bytes big. In both cases, key periods of 25 ms and less could be achieved, maintaining a stable data connection. This, using the recommended key length of 256 bit, surpasses the goal of 12,500 key bits per second (the currently maximal quantum key distribution rate under ideal circumstances), even though (deliberately) legacy equipment and a less-than-ideal network environment was used. Comparison of the performance shows a (expectable) higher data transfer period of QKDIPsec and unencrypted traffic, but no significant difference to traditional IPsec. Only the packet losses on a simultaneously running ping test were a few percentage points higher (mainly in the WAN environment).

TABLE I. Performance Test Results

LAN			
Setting	A→B	B→A	Ping
unencrypted	6/6/7/6	7/9/7/8	100%
AES-256 CCM			
standard IPsec	14/14/16/15	17/18/26/18	100%
50 ms	8/10/8/9	14/16/16/16	100%
25 ms	10/9/8/8	14/15/17/16	100%
20 ms	9/9/9/9	11/16/17/12	100%
AES-256 CBC			
20 ms	9/7/7	11/13/17	100%
Blowfish-448			
20 ms	14/9/7	15/13/14	99%
WAN			
Setting	A→B	B→A	Ping
unencrypted	10/10/10/10	9/7/6/7	99%
AES-256 CCM			
standard IPsec	11/11/11/11	11/5/6/5	99%
50 ms	14/10/11/13	6/5/5/5	95%
25 ms	10/11/10/10	6/7/6/7	94%
20 ms	12/11/13/10	9/5/6/6	98%
AES-256 CBC			
20 ms	10/11/11	9/7/8	100%

To verify the key changes, a network sniffer, Eve, was keeping track of the actual SPI changes of the packets transmitted between Alice and Bob. Table II shows a random sample of key change periods in milliseconds during the above mentioned LAN 20ms AES-256-CCM test. Within this table, the first column shows the key change times for data (ESP) packets from Alice to Bob while the second shows the opposite direction. As the recorded data contains one file copy from Alice to Bob (in the first half of the record) and one vice versa (in the second half), one randomly chosen sample of five consecutive key changes for each direction and from each half is chosen. This form of sample choosing from different phases and directions of the communication session and averaging them compensates inaccuracies, induced by the pause between key change and respective next following packet, which become greater the less traffic is sent. As the receiver only acknowledges received data and, therefore, sends significantly less packets, the vagueness of the non-averaged results is greater when receiving. The total average of all four of these averaged values is 0.020495 ms, which is approximately 2.5% above 20 ms per key change. This may be explained by the send and receive overhead for processing the key change messages, for the period is determines only the sleeping duration of a sender thread.

Because of the lower amount of traffic (due to the lower speed) and higher latency such exact time readings are not possible in the WAN environment. Therefore, the measurement method was changed to averaging a sample set of 20 key change periods, using the same random choosing as above. With approximately 0.2475, the total averaged result lies significantly higher (approximately 19%) than the one of the LAN setting. One possible explanation for this behavior is the latency in this environment.

	A→B		B→A	
	1st	2nd	1st	2nd
LAN	0.0220	0.0216	0.0208	0.0203
	0.0187	0.0204	0.0197	0.0235
	0.0145	0.0216	0.0203	0.0176
	0.0195	0.0243	0.0204	0.0197
	0.0225	0.0180	0.0207	0.0238
Ø	0.0194	0.0212	0.0204	0.0210
WAN ∑ 20	0.5201	0.4899	0.4302	0.5397
Ø	0.0260	0.0245	0.0215	0.0270

TABLE II. Network Sniffing Results

Additionally, the recovery behavior was tested by letting the master deliberately omit key change notifications through the introduction of *if* clauses within the sending routine, while again running ping tests and file copies. Omitting single key change messages (and, thus, testing the recovery mechanism) yield in no measurable impact on the connection (along with 100% of successful pings). Also, by the same method of omitting key change requests, but this time surpassing the recovery queue size, the reset procedure was tested. The queue size was set to 50 and *Alice* was programmed to omit 50 sending key change messages after 200 sent ones. Expectedly, *Bob* initiated a reset procedure during the hiatus, resulting in a cycle of 200 key changes and a subsequent reset. Despite these permanent reset-induced interruptions, bidirectional ping tests only yielded insignificant losses (99.74% from *Alice* to *Bob* and 99.36% vice versa). Furthermore, a file copy in both directions was still possible.

These proof of concept tests show that using IPsec with appropriate key management is able to overcome the bandwidth restrictions of QKD, even when operating the data channels in less-than-ideal conditions. Furthermore, this paper presents an approach to provide QKD-secured links with high speeds meeting the bounds discussed in Section II, including a suitable performant and fault-tolerant key synchronization protocol (the *rapid rekeying protocol*) and a corresponding software solution running under Linux (*QKDIPsec*), that is to be integrated as a module into the AIT QKD software.

References

[1] H. Zbinden, H. Bechmann-Pasquinucci, N. Gisin, and G. Ribordy, “Quantum cryptography,” Applied Physics B, vol. 67, no. 6, 1998, pp. 743–748.

[2] M. A. Nielsen and I. L. Chuang, Quantum Computation and Quantum Information, ser. Lecture Notes in Physics. Cambridge: Cambridge University Press, 2000.

[3] A. Treiber et al., “A fully automated entaglement-based quantum cryptography system for telecom fiber networks,” New Journal of Physics, no. 11, April 2009, p. 045013.

[4] F. Xu et al., “Field experiment on a robust hierarchical metropolitan quantum cryptography network,” Chinese Science Bulletin, vol. 54, no. 17, 2009, pp. 2991–2997.

[5] S. Kent and K. Seo, “RFC4301: Security Architecture for the Internet Protocol,” 2005.

[6] E. Barker, W. Barker, W. Burr, W. Polk, and M. Smid, “Recommendation for Key Management Part 1: General (Revision 3 - NIST Special Publication 800-57),” 2012, retrieved at July 10, 2015. [Online]. Available: http://csrc.nist.gov/publications/nistpubs/800-57/sp800-57_part1_rev3_general.pdf

[7] A. Poppe et al., “Practical quantum key distribution with polarization entangled photons,” Optics Express, vol. 12, no. 16, 2004, pp. 3865–3871.

[8] Internet Assigned Numbers Authority, “IPSEC ESP Transform Identifiers,” 2012, retrieved at July 10, 2015. [Online]. Available: http://www.iana.org/assignments/isakmp-registry/isakmp-registry.xhtml#isakmp-registry-9

[9] J. Kang, K. Jeong, J. Sung, S. Hong, and K. Lee, “Collision Attacks on AES-192/256, Crypton-192/256, mCrypton-96/128, and Anubis,” Journal of Applied Mathematics, vol. 2013, 2013, p. 713673.

[10] P. Hoffman, “RFC4308: Cryptographic Suites for IPsec,” 2005.

[11] D. A. McGrew, “Impossible plaintext cryptanalysis and probable-plaintext collision attacks of 64-bit block cipher modes.” IACR Cryptology ePrint Archive, vol. 2012, 2012, p. 623.

[12] “ECRYPT II Yearly Report on Algorithms and Keysizes (2011-2012),” 2012, retrieved at July 10, 2015. [Online]. Available: http://www.ecrypt.eu.org/ecrypt2/documents/D.SPA.20.pdf

[13] E. Freeman, E. Robson, B. Bates, and K. Sierra, Head First Design Patterns. Sebastopol: O’Reilly, 2004.

BB84 Quantum Key Distribution with Intrinsic Authentication

Stefan Rass
Department of Applied Informatics, System Security Group
Universität Klagenfurt, Universitätsstrasse 65-67
9020 Klagenfurt, Austria
email: stefan.rass@aau.at

Sandra König, Stefan Schauer
Digital Safety & Security Department
Austrian Institute of Technology, Klagenfurt, Austria
email: {sandra.koenig, stefan.schauer}@ait.ac.at

***Abstract*—We describe a method to authenticate the qubit stream being exchanged during the first phases of the BB84 quantum key distribution without pre-shared secrets. Unlike the conventional approach that continuously authenticates all protocol messages on the public channel, our proposal is to authenticate the qubit stream already to verify the peer's identity. To this end, we employ a second public channel that is physically and logically disjoint from the one used for BB84. This is our substitute for the otherwise necessary assumption on the existence of pre-shared secrets. Shifting authentication to the first phase of BB84 spares bandwidth during public discussion and thus makes the overall protocol also somewhat more efficient.**

Keywords*–*Quantum Key Distribution; Authentication; BB84

I. Introduction

It is a well recognized requirement of any quantum key distribution protocol to employ an authenticated public channel for the key distillation. Traditionally, such authentication utilizes universal hashing [1] to continuously attach message authentication codes (MACs) to all protocol messages. This continuous authentication [2] shall thwart person-in-the-middle attacks by an eavesdropper sitting in between Alice and Bob, running BB84 [3] with both of them. In that sense, quantum key distribution does not really create keys from nothing, but is rather a method of key expansion. The question discussed in this work relates to whether we can cast BB84 into a protocol that in fact *does* create keys from nothing, while retaining the security of "conventional BB84".

To this end, observe that it may already be sufficient for Alice to verify Bob's identity, if she can somehow verify that Bob is really the person from which her received qubit stream originated. One possibility to do so is to ask Bob for the way in which he created the stream, say as a pseudorandom sequence, so as to prove his identity. Of course, it is neither viable nor meaningful in our setting to let Bob create his entire qubit stream pseudorandomly, but it may indeed be useful to have him embed pseudorandom bits at a priori unknown places, while leaving the rest of the stream truly random. Alice, in an attempt to verify Bob as the "owner" of the qubit stream, may ask Bob for the seeds to recover the pseudorandom bits and their positions. An eavesdropper, on the other hand, cannot reasonably pre-compute Bob's response to Alice's inquiry, if the pseudorandom bits cannot be recognized (distinguished from) the truly random bits. While this apparently induces a flavour of computational security (indistinguishability of pseudorandom from really random), we can almost avoid threats by computationally unbounded adversaries. To see why, assume that the pseudorandom sequence originates via iterative bijective transformations from a uniformly distributed and truly random seed. If so, then all pseudorandom bits will themselves enjoy a uniform distribution. As being embedded inside another sequence of independent uniformly distributed bits, the distribution of the pseudorandom bits is identical to that of the truly random bits. Despite the correlation that inevitably exists among the pseudorandom bits, the distributions are nevertheless indistinguishable, except in case when the positions of the pseudorandom bits are known a priori. However, since these positions are chosen secretly and independently of any publicly available information, the attacker has no hope better than an uninformed guess about which positions matter.

Organisation of the paper: The following Sections I-A and I-B give details on BB84 to the extent needed in the following, and relate the proposal to other solutions in the literature. Section II expands the technique how we embed pseudorandom bits into the qubit stream during BB84. Section III discusses the security of our modified version of BB84, and Section IV draws conclusions.

A. BB84 at a Glance

The BB48 protocol has first been presented by Bennett and Brassard [3]. It allows two communication parties, Alice and Bob, to generate a classical key between them by using the polarization of single photons to represent information. Therefore, Alice is in possession of a single photon source and prepares the photons randomly according to the horizontal/vertical basis (Z-basis) and the diagonal basis (X-basis), i.e., for each photon she prepares one of states $\{|0\rangle, |1\rangle\}$ and $\{|x+\rangle, |x-\rangle\}$, respectively. After Alice choses the basis, the qubit is sent to Bob, who performs a measurement on it. Since Bob does not know which basis Alice used for the preparation he does not know which measurement basis he should use and thus he will not be able to retrieve the full information from each qubit. Hence, the best strategy for him is to randomly choose between the Z- and X-basis for his measurement himself. In this case Bob will choose the correct basis half of the time – but he does not know in which cases he has guessed right. Thus, Alice and Bob compare the choice of their bases in public after Bob measured the last qubit.

During the *sifting phase* [4], Alice and Bob eliminate their measurement results for those measurements where they used different bases. The remaining measurement results are converted into classical bits using the mapping

$$\begin{aligned} \{|0\rangle, |x+\rangle\} &\longrightarrow 0 \\ \{|1\rangle, |x-\rangle\} &\longrightarrow 1. \end{aligned} \tag{1}$$

At this stage, Alice and Bob should have identical classical bit strings if the channel is perfect (noiseless channel, no eavesdropper). In reality, a certain error rate is introduced in the protocol due to physical limitations (lossy and noisy channels, imperfect devices, no single photon sources, etc.). To estimate this error rate, Alice and Bob publicly compare a fraction of their results in public to check whether they are correlated. Then, classical error correction protocols are used to identify and eliminate the differences in their bit strings. Such a procedure that has been heavily used for error correction is the *CASCADE* algorithm first introduced by Bennett et al. [5]. Due to the fact that Alice and Bob publicly compare some information during the error correction, an adversary is able to obtain further information about the secret bit string (assuming Eve's presence has not been detected during error correction). Therefore, a last process called *privacy amplification* [6] performed by Alice and Bob uses *strongly-universal*$_2$ *hash functions* (as presented in [7] and recently discussed in [8]) to minimize the amount of information leaked to the adversary. After all, the security of QKD protocols has been discussed in depth and various security proofs have been provided, for example, in [9] or [10]. A main result of these proofs shows that Alice and Bob are still able to establish a secret key, if the error rate is below a maximum value of $\simeq 11\%$ [9].

B. Related Work

There have been several approaches to replace the authentication protocol for the classical channel by quantum approaches. For example, an authentication scheme is presented in [11], which provides an increased conditional entropy for the seed of the adversary and which is optimized for scenarios where the shared symmetric key used in the authentication becomes extremely short.

Other protocols entirely eliminate the classical channel thus also eliminating the need for classical authentication [12]. Such protocols make use of quantum authentication, a topic which has been studied for more than 15 years and which has already been formally defined in 2002 [13]. Quantum authentication protocols perform the task of authentication with little or no help of classical cryptography solely using quantum mechanical sources. Hence, some of these protocols combine QKD protocols with authentication [14] or use quantum error correction for the authentication of the communication parties [15]. Other quantum authentication protocols also use entanglement as a source for authentication (e.g., [16][17][18] to name a few). Entangled states consist of two or more particles which have the specific property that they give completely correlated results when the respective particles are measured separately. As it has been shown by Bell [19], as well es Clauser et al. [20], this correlation can be verified if the measurement results violate some special form of inequalities. In some QKD protocols, for example the Ekert protocol [21], this argument is used to generate a secure key, but these protocols still require an authenticated classical channel (cf. [21]).

II. Assembling Authentication into the Protocol

In a standard person-in-the-middle scenario, we have Eve sitting in between Alice and Bob, executing BB84 with both of them simultaneously.

Alice and Bob, to authenticate one another, make contact *out of band*, by contacting the other on a physically and logically separate channel that Eve has not intercepted. In that sense, we augment the usual picture of BB84 by another channel, shown dashed in Figure 1.

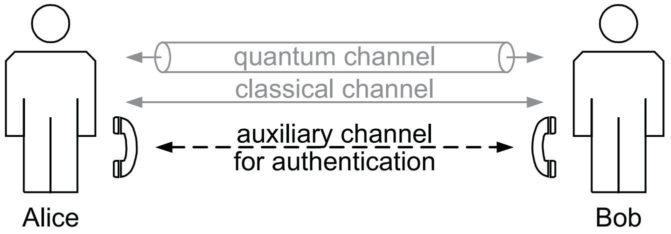


Figure 1. Channel configuration of our enhanced protocol

The key point here is that during the public discussion phase of BB84, Alice and Bob both reveal to each other their entire random sequence of polarization settings, along which their – so far private – random sequences are disclosed. Within these private random sequences, Alice will embed a pseudorandom subsequence that is indistinguishable from the truly random rest of the sequence, but for which she can tell Bob the way in which she constructed the bits and their positions. Our intuition behind this is that Alice, running BB84 with Eve, and Eve in turn running BB84 with Bob, Eve will not know (nor can determine) which of the transmitted bits are pseudorandom, and which are not. In turn, she cannot reproduce or relay these specific bits to her communication with Bob, in order to mimic Alice's behavior correctly.

Upon authentication, which happens after the public discussion phase and before the final key is distilled, Bob will get the information required to reproduce Alice's pseudorandom sequence on his own. If he were talking to Eve instead, his recorded bitstream will – with a high likelihood – not match what he received from Eve, thus revealing her presence.

Now, let us make this more rigorous. In the following, let $|x|$ denote the bitlength of a string x, and let $t \in \mathbb{N}$ be a security parameter. By the symbol $x \overset{r}{\leftarrow} \Omega$, we denote a uniformly random draw of an element x from the set Ω. Let $\mathcal{H} = \left\{ H_k : \{0,1\}^t \to \{0,1\}^t \,|k \in \{0,1\}^t \right\}$ be a family of *permutations*, which will act as uniform hash-functions in our setting (note that our scenario permits this exceptional assumption, as our goal is not as usual on hashing arbitrarily long strings, but on producing pseudorandom sequences by iteration). Furthermore, let m be an integer that divides 2^t.

Under this setting, let us collect some useful observations: take $x \overset{r}{\leftarrow} \{0,1\}^t$, then for any k, the value $H_k(x)$ must again be uniformly distributed over $\{0,1\}^t$, since H_k is a permutation. Likewise, since m divides 2^t, the value $H_k(x) \bmod m$ is uniformly distributed over $\{0, 1, \ldots, m-1\}$.

To embed authentication information in her bit stream, Alice secretly chooses two secret values $k_v, k_p \overset{r}{\leftarrow} \{0,1\}^t$ define a permutation H_{k_v} on $\{0,1\}^t$ and a function $h_k(x) := 1 + [H_{k_p}(x) \mod m]$ on $\{1, 2, \ldots, m\}$. Using these two functions, she produces a pseudorandom sequence of *values* $v_{n+1} = H_{k_v}(v_n)$ and another (strictly increasing) pseudorandom sequence of *positions* $p_{n+1} = p_n + h_{k_p}(p_n)$, with starting values $v_0, p_0 \overset{r}{\leftarrow} \{0,1\}^t$.

Within the first phase of BB84, i.e., when the randomly polarized qubits are being transmitted, Alice uses the pseudorandom information $f(v_i)$ whenever the p_i-th bit is to be transmitted, and true randomness otherwise. In other words, Alice constructs the bitstream

$$(b_n)_{n\in\mathbb{N}} = (b_0, b_1, \ldots, b_{p_i-1}, b_{p_i} = f(v_i), b_{p_i+1}, \ldots) \tag{2}$$

with truly random b_i whenever $i \notin \{p_0, p_1, \ldots\}$ and inserts a pseudorandom value v_i at each position p_i for $i = 1, 2, \ldots$. This sequence determines the respective qubit stream upon polarizing photons according to $(b_n)_{n\in\mathbb{N}}$.

A. *Authentication*

To authenticate, Bob calls Alice on a separate line and asks for k_p, k_v, v_0, p_0, which enables him to reproduce the pseudorandom sequence and bits and to check if these match what he has recorded. He accepts Alice's identity as authentic if and only if all bits that he recorded match what he expects from the pseudorandom sequence. The converse authentication works in the same way.

B. *The Auxiliary Public Channel*

We stress that the auxiliary public channel does not need to be confidential. However, some sort of authenticity is assumed, but without explicit measures for it. This is because authenticity in our proposal relies on the assumption that the adversary is unable to intercept *both* public channels at the same time (otherwise, a person-in-the-middle attack is impossible to counter in the absence of pre-shared secrets).

The assumption of an auxiliary public channel puts security to rest on Eve not intercepting now two public channels simultaneously. If more such channel redundancy is available, then known techniques of multipath transmission allow to relax our assumption towards stronger security (by enforcing Eve to intercept > 2 paths in general). We believe this approach to practically impose only mild overhead, since many reference network topologies and multi-factor authentication systems successfully rely on and employ multiple independent and logically disjoint channels, at least for reasons of communication infrastructure availability. Suitable multipath transmission techniques [22] are well developed and successfully rely on exactly this assumption (although pursuing different goals [23]). Moreover, a common argument against multipath transmission (which technically offers an entirely classical alternative to quantum key distribution with very similar security guarantees) that relates to the blow-up of communication overhead does not apply to our setting here. The amount of information being exchanged over the auxiliary (multipath) channel is very small, thus making the additional overhead negligibly small. Therefore, the only physical obstacle that remains is a topology permitting the use of multiple channels; however, many physical network reference topologies are at least bi-connected graphs and thus offer the assumed additional channel (besides the usually valid assumption on the co-existence of independent communication infrastructures besides the quantum network).

III. Security

First, observe that endowing Eve with infinite computational power could essentially defeat any form of authentication, since Eve in that case could then easily intercept Alice and Bob's communication by a two-stage attack: First, she would let Alice and Bob do a normal run of BB84, sniffing on the authenticated public discussion and doing passive eavesdropping to make Alice and Bob abort the protocol and abandon the key. Before Alice and Bob restart again, Eve can – thanks to unlimited computing power – extract or simply guess-and-check the authentication secret, so as to perfectly impersonate Alice and Bob as person-in-the-middle during their next trial to do BB84. If Alice and Bob decide to use another authentication secret this time, Eve will fail the authentication but will have further data to learn more authentication secrets, until Alice and Bob eventually run out of local keys. Thus, Eve has a good chance to succeed ultimately.

Even if a universal hash function is in charge, the universality condition and the fact that strings of arbitrary length are hashed, both guarantee the existence of more than one possible key (hashes) that would produce the given result. Thus, the residual uncertainty about the authentication secret remains strictly positive. However, this residual uncertainty is not necessarily retained in cases where consistency with three or more MACs is demanded.

Therefore, it appears not too restrictive to assume that Eve cannot recognize the pseudorandom part in $(b_n)_{n\in\mathbb{N}}$ from the truly random portion, as neither the number nor the position of the pseudorandom bits is known. In other words, if N bits have been used, then Eve would have to test all 2^N subsets against their complements to decide which bits to pass through in either direction. However, even if she succeeds and recognizes which bits are the pseudorandom ones and how they have been created (i.e., if she finds the proper keys and preimages to the hash-values), this information becomes available too late, as the relevant protocol phase has been completed by this point.

Let us compute the likelihood for Alice to tell Bob the correct values, although Bob ran BB84 with Eve who impersonated Alice. Hence, the chances for Eve to remain undetected equal the likelihood for Alice's and Bob's pseudorandom sequences to entirely match by coincidence. We compute this probability now.

Let $X_1, \ldots, X_n$ be the random variables (position *and* value) corresponding to Alice's pseudorandom part in $(b_n)_{n\in\mathbb{N}}$. Likewise, let $y_1, \ldots, y_n$ be what Bob expects these values to be upon Alice's response to his authentication request. Define the random indicator variable $\chi_k = 1 :\Longleftrightarrow X_k = y_k$, for $1 \leq k \leq n$. Bob buys Alice's claimed identity if and only if $\sum_{k=1}^{n} \chi_k = n$. Hence, we look for a tail bound to $S_n := \sum_{k=1}^{n} \chi_n$ in terms of n.

By construction, the sequence $X_1, \ldots, X_n$ is identically but not independently distributed. More precisely, each realization x_k of X_k points to a position p_k and value $v_k = b_{p_k}$ expected at this position, where position and value are stochastically independent.

So, let us compute the likelihood that Bob finds the expected bit at the told position, i.e.,

$$\Pr[X_k = y_k] = \mathbb{E}[\chi_k] = \Pr[b_{p_k} = v_k] \tag{3}$$

Since each b_i in the sequence $(b_i)_{i=1}^{n}$ is uniformly distributed irrespectively of its particular position, we get $\Pr[b_{p_k} = v_k] = 1/2$. Hence, as $\mathbb{E}[\chi_k]$ is bounded within $[0, 1]$ and the expectations of all $\mathbb{E}[\chi_k]$ are independent (although the χ_k's themselves are indeed dependent as emerging from a deterministic process), we can apply Smith's version [24] of the Hoeffding-bound to obtain

$$\Pr[S_n - \mathbb{E}[S_n] \geq \varepsilon] \leq \exp\left(-\frac{2\varepsilon^2}{n}\right) \tag{4}$$

Applied to the event $S_n \geq \varepsilon + \mathbb{E}[S_n] = n$ and considering $\mathbb{E}[S_n] = \sum_{k=1}^{n} \mathbb{E}[\chi_k] = n/2$ we may set $\varepsilon = n/2$ to conclude that a pseudorandom sequence constructed from random, i.e., incorrect, authentication secrets, will make Bob accept with likelihood

$$\Pr[\text{all } X_n \text{ match}|\text{incorrect seeds}] = \Pr[S_n \geq n] \leq e^{-n/2}. \tag{5}$$

Now, we can compute the overall probability of a successful impersonation from the law of total probability. Eve will successfully convince Bob to be Alice, if any of the following two events occur:

E_1: She correctly guesses the authentication secrets, in which case Bob's reconstructed pseudorandom sequence matches his expectations. Thus, $\Pr[\text{all } X_n \text{ match}|\text{correct seeds}] = 1$, obviously. However, $\Pr[E_1] = 2^{-O(t)}$, since the authentication secrets are chosen independently at random and have bitlength t (implied by the security parameter).

E_2: She incorrectly guesses the authentication secrets, and thus presents a "random" pseudorandom sequence to Bob. The likelihood of success is bounded by (5), and the likelihood for E_2 to occur is $1 - 2^{-O(t)}$.

The law of total probability then gives

$$\begin{aligned} \Pr[\text{Bob accepts}] &= \Pr[\text{all } X_n \text{ match}] = && (6)\\ &= \Pr[\text{all } X_n \text{ match}|E_1]\Pr[E_1] \\ &+ \Pr[\text{all } X_n \text{ match}|E_2]\Pr[E_2] && (7)\\ &\leq e^{-n/2}(1-2^{-O(t)}) + 2^{-O(t)} \leq 2^{-O(t+n)}, && (8)\end{aligned}$$

where n is the number of pseudorandom bits embedded, and t is the security parameter (bitlength of authentication secrets).

IV. Conclusion

Authentication is a crucial issue for quantum key distribution and can be tackled in several ways. Traditionally, this matter is handled by authentication based on strong symmetric cryptography, which makes shared secrets necessary in the standard setting. These shared secrets can, however, be replaced by assumptions on the availability of additional communication channels, similarly as in multipath communication. Indeed, by having the peers in a BB84 protocol embed pseudorandomness in their qubit stream, we can use out of band authentication in a straightforward form to secure a BB84 execution. Our treatment here so far does not account for measurement errors, say when a pseudorandom qubit goes lost (recovery from measurement errors may be easy upon simply discarding lost qubits from the check; at the cost of taking more pseudorandom bits accordingly), or discusses applications to other forms of quantum key distribution. Details, issues and implications of such modifications in other protocols are to be discussed in future work.

References

[1] D. R. Stinson, "Universal hashing and authentication codes," in CRYPTO '91: Proceedings of the 11th Annual International Cryptology Conference on Advances in Cryptology. London, UK: Springer, 1992, pp. 74–85.

[2] G. Gilbert and M. Hamrick, "Practical quantum cryptography: A comprehensive analysis (part one)," 2000, [retrieved: june, 2015]. [Online]. Available: http://arxiv.org/abs/quant-ph/0009027

[3] C. Bennett and G. Brassard, "Public key distribution and coin tossing," in IEEE International Conference on Computers, Systems, and Signal Processing. Los Alamitos: IEEE Press, 1984, pp. 175–179.

[4] B. Huttner and A. Ekert, "Information Gain in Quantum Eavesdropping," J. Mod. Opt., vol. 41, no. 12, 1994, pp. 2455–2466.

[5] C. H. Bennett, F. Bessette, G. Brassard, L. Salvail, and J. Smolin, "Experimental Quantum Cryptography," J. Crypt., vol. 5, no. 1, 1992, pp. 3–28.

[6] C. H. Bennett, G. Brassard, and J. M. Robert, "Privacy Amplification by Public Discussion," SIAM Journal of Computing, vol. 17, no. 2, 1988, pp. 210–229.

[7] M. N. Wegman and J. L. Carter, "New Hash Functions and their Use in Authentication and Set Equality," Journal of Computer and System Science, vol. 22, 1981, pp. 265–279.

[8] T. Tsurumaru and M. Hayashi, "Dual Universality of Hash Functions and Its Applications to Quantum Cryptography," IEEE Transactions on Information Theory, vol. 59, no. 7, 2013, pp. 4700–4717.

[9] P. Shor and J. Preskill, "Simple Proof of Security of the BB84 Quantum Key Distribution Protocol," Phys. Rev. Lett., vol. 85, no. 2, 2000, pp. 441–444.

[10] R. Renner, "Security of quantum key distribution," Ph.D. dissertation, Dipl. Phys. ETH, Zurich, Switzerland, 2005.

[11] F. M. Assis, A. Stojanovic, P. Mateus, and Y. Omar, "Improving Classical Authentication over a Quantum Channel," Entropy, vol. 14, no. 12, 2012, pp. 2531–2549.

[12] N. Nagy and S. G. Akl, "Authenticated quanntum key distribution without classical communication," Parallel Processing Letters, vol. 17, no. 03, 2007, pp. 323–335.

[13] H. Barnum, C. Crepeau, D. Gottesman, A. Smith, and A. Tapp, "Authentication of Quantum Messages," in Proceedings of the 43rd Annual IEEE Symposium on the Foundations of Computer Science (FOCS'02). IEEE Press, 2002, pp. 449–458.

[14] M. Dušek, O. Haderka, M. Hendrych, and R. Myska, "Quantum Identification System," Phys. Rev. A, vol. 60, no. 1, 1999, pp. 149–156.

[15] J. G. Jensen and R. Schack, "Quantum Authentication and Key Distribution using Catalysis," quant-ph/0003104 v3, 2000, [retrieved: june, 2015]. [Online]. Available: http://arxiv.org/abs/quant-ph/0003104

[16] H. N. Barnum, "Quantum Secure Identification using Entanglement and Catalysis," quant-ph/9910072 v1, 1999, [retrieved: june, 2015]. [Online]. Available: http://arxiv.org/abs/quant-ph/9910072

[17] Y.-S. Zhang, C.-F. Li, and G.-C. Guo, "Quantum Authentication using Entangled State," quant-ph/0008044 v2, [retrieved: june, 2015]. [Online]. Available: http://arxiv.org/abs/quant-ph/0008044

[18] M. Curty and D. J. Santos, "Quantum Authentication of Classical Messages," Phys. Rev. A, vol. 64, no. 6, 2001, p. 062309.

[19] J. Bell, "On the Einstein Podolsky Rosen Paradox," Physics, vol. 1, 1964, pp. 403–408.

[20] J. F. Clauser, M. A. Horne, A. Shimony, and R. A. Holt, "Proposed Experiment to Test Local Hidden-Variable Theories," Phys. Rev. Lett., vol. 23, no. 15, 1969, pp. 880–884.

[21] A. Ekert, "Quantum Cryptography Based on Bell's Theorem," Phys. Rev. Lett., vol. 67, no. 6, 1991, pp. 661–663.

[22] M. Fitzi, M. K. Franklin, J. Garay, and S. H. Vardhan, "Towards optimal and efficient perfectly secure message transmission," in 4th Theory of Cryptography Conference (TCC), ser. Lecture Notes in Computer Science LNCS 4392, S. Vadhan, Ed. Springer, 2007, pp. 311–322.

[23] H. Han, S. Shakkottai, C. V. Hollot, R. Srikant, and D. Towsley, "Multipath TCP: a joint congestion control and routing scheme to exploit path diversity in the internet," IEEE/ACM Trans. Netw., vol. 14, December 2006, pp. 1260–1271.

[24] W. D. Smith, "Tail bound for sums of bounded random variables," scorevoting.net/WarrenSmithPages/homepage/imphoeff.ps, April 2005, [retrieved: june, 2015].

An Improved Hirata Algorithm for Quantum Circuit LNN Conversion

Angel Amarilla, Joaquín Lima, Benjamín Barán
Facultad Politécnica
Universidad Nacional de Asunción
San Lorenzo, Paraguay
e-mail: fangelith@gmail.com, joaquin.lima@pol.una.py, bbaran@pol.una.py

***Abstract*—Hirata et al. proposed an efficient technique for converting general quantum circuits to Linear Nearest Neighbor architecture where an optimal transformed circuit is calculated from the original circuit. However, for some circuits, this algorithm still requires a considerable amount of running time for the conversion. Therefore, in this paper, additional techniques based on *Dynamic Programming* and *Branch & Bound* are proposed in order to improve the running time. Several test circuits from the state of the art have been tested. Experimental results demonstrate the effectiveness of the proposed improvements to reduce the original running time without any loss in solution quality measured as the number of SWAP gates that have been added.**

Keywords*—Quantum Computation; Quantum Circuit Conversion; Linear Nearest Neighbor; Dynamic Programming; Branch & Bound.

I. Introduction

A *quantum computer* is a device whose operation is based on the principles of quantum mechanics. Essentially, these devices allow the implementation of State Models based on performing *gates* operations over qubits, allowing algorithms that may potentially be exponentially faster than their classic counterparts [1]. Currently, these algorithms are implemented in *quantum circuits*. The theoretical design of quantum circuits generally assume the possible interaction of any pair of qubits. However, today quantum computers can only make operations between physically adjacent qubits [2][3]. The architecture of quantum circuits that only consider the interaction of adjacent qubits is called Linear Nearest Neighbor (LNN) [4].

Hirata et al. [5] present a traditional computing algorithm for the conversion of arbitrary quantum circuits to the LNN architecture. This technique is based on inserting *SWAP gates* within the original circuit in order to leave all operations between adjacent qubits. The SWAP gates are inserted taking into account the conversion efficiency of up to w subsequent gates, being w a *depth value* that defines the local search intensity. An appropiate depth value represents a trade-off between the conversion time and the quality of the solutions. Thus, a lower w value might result in a smaller solution quality while a higher w would result in a large calculation time.

This paper proposes two improvements that reduces the running time of Hirata algorithm for the same w values without loss of solution quality. The first approach is based on *Dynamic Programming* [6], in which the solution of circuit patterns are stored and reapplied when these patterns appear again. The second improvement is based on *Branch & Bound* [7], and reduces the time of exploration of the search tree when the next gates are considered for the conversion, avoiding path exploration that will not provide an improvement.

The rest of this work is organized as follows. The next section introduces quantum circuits. Section III presents the LNN Architecture. Section IV discusses the conversion of quantum circuits to an LNN Architecture. Later, in Section V the proposed methods are exposed. Finally, Section VI presents experimental results and the conclusions of this work.

II. Quantum Circuits

A quantum bit or *qubit* [4] is modeled as a mathematical entity that can hold two basic states $|0\rangle$ and $|1\rangle$, which are analogous to states 0 and 1 of classics bits, and also can hold a lineal combination of the basic states: $\alpha|0\rangle + \beta|1\rangle$, called superposition of states [8].

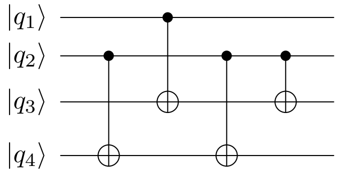


Figure 1. Example of a general quantum circuit with 4 qubits and 4 gates.

On other hand, *quantum gates* are capable of taking as input the states of a given number of qubits and to change the states of these qubits in a desired manner. In practice, a number of consecutive quantum gates conforms a *quantum circuit*. Figure 1 illustrates a quantum circuit, where the horizontal lines represent qubits being operated by quantum gates, which are shown transversely, taking as input one or more qubits. Thus, Figure 1 shows a circuit of 4 qubits and 4 gates.

III. Lineal Nearest Neighbor Architecture

In designing a quantum circuits, it is generally supposed that the quantum gates can operate any pair of qubits for example in [9] and [10]. However, quantum computers that allow a real implementation of quantum circuits with current technology, may not support the interaction of arbitrary further qubits. Indeed, some quantum architectures require circuits in where its quantum gates can only operate qubits that are physically adjacent [2][3]. The architecture of quantum circuits in which only adjacent qubits interact through a gate is called LNN [4].

A quantum gate of special attention is the *SWAP gate*, that is capable of taking two input qubits and interchange their states. The operation of a SWAP gate is given by $swap(|a, b\rangle) =$

ISBN: 978-1-61208-431-2

$|b, a\rangle$. The representation of a SWAP gate in quantum circuits is given by the symbol × on two adjacent qubits as shown in Figure 2.

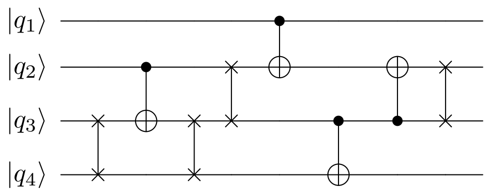


Figure 2. LNN circuit that is obtained adding four SWAP gates to circuit of Figure 1.

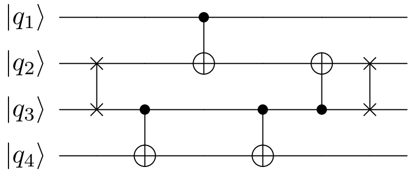


Figure 3. LNN circuit that is obtained adding two SWAP gates to circuit of Figure 1.

In general, quantum circuits can be converted to an LNN architecture by inserting additional SWAP gates within them. Figures 2 and 3 show two LNN circuits that are obtained by inserting additional SWAP gates in the general quantum circuit of Figure 1. It is interesting to note that in Figure 2 four new SWAP gates have been added to the original circuit, while in Figure 3 only two new SWAP gates are needed for the same task. In those cases, the insertion of a smaller number of new SWAP gates is preferred.

IV. LNN CONVERTION OF QUANTUM CIRCUITS

The conversion of a general quantum circuit to LNN architecture is defined by Hirata [5] as:

- **Input**: A general quantum circuit.
- **Output**: An equivalent LNN quantum circuit.
- **Objective**: Minimize the number of added SWAP gates.
- **Restriction:** the equivalent circuit output should have all qubits in the same original order.

$$\begin{array}{cccc} |q_1\rangle & |q_4\rangle & |q_2\rangle & |q_3\rangle \\ |q_2\rangle & |q_1\rangle & |q_4\rangle & |q_3\rangle \\ |q_2\rangle & |q_3\rangle & |q_1\rangle & |q_4\rangle \end{array}$$

Figure 4. Permutations considered for the Hirata algorithm.

One possible strategy to solve this problem is to convert each original gate inserting the smaller possible number of new SWAP gates. After the conversion of an original gate it must be proceed to converting the next original gate, taking into account the new order of the qubits according to the SWAP gates inserted before. Finally, after the conversion of the last original gate a small number of new SWAP gates it must be inserted in order to get the order of qubits given by the original circuit. This algorithm is known as *Greedy Strategy* [5], which is always able to find reasonable (sub-optimal) solutions.

V. HIRATA ALGORITHM

Hirata et al. [5] proposed an efficient algorithm to convert general quantum circuits to an LNN architecture.

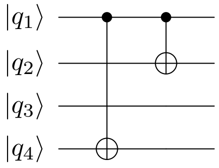


Figure 5. Quantum circuit of example.

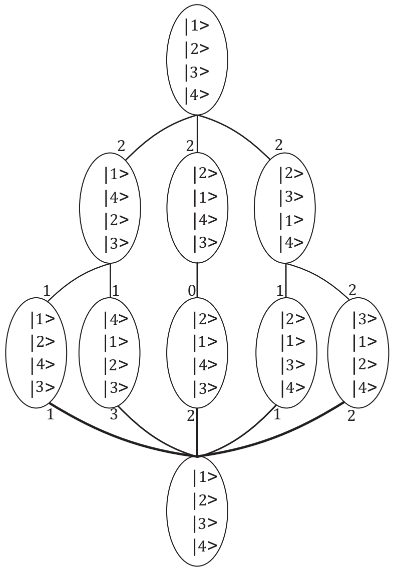


Figure 6. Whole search Tree for the quantum circuit of Figure 5.

For each pair of non-adjacent operated qubits that have m intermediate qubits, Hirata algorithm take into account $(m+1)$ possible permutations. For example, for the case of converting the first gate of the circuit of Figure 5, the qubits to be operated are $|q_1\rangle$ and $|q_4\rangle$, the value of m is 2 and the permutations considered by the algorithm are listed in the Figure 4. Note that in all cases, $|q_1\rangle$ and $|q_4\rangle$ are become neighbors, preserving the original order between them.

In Hirata algorithm [5], each considered permutation is called a *candidate*. Each candidate represents a possible reorder of the qubits to convert the current gate and it is a part of one or more solutions. In Figure 6, the conversion of the first gate of the quantum circuit from Figure 5 have 3 candidates. The numerical value over each candidate represents its required number of SWAP gates. Thus, the cost for each complete solution is given by the sum of these numerical values. The desired solution is the one with the lowest cost, that is illustrated in Figure 6 by the central path, with a final cost of $2 + 0 + 2 = 4$ additional swap gates.

```
swaps  = 0
l_order = {1,2,...,n}
c_order = l_order
w = local search deep
K = total of original gates
procedure HirataAlgorithm() {
  for i=1 to K do
    val_func_min = $\infty$
    S = $\emptyset$
    for j = 1 to candidates(i) do
      val_func=local_search(n_$order_{ij}$,w)
           +calc_swap(c_order,n_$order_{ij}$)
           +$\frac{c_k}{K+1-i}$calc_swap(n_$order_{i,j}$,$l\_order$)
      if (val_func < val_func_min) then
        val_func_min = val_func
        S = {n_$order_{i,j}$}
      else if (val_func = val_func_min) then
        S = S $\cup$ n_$order_{i,j}$
      end if
    end for
    $S_p$ = random(1,$|S|$)
    swap = swap + calc_swap(c_order, $S_p$)
    c   _order = $S_p$
  end for
  swap = swap + calc_swap(c_order,l_order)
  return swap
end procedure HirataAlgorithm
```

Figure 7. Hirata algorithm from LNN convertion.

Hirata algorithm [5], presented in Figure 7, performs the selection of a candidate for each gate considering:

- a Local Search procedure that evaluates the quality of the candidates orders (indicated by n_order) considering the following w gates;
- the amount of SWAP gates to be added to obtain the candidate order (n_order) from the current order (c_order);
- and, the cost of converting the candidate order (n_order) to the original order (l_order) of the quantum circuit, weighted by a value that takes precedence towards the end of the circuit.

VI. Proposed improvements

This section presents two improvements for the algorithm of Hirata et al. [5]: one based on Dynamic Programming [6] [11] and the other based on Branch & Bound [7].

A. Branch & Bound improvement

This improvement is applied in the local search procedure to prevent the exploration of branches of the search tree that will not lead to a better solution than the one in search [7].

Figure 8 illustrates a complete *local search tree* corresponding to candidate $|q_1q_2q_4q_3\rangle$. Initially there is no *selected solution*. The calculation begins considering the first candidate that is considered as the selected one. For example, consider as the *first selected solution* the candidate most in the left of the tree (see Figure 8) whose evaluation equals to 4 at the end of the local search tree. Note also that the height of local search tree is defined by $w = 2$.

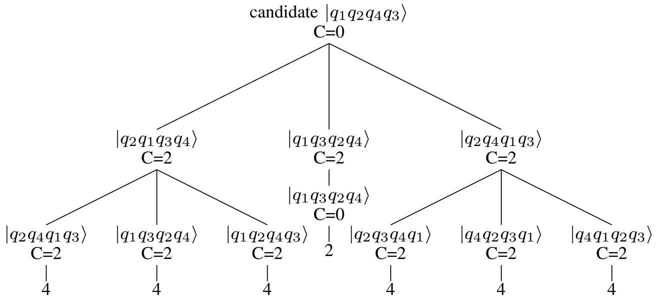


Figure 8. Local search tree corresponding only to candidate $|q_1q_2q_4q_3\rangle$ of the first gate of Figure 1 circuit.

Afterward, the algorithm proceed in evaluating the *next candidate*, whose evaluation culminate only if it is not worse than the current selected solution. This is determined taking into account the amount of SWAP gates accumulated during the search tree exploration. If the amount of SWAP gates corresponding to the current candidate is larger than the amount of the current selected solution, then the exploration is stopped and the candidate is discarded.

When the evaluation of a candidate ends with a lower amount of SWAP gates than the corresponding amount of the current selected solution, the candidate is chosen as the *new selected solution*. Consider again the search tree of Figure 8. Initially, the selected solution is the candidate most in the left, that is subsequently replaced by the candidate of the center.

With this improvement, the local search procedure can find the same solution as the original local search algorithm, but with the advantage of not fully evaluating the search tree, reducing this way its processing time. Figure 9 illustrates the search tree that is explored when the Branch & Bound improvement is applied. Note that it is not necessary to fully evaluate the final candidate

The local search procedure with the improved proposal of Branch & Bound is presented in Figure 10.

B. Dynamic Programming improvement

Dynamic Programming [6] is a technique that solves a problem P based on recursively solving all sub-problems S_i therein. Each sub-problem S_i is solved only once and its solution is saved in a table $T[S_i]$. When the sub-problem S_i is found once again, the solution saved in $T[S_i]$ is reapplied; therefore, the time and effort spent in solving S_i is saved.

The proposed improvement apply Dynamic Programming [6] in the Hirata algorithm considering as a sub-problem S_i each current order of qubits c_order and its w corresponding successive pairs of qubits.

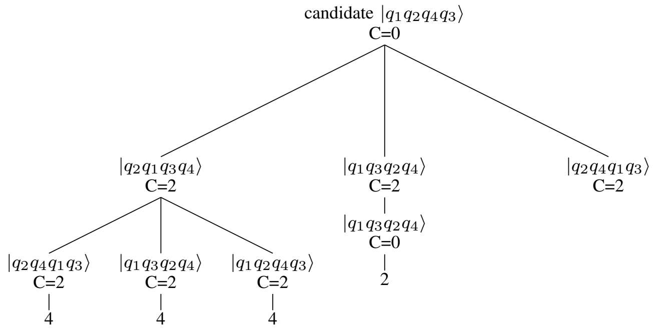


Figure 9. Local search tree explored for candidate $q_1q_2q_4q_3$ of the first gate of circuit of Figure 1 when considering the Branch & Bound improvement.

```
K = total of original gates
c_swap = SWAP gates in current solution
min_swap = SWAP gates in candidate solution
k = current gate in convertion
i = current gate in local search
w = deep value

procedure local_search(c_swap,i,c_order,k)
  if (i > K) or (i > k + w) then
    if c_swap < min_swap then
      min_swap = c_swap
    end if
  else
    for j = 1 to (m+1) candidates do
      n_swap=c_swap + calc_swap(c_order,n_$order_{i,j}$)
      if n_swap < min_swap then // Branch&Bound
        min_swap =
          local_search(n_swap,i+1,n_$order_{i,j}$,k)
      end if
    end for
  end if
  return min_swap
end procedure local_search
```

Figure 10. Local Search procedure with Branch & Bound.

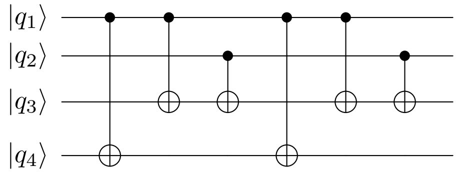


Figure 11. Example circuit for illustrate the Dynamic Programming improvement

For example, consider the circuit of Figure 11 and $w = 2$. The sub-problem in the first gate is as follows:

- c_order_1 = $q_1q_2q_3q_4$
- w next pairs of qubits = $[(q_1, q_3), (q_2, q_3)]$

Considering that the calculated solution for this sub-problem adds the SWAP gates between (q_1, q_2) and between (q_3, q_4) before the first gate, as shown in Figure 12; then, the corresponding input in the table T of patterns is as follows:

$$T[q_1q_2q_3q_4, (q_1, q_3), (q_2, q_3)] = (q_1, q_2), (q_3, q_4) \qquad (1)$$

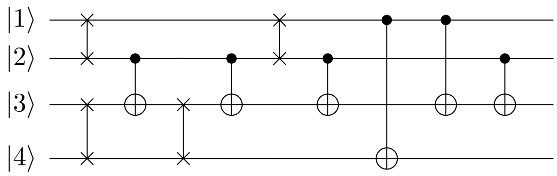


Figure 12. LNN conversion of the circuit 11 for the first 3 gates

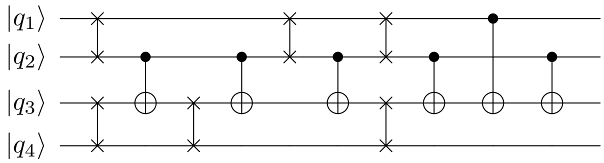


Figure 13. LNN conversion of the circuit 11 for the gate $k = 4$

Later, this sub-problem could need to be considered again. For example, when the conversion reaches up to the fourth gate, as shown in Figure 12. Here, the solution of equation (1) saved in the Table T is reapplied without recalculating the same pattern, as shown in Figure 13.

VII. Experimental Results

A. Combination of the proposed improvements

This section presents the experimental results obtained by the implementation of both proposed improvements in the original Hirata algorithm [5]. Nine test circuits were considered: three Shor circuits, three Modmulti circuits and three random circuits, also used in Hirata et al. [5].

For the performed experiments, a computer with 3 GHz Dual Core i5 processor, 8 GB RAM and the implementation of the algorithms in Java 7 were set. The values of $w \in \{8, 9, 10\}$ have been considered. Given the randomness of the algorithm in case of ties, each test circuit has been converted 10 times using each algorithm:

- H: the original Hirata algorithm,
- H-BB-PD: Hirata algorithm with both proposed improvements (Dynamic Programming and Branch & Bound);
- H-PD: Hirata algorithm with Dynamic Programming; and
- H-BB: Hirata algorithm with Branch & Bound.

Thus, a total of 360 runs were considered, obtaining in each test the following metrics:

- the average running time ($\overline{T}$) and its standard deviation ($\sigma(T)$); and
- the average number of SWAP gates of the calculated solution ($\overline{S}$) and its standard deviation ($\sigma(S)$).

A comparison of results of the original Hirata algorithm (H) versus the same algorithm with both proposed improvements (H-BB-PD) are shown in table I. In general, it can be observed that proposed improvements can achieve a significant reduction in running time without an appreciable difference with respect to the number of calculated SWAP gates by the original Hirata Algorithm.

On the other hand, another considered experiment consists in taking into account both the elapsed running time and the total number of SWAP gates after each 50 processed original gates. Thus, Figure 14 presents a comparison of the

TABLE I. RESULTS OF THE ORIGINAL HIRATA ALGORITHM COMPARED TO THE PROPOSED IMPROVED ALGORITHM.

		W=8		W=9		W=10	
		Time (ms.)	SWAPs	Time (ms.)	SWAPs	Time (ms.)	SWAPs
Circuit	Algorithm	$\overline{T}/\sigma(T)$	$\overline{S}/\sigma(S)$	$\overline{T}/\sigma(T)$	$\overline{S}/\sigma(S)$	$\overline{T}/\sigma(T)$	$\overline{S}/\sigma(S)$
Shor 3	H	721 / 125.79	1876 / 34.78	1255 / 138.00	1870 / 25.64	2674 / 235.20	1880 / 0.0
	H-BB-PD	92 / 49.78	1876 / 34.78	106 / 12.59	1870 / 25.64	146 / 10.62	1880 / 0.0
Shor 5	H	21780 / 1375.69	11346 / 110.69	65072 / 3770.91	10959 / 0.0	194128 / 14956.65	11968 / 233.01
	H-BB-PD	765 / 111.05	11346 / 110.69	1484 / 324.06	10959 / 0.0	2380 / 301.39	11968 / 233.01
Shor 6	H	75134 / 3033.71	21000 / 520.83	244525 / 8562	21184 / 306	769350 / 47910	21198 / 253
	H-BB-PD	2636 / 491.29	21000 / 520.83	3752 / 799	21184 / 306	7004 / 1711	21198 / 253
Random 500	H	2004 / 153.46	748 / 1.15	5686 / 159.28	742 / 0.0	18419 / 482.62	752 / 3.02
	H-BB-PD	270 / 31.47	748 / 1.15	480 / 44.69	742 / 0.0	915 / 90.36	752 / 3.02
Random 1000	H	4076 / 202.39	1492 / 12.08	11914 / 394.75	1464 / 1.05	38652 / 271.91	1462 / 1.70
	H-BB-PD	563 / 55.94	1492 / 12.08	1000 / 103.17	1464 / 1.05	1866 / 147.84	1462 / 1.70
Random 2000	H	8920 / 126.29	3084 / 4.94	26650 / 659.38	3000 / 1.63	86859 / 868.22	3004 / 6.60
	H-BB-PD	1211 / 86.96	3084 / 4.94	2149 / 156.11	3000 / 1.63	4111 / 159.87	3004 / 6.60
Modmulti 3	H	143 / 12.56	322 / 7.73	307 / 31.82	320 / 6.82	740 / 78.67	328 / 1.15
	H-BB-PD	32 / 5.07	322 / 7.73	67 / 7.54	320 / 6.82	90 / 12.38	328 / 1.15
Modmulti 4	H	663 / 64.13	648 / 2.00	1714 / 113.42	654 / 3.89	5014 / 224.46	654 / 7.73
	H-BB-PD	142 / 19.94	648 / 2.00	256 / 16.39	654 / 3.89	425 / 35.33	654 / 7.73
Modmulti 5	H	2444 / 308.53	1138 / 10.80	6195 / 152.92	1124 / 1.94	19662 / 1022.59	1180 / 2.83
	H-BB-PD	535 / 54.45	1138 / 10.80	803 / 76.47	1124 / 1.94	1323 / 182.54	1180 / 2.83

performance of both algorithms for a value of $w = 8$ in test circuit Shor 6. It can be observed that the proposed improvements allow an equivalent solution in a shorter running time.

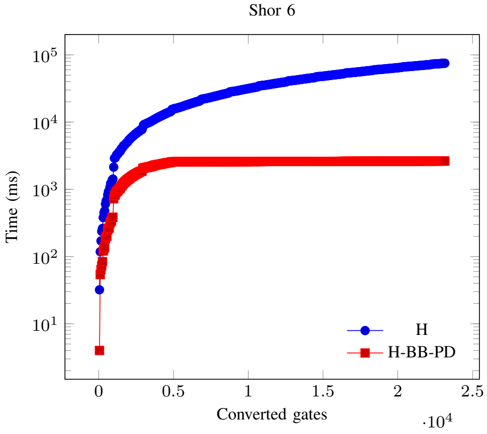


Figure 14. Evolution of conversion versus the conversion time between the original algorithm and both improvements.

B. Comparison between both methods

In order to determinate the individual advantage of using both improvement proposals, each of these two improvements were separately applied and then compared to the original Hirata algorithm [5]. Table II shows the normalized results of this experiment, using the notation:

$$\overline{T}_{norm} = \frac{\overline{T}_{alg}}{\overline{T}_H} \tag{2}$$

where $\overline{T}_{norm}$ is the normalized running time, $\overline{T}_{alg}$ is the value of the average running time of the considered algorithm, $\overline{T}_H$ is the average running time for the original Hirata algorithm. Thus, lower values of $\overline{T}_{norm}$ are preferred.

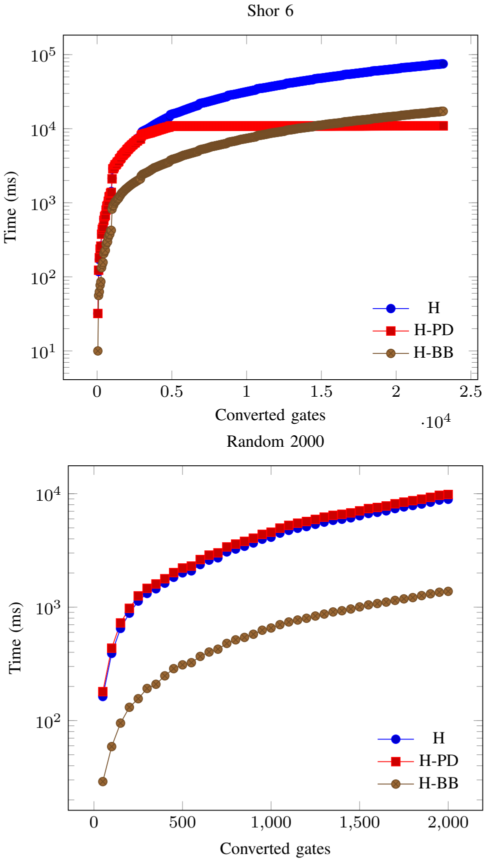


Figure 15. Evolution of conversion versus the conversion time for each improvement.

TABLE II. NORMALIZED COMPARISON BETWEEN IMPROVEMENTS.

		H-BB	H-DP	H
Benchmark	w	$\overline{T}_{norm}$	$\overline{T}_{norm}$	$\overline{T}_{norm}$
Shor 3	8	0,2927	0,3536	1
	9	0,2153	0,3593	1
	10	0,1622	0,3144	1
Shor 5	8	0,2012	0,1505	1
	9	0,1287	0,1509	1
	10	0,0730	0,1426	1
Shor 6	8	0,2069	0,1314	1
	9	0,1006	0,1177	1
	10	0,0565	0,1303	1
Random 500	8	0,1342	0,9017	1
	9	0,0818	0,8994	1
	10	0,0482	0,8912	1
Random 1000	8	0,1418	0,8929	1
	9	0,0833	0,8989	1
	10	0,0489	0,8880	1
Random 2000	8	0,1381	0,9863	1
	9	0,0819	1,0288	1
	10	0,0494	1,0386	1
Modmulti 3	8	0,2917	0,9375	1
	9	0,2016	0,8638	1
	10	0,1233	0,8657	1
Modmulti 4	8	0,2280	0,8297	1
	9	0,1578	0,9185	1
	10	0,0871	0,8961	1
Modmulti 5	8	0,2270	0,8996	1
	9	0,1241	0,8661	1
	10	0,0658	0,8738	1

Considering the improvement based on the Dynamic Programming technique, it can be seen that the test circuits Shor 3, Shor 5 and Shor 6 were converted in a shorter running time because several subcircuits are repeated during calculation. However, when there are few or no subcircuits that repeat within the original circuit to convert, such as in Random and Modmulti circuits, this improvement can not reach a better running time compared to the original Hirata algorithm. Thus, this method is only effective in circuits with repetitive patterns.

On the other hand, considering the Branch & Bound improvement, the results in all test circuits show a significant decrease in running time respect to the original Hirata algorithm. Therefore, this improvement is of general application in contrast to the Dynamic Programming approach which is more selective in its applicability.

There also have been taken samples to show the evolution of the resolution time versus the original number of converted gates. Figure 15 confirms that Dynamic Programming improvement only contribute to a reduction of the running time in circuits with repetitive patterns, such as in Shor circuits. On the other hand, all the plots in Figure 15 show a shorter running time when the improvement based on the Branch & Bound technique is applied. This result confirms the general application of the Branch & Bound improvement and the advantage of using it in almost any case.

VIII. CONCLUSIONS

In this work, it has been presented two proposals that improve the running time of the original Hirata algorithm [5] for the conversion of arbitrary quantum circuits to LNN architecture without any loss of quality in the calculated solutions.

The first improvement based on Dynamic Programming [6], proved to be effective converting circuits with repetitive constructions as Shor circuits, avoiding the recalculation of patterns that repeats; however, the improvement is only effective in such circumstances, and not in every studied circuit.

On the other hand, the second improvement based on Branch & Bound [7], has proved to be an improvement of general application with a positive effect over the original algorithm in all studied cases.

In general, the experimental results demonstrates that applying the two proposed improvements can achieve the same quality results that the original Hirata algorithm with a smaller running time without any loss of solution quality. Therefore, both proposals can be considered in order to implement an efficient version of the Hirata algorithm.

ACKNOWLEDGMENT

Special thanks to the Nara Institute of Science and Technology (NAIST) in the person of Yasuhiko Nakashima, Hiroyuki Seki, Masaki Nakanishi, Shigeru Yamashita and Yuichi Hirata, for all help and for providing the test circuits.

REFERENCES

[1] E. Strubell, "An introduction to quantum algorithms," COS498 Chawathe Spring, 2011.

[2] H. Häffner et al., "Scalable multiparticle entanglement of trapped ions," Nature, vol. 438, no. 7068, 2005, pp. 643–646.

[3] M. Laforest, D. Simon, J.-C. Boileau, J. Baugh, M. J. Ditty, and R. Laflamme, "Using error correction to determine the noise model," Physical Review A, vol. 75, no. 1, 2007, p. 012331.

[4] M. A. Nielsen and I. L. Chuang, Quantum computation and quantum information. Cambridge university press, 2010.

[5] Y. Hirata, M. Nakanishi, S. Yamashita, and Y. Nakashima, "An efficient conversion of quantum circuits to a linear nearest neighbor architecture," Quantum Info. Comput., vol. 11, no. 1, Jan. 2011, pp. 142–166.

[6] S. Dreyfus, "Richard bellman on the birth of dynamic programming," Operations Research, vol. 50, no. 1, 2002, pp. 48–51.

[7] J. Clausen, "Branch and bound algorithms-principles and examples," Department of Computer Science, University of Copenhagen, 1999, pp. 1–30.

[8] R. P. Feynman, "Quantum mechanical computers," Foundations of physics, vol. 16, no. 6, 1986, pp. 507–531.

[9] P. W. Shor, "Algorithms for quantum computation: Discrete logarithms and factoring," in Foundations of Computer Science, 1994 Proceedings., 35th Annual Symposium on. IEEE, 1994, pp. 124–134.

[10] S. A. Cuccaro, T. G. Draper, S. A. Kutin, and D. P. Moulton, "A new quantum ripple-carry addition circuit," arXiv preprint quant-ph/0410184, 2004.

[11] R. E. Bellman and S. E. Dreyfus, Applied dynamic programming. Rand Corporation, 1962.

Charge Qubits in Doped Quantum Dots : Effects on Computation and Coherence

[1*] Thierry Ferrus, [1]Tsung-Yeh Yang, [2]Yu Yamaoka, [2]Tomohiro Kambara, [2]Tetsuo Kodera, [2]Shunri Oda
and [1]David Arfon Williams
[1]Hitachi Cambridge Laboratory, J J Thomson avenue, CB3 0HE, Cambridge, United Kingdom
[2]Department of Physical Electronics, Quantum Nanoelectronics Research Centre, Tokyo Institute of Technology,
Tokyo, 152-8552 Japan
* Email: taf25@cam.ac.uk

***Abstract*—Doping in quantum information architectures is often presented as a source of decoherence for the qubit to study. However, high doped material also present additional properties that could be used in quantum computation. Here, we show that the application of GHz photons to a doped double quantum dot allows a pair of active dopant to be selected and gate operation to be implemented with potential coherence time exceeding 200 μs.**

***Index Terms*—Quantum computing; silicon; charge qubit; quantum dots**

I. INTRODUCTION

In a previous paper, we did present a scalable and industrially compatible architecture for performing quantum computation [1]. In this proposal, qubit states are defined by the spatial location of an extra charge in an highly doped double quantum dot (IDQD) [2][3]. The structure is realized by implanting at high dose phosphorous atoms into the 40-nm thick silicon layer of a silicon-on-insulator (SOI) wafer to obtain an effective doping density of about 3 10^{19} cm^{-3}. Full electrical insulation of the qubit is then achieved by etching away any conducting material between the various structures (gates, detector and qubit), a method leading to a significant decrease in the electron temperature of the qubit [4]. The device is further encapsulated by a 12-nm thick thermal silicon oxide with a final value for the dot diameter reaching about 60 nm. Doping by phosphorous atoms and the realization of constrictions in order to define the quantum dot tunnel barriers offer a substitutionary method to standard Metal-Oxide-Semiconductor (MOS) device fabrication, enhancing the scalability possibilities.

Despite attractiveness for manufacturing purposes, doped architectures are rarely used in practical qubit implementations and, most of the research on charge or spin qubits concentrates on the use of fin field effect transistors (FinFETs) or nanoscale transistor technology pushing the well know Moore law [5] below the 100-nm scale [6]. One obvious justification is the presence of dopants that are often associated with localization effects, especially at interfaces, causing random telegraph signal events and contributing to electronic noise. Ensemble of donors may also be difficult to manipulate due to electron-electron interaction and glassy behavior that may arise near the metal-to-insulator transition [7], whereas controlling single donors has not been possible until recently [8].

In this paper, we would like to discuss some major differences in operation between doped and undoped qubits, showing in particular that dopants are not necessarily detrimental to qubit operations if an adapted control is designed. We also discuss some surprising striking consequences on the value of coherence time in doped quantum dots and give some insight for the realizing spin qubits in a doped environment.

In Section II, we first discuss the realization charge qubit states in semiconductors. We then describe the differences in pulsing experiments between a doped (Sec. III. A.) and undoped device (Sec. III. B.), bringing the notion of $dopant\ degeneracy$. Effects on coherence are discussed in Section IV before some conclusions and comments on scalability are drawn.

II. CHARGE QUBIT STATES IN SILICON

Within the large number of qubit implementations, charge qubits are one of the simplest ways to realize quantum computation in semiconductors and, in particular, silicon. There are two possible qubits that could be implemented. In the first case, localized states are used and the $\mid 0\rangle$ and $\mid 1\rangle$ are defined by the spatial location of a charge in a double quantum dot. In this case, the quantum gate $\mid \sigma_x\rangle$ is simply implemented by the tunneling of an electron between the two dots across the tunnel barrier. Another possibility is to use hybridized states so that qubits states are defined by the bonding-antibonding states, similarly to the hydrogen molecule model. In both cases, there is the need for confinement. There are currently two different methods for achieving this.

The first is to use a standard Metal-Oxide-Semiconductor layer in order to create a two-dimensional electron gas at the Si-SiO$_2$ interface (Fig. 1a). A change in the voltage applied to the top metal gate will bend the silicon conduction band, accumulating electrons at the interface. This allows an easy control of the electron density and the single electron regime can potentially be attained in nanoscale structures and a double dot system can be created from corner states [9]. Confinement itself is then attained by shaping the gate around a predefined etched silicon dot or wire and then realizing a FinFET structure or by reducing the gate length to several tens of nanometers. In some structures, an additional back-gate can be used and gives the ability to control independently the electron-electron interaction and the disorder strength [10]. However, these types of devices are in general very susceptible and sensitive to defects, in particular Pb centers [11] resulting from the lattice mismatch between silicon and silicon oxide. Reliability of quantum operations also relies on a high quality oxide and, in the case of SiO_2, it should be free of fast diffusers like ions (K^+, Na^+...) or heavy metals (Cu, Au, Cr...) as this induces localized states at the interface [12]. This is potentially a problem for quantum computing schemes as this type of traps tend to localize spins 1/2 at the Si-SiO_2 interface.

The other possible method consists in implanting atoms at high dose and at energies exceeding few tens of keV that will substitute themselves to silicon atoms. In the case of phosphorous, the atom is pentavalent whereas silicon atoms have four covalent bonds (Fig. 1b). This leaves an extra free electron in the lattice and, at high concentration, this method provides a sufficiently high electron density so that conduction occurs in the device despite natural silicon being intrinsic and insulating at low temperatures. In this case, quantum dots are defined by etching away parts of the doped material and by realizing two constrictions where tunnel barriers are most likely to be formed due to the local modification of the electrostatic potential (Fig. 1c). It is important to realize that, some localized states will still be present even at very high concentration. This is the results of i) dielectric screening that enhance electron trapping around Pb centers at the Si-SiO_2 interface as well as phosphorous atoms that have diffused into the oxide during the device process, ii) partial electron screening at the center of the dot due to non-uniform doping and long-range disorder due to distant traps. Density of states for these localized centers is in general small but influence greatly transport properties. Here, most of the noise source arises from defects at the edge of the structure where non-(100) surfaces are present. Edge localization could be important and is responsible for the commonly observed aperiodic conductivity background in quantum dot transport measurements [2].

III. Pulsing Experiments and the Dopant Degeneracy

To operate a qubit, a set of quantum gates have to be defined so that the Bloch sphere could be entirely accessible and a pulsing scheme have to be devised. The latter is generally implemented by a series of DC voltage pulses applied to one or several gates or a series of microwave pulses at given frequencies and powers. For both methods, there is conceptually a major difference between doped and undoped devices.

A. Undoped devices

In a MOS-based quantum dot, quantum levels are well defined and their energies in the absence of gate or source-drain biases are generally well approximated within the constant interaction model [13]:

$$E = \sum_{i}^{n} n_i \varepsilon_i + n^2 e^2 / 2C \qquad (1)$$

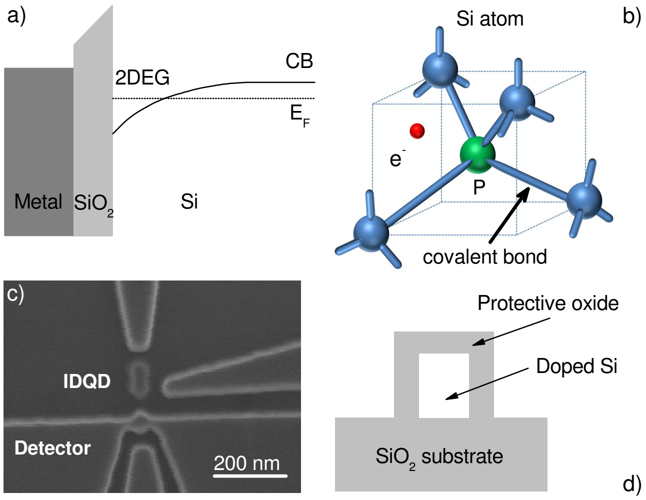


Fig. 1. a) 2D electron gas formation in a MOS structure. b)Structure of the silicon lattice after implantation with phosphorous atoms. c) SEM image showing an IDQD with its charge detector. d) Typical structure of a doped device.

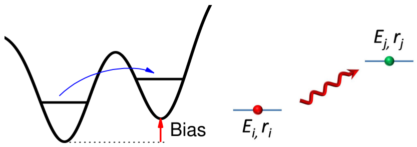


Fig. 2. Photon assisted tunneling in undoped devices.

where ϵ_i and n_i are respectively the single particle state energies due to the quantum dot energy quantization and the occupation number of the level i, n the quantum number and C the capacitance of the quantum dot that is linked to its diameter R by $C = 4\pi\epsilon_0 R$ where $\epsilon_0 \sim 11.7$ is the dielectric constant in silicon.

In the case of a DC pulse, the voltage at the end of the gate modifies the electric potential of the double dot. Quantized energies then acquire an additional term proportional to the pulse amplitude and dependent on the capacitive coupling between the gate and the qubit. This leads to a change in the structure of the levels that modifies the electron tunneling between the two dots (localized states) or modifies the double dot wavefunction (hybridized states).

If photons are to be used for realizing quantum operations, then the relative position of the levels of each dot remains unchanged and gets determined by the predefined values of the gate voltage. Photon assisted tunneling [14] is then used to perform gate operations. The resonant condition is obtained when the photon energy matches the difference in level energies. This effect is much less disturbing for the double dot as levels are not modified during the operation and consequently, this method is less sensitive to the eventual surrounding traps (Fig. 2)

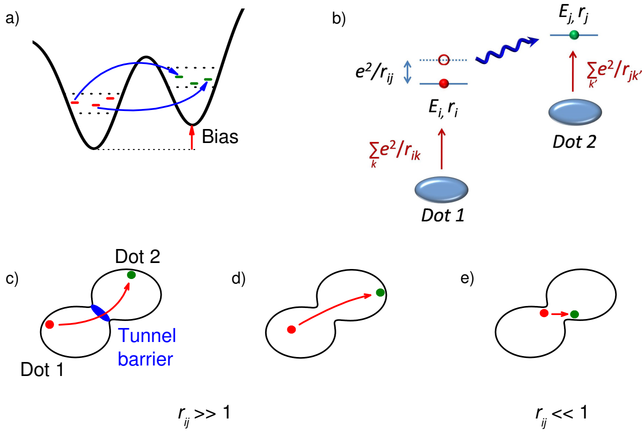


Fig. 3. Photon assisted tunneling in doped devices : a) non-equivalent tunneling, b) Microwave induced tunneling, c)-d) Possible site pairs involved.

B. Doped devices

In doped devices, the situation is far more complex. The presence of a narrow band of levels and surrounding traps lift the degeneracy that would have prevailed in the absence of disorder. Consequently, not all tunneling events remain equivalent (Fig. 3a). This makes a substantial difference with the undoped version of the device. It is important to notice that degeneracy should be recovered in principle if the temperature exceeds the average level energy spacing within the band but, this is not observed in practice. The reason comes from the bad thermalization of these structures by the phonon bath due to the phonon wavelength getting smaller than the dot dimensions ($\lambda_\gamma \sim 26$ nm in highly doped silicon at 4.2 K [15]). As a result, excitations with energies below $k_{\mathrm{B}}T$ can still be observed if the process involves a timescale shorter than the thermalization time and the longitudinal relaxation time T_1.

One consequence on DC pulsing, is that the pair of sites involved in the tunneling may not be the same if the experiment is repeated under the same conditions despite a single charge being transferred across the tunnel barrier each time. Although semi-classical operations like quantum-cellular-automata could be preformed with high probability, coherence may be affected in the quantum case. However, the situation is much different with photons. Indeed, it has been proposed recently [16] that, in doped devices, microwave photons can both select a pair of sites by their energy difference but also by their spatial location, so that the resonant condition differs sensibly from the usual photon assisted tunneling's by the addition of a interaction term (Fig. 3b):

$$h\nu = E_j - E_i - e^2/(4\pi\epsilon_0 r_{ij})\exp(-r_{ij}/\lambda_{\mathrm{TF}}) \qquad (2)$$

$\lambda_{\mathrm{TF}} \sim n_{\mathrm{e}}^{-1/2}$ is the Thomas-Fermi screening length and takes into account the presence and influence of the other electrons at a density of n_{e}. E_i and E_j are the respective level energies of site i and j whereas r_{ij} is the distance between the two sites.

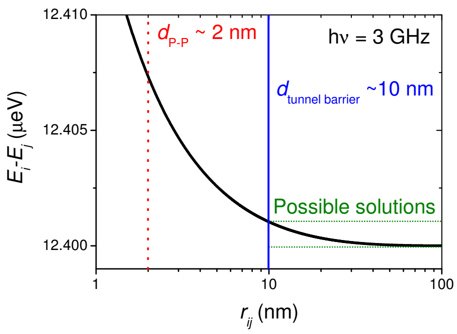


Fig. 4. Solutions of Eq. 2 for a 10 nm tunnel barrier and hν = 3 GHz.

If the two sites are spatially distant, the Coulomb term becomes negligible and resonance frequencies in the GHz range imply a substantial difference in levels energies. This situation can be encountered mostly at the edge of the structure where localized states and disorder are present and so, where the density of states has long tails (Fig. 3c). At the center of the dots, the individual wavefunctions overlap quite significantly leading to an effective screening by other surrounding electrons. Consequently, strong localization seems unlikely. However, the presence of acceptors (for example Boron that is used in p-type silicon wafer as background doping) that are trivalent and can consequently create a vacant site in silicon, can locally affect electron screening and weak localization is possible around such a defect. This would allow electrons to tunnel between two sites at the center of the dot (Fig. 3d). Indeed, recent simulations have shown that, in the case of trap formation at the center of the dot, electron screening is only partial.

On the contrary, if the sites are spatially close then Coulomb interaction is strong and sites involved are most likely located at or near the tunnel barriers (Fig. 3e).

C. Site pair selection and effect on coherence

Microwave pulsing is particularly adapted to doped structures due to its ability to select pairs of sites within a highly doped environment. However, for quantum computation, high fidelity has to be achieved and coherence time should be as long as possible, so the active pair has to be isolated from all the other possible pairs for at least T_1. The apparent difficulty in achieving this comes from the fact that the solution of equation 2 is, a priori, not unique and that the number of possible pairs depends both on the trap distribution and the disorder strength, e.g. the shape of the density of states of the localized states. Figure 4 shows possible solutions of Eq. 2 allowing charge tunneling between the two IDQD dots separated by a 10-nm tunnel barrier in the case hν = 3 GHz. This is in contrast with experimental findings where the uniqueness of the pair and long coherence times have been previously reported [16],[17].

In the case of a double quantum dot, the two high electron density regions are separated by a tunnel barrier and, to realize a σ_x rotation, a pair of sites have to be found on either sides of the tunnel barrier. With the exception of localized states at the edges of the structure or lattice defects, these pairs are mostly non-interacting owing to the large distance between the two sites and screening in the high density regions. However, we can show that single electron tunneling between the two IDQD dots is still possible in the case of weakly localized and screened states. In this case, the situation is indeed very similar to how microwaves can select a resonant pair of Rydberg atoms within a large ensemble and how the population of excited states is limited, in practice [18]. In the present case, microwaves can be used to excite one electron from one site i in one of the dot to a higher level (excited states or virtual states) where Coulomb interaction with the other vacant site j from the other dot is unscreened. This microwave-induced interaction adds the Coulomb term e^2/r_{ij} that was initially screened. The other possible pairs (k, k') now becomes non-resonant allowing many oscillations to occur between the two sites before loss of coherence. This phenomenon is well known as $dipole\ blockade$. At sufficiently short distances, dipole interaction may be non-negligible and additional terms in $1/r_{ij}^3$ may need to be taken into account. Considering that the limiting factor will either be the electron-phonon coupling strength or the timescale for interaction with the surrounding electrons, both being respectively weak and long in this type of device, the coherence time could potentially be as large as $2T_1$. Yet, the experimental value of T_1 is surprisingly large and can exceed several 100 μs [19] due to the glassy behavior of doped materials. Consequently, it is possible to envision that the coherence time T_2 could be as large as 200 μs despite the very high doping concentration. Current experiments are actually ongoing to investigate this possibility.

IV. Conclusion

We have shown that doped double quantum dots can offer interesting opportunities for optical manipulation in the microwave range by allowing to address a single pair of donors individually located in each dot of a double quantum dot structure. Surprisingly, the process is coherent with a large coherence time due to microwave-induced blockade for the other possible states. Such a selection process for a pair of sites can thus potentially be extended to the realization of spin qubits in a dopant-rich environment.

Acknowledgment

This work was partly supported by JSPS KAKENHI Grants-in-Aid (Nos. 26709023, 26630151, and 26249048), and the Project for Developing Innovation Systems of the Ministry of Education, Culture, Sports, Science and Technology (MEXT) of Japan.

References

[1] T. Ferrus, A. Rossi, A. Andreev, P. Chapman, and D. A. Williams, *Quantum Computing with Charge States in Silicon : Towards a Leadless Approach*, 5th International Conference on Quantum, Nano and Micro Technologies, 2011, pp 41.

[2] T. Ferrus, A. Rossi, M. Tanner, G. Podd, P. Chapman, and D. A. Williams, *Detection of charge motion in a non-metallic silicon isolated double quantum dot*, New J. Phys., vol 13, no 10, 2011, pp 103012.

[3] A. Rossi, T. Ferrus, G. J. Podd, and D. A. Williams, *Charge Detection in Phosphorus-doped Silicon Double Quantum Dots*, Appl. Phys. Lett., vol 97, 2010, pp. 223506.

[4] A. Rossi, T. Ferrus, and D. A. Williams, *Electron temperature in electrically isolated Si double quantum dots*, Appl. Phys. Lett., vol 100, no 13, 2012, pp. 133503.

[5] G. Moore, *Cramming More Components onto Integrated Circuits*, Electronics Magazine, vol. 38, No. 8, 1965

[6] H. Sellier, et al, *Transport Spectroscopy of a Single Dopant in a Gated Silicon Nanowire*, Phys. Rev. Lett., vol 97, 2006, pp 206805; V. Deshpande, et al, *Scaling of Trigate Nanowire (NW) MOSFETs Down to 5 nm Width: 300 K Transition to Single Electron Transistor, Challenges and Opportunities*, Solid-State Device Research Conference (ESSDERC), 2012 Proceedings of the European , 2012, pp.121,124.

[7] S. Bogdanovich and D. Popovic, *Onset of Glassy Dynamics in a Two-Dimensional Electron System in Silicon*, Phys. Rev. Lett., vol 88, 2002, pp 236401.

[8] J. J. Pla, et al, *High-fidelity readout and control of a nuclear spin qubit in silicon*, Nature, vol 496, no 7445, 2013, pp 334-338.

[9] M. F. Gonzalez-Zalba, S. Barraud, A. J. Ferguson and A. C. Betz, *Probing the limits of gate-based charge sensing*, Nature Communications, vol 6, 2015, pp 6084.

[10] A. Lewalle, et al, *Relative importance of the electron interaction strength and disorder in the two-dimensional metallic state*, Phys. Rev. B, vol 66, 2002, pp 075324.

[11] G. J. Gerardi, E. H. Poindexter, P. J. Caplan and N. M. Johnson, *Interface traps and Pb centers in oxidized (100) silicon wafers*, Appl Phys Lett, vol 49, 1986, pp 348-350.

[12] T. Ferrus, R. George, C. H. W. Barnes and M. Pepper, *Disorder and electron interaction control in low-doped silicon metal-oxide-semiconductor field effect transistors*, Appl. Phys. Lett., vol 97, no 14, 2010, pp 142108.

[13] C. W. J. Beenaker, Theory of Coulomb-blockade oscillations in the conductance of a quantum dot, Phys. Rev. B 44, vol 4, 1991, pp 1646-1656. D. V. Averin, A. N. Korotkov, and K. K. Likharev, Theory of single-electron charging of quantum wells and dots, Phys. Rev. B 44, vol 12, 1991, pp 6199-6211. S. Oda and D. Ferry, *Silicon nanoelectronics*, Taylor and Francis, CRC Press, 2006, pp 159-160.

[14] L. P. Kouwenhoven, et al, *Photon-assisted tunneling through a quantum dot*, Phys. Rev. B, vol 50, 1994, pp 2019-2022; G. Platero and R. Aguado, *Photon-assisted transport in semiconductor nanostructures*, Physics Reports, vol 395, 2004, pp 1-157.

[15] J. Seyler and M. N. Wybourne, *Acoustic waveguide modes observed in electrically heated metal wires*, Phys. Rev. Lett., vol 69, 1992, pp 1427-1430.

[16] T. Ferrus, et al, *GHz photon-activated hopping between localized states in a silicon quantum dot*, New J. Phys., vol 16, no 1, 2014, pp 013016.

[17] J. Gorman, D. G. Hasko, and D. A. Williams, *Charge qubit operation of an isolated double quantum dot*, Phys. Rev. Lett., vol 95, 2005, pp 090502.

[18] M. D. Lukin, et al, *Dipole Blockade and Quantum Information Processing in Mesoscopic Atomic Ensembles*, Phys. Rev. Lett., vol 87, 2001, pp 037901; E. Urban, et al, *Observation of Rydberg blockade between two atoms*, Nat. Phys., vol 5, 2009, pp 110; M. Weidemuller, *Rydberg atoms: There can be only one*, Nature Physics 5, 2009, pp 91-92.

[19] T. Ferrus, et al *Cryogenic instrumentation for fast current measurement in a silicon single electron transistor*, J. Appl. Phys., vol 106, 2009, pp. 033705.

A Parallel Approach to Convert Quantum Circuits to an LNN Architecture

Edgar Meza, Joni Fernández, Bengamín Barán, Joaquín Lima

Universidad Nacional de Asunción
San Lorenzo, Paraguay
Email: edgar.meza.franco@gmail.com, jonifernandezc@gmail.com
Email: bbaran@pol.una.py, joaquin.lima@pol.una.py

***Abstract*—This paper describes four algorithms implemented to solve the problem of converting general quantum circuits to a *Linear Nearest Neighbor* (LNN) architecture. All the implemented algorithms are based on the HIRATA II algorithm and consider two improvements: (i) the use of parallel computing, and (ii) branch & bound technique. The proposed parallel algorithms are tested with the largest test circuit presented in the work of Hirata et al., this circuit correspond to Shor's factorization algorithm (named as Shor10 circuit). Experimental results show a speedup of an order of magnitude from hours to seconds, improving slightly the quality of the converted circuit, measured as the number of inserted swap gates.**

Keywords*— *LNN architecture; Quantum Circuits; Parallel Computing.

I. Introduction

The design of a general quantum circuit allows the interaction of non-adjacent qubits; however, the current technology may not allow the interaction between non-adjacent qubits [13]. Therefore, quantum circuits require an architecture that facilitates implementation. *Linear Nearest Neighbor* (LNN) architecture [16] facilitates the implementation of quantum circuits. The conversion of a general quantum circuit to an LNN architecture is a hard task for conventional heuristics.

Among several alternatives, HIRATA II algorithm [11] provides a general conversion scheme applicable to non-trivial quantum circuits. Therefore, this paper proposes several parallel versions of HIRATA II algorithm. Proposed parallel versions implement a branch and bound scheme [19] to improve algorithm performance.

The algorithms implemented in this work are tested with the largest circuit considered in the work of Hirata et al. [11] to prove the advantage of the proposed alternatives.

Parallel computing seems an interesting alternative for this work given the size of computation, and availability of multi-core processors in today servers. This way, the studied problem may be solved considerable faster with a cluster of computers or even a multi-core Central Processor Unit (CPU) as far as the problem can be efficiently partitioned in smaller subproblems.

This work is organized as follows: Section II presents the general conversion problem while Section III describes HIRATA II algorithm. Then, the implemented algorithms are presented in Section IV while experimental results are presented in Section V. Finally, conclusions are left for Section VI, where future works are also presented.

II. Conversion of general quantum circuits to an LNN architecture

There are already known methods to convert a quantum circuit to an LNN architecture. Fowler et al. [5] describe a construction scheme of quantum circuits in LNN architecture. On the other hand, Hirata et al. [11] present a general conversion scheme applicable to any quantum circuit that is considered in this work for parallelization.

The process of converting a quantum circuit to LNN architecture involves the insertion of SWAP gates to the original circuit to change the order of the qubits in such a way that needed gates only opperate on neighboring qubits.

The conversion of a general quantum circuit to the LNN architecture is defined by Hirata et al. [11] as:

- **Input:** a general quantum circuit, composed of N qubits and K gates.
- **Output:** an equivalent LNN quantum circuit.
- **Objective:** to minimize the total SWAP gates added.
- **Restriction:** the equivalent circuit output should have all qubits in the same original order.

Quantum gates are in LNN architecture if the qubits necessary to operate a gate are adjacent. A quantum circuit is in a LNN architecture when all its gates are LNN.

Circuit in Figure 1 represents a quantum circuit that is not in an LNN architecture. The circuits in Figures 2 and 3 are in LNN architecture and they are equivalent to the circuit in Figure 1. Moreover, the circuit in Figure 3 represents the best solution of the two alternatives because it adds fewer SWAP gates to the original circuit. In fact, the number of inserted SWAP gates needed to convert a general quantum circuit to an LNN architecture is here considered as the main quality indicator of the convertion process.

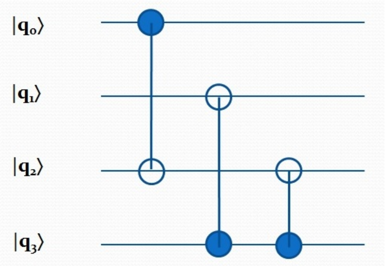


Figure 1. A quantum circuit not LNN

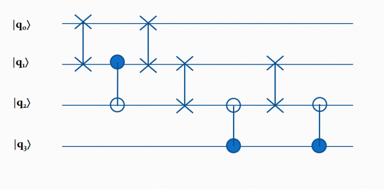


Figure 2. A quantum circuit equivalent to the one presented in 1 in an LNN architecture

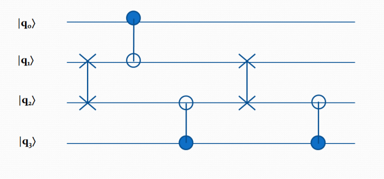


Figure 3. LNN quantum circuit equivalent to circuit of 1 with fewer SWAP gates

III. HIRATA II Algorithm

Hirata et al. [11] presented the HIRATA II algorithm to reduce the number of candidates to be evaluated in the convertion of a general quantum circuit to on LNN architecture with respect to classic heuristics as greedy algorithms.

As explained in detail in [11], HIRATA II algorithm defines an objective function f to be minimized wich can be evaluated for each candidate solution. The objective function to evaluate candidates is given by:

$$f(n_{i,j}) = f_1(n_{i,j}) + f_2(n_{i,j}) + f_3(n_{i,j}) \qquad (1)$$

with:

$$f_1 = local\_search_w(n_{i,j}, w),$$

$$f_2 = calc\_swaps(c\_order, n_{i,j}) \quad and$$

$$f_3 = \frac{c_k}{k-i+1} calc\_swaps(n_{i,j}\; l\_order)$$

where $n_{i,j}$ represents the current candidate j for the current gate i; c_order is a list that represents the current order of the qubits while l_order is a list representing the initial order of the qubits. $local\_search_w(n_{ij})$ represents the lowest cost of converting the following w gates if j is chosen. $calc\_swaps(c\_order, n_{ij})$ is the number of SWAP gates necessary to obtain n_{ij} from c_order. $\frac{c_k}{k-i+1} calc\_swaps(n_{ij}\; l\_order)$ represents an estimation of the cost necessary to re-order the final order to the original order (see restriction in Section I). This term receives more preponderance in the conversion when the process progresses and it gets closer tho the end. The ck constant is chosen a priori.

IV. Implemented Algorithms

In this paper, three algorithms are proposed based on the original HIRATA II algorithm. The H2-S version is a sequential algorithm that implements a branch and bound technique in the evaluation of f_1 needed for the calculation of objective function (2).

```
Algorithm 1 HIRATA II
Require: $N$,$K$,$w$
$l\_order \leftarrow \{0, 1, 2, ..., N-1\}$
$c\_order \leftarrow l\_order$
$swaps \leftarrow 0$
$i \leftarrow 0$
$MIN \leftarrow \infty$
while $i < K$ do
    $n \leftarrow makeCandidates(c\_order, i)$
    $j \leftarrow 0$
    while $j \leq |n| - 1$ do
        $AUX \leftarrow evaluate\_objective\_function()$
        if $AUX < MIN$ then
            $S \leftarrow \emptyset$ // List of candidates $n_{ij}$, the amount of
            swap gates was minimal.
            $S \leftarrow S \cup n\_ij$
            $MIN \leftarrow AUX$
        else if $AUX = MIN$ then
            $S \leftarrow S \cup n_{ij}$
        end if
        $j + 1$
    end while
    $S_a \leftarrow random(S)$ // Candidate $n_{ij}$ the set S chosen
    randomly.
    $swaps \leftarrow swaps + calc\_swap(c\_order, S_a)$
    $n \leftarrow \emptyset$
    $S \leftarrow \emptyset$
    $c\_order \leftarrow S_a$
    $i \leftarrow i + 1$
end while
$swaps \leftarrow swaps + calc\_swap(c\_order, l\_order)$
return $swaps$
```

Figure 4. Hirata II sequential algorithm

All parallel algorithms implement a branch and bound technique. Figure 7 shows an example of the number of nodes evaluated by the original algorithm. Figure 8 shows a decrease in the number of evaluated nodes when implementing a branch and bound technique.

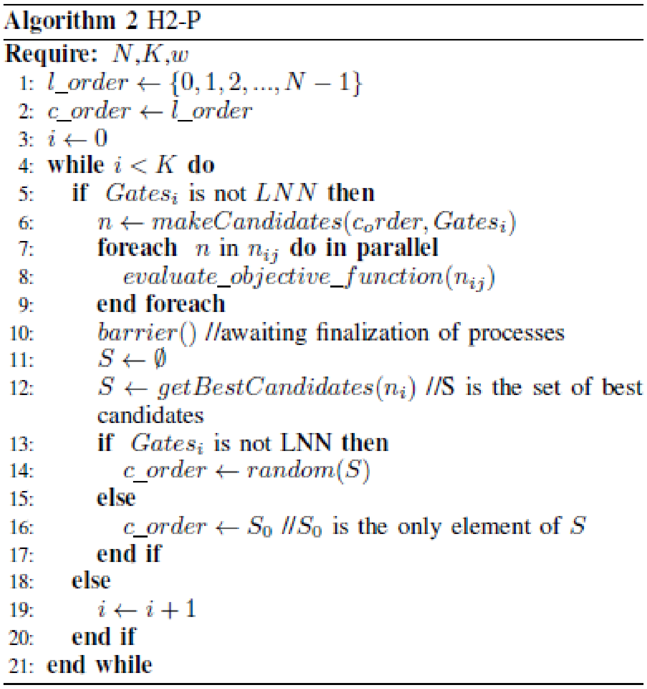
Algorithm H2-P is a parallel version based on the scheme of task division. A task is an evaluation of a candidate, i.e. the calculation of the term f_1 of the objective function given by (2).

H2-X algorithm is a hybrid parallel version scheme based on task division and problem partitioning. This algorithm first divides the problem into X parts. Then, the algorithm is applied in parallel to each part of the problem. In what follows, two values of X are used: $X = 2$ and $X = 5$.

Following Frutos suggestion [6], objective function is modified to use different weights as follows:

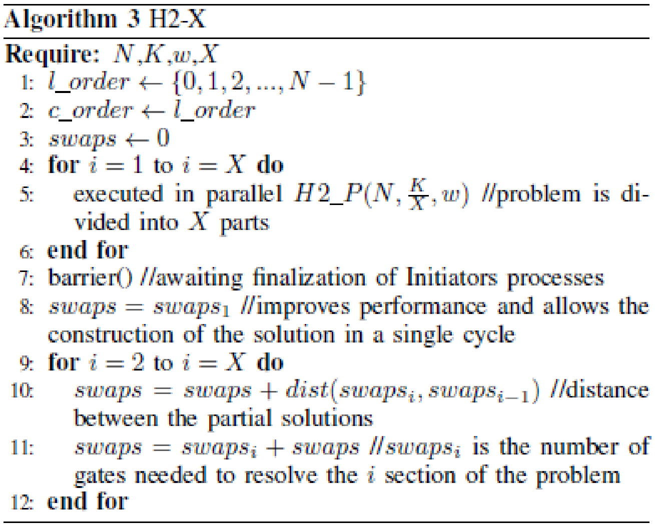
$$f(n_{i,j}) = P_1 * f_1(n_{i,j}) + P_2 * f_2(n_{i,j}) + P_3 * f_3(n_{i,j}) \qquad (2)$$

where P_1, P_2 and P_3 are weights satisfying the relation $P_1 + P_2 + P_3 = 1$. This paper only considers the special cases presented in Table I.



```
Algorithm 2 H2-P
Require: $N$,$K$,$w$
$l\_order \leftarrow \{0, 1, 2, ..., N-1\}$
$c\_order \leftarrow l\_order$
$i \leftarrow 0$
while $i < K$ do
    if $Gates_i$ is not $LNN$ then
        $n \leftarrow makeCandidates(c_order, Gates_i)$
        foreach  $n$ in $n_{ij}$ do in parallel
            $evaluate\_objective\_function(n_{ij})$
        end foreach
        $barrier()$ //awaiting finalization of processes
        $S \leftarrow \emptyset$
        $S \leftarrow getBestCandidates(n_i)$ //S is the set of best candidates
        if  $Gates_i$ is not LNN then
            $c\_order \leftarrow random(S)$
        else
            $c\_order \leftarrow S_0$ //$S_0$ is the only element of $S$
        end if
    else
        $i \leftarrow i + 1$
    end if
end while
```

Figure 5. First proposed parallel algorithm H2-P



```
Algorithm 3 H2-X
Require: $N$,$K$,$w$,$X$
$l\_order \leftarrow \{0, 1, 2, ..., N-1\}$
$c\_order \leftarrow l\_order$
$swaps \leftarrow 0$
for $i = 1$ to $i = X$ do
    executed in parallel $H2\_P(N, \frac{K}{X}, w)$ //problem is divided into $X$ parts
end for
barrier() //awaiting finalization of Initiators processes
$swaps = swaps_1$ //improves performance and allows the construction of the solution in a single cycle
for $i = 2$ to $i = X$ do
    $swaps = swaps + dist(swaps_i, swaps_{i-1})$ //distance between the partial solutions
    $swaps = swaps_i + swaps$ //$swaps_i$ is the number of gates needed to resolve the $i$ section of the problem
end for
```

Figure 6. Second parallel algorithm proposed H2-X

TABLE I. COMBINATION OF P_1, P_2 Y P_3 USED IN THE REPORTED TESTS

P_1	P_2	P_3	Comments
1	0	0	only $f_1()$ is use
0	1	0	only $f_2()$ is use
0	0	1	only $f_3()$ is use
1/3	1/3	1/3	HIRATA II
0.5	0.25	0.25	$f_1()$ is more important
0.25	0.5	0.25	$f_2()$ is more important
0.25	0.25	0.5	$f_3()$ is more important

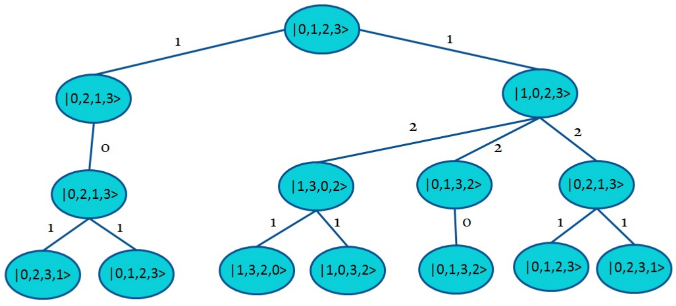


Figure 7. Nodes evaluated with the original $local\_search$ function

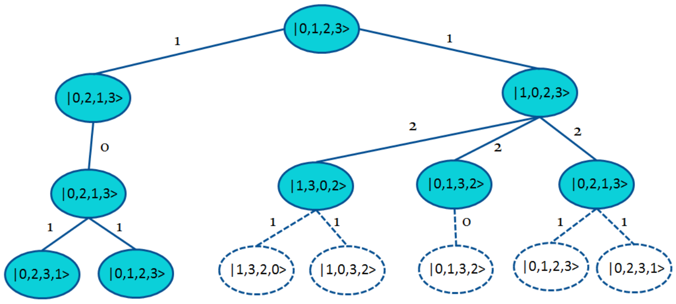


Figure 8. Nodes evaluated with the $local\_search$ function and branch and bound technique

Clearly, combination of weights where $P_1 = P_2 = P_3 = \frac{1}{3}$ corresponds to the original optimization function proposed in [11].

The nature of the parallel algorithm conversion requires the development of a communication protocol as OpenMPI [7] used in this work. The communication protocol has an Initiator, a Router and Worker processes:

- **Initiator:** the Initiator process determines the extent of the problem;
- **Router:** the Router process manages the pool of worker processes;
- **Worker:** worker processes run tasks, and even the Initiator process is also a worker process.

This protocol avoids the scheme "master / slave" implementing a scheme of *pool of workers*.

Table II summarizes the implemented algorithms and techniques used for each algorithm.

TABLE II. COMPARISON ALL IMPLEMENT ALGORITHM

Algorithm	Branch and Bound	Parallelization of calculation of objective function	Circuit partitioning	Comments
Hirata II	No	No	No	Hirata et al. [11]
H2-S	Yes	No	No	[11] with branch and bound: Algorithm (1)
H2-P	Yes	Yes	No	Objective function calculated in parallel
H2-X	Yes	Yes	Yes	Algorithm (2)

TABLE III. RESULTS OBSERVED FOR SPECIAL WEIGHTS IN THE SHOR10 CIRCUIT

Weight	Alg.	w=10		w=12		w=14	
		$\overline{swaps}$ / $\sigma\left(swaps\right)$	$\overline{t_{seg}}$ / $\sigma\left(t_{seg}\right)$	$\overline{swaps}$ / $\sigma\left(swaps\right)$	$\overline{t_{seg}}$ / $\sigma\left(t_{seg}\right)$	$\overline{swaps}$ / $\sigma\left(swaps\right)$	$\overline{t_{seg}}$ / $\sigma\left(t_{seg}\right)$
1/3; 1/3 ;1/3	H2-S	163615,4 / 2552,19	136,851 / 2,80	150355,8 / 1050,52	558,765 / 2,28	148364,1 / 982,27	2656,19 / 2,32
	H2-P	153674,2 / 1954,26	43,694 / 1,93	140800,8 / 779,10	177,602 / 3,91	141167,4 / 761,14	745,632 / 40,03
	H2-X2	154718,8 / 2134,27	22,073 / 0,87	142723,4 / 902,84	88,617 / 4,25	141018,2 / 823,06	365,461 / 28,01
	H2-X5	152410,6 / 1709,77	10,448 / 1,21	145624,6 / 1583,82	41,144 / 2,35	145109,2 / 1498,90	147,06 / 8,44
0.5 ;0.25; 0.25	H2-S	145609 / 749,12	98,886 / 1,21	139809,2 / 633,56	388,584 / 1,63	140433,6 / 513,01	3186,251 / 4,08
	H2-P	154321,8 / 1612,67	44,781 / 1,28	140918,8 / 799,92	175,399 / 5,54	140989,6 / 580,96	744,693 / 56,84
	H2-X2	154944,6 / 1650,27	22,191 / 0,65	142394,6 / 1284,04	88,942 / 3,99	141539,6 / 466,87	359,981 / 21,28
	H2-X5	152727,8 / 937,09	9,453 / 0,43	144842,4 / 1704,23	38,127 / 1,46	144829 / 1072,02	140,265 / 14,25
0.25 ;0.5 ;0.25	H2-S	171704,6 / 625,71	146,504 / 2,25	171464,4 / 617,81	570,564 / 1,96	165006,2 / 1344,39	2039,95 / 3,02
	H2-P	154272,6 / 2316,71	43,648 / 2,07	140634,4 / 458,33	178,132 / 4,24	141376,8 / 416,88	768,926 / 78,55
	H2-X2	154401,6 / 1717,95	22,29 / 0,95	142341,4 / 639,11	92,126 / 3,39	141263,6 / 389,22	366,553 / 19,26
	H2-X5	153910,7 / 1978,02	9,881 / 0,74	143864,7 / 931,65	39,732 / 2,36	145080,2 / 1535,01	137,357 / 15,53
0.25 ;0.25 ;0.5	H2-S	162204,9 / 1729,59	118,83 / 1,36	153733,5 / 1042,29	685,323 / 1,79	157097,5 / 464,64	3097,569 / 2,31
	H2-P	153339,6 / 1755,63	43,842 / 2,36	140970 / 810,46	178,819 / 3,85	141207,2 / 757,31	718,053 / 56,23
	H2-X2	155090,2 / 1216,85	21,899 / 0,83	142839,2 / 652,68	90,129 / 3,60	141261 / 715,41	357,025 / 29,05
	H2-X5	153845,9 / 1507,76	9,774 / 0,58	145869,5 / 1366,39	45,245 / 1,98	144738,2 / 1066,47	148,415 / 15,39
1 ;0 ;0	H2-S	141604,1 / 327,71	101,027 / 0,98	136209,1 / 699,13	379,494 / 2,56	136456,7 / 396,55	3515,705 / 3,51
	H2-P	153671,4 / 1582,54	44,062 / 2,50	140613,4 / 567,89	176,709 / 4,68	141174 / 688,45	730,881 / 30,94
	H2-X2	154514,4 / 1738,62	21,916 / 0,94	142319,8 / 706,41	91,789 / 3,41	140975,6 / 548,79	361,095 / 29,26
	H2-X5	153853,9 / 1720,86	9,758 / 0,59	144962,9 / 1528,26	42,392 / 2,88	144832,2 / 1110,51	146,448 / 15,09
0 ;1 ;0	H2-S	256266,7 / 1113,46	454,973 / 1,94	258657,4 / 661,86	1851,703 / 2,12	259102,6 / 539,50	5851,167 / 5,14
	H2-P	272585,9 / 2034,55	189,944 / 13,12	274059,9 / 1370,88	700,717 / 19,88	282117,3 / 918,42	2442,572 / 72,16
	H2-X2	270386,3 / 3346,48	93,088 / 5,22	278743,7 / 1706,26	351,268 / 18,94	281370,6 / 2738,40	1454,563 / 132,50
	H2-X5	270821,2 / 1947,30	43,842 / 3,70	276104,8 / 1870,61	183,48 / 8,10	282341,2 / 3062,77	632,463 / 10,58
0 ;0 ;1	H2-S	252368,7 / 930,42	408,199 / 1,38	257750 / 625,07	2216,577 / 3,39	257017,4 / 289,33	5996,26 / 2,80
	H2-P	274425 / 1725,57	195,747 / 7,80	277003,2 / 3693,99	721,374 / 9,74	281502,6 / 2514,29	2518,742 / 38,91
	H2-X2	271276,2 / 1927,95	91,79 / 2,09	277446,3 / 1016,30	355 / 15,92	283420,9 / 2629,49	1.457 / 120,07
	H2-X5	271754,1 / 3659,20	41,78 / 3,27	276948,9 / 1597,44	186,845 / 1,44	283434,3 / 1688,50	625 / 12,03

V. EXPERIMENTAL RESULTS

The largest circuit presented in [11] is used in this work, this circuit correspond to Shor's factorization algorithm [14]. It is composed of 24 qubits and 132,204 quantum gates (named as Shor10 circuit). Given that the algorithms are probabilistic when there is a tie, experiments are run 10 times for each algorithm implemented considering three values of w, giving a total of 10x7x4x3 = 840 experimental runs.

Algorithm H2-S was run in a computer with Intel processor I7 Quadcore 2.3 GHz and 16 GB of Random Access Memory (RAM). On the other hand, parallel algorithms were executed on a cluster of computers with Intel processor I5 Quadcore 2.3 GHz and 4 GB of RAM. It is important to note that the equipment used for the execution of H2-S is clearly better than the one used for the parallel implementations.

TABLE IV. SPEEDUP AND QUALITY MEASURE COMPARISON

Weight	Algorithm	w=10		w=12		w=14	
		Q	S_p	Q	S_p	Q	S_p
1/3;1/3;1/3	H2-S	1,00	1,00	1,00	1,00	1,00	1,00
	H2-P	0,94	3,13	0,94	3,15	0,95	3,56
	H2-X2	0,95	6,20	0,95	6,31	0,95	7,27
	H2-X5	0,93	13,10	0,97	13,58	0,98	18,06
0.5;0.25;0.25	H2-S	1,00	1,00	1,00	1,00	1,00	1,00
	H2-P	1,06	2,21	1,01	2,22	1,00	4,28
	H2-X2	1,06	4,46	1,02	4,37	1,01	8,85
	H2-X5	1,05	10,46	1,04	10,19	1,03	22,72
0.25;0.5;0.25	H2-S	1,00	1,00	1,00	1,00	1,00	1,00
	H2-P	0,90	3,36	0,82	3,20	0,86	2,65
	H2-X2	0,90	6,57	0,83	6,19	0,86	5,57
	H2-X5	0,90	14,83	0,84	14,36	0,88	14,85
0.25;0.25;0.5	H2-S	1,00	1,00	1,00	1,00	1,00	1,00
	H2-P	0,95	2,71	0,92	3,83	0,90	4,31
	H2-X2	0,96	5,43	0,93	7,60	0,90	8,68
	H2-X5	0,95	6,60	0,95	15,15	0,92	20,87
1;0;0	H2-S	1,00	1,00	1,00	1,00	1,00	1,00
	H2-P	1,09	2,29	1,03	2,15	1,03	4,81
	H2-X2	1,09	4,61	1,04	4,13	1,03	9,74
	H2-X5	1,09	10,35	1,06	8,95	1,06	24,01
0;1;0	H2-S	1,00	1,00	1,00	1,00	1,00	1,00
	H2-P	1,06	2,40	1,06	2,64	1,09	2,40
	H2-X2	1,06	4,89	1,08	5,27	1,09	4,02
	H2-X5	1,06	10,38	1,07	10,09	1,09	9,25
0;0;1	H2-S	1,00	1,00	1,00	1,00	1,00	1,00
	H2-P	1,09	2,09	1,07	3,07	1,10	2,38
	H2-X2	1,07	4,45	1,08	6,24	1,10	4,12
	H2-X5	1,08	9,77	1,07	11,86	1,10	9,59

Table III describes the mean values and standard deviations observed for different weight combinations. In particular, Table III shows that H2-X5 is strictly better in running time and number of SWAP gates than the sequential version of HIRATA II algorithm (found when all weights are equal $\frac{1}{3}$). Even more, H2-X5 has a shorter running time than any other implemented algorithm, proving its effectiveness. At the same time, all implemented parallel algorithms proved to be quite competitive with respect to the sequential algorithm H2-S.

The standard deviation observed in Table III is small in general, showing that no peaks are seen neither in the running time nor in the necessary amount of quantum gates.

Table IV shows the SpeedUp (S_p) [12] and quality (Q) calculated for each parallel algorithm with respect to algorithm H2-S. Clearly, parallel algorithms have the highest acceleration and H2-X5 is the best parallel alternative.

Finally, it can be noticed in Table III that the worst results were obtained when $P_1 = 0$, confirming same conclusion reported in [6].

VI. Conclusion and Future Works

Experimental results show a speedup of an order of magnitude with respect to the resolution times (from hours to seconds), even improving slightly the quality of the converted circuit, measured by the number of inserted swap gates.

Experiments corroborate the importance of the term f_1 in (1) and (2) for the quality of results. Combinations where $P_1 = 0$ obtained the worst experimental results, confirming Frutos conclusion [6].

The communication protocol designed and implemented in this work using Open MPI seems very efficient and it can be used in building other non-trivial parallel algorithms.

For future work, the authors are working on the following improvements: (i) apply meta-heuristic techniques, possibly based on a strategy of task division, using the communication protocol designed and implemented in this work; (ii) modify HIRATA II algorithm to avoid the randomness in the selection of candidates in tie situations and (iii) apply further partitioning to solve a give partition to increase the potential of using parallelism in new multi-core cluster of cumputers.

Acknowledgment

Special thanks to the Nara Institute of Science and Technology (NAIST) in the person of Yasuhiko Nakashima, Hiroyuki Seki, Masaki Nakanishi, Shigeru Yamashita and Yuichi Hirata, for all help and for providing the test circuits.

References

[1] P. Benioff, ”The computer as a physical system: A microscopic quantum mechanical hamiltonian model of computers as represented by turing machines,” Journal of Statistical Physics, vol. 22, no. 5, pp. 563–591, 1980.

[2] H. Charles Bennett, et al. ”Teleporting an unknown quantum state via dual classical and Einstein-Podolsky-Rosen channels,” Physical Review Letters, vol. 70, no. 13, p. 1895, 1993.

[3] P. A. Maurice Dirac, ”A new notation for quantum mechanics,” In Mathematical Proceedings of the Cambridge Philosophical Society, vol. 35, Cambridge Univ Press, pp. 416-418, July 1939.

[4] R. P. Feynman, (1982). ”Simulating physics with computers,” International journal of theoretical physics, vol. 21, no. 6, pp.467-488, 1982.

[5] A. G. Fowler, S. J. Devitt and L. C. Hollenberg, ”Implementation of Shor’s algorithm on a linear nearest neighbour qubit array,” arXiv preprint quant-ph/0402196, 2004.

[6] L. Frutos, ”Conversion de circuitos cunticos a la arquitectura LNN,” Enginering Thesis at Universidad Nacional de Asuncion - Facultad Politecnica, 2012.

[7] E. Gabriel, et al. ”Open MPI: Goals, concept, and design of a next generation MPI implementation,” In Recent Advances in Parallel Virtual Machine and Message Passing Interface, pp. 97-104, Springer, 2004.

[8] R. L. Graham, T. S. Woodall and J. M. Squyres, ”Open MPI: A flexible high performance MPI,” In Parallel Processing and Applied Mathematics, pp. 228-239, Springer, 2006.

[9] W. Gropp, E. Lusk and R. Thakur, ”Using MPI-2: Advanced features of the message-passing interface,” MIT press, 1999.

[10] L. K. Grover, ”A fast quantum mechanical algorithm for database search,” In Proceedings of the twenty-eighth annual ACM symposium on Theory of computing, pp. 212-219, ACM, 1996.

[11] Y. Hirata, M. Nakanishi, S. Yamashita and Y. Nakashima, ”An efficient method to convert arbitrary quantum circuits to ones on a linear nearest neighbor architecture,” In Quantum, Nano, and Micro Technologies, First International Conference, pp. 26-33, IEEE, 2009.

[12] V. Kumar, A. Grama, A. Gupta and G. Karypis, ”Introduction to parallel computing: design and analysis of algorithms,” Benjamin/Cummings Publishing Company Redwood City, CA, 1994.

[13] M. A. Nielsen and I. L. Chuang, ”Quantum computation and quantum information,” Cambridge university press, 2010.

[14] P. W. Shor, ”Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer,” SIAM journal on computing, vol. 26, no. 5, pp. 1484-1509, 1997.

[15] M. Snir, ”MPI the Complete Reference: The MPI core,” vol. 1, MIT press, 1998.

[16] S. Yamashita and I. L. Markov, ”Fast equivalence-checking for quantum circuits,” In Proceedings of the 2010 IEEE/ACM International Symposium on Nanoscale Architectures, pp. 23-28, IEEE Press, 2010.

[17] R. L. Rivest, A. Shamir and L. Adleman, ”A method for obtaining digital signatures and public-key cryptosystems,” Communications of the ACM, vol. 21, no. 2, pp. 120-126, 1978.

[18] D. Deutsh, ”Quantum theory, the Church-Turing principle and the universal quantum computer,” The Royal Society, 1985.

[19] J. Clausen, ”Branch and bound algorithms-principles and examples,” Department of Computer Science, University of Copenhagen, 1999.