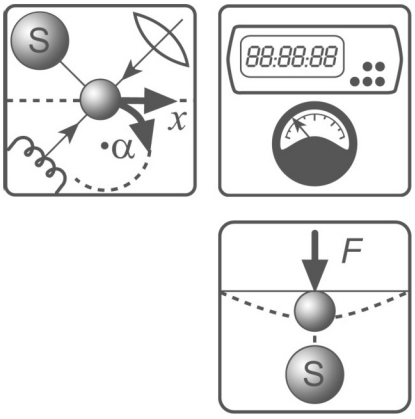
ADVANCES IN SENSORS: REVIEWS 3

Sergey Y. Yurish
Editor

Sensors, Transducers, Signal Conditioning and Wireless Sensors Networks



IFSA

Sensors, Transducers, Signal Conditioning and Wireless Sensors Networks

Sergey Y. Yurish
Editor

Sensors, Transducers, Signal Conditioning and Wireless Sensors Networks

Book Series: Advances in Sensors: Reviews, Vol. 3

International Frequency Sensor Association Publishing

Sergey Y. Yurish
Editor

Sensors, Transducers, Signal Conditioning and Wireless Sensors Networks
Advances in Sensors: Reviews, Vol. 3

E-mail (for orders and customer service enquires): ifsa.books@sensorsportal.com

Visit our Home Page on http://www.sensorsportal.com

ISBN: 978-84-608-7705-9
BN-20160415-01
BIC: TJFC

Acknowledgments

As Editor I would like to express my undying gratitude to all authors, editorial staff, reviewers and others who actively participated in this book. We want also to express our gratitude to all their families, friends and colleagues for their help and understanding.

Contents

Contributors

Abdelrahman Abuarqoub Faculty of Information Technology, Middle East University, Amman, Jordan

Boris G. Ageev V. E. Zuev Institute of Atmospheric Optics SB RAS, Tomsk 634021, Russia Omar Aldabbas Faculty of Engineering Al-Balqa Applied University, Amman, Jordan

Fayez Al-Fayez School of Computing, Mathematics and Digital Technology, Manchester Metropolitan University, Manchester, UK

Almuatasim Alomari Department of Physics, Chemistry and Mathematics, Alabama A&M University, Normal, AL 35762 USA

Laurent Amanton University of Le Havre, BP 1123, 76063 Le Havre Cedex, France

P. Arulmozhivarman School of Electrical Engineering, VIT University, Vellore, 632014, India

Ashok Batra Department of Physics, Chemistry and Mathematics, Alabama A&M University, Normal, AL 35762 USA

Marko Beko Universidade Lusófona de Humanidades e Tecnologias, Lisbon, Portugal CTS, UNINOVA – Campus FCT/UNL, Caparica, Portugal

Abderrahmen Belfkih University of Le Havre, BP 1123, 76063 Le Havre Cedex, France

Manabendra Bhuyan
Department of Electronics and Communication Engineering, Tezpur University, India

Zhuangqi Cao Department of Physics and Astronomy, Jiaotong University, Shanghai 200240, China

Rui Dinis Instituto de Telecomunicações, Lisbon, Portugal, DEE/FCT/UNL, Caparica, Portugal

Claude Duvallet University of Le Havre, BP 1123, 76063 Le Havre Cedex, France

Aleksandr N. Gruzdev M. A.Obukhov Institute of Atmospheric Physics RAS, Moscow 119017, Russia

Xiaowei Guo University of Electronic Science and Technology of China

Mohammad Hammoudeh School of Computing, Mathematics and Digital Technology, Manchester Metropolitan University, Manchester, UK

Eduardo Ismael Hernández Engineering Department, UPAEP University, Puebla, Mexico

Frédérique Jacquet Clermont Université / LIMOS CNRS - Complexe scientifique des Cézeaux, 63172 Aubière cedex, France

Prateek Jain School of Electrical Engineering, VIT University, Vellore, 632014, India

Volodymyr Y. Kondratenko Pacific Metrics, Lakewood, CO, USA

Yuriy P. Kondratenko Petro Mohyla Black Sea State University, Ukraine; [2]Cleveland State University, Cleveland, OH, USA

Rui Krause Chemistry Department, Rhodes University, P.O. Box 94, Grahamstown, South Africa

Lina M. Pestana Leão Brito M-ITI (Madeira Interactive Technologies Institute), Exact Sciences and Engineering Center, University of Madeira (UMa), Portugal

Jong-Ha Lee Department of Biomedical Engineering, Keimyung University School of Medicine, Daegu, South Korea

Michel Misson Clermont Université / LIMOS CNRS - Complexe scientifique des Cézeaux, 63172 Aubière cedex, France

Hardik Nagrecha School of Electrical Engineering, VIT University, Vellore, 632014, India

El Hadji Malick Ndoye Clermont Université / LIMOS CNRS - Complexe scientifique des Cézeaux, 63172 Aubière cedex, France, Laboratoire d'Informatique, Université Cheikh Anta Diop de Dakar (UCAD), B. P. 5005 Dakar-Fann, Sénégal

Nokuthula Ngomane Chemistry Department, Rhodes University, P.O. Box 94, Grahamstown, South Africa, Ecoplanet Bamboo Laboratory, Kowie Bamboo Farm, Martindale Road, Bathurst, 6166, South Africa

Ibrahima Niang Laboratoire d'Informatique, Université Cheikh Anta Diop de Dakar (UCAD), B. P. 5005 Dakar-Fann, Sénégal

Saurav Paul School of Electrical Engineering, VIT University, Vellore, 632014, India

Yurii N. Ponomarev V. E. Zuev Institute of Atmospheric Optics SB RAS, Tomsk 634021, Russia

Nagabhushan Raju Konduru Sri Krishnadevaraya University, Ananthapuramu, Andhra Pradesh, India

Prabhu Ramanathan School of Electrical Engineering, VIT University, Vellore, 632014, India

Sudha Ramasamy School of Electrical Engineering, VIT University, Vellore, 632014, India

Laura Margarita Rodríguez Peralta Engineering Department, UPAEP University, Puebla, Mexico

Tapio Saarelainen Army Academy, Väinö Valveen katu 4, 53900 Lappeenranta, Finland Lappeenranta University of Technology, PL 20, 53851, Lappeenranta, Finland

Bruno Sadeg University of Le Havre, BP 1123, 76063 Le Havre Cedex, France

Valeria A. Sapozhnikova, Institute of Monitoring of Climatic and Ecological Systems, Tomsk 634055 Russia

Dmitry A. Savchuk Institute of Monitoring of Climatic and Ecological Systems, Tomsk 634055 Russia

Eduard A. Shvets National University of Shipbuilding, Mykolaiv, Ukraine

Manorama Singh Department of Chemistry, Guru Ghasidas Vishvidyalaya, Bilaspur, (CG)-495009, India

Chandra Mohan Reddy Sivappagari JNTUA College of Engineering Pulivendula, J.N.T.University Anantapur, Andhra Pradesh, India

Durlav Sonowal Department of Electronics and Communication Engineering, Tezpur University, India

Dan Mihai Ştefănescu Romanian Measurement Society, Romania

Rao Tatavarti Gayatri Vidya Parishad College of Engineering, Visakhapatnam, 530048, India

Venkataraman Thangadurai Department of Chemistry, University of Calgary, 2500 University Drive NW, Calgary, AB, T2N 1N4, Canada

Ida Tiwari Department of Chemistry, Faculty of Science, Banaras Hindu University, Varanasi (UP)-221005, India

Slavisa Tomic ISR/IST, LARSyS, Lisbon, Portugal

Nelson Torto Advanced Materials Division, Nanoscience and Nanotechnology, 200 Malibongwe Drive 2125, Randburg, South Africa, Botswana Institute for Technology Research and Innovation, Private Bag 0082, Plot 50654, Machel Drive, Gaborone, Botswana

Milan Tuba Megatrend University, Faculty of Computer Science, Belgrade, Serbia

Sibulelo Vilakazi Advanced Materials Division, Nanoscience and Nanotechnology, 200 Malibongwe Drive 2125, Randburg, South Africa

Xianping Wang College of Physics and Communication Electronics, Jiangxi Normal Univeristy, Key Laboratory of Optoelectronic and Telecommunication of Jiangxi Province, Nanchang 330022, China

Werner Weppner Institute of Inorganic Chemistry, Christian-Albrechts University, Max-Eyth-Str. 2, D-24118, Kiel, Germany

Cheng Yin College of Computer and Information Engineering, Hohai University, Changzhou 213022, China

Yongming Zhu Faculty of Applied Chemistry, Harbin Institute of Technology (Weihai), 264209 Weihai, China

Preface

I am very happy to present the third volume from the Advances in Sensors: Reviews book Series started by the IFSA Publishing in 2016. The 1st and 2nd volumes were published in 2012 and 2013 respectively. The very rapid progress in sensor research, applications and technologies are observed during the last few years.

The third volume titled 'Sensors, Transducers, Signal Conditioning and Wireless Sensors Networks' contains nineteen chapters with sensor related state-of-the-art reviews and descriptions of latest achievements written by 55 authors from academia and industry from 19 countries: Botswana, Canada, China, Finland, France, Germany, India, Jordan, Mexico, Portugal, Romania, Russia, Senegal, Serbia, South Africa, South Korea, UK, Ukraine and USA.

This book ensures that our readers will stay at the cutting edge of the field and get the right and effective start point and road map for the further researches and developments. By this way, they will be able to save more time for productive research activity and eliminate routine work.

Built upon the series Advances in Sensors: Reviews - a premier sensor review source, it presents an overview of highlights in the field and becomes. Coverage includes current developments in physical sensors and transducers, chemical sensors, biosensors, sensing materials, signal conditioning energy harvesters and wireless sensor networks. Sure, we would have liked to include even more topics, but it is difficult to cover everything due to reasonable practical restrictions. With this unique combination of information in each volume, the Advances in Sensors book Series will be of value for scientists and engineers in industry and at universities, to sensors developers, distributors, and users.

Like the first two volumes of this book Series, the third volume also has been organized by topics of high interest. In order to offer a fast and easy reading of the state of the art of each topic, every chapter in this book is independent and self-contained. All chapters have the same structure: first an introduction to specific topic under study; second particular field description including sensing applications. Each of chapter is ending by well selected list of references with books, journals, conference proceedings and web sites.

Like the first two volumes, the third volume is also published in two formats: electronic (pdf) with full-color illustrations and print (paperback). I shall gratefully receive any advices, comments, suggestions and notes from readers to make the next volumes of Advances in Sensors: Reviews book Series very interesting and useful.

Dr. Sergey Y. Yurish

Barcelona, Spain

Editor
IFSA Publishing

Chapter 1
Advances in Intelligent Force Transducers

Dr. Dan Mihai Ştefănescu

1.1. Introduction and Terminology

At present there are two opposite progress ways in force measurement [1]:

- Providing the transducer with electronic and data processing capabilities, "black box", the force transducer having three ports (input, interrogation and output), capable to perform a great part of the intelligent system functions;

- Simplifying the transducer configuration, reduced to a Wheatstone bridge, "white box", and using the interfaces (high efficiency and performance data acquisition, conversion and control boards) in order to allow the digital connection to any computer [2].

Which is better: Intelligent Force Transducers (IFT) or Intelligent Force Sensors (IFS)? Let us use the correct terms [3], in accordance with the international terminology of Metrology! *Sensor* is the primary sensitive element that enables a first conversion of the measurand applied to its input (for instance, the strain gauge – abbreviated SG – bonded on the elastic element – EE), while the *transducer* is the assembly in which the *sensor* is mechanically installed and electrically connected (Fig. 1.1,a). In general, Americans and Asians give preference to Force Sensors, while Europeans prefer Force Transducers. A unanimously accepted term is Intelligent Force Sensing, giving the specificity of this work.

Force transducers (FTs) can be viewed as energy and information converters at the same time. The Wheatstone bridge (Fig. 1.1,b) is a single device that combines the two transduction processes, symbolized by the merged symbol, *PM*:

- *Mechanical-to-electrical conversion*, i.e. relative variation of length (or specific deformation) to relative variation of resistance (its intrinsic resistivity);
- *Signal conditioning* by the Wheatstone bridge.

Dr. Dan Mihai Ştefănescu
Romanian Measurement Society, Romania

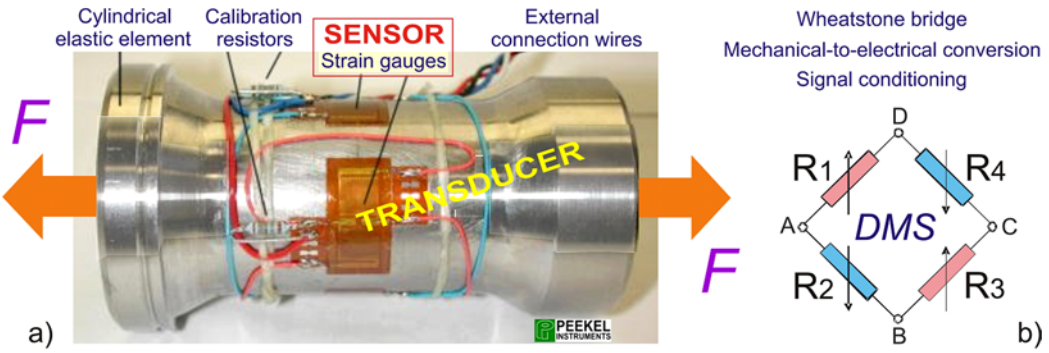


Fig. 1.1. Force transducer including strain gauge sensors.

The entire process could be synthetically defined as an information *Type CPM (conversion by property modulation)*, in this case *strain sensing*, which is elastoelectric.

As a parenthesis, specific differences between terms *intelligent* and *smart* are well presented by Yurish [4]. Returning to our subject, intelligence is required in all stages of implementing a high performance force measuring system:

- Selecting the electric measuring method / principle, for which a useful systematic presentation may be found in [5];
- Designing the transducer following an iterative process, where the physico-mathematical modeling (e.g. finite element method – FEM) is joined with experimental tests, in view of adopting those elastic structures benefiting from symmetry / antisymmetry;
- Choosing the electronic components of the measuring channel, together with some typical examples of signal conditioners and displays;
- Signal processing and information technology [6] in a single or multi-channel / multi-sensor system, with various facilities such as self-monitoring and self-calibration;
- Transition to virtual instrumentation.

1.2. Intelligent Design of Force Transducers

Brignell [7] has put forward the intelligent principles of refined structural compensation by the symmetry of design in both electrical (strain gauges in Wheatstone bridge) and mechanical (elastic element of transducer) sub-systems, both forming an indestructible monolith, a perfect example of electromechanical integration:

- An elastic structure composed of multiple and symmetrical flexible elements minimizes the effect of load eccentricity;

- Strain gauges connected in full Wheatstone bridge give the maximum differential output with linearity (plus thermal compensation), in contrast to a simple strain gauge.

An original multicomponent balance for aircraft models tested in a supersonic wind tunnel [8] is shown in Fig. 1.2. The axial force is measured by four short lateral arms close to the middle of the balance interior. The other five components are measured in two symmetrical sections, each consisting of a casing with three beams. Complex mechanical and electrical analysis has been carried out in order to find the optimum method of attaching and connecting the SGs to the wind tunnel balance aiming at increasing the sensitivity and compensating for disturbing influences.

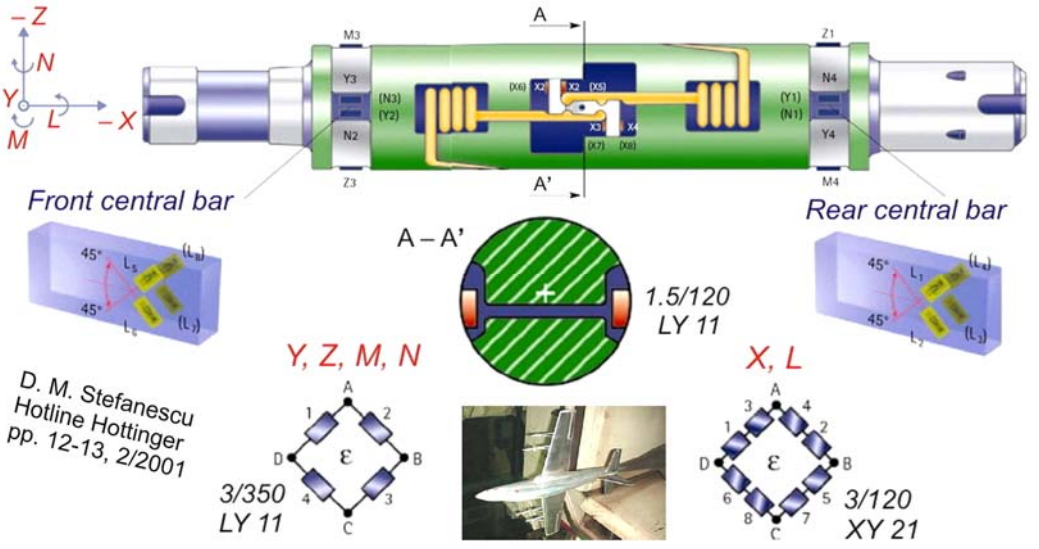


Fig. 1.2. The force (X, Y, Z) and torque (L, M, N) system is applied to a cylindrical elastic element with vertical and horizontal slots, that sets bounds to the strain gauges location near the central section A-A' and towards the two ends. The Wheatstone bridges have 8 SGs for the X and L components, and 4 SGs for the Y, Z, M and N components, respectively. Finally, the tensometric balance is inserted in the airplane model being subject to aerodynamic loads in the wind tunnel.

Force Sensing Resistors (FSRs) are easier to integrate into medical devices as compared to the strain gauges due to their thin, flexible nature [9]. They are based on varying the contact area when an elastomer (e.g. rubber) is squeezed between two conductive plates with different pattern of shunts.

The human body is a sensitive object, so the amount of force applied by a doctor or medical instrument is a key factor for consideration. Force feedback is an increasingly valuable feature and these intelligent tools (e.g. Tekscan surgical gripper, Fig. 1.3), help to eliminate guesswork, create more objective feedback, and improve patient outcomes.

A parametric measurement system competing the resistive one is the capacitive one [10]. iLoad Flex from Loadstar Sensors [11] are designed for applications where size is of

critical importance. They are based on patented capacitive force sensing technology which offers unprecedented sensitivity in a rugged and tiny package. These sensitive devices can be built in a single sensor format or as an array of sensors that can be used to measure pressure (as distributed force) on a flexible surface. The high level analog (0…5 V DC) or digital Serial / USB outputs, makes it easy to incorporate into OEM products or into test & measurement and process control applications. The load cell appears on the PC as a virtual COM port.

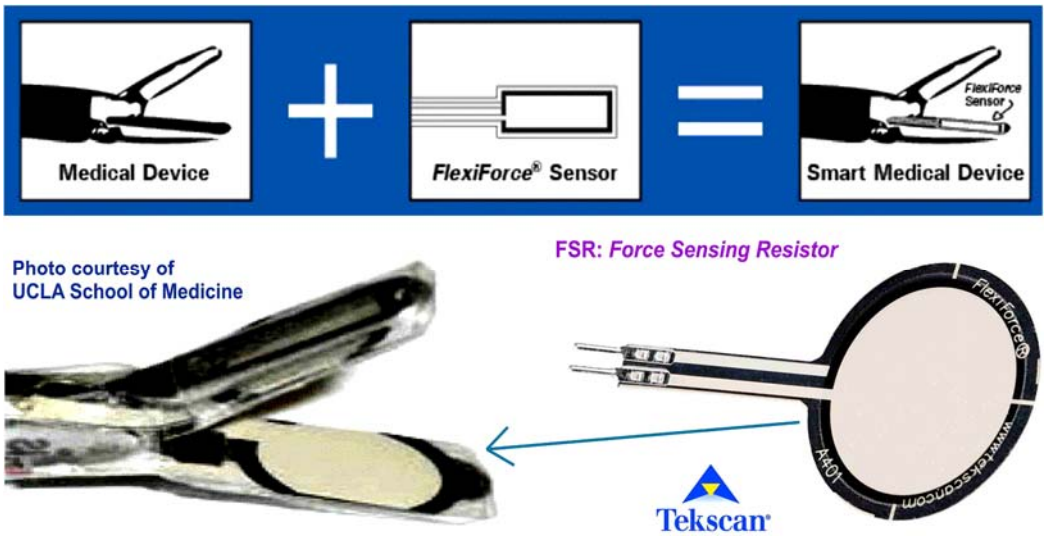


Fig. 1.3. Intelligent surgical gripper equipped with force sensing resistors.

1.3. Electrical Methods for Intelligent Force Measurement

Other electrical methods / principles for measuring of force and related quantities, besides the parametrical ones (resistive and capacitive) are: electromagnetic force compensation (EMFC), vibrating wire, piezoelectric and a combination including optical at a complex catheter.

1.3.1. Electromagnetic Force Compensation Weighing Cells

The weighing cells made by Sartorius [12] function according to the EMFC principle (Fig. 1.4,a) and feature a patented, completely monolithic design. The function of the parallel guides (2, 3) is to intercept the additional torque generated when a sample is off-centered on the weighing pan (1). All moving parts of the system rest on rugged joint flexures (4, 5), an excellent compromise between elasticity and rigidity. The force applied by the mass of a sample being weighed is compensated by the magnetic force of a coil (6) through which current flows. This coil is located in the gap of a permanent magnet (7), replacing the second pan with weights or the switching weights on the conventional balance. A precision resistor (8), integrated into the control loop, converts the current into

a voltage, proportional to the load on the pan. This signal is transmitted to the analog-to-digital converter (9), then filtered and digitally compensated by the microprocessor (10) and shown on a display (11). This rugged technology delivers the highest weight resolution along with the greatest accuracy and shortest response times.

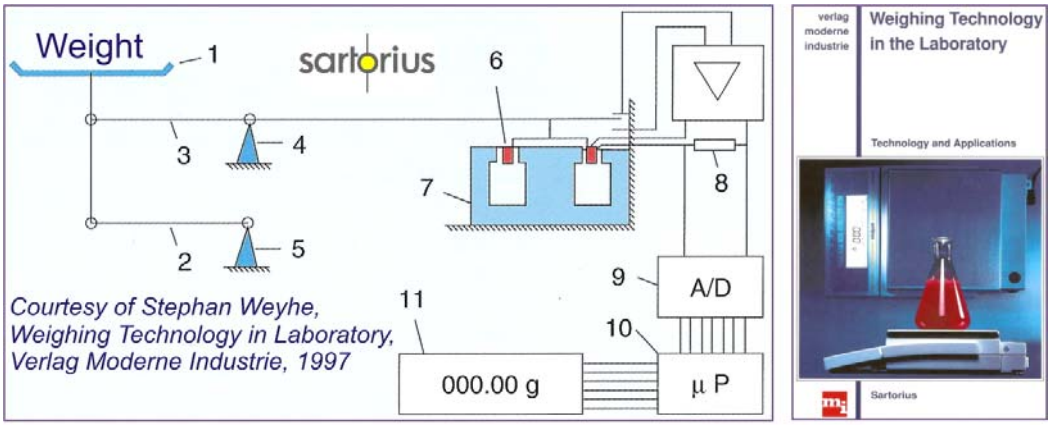


Fig. 1.4. Working principle of electromagnetic force compensation weighing cells (a), and associated documentation edited by Sartorius (b).

1.3.2. Smart Force Transducer with Vibrating Wire

The K-Tron Smart Force Transducer Weighing Technology [13] is illustrated in Fig. 1.5,a. The applied load changes the vibrating wire resonant frequency (within the 10 – 15 kHz range) and the resulting signal is converted to a square wave. In the weighing system SFT2 (Fig. 1.5,b) the digital filtration avoids the interference noises in the equipment; there are also thermal compensation circuits and incorporated micro-processors providing the maximum resolution, without the need for *in situ* recalibration.

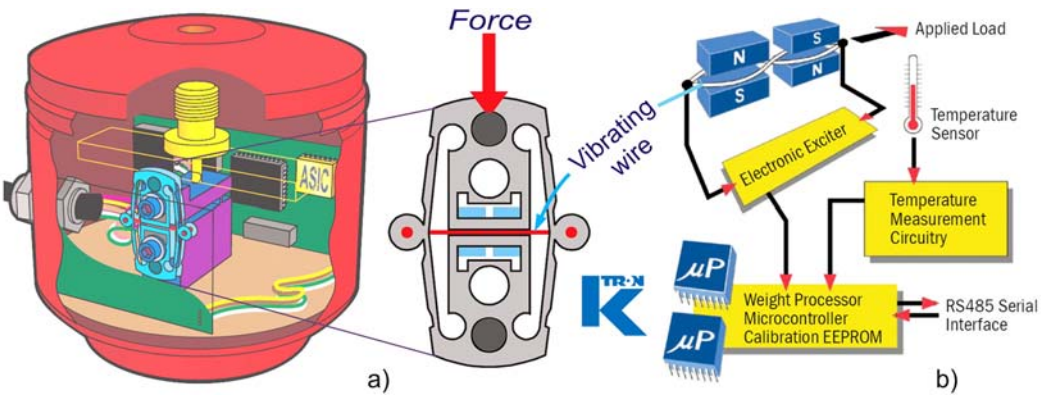


Fig. 1.5. K-Tron Smart Force Transducer Weighing Technology: a) vibrating wire assembly; b) intelligent electronic circuitry.

Other important features of the smart vibrating wire transducer are:

- 100 % digital design, no calibration required;
- Resolution of 1:4,000,000 in 80 ms;
- Weight captured and linearized 112 times per second;
- On-board microcontroller, custom frequency processor, calibration memory and voltage regulator.

1.3.3. Classical and Differential Piezoelectric Force Transducers

In Russian patents [14], beyond the "secretive" description and the exciting term "Intellectual Force Sensor", it is about piezoelectric crystals in a special mount allowing opposed variations of frequencies, then measured differentially. The working principle of the force transducer shown in Fig. 1.6,a: Under the action of the measured force (weight) F on the housing (1), in the form of a coupled double beam with three round cavities, there is a change in frequency of the excitation piezoplates (2), precisely located inside the central cavity, which is generated by the driving circuits (17). Their frequencies (f_1 and f_2) come to the adder (18), the signal of which comes in the differential frequency detector (19), its readout being proportional with the applied load.

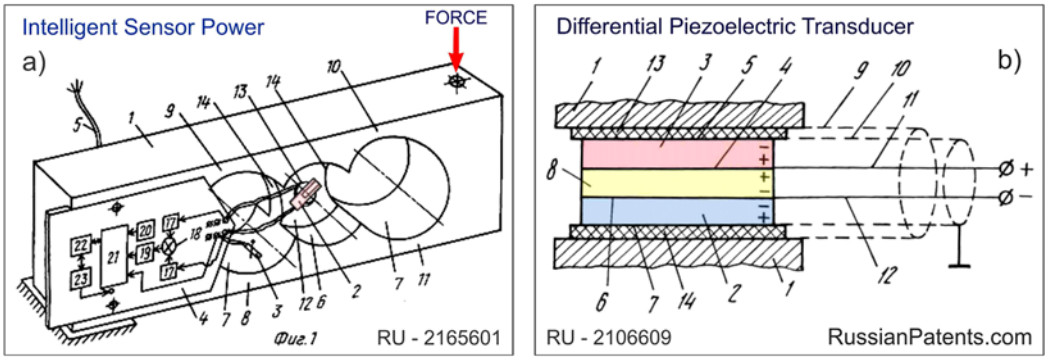


Fig. 1.6. Two Russian patents of differential piezoelectric force transducers.

A special converter (Fig. 1.6,b) can be provided by the scheme of piezoelectric elements loading in the mode of axial tension-compression, shear or bending, but regardless of this, the vectors of polarization of the piezoelectric elements (2 and 3) must be directed opposite to the vector of polarization of the piezoelectric-insulator (8). Using electrodes 5 and 7 of the piezoelectric elements are removed by the electric charges of different polarity but equal in amplitude and therefore, when combined, on the internal screen 10 (which is connected with a common point of the differential amplifier), creates a sustainable zero signal, measuring electrodes remaining only 4 and 6. Using a multilingual expression, "the invention relates to izmeritelnaya technique": a differential piezoelectric

transducer for measuring forces, pressures or accelerations in low-frequency dynamic processes (in the range up to 50 Hz), in terms of the high-frequency interference (over 1 kHz), such as nuclear power plants under extreme operating conditions.

In other application piezoceramic chips are used to measure both normal force and AE during grinding process [15]. Small and inexpensive piezoceramic chips (about $10 \times 3 \times 2$ mm^3) are embedded in the aluminum alloy wheel core of 344 mm diameter and 152 mm bore. Their wide frequency response, from a few Hz to several hundred kHz, makes them suitable for measuring both normal force (macroscopic attribute) and acoustic emission (microscopic attribute).

1.3.4. Electro-optical Catheter

Novel microtactile force sensors have been developed to measure the frictional force and contact force between blood vessel and the side of the catheter [16]. The prototype developed at the Kagawa University, Japan is shown in Fig. 1.7. Pressure sensitive rubbers ($4.0 \times 4.0 \times 0.5$ mm^3) are fixed on the side wall of catheter by a linking shape. They could be piezoresistive [17] or piezoelectric [18].

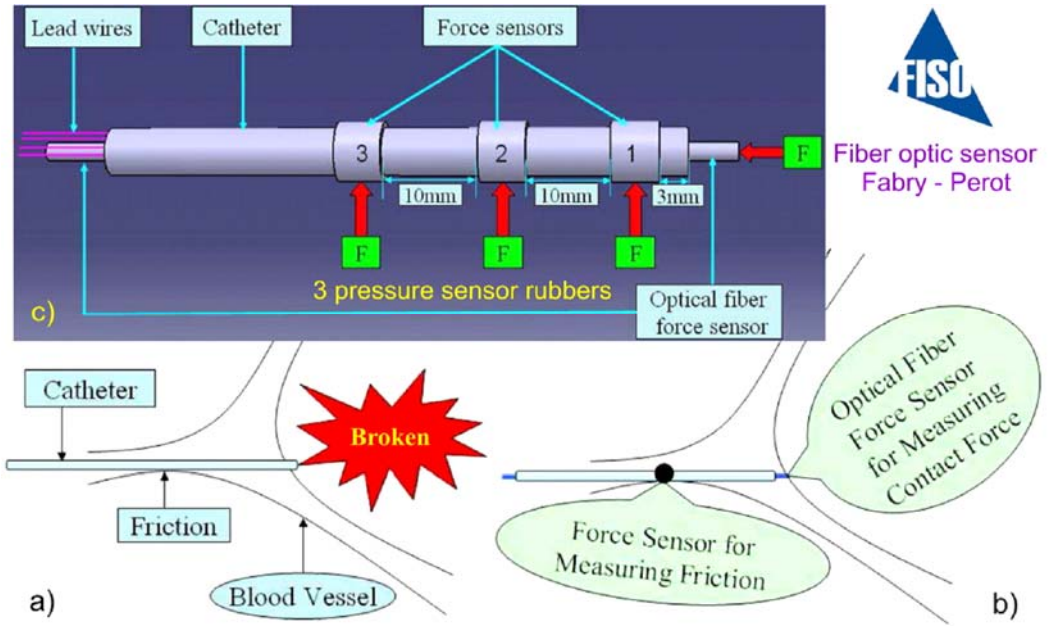


Fig. 1.7. Operating difference between a classical catheter and one tooled with frictional and contact sensors (a), and electro-optical catheter component pieces (b).

An optical fiber serves as guide wire to lead the catheter for inserting and rotating and, finally, to measure front end force of the catheter. The FOP-M fiber optic pressure transducer made by FISO Technologies Inc. is based on proven Fabry-Perot interferometer technology. Pressure creates a variation in the length of the cavity and the

fiber optic signal conditioners can consistently and accurately measure the cavity length with high accuracy under all adverse conditions of temperature, EMI, humidity and vibration.

1.4. Intelligent Force Measurement Channels

1.4.1. Signal Conditioners

A survey of traditional resistance-based sensor conditioning techniques, going from simple analogue op amps to sophisticated digital solutions is presented in [19].

The precision measurement equipment (PME) range of HBM products offers different types of modules for the determining of all measured quantities relevant to processing, such as force and associated quantities. DMP series is widely used at high-precision measurements, as "reference-instrument" [20]. The four main principles used to reach the accuracy class 0.0005 are: Wheatstone bridge (for parametric transducers: R, L, C), six wire circuit, carrier frequency, and symmetric voltage architecture.

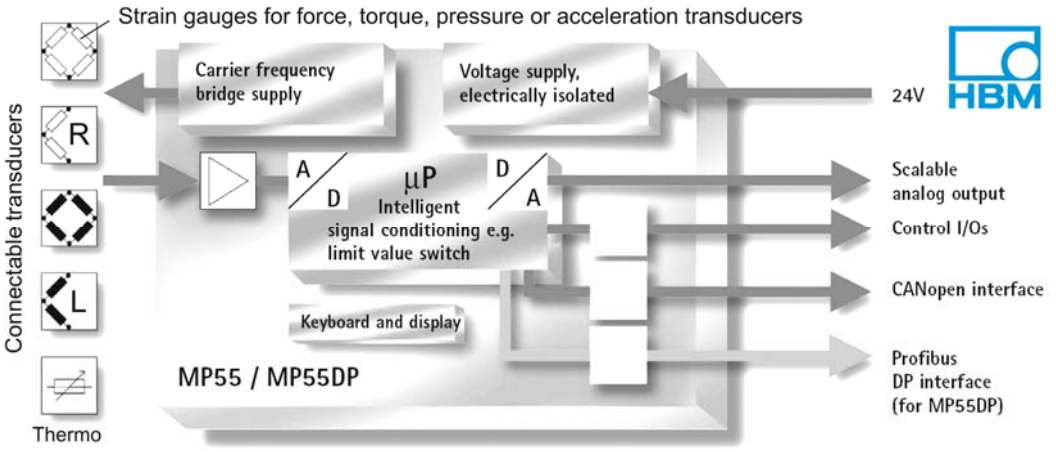


Fig. 1.8. MP55DP intelligent module amplifier with scalable analogue output and digital interface for measuring mechanical quantities: force, torque, pressure, and so on.

Precision load cells have been developed specifically to be used in dynamic weighing processes by means of fast intelligent transducers (FITs) [21].

For the major part of force transducer types the Quantum X signal conditioner is adequate [22]. All force transducers presented in the previous subchapters may be connected to this universal signal conditioner: resistive and capacitive (output voltages), EMFC – i.e. electrodynamic force balance (output current), vibrating wire and piezoelectric (frequency measurements), and optical (impulse counting).

For the full measurement channel the intelligent attribute is given, among others, by the metrological self-check feature. This is interpreted as "an automatic check of the metrological health of the measuring system within an operation process, which is carried out by using embedded hardware and software" [23]. A typical representation is given in Fig. 1.9.

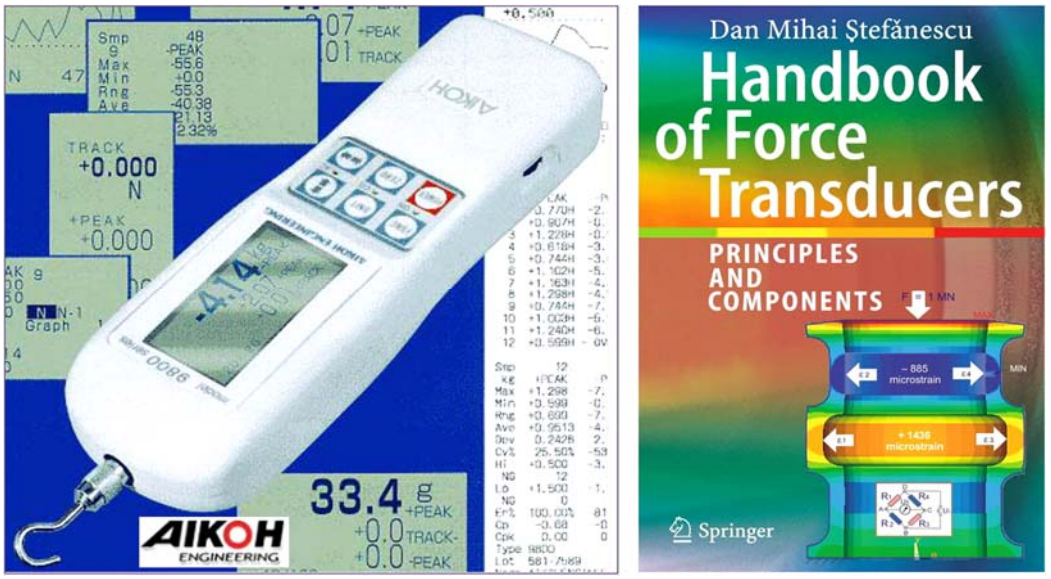


Fig. 1.9. Sensor or transducer with built-in self-test or self-calibration unit.

1.4.2. Digital Displays

Some latest achievements for the force measuring chains include the digital display (in engineering units) within the strain gauged transducers (Fig. 1.10). The Aikoh 9800 series push pull force gauges use a precision force transducer to measure the compressive and tensile forces applied to the force probe [24]. They have a lot of applications in connection with high precision motorized testing machines. The unit displays real time force data on the LCD screen. The instruments can be set up to record a series of measurement results and to tabulate and plot peak and average values. High and low limit values can be set up for quality control and inspection operators.

Other components, as well as measuring methods and principles, may be found in [5], with the final chapter dedicated to intelligent transducers.

1.5. Intelligent Force Sensing Applications

The most complex intelligent applications in force measurements may be found in the areas of Robotics, Wireless Force Sensing and Virtual Instrumentation.

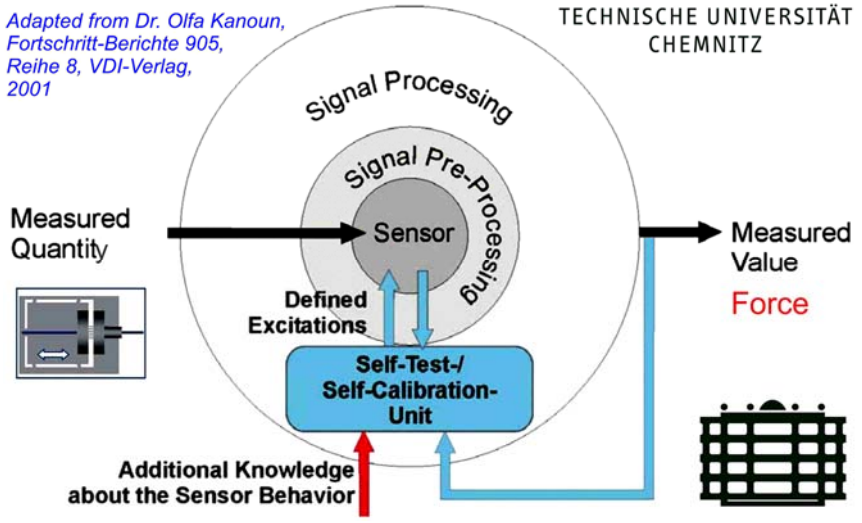


Fig. 1.10. Japanese push pull gauge for weighing applications. One can observe the wide variety of presentation for the weighing documents.

1.5.1. Intelligent Robots

Interactive, intelligent robots can be considered both sensitive force transducers and swift actuators [25]. Typical examples are some achievements of top companies such as FANUC (Asia), Adept (North America) and ABB (Europe).

The FANUC robot shown in Fig. 1.11,a may be seen as a 6-axis force moment (often called torque) transducer, similar to multicomponent tensometric balances (Fig. 1.2), while notations are different and the six components are given names specific to the particular application.

Fig. 1.11. Intelligent force sensing robots: FANUC representation of loading with 3 forces and 3 torques (a) and multicomponent force transducer mounted into Adapt smart robot (b).

The generic term *Intelligent Force Sensing System* is used by Adept (Fig. 1.11, b), avoiding the Force Sensor vs. Force Transducer "dilemma". There is a tight integrated hardware and software package that allows controlled robots to react to sensed forces and moments [26]. As a result, it reduces force overshoot and robot stopping time when forces or moments exceed preset thresholds, so the assembling operations can be performed with higher speed and accuracy.

Part of *hardware*, the force transducer is a device that detects forces and moments applied by the end-of-arm gripper by means of strain gauges mounted on internal flexing elements. The electronic readings are transmitted to a Smart Controller EX motion controller and converted into useful force readings using a specific calibration matrix. The *software* instructions execute a full range of customized functions, including force-guided identification of the geometry of the most common contact configurations and controlling force operating modes.

Robots get smart with real-time feedback [27]. Integrated Force Control handles process variations with human sensitivity, improving performance, and reducing programming time. It makes robots more intelligent and able to handle, as a human, delicate items with real-time external inputs.

At a microforce scale an example on how to transfer flexible electronic skin to the next generation of humanoid robots at the Institute for Cognitive Systems, Technische Universität München (Germany) is given in [28]. It is about the intelligent integrating (cognitively and mechatronically) of discrete force cells into multi-modal self-organizing robotic skin.

Robots can assist people with motor impairments to perform activities of daily living (ADLs) that often involve contact with a person's face [29]. In a recent application at the Healthcare Robotics Lab, Georgia Institute of Technology, Atlanta, USA, people have the ability to control the forces they apply, and receive feedback from both sides of each contact (i.e. feedback from the hand that is holding the tool and the location on the body where the tool is being used) (Fig. 1.12). So, it is reasonable to assume that able-bodied people tend to apply forces to themselves that they find comfortable, safe, and effective. By emulating these forces, a robot equipped with a six-axis force / torque transducer (ATI Nano 25) could apply such target forces (till 10 N) for various human tasks.

Another application in which the robot "copies" the human skill is the smart glove. While *intelligence* refers to the functional aspect, the term *smart* is connected to technology, such as comes out of a recent achievement, a novel smart glove technology as a biomechanical monitoring tool [30]. The authors of this development in Virtual Reality (VR) technology conclude: "Data gloves have been proven to be a viable replacement for the universal goniometer and can offer unbiased finger joint ROM measurement."

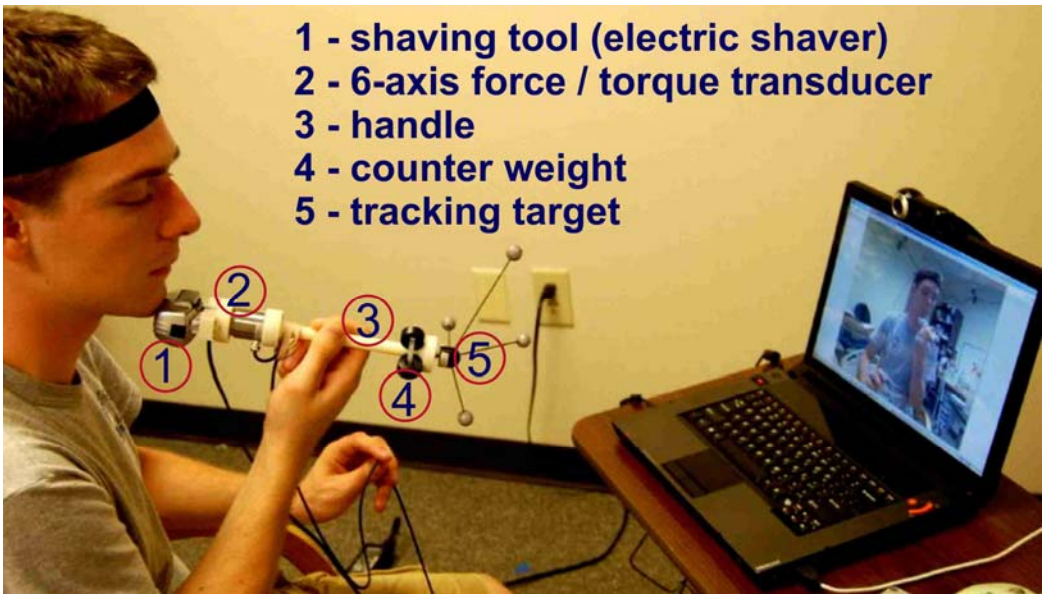


Fig. 1.12. Assistive robot with models of contact forces from able-bodied face shaving.

1.5.2. Wireless Force Sensing

Mukhopadhyay [31] links Intelligent Sensing first of all with Wireless Sensors. Varghese et al. [15] have used a DSP (digital signal processing)-based telemetric data acquisition and transmission module from the sensor-integrated "intelligent" grinding wheel to the remote receiver at 900 MHz bandwidth.

For bridges in service steel cable forces are an important parameter for their safety assessment [32]. This is monitored by measuring the transverse vibration frequencies of cables by means of a wireless sensing unit model with a powerful computational core and low power consumption characteristics (Fig. 1.13). The software is embedded in the microcontroller core and structured using a multi-layer architecture. At the lowest layer is the real-time operating system (OS) which directly operates the wireless sensor hardware. Software for managing sensors data resides on a second tier, using algorithms exploiting the intelligent characteristics of the wireless sensor.

1.5.3. Virtual Instrumentation – TensoDentar

Another author's application, called TensoDentar, is described in [33]. An original strain gauged force transducer, with graphical systems engineering LabVIEW and virtual instrumentation (Fig. 1.14), are used to evaluate and compare the efficiency of sandblasting and plasma treatment on complete denture base materials.

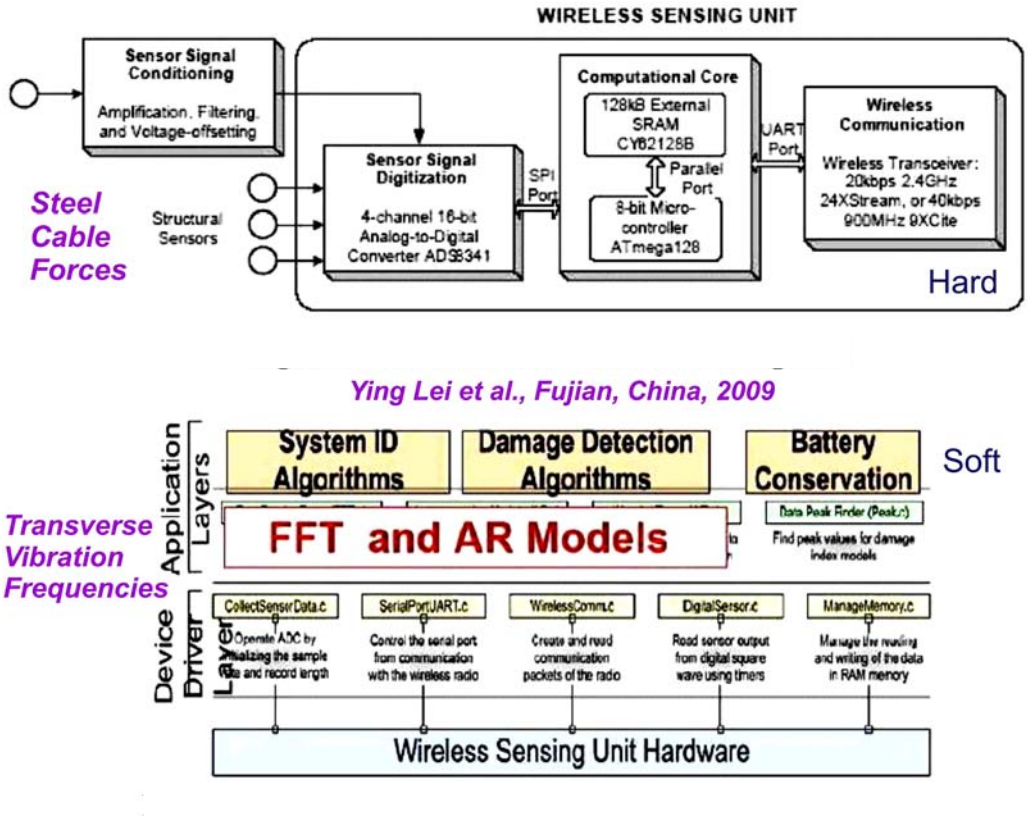


Fig. 1.13. Wireless force sensing with three main functional modules: sensor interface, computational core (with embedded software) and wireless transceiver.

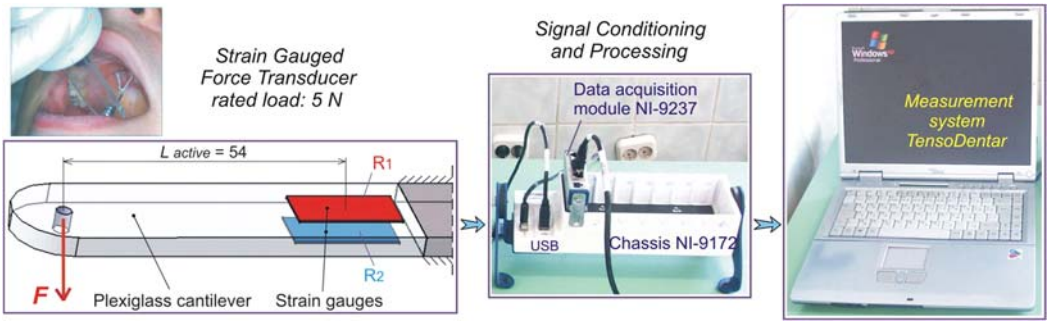


Fig. 1.14. Block diagram of TensoDentar: force transducer (cantilever beam with strain gauges), data acquisition hardware ("architecture" of National Instruments integrated solutions) and laptop (with its associated software LabWindows).

1.6. Conclusions

Aspects of using intelligence in various stages of implementing high performance force and related quantities measuring systems and transducers were presented. Strain gauge Wheatstone bridge connected force transducers were emphasized. Other intelligent force transducer types were also described: electromagnetic, vibrating wire, piezoelectric and an electro-optic combination. A series of intelligent applications were presented: robots, wireless systems and virtual instrumentation (an original device, TensoDentar, for evaluation of dislodging forces for complete dentures or palatal plates).

This chapter represents a development of the presentation delivered as Keynote Speaker at SEIA'2015 [34].

Acknowledgements

The author expresses his gratitude to Dr. Sergey Yurish, IFSA Publishing, and Dr. Amin Daneshmand Malayeri, ASAME, as well as to the Romanian compatriots, Dana Lepinzan, Mircea Duşa and Răsvan Stănescu, for their valuable support related to attending the SEIA International Conference in Dubai, UAE, 21 – 22 November 2015.

References

[1]. J. Bryzek, Evolution of smart transducer technology, *Sensors'95 Kongreßband*, Nürnberg, Germany, 9-11 May 1995, pp. 45-50.
[2]. P. Zecchin (Chair), Digital Load Cells – A Comparative Review of Performance and Application, *The Institute of Measurement and Control,* London, 2003.
[3]. D. M. Ştefănescu, Energetical and informational aspects in transducers, in *Proceedings of the 20th IMEKO TC4 International Symposium on Electric and Electronic Measurement for the Economic Upturn*, Benevento, Italy, 15-17 September 2014, pp. 463-468.
[4]. S. Y. Yurish, Sensors: Smart vs. intelligent, *Sensors & Transducers*, Vol. 114, Issue 3, March 2010, pp. I-VI.
[5]. D. M. Ştefănescu, Handbook of Force Transducers – Principles and Components, *Springer-Verlag Berlin and Heidelberg*, 2011.
[6]. H.-R. Tränkler, O. Kanoun, Some contributions to sensor technologies, in *Proceedings of the Conference on Sensors and Systems*, Saint Petersburg, Russia, 24 – 27 June 2002.
[7]. J. B. Brignell, N. M. White, Intelligent Sensor Systems, *Institute of Physics Publishing, Bristol and Philadelphia, PA*, 1994.
[8]. D. M. Ştefănescu, Strain gauges in a wind-tunnel application, in News from the world of test and measurement, *Hotline Hottinger Baldwin Messtechnik GmbH, Darmstadt*, Issue 2, 2001, pp. 12-13.
[9]. Improving medical devices with force sensing technology, *Tekscan, South Boston, MA, USA*, PDF created on 26 February 2015.
[10]. W. Y. Du, S. W. Yelich, Resistive and capacitive based sensing technologies, in Modern Sensors, Transducers and Sensor Networks (Book Series: Advances in Sensors: Reviews, Vol. 1), S. Yurish (Ed.), *International Frequency Sensor Association (IFSA) Publishing*, 2012.

[11]. Intelligent force measurement using revolutionary capacitive sensors, *Loadstar Sensors, Fremont, CA, USA,* Accessed 8 November 2015. (http://www.loadstarsensors.com/force-measurement-capacitive-sensors.html).
[12]. St. Weyhe, Weighing Technology in the Laboratory: Technology and Applications, *Sartorius + Verlag Moderne Industrie, Landsberg/Lech, Germany*, 1997.
[13]. K-Tron smart force transducer weighing technology for gravimetric feeding, batching and metering, Leaflet F-700012-en, *K-Tron International, Inc.,* Pitman, NJ, USA, May 2009.
[14]. Intelligent sensor power – *RussianPatents.com.* (russianpatents.com/patent/216/2165601.html).
[15]. B. Varghese, S. Pathare, R. Gao, C. Guo, S. Malkin, Development of a sensor-integrated, intelligent, grinding wheel for in-process monitoring, *Annals of the CIRP* (The International Academy for Production Engineering), Vol. 49, Issue 1, 2000, pp, 231-234.
[16]. S. X. Guo, J. Guo, N. Xiao, T. Tamiya, Robotic catheter operating systems for endovascular neurosurgery, in Explicative Cases of Controversial Issues in Neurosurgery, F. Signorelli (Ed.), *InTech (Open Science – Open Minds)*, 2012, pp. 457-478.
[17]. A. Nikonovas, A. J. Harrison, S. Hoult, D. Sammut, The application of force-sensing resistor sensors for measuring forces developed by the human hand, in *Proceedings of the Institute of Mechanical Engineering [H]*, *SAGE Publications, UK*, Vol. 218, Issue 2, 2004, pp. 121-126.
[18]. D. Castano Cano, M. Grossard, A. Hubert, Multi-axis force sensing with pre-stressed resonant composite plates: An alternative to strain gauge force sensors, in *Proceedings IEEE / ASME International Conference on Advanced Intelligent Mechatronics (AIM'14)*, 8-11 July 2014, Besançon, France, pp. 1-7.
[19]. D. Cheeke, Sensor signal conditioning, *Sensors & Transducers Journal*, Vol. 82, Issue 8, August 2007, pp. 1381-1388.
[20]. A. Schäfer, The ultra-precision instrument DMP 41 – first experiences & appropriate filter settings, in *Proceedings of the IMEKO 22nd TC3, 12th TC5 and 3rd TC 22 International Conferences*, Cape Town, Republic of South Africa, 3-5 February 2014, Paper 127.
[21]. U. Milz, Soybean oil and baby milk split-second bottling with FIT® fast intelligent transducers, *Hotline Hottinger*, Issue 1, 2002, p. 27.
[22]. J. Boersch, QuantumX – the new, multi-functional amplifier system, *Hotline Hottinger*, Issue 2, 2007, pp. 4-5.
[23]. R. Taymanov, K. Sapozhnikova, I. Danilova, I. Druzhinin, Multi-channel intelligent measuring systems, in *Proceedings of the XXI IMEKO World Congress Measurement in Research and Industry*, Prague, Czech Republic, 30 August – 4 September 2015, pp. 34-39.
[24]. Aikoh 9800 digital push / pull force gauge, *Aikoh Engineering, Tokyo, Japan*, Accessed 15 August 2015. (www.microphotonics.com).
[25]. G. L. Xiong, H. C, Chen, R. H. Zhang, H. Zhang, J. B. Huang, Development of intelligent force sensor system for interactive robot, in *Proceedings of the International Conference on Mechanical Engineering and Technology (ICMET 2011)*, London, UK, 24-25 November 2011, pp. 505-510.
[26]. Adept intelligent force sensing system, User's Guide, 14155-000 Rev. A, *Adept Technology, Inc., Pleasanton, CA, USA*, September 2014.
[27]. ABB Robotics' integrated force control, *ABB, Zürich, Switzerland*, Accessed on 4 November 2015. (http://foundrymag.com/finishingmro/robots-get-smart-real-time-feedback).
[28]. Ph. Mittendorfer, G. Cheng, From a multi-modal intelligent cell to a self-organizing robotic skin, *Technische Universität München, Germany,* PDF created on 11 March 2013. (See also http://www.ics.ei.tum.de).
[29]. K. P. Hawkins, C.-H. King, T. L. Chen, C. C. Kemp, Informing assistive robots with models of contact forces from able-bodied face wiping and shaving, *Georgia Institute of Technology, Atlanta, USA*, PDF created on 23 April 2013.

[30]. B. O'Flynn, J. T. Sanchez, S. Tedesco, B. Downes, J. Connolly, J. Condell, K. Curran, Novel smart glove technology as a biomechanical monitoring tool, *Sensors & Transducers*, Vol. 193, Issue 10, October 2015, pp. 23-32.

[31]. S. C. Mukhopadhyay, Intelligent Sensing, Instrumentation and Measurements. *Springer-Verlag Berlin Heidelberg*, 2013.

[32]. Y. Lei, W. A. Shen, Y. Song, Y. Wang, J. P. Lynch, Intelligent sensors with application to the identification of structural modal parameters and steel cable forces, *Xiamen University, Fujian, China*, PDF created on 9 October 2009.

[33]. D. M. Ştefănescu, A.-T. Farcaşiu, A. Toader, Strain gauge force transducer and virtual instrumentation used in a measurement system for retention forces of palatal plates or removable dentures, *IEEE Sensors Journal*, Vol. 12, Issue 10, 2012, pp. 2968-2973.

[34]. D. M. Ştefănescu, New achievements in the field of intelligent force transducers, in *Proceedings of the International Conference on Sensors Engineering and Electronics Instrumental Advances (SEIA' 2015),* Dubai, UAE, 21-22 November 2015, pp. 28-31.

Chapter 2
Intelligent Slip Displacement Sensors in Robotics

Yuriy P. Kondratenko, Volodymyr Y. Kondratenko and Eduard A. Shvets

2.1. Introduction

Intelligent properties are important for modern robots to gain experience and to adapt to real non-stationary working environment for various missions executing. Updated intelligent robots pose high dynamic characteristics and effectively function under determinate conditions. The robot control problem is more complex in the uncertain environment, as robots are deficient of flexibility [1, 2]. Supplying robots with effective sensor systems provides essential extension of their functional and technological feasibility [3, 4]. For example, robot may often encounter a problem of gripping and holding i-th object in the process of manipulations with the required clamping force P_i^r excluding its deformation or mechanical injury, $i = 1...n$.

To successfully solve the current tasks the robot should possess the capability to recognize the objects by means of their own sensory systems. Besides, in some cases the main parameter due to which robot can distinguish objects of the same geometric form is their mass $m_i, (i = 1...n)$. The robot sensor system should identify the mass m_i of each i-th manipulated object in order to identify a class (set) an object refers to. The sensor system should develop required clamping force P_i^r corresponding to mass value m_i, as

$$P_i^r = f\left(m_i\right). \tag{2.1}$$

Such current data may be used when the robot functions in dynamic or random environments. For example, there may be a situation when the robot should identify for any type of objects their unknown parameters and location in robot's working zone. The

Yuriy P. Kondratenko
Petro Mohyla Black Sea State University, Ukraine
Cleveland State University, Cleveland, OH, USA

visual sensor system may not always be utilized, in particular, in poor vision conditions. Furthermore, in such cases when the robot manipulates with an object of variable mass $m_i(t)$, its sensor system should provide the appropriate change of a clamping force value $P_i^r(t) = f\left[m_i(t)\right]$ for gripper fingers. This information can be also used for the robot control algorithm correction, since a mass of the robot's arm last component and its summary inertia moment vary.

One of the modern approaches to solve the identification problem concerning the mass m_i of grasped objects and producing the required clamping force P_i^r is in the development of tactile sensor systems based on object slippage registration while slipping between the gripper fingers. Some sensitive systems based on the slip displacement sensors were considered in [5, 6], but the random robot environment very often requires the development of new robot sensors and sensor systems for increasing speed of operations, the growth of positioning accuracy or the desired path-following precision [3].

2.2. Methods and Sensors for Detection of Slip Displacement Signals in Intelligent Robotics

2.2.1. Analysis of the Main Methods for Slip Displacement Signals Detection

Thus, the task of the slippage signals registration between robot fingers for manipulated objects stands in connection with [3, 7]:

a) The necessity of the required force creation being adequate to the object mass value;

b) The recognition of objects;

c) The necessity of robot gripper's trajectory and control algorithm correction.

As usual, the slippage signals detection in robotic systems is accomplished either in the trial motion or in the regime of continuous lifting of the robot arm.

The idea of a trial motion regime comprises the process of the iterative increase in the compressive force value if the slippage signal is being detected.

The choice of the method of slip displacement data acquisition depends on robot's purpose, the salient features of its functioning medium, the requirements of its speed of response and performance in terms of an error probability.

Let's consider the main methods of slip displacement data acquisition [3, 4, 7–13]:

The method of rolling motion detection. The method is characterized by transducing the object displacements in the vertical direction at slipping to the rolling motion of a sensitive element. A slip displacement signal is detected at rolling of a cylinder roller with elastic covering and a large friction coefficient. The roller's rolling motions may be converted to

an electric signal by means of photoelectric or magnetic transducers, containing a permanent magnet on a movable roller, and in case of a magnetic head being placed on a slippage sensor.

The method of the configuration changes detection in the sensitive elements. The essence of the method incorporates the measurement of the varying parameters when the elastic sensitive element configuration changes. The sensitive elements made of conductive rubber afford coating of the object surface protruding above the gripper before the trial motion. When the object is displacing from the gripper the configuration changes, the electrical resistance of such sensitive elements changes accordingly confirming the existence of slippage.

The method of fixing the sensitive elements on the object. The method is based on fixing the sensitive elements on the surface of the manipulated objects before the trial motions with the subsequent monitoring of their displacement relative to the gripper at slipping.

The method of vibration detection. This method is based on a principle of the vibration detection in the sensing element when the object is slipping. To implement the method mentioned such sensing elements may be adopted: a sapphire needle interacting with the crystal receiver or a rod with a steel ball, connected with the electromagnetic vibrator.

The method of pressure re-distribution detection. The method relies on the detection of a distribution change in pressure between gripper fingers at the object slippage and is based on the physiological sensibility function of man's skin. The pressure transducers serve as nerves and are surrounded by elastic substance as in human body.

The method of impact-sliding vibrations detection. A core of the method implies the detection of liquid impact-sliding vibrations when the object is slipping. An acrylic disk with cylinder holes is used in the slip displacement sensor, realizing the method under consideration. A rubber gasket made in the form of a membrane protects one end of the disk, and a pressure gauge is installed on another end. The hole is filled with water in the way that its pressure exceeds slightly the atmospheric pressure. While the motion of the object is in contact with a membrane the impact-sliding vibrations are appearing and, therefore, inducing impulse changes in water pressure imposed by a static pressure.

The method of acceleration detection. This method is based on the measurement of accelerations of the sensitive element motion by the absolute acceleration signal separation. The slip displacement sensor comprising two accelerometers can be used in this case. One of the accelerometers senses the absolute acceleration in the gripper, another responds to the acceleration of the sensitive plate springing when the detail is slipping. The sensor is attached to the computer identifying the slip displacement signal by comparing the output signals of both accelerometers.

The method of the interference pattern changes detection. This method involves the conversion of the intensity changes reflected from the moving surface of the interference pattern. The intensity variation of the interference pattern is converted to a numerical code,

the auto-correlation function is computed and it achieves its the peak at the slip displacement disappearance.

The method of data acquisition by means of the photoelastic effect. An instance representing this method may be illustrated by a transducer, in which under the applied effort the deformation of sensitive leather produces the appearance of voltage in the photoelastic system. The object slippage results in the change of the sensitive leather deformation being registered by the electronic visual system. The photosensitive transducer is a device for the transformation of interference patterns into the form of a numerical signal. The obtained image is of binary character, each pixel gives a bit of information. The binary representation of each pixel enables to reduce the time of processing.

The method of data acquisition based on the friction detection. The method involved ensures the detection of the moment when the friction between the gripper fingers and the object to be grasped goes over from friction of rest to dynamic friction.

The method based on recording oscillatory circuit parameters. The method bases on a change in the oscillatory circuit inductance while the object slippage. The inductive slip sensor with a mobile core, stationary excitation winding and solenoid winding being one of the oscillatory circuit branches implements the method. The core may move due to the solenoid winding. The reduction of the solenoid winding voltage is leading to the process of lowering, the core is lowering under its own weight from the gripper center onto the object to be grasped. The oscillatory circuit induces the forced oscillations with the frequency coinciding with the frequency of excitation in the excitation winding.

The method of video signal detection. The basis of this method constitutes a change in detection and ranging of patterns or video pictures as an indication of the object slippage. The slip displacement detection is accomplished by means of the location sensors or visual sensors based on a laser source that has either a separated and reflecting beam, or a vision with a non-coherent beam of light conductors for picture lighting and a coherent beam for image transmission.

Analysis of last publications on tactile and slip displacement sensors shows [3, 4, 7, 12, 13] that researchers develop a lot of new solutions using well-known and new design methods.

2.2.2. Main Requirements for Real-Time Detection of Slip Displacement Signals

Frequently the handling operations require a compressive force being exerted through the intermediary of the robot's sensing system in the continuous hoisting operation. This regime shows a simultaneous increase in the compressive force while the continuous hoisting operation and lifting of the gripper in the vertical direction accompanied by the slip displacement signal measurement. When the slippage signal disappears the compressive force does not increase and, therefore, the operations with the object are accomplished according to the robot's functioning algorithm. To realize the trial motion

regime and the continuous hoisting operation being controlled in real time, the stringent requirements to the parameters should be met [4, 7], in particular:

a) The response time between the moment of slippage emergence and the moment when the gripper fingers begin to increase the compressive force;

b) The time of the sliding process including the moments between the emergence of sliding and its disappearance;

c) The minimal object displacement detected by the slip signal.

The problem of raising the sensors response in measuring the slip displacement signals is tackled by improving their configuration and using the measuring circuit designs with high resolving power.

2.2.3. The Trends in Slip Sensors Design and Their Modern Modifications

The choice of a slip displacement detection method involves the multi-criterion approach taking into account the complexity of implementation, the bounds of functional capabilities, mass values and overall dimensions, reliability and cost.

In this chapter, the authors consider a few instances of updating the measurement systems. To suit requirements of increasing the noise immunity of the vibration measurement method a modified method has been developed. The modified method is founded on the measurement of the sensitive element angular deviation occurring at the object slippage.

The second method serves as a basis for developing the distributed slip sensors of the matrix type, for example, a two-dimensional tactile slip displacement sensor (SDS) that has a form of a round piezooptic beamconductive plate. Another modification of the matrix type slip sensor is a sensitized module constructed in the thin conducting film in the form of 12 contactors joined in a measurement circuit.

The sensor for detection of slip displacement signal with optical measurement system [14] is represented in Fig. 2.1.

The need for rigid gripper orientation before a trial motion has been causing the development of the slip sensor based on a cylinder roller with an additional load which has two degrees of freedom [11].

The sensitive element of new developed by authors [3] sensor has the form of a ball with light-reflecting sections disposed in a staggered order thus providing the slippage detection by the photo-method. The ball is arranged in the sensor's space through the spring-loaded slides, each slide is connected with the surface of the gripper's space by an elastic element made of conductive rubber.

The ball motion is secured by friction-wheels and is measured with the aid of incremental transducers in another modification of the slip sensor with the ball acting as a sensitive

element. The ball contacts with the object through the hole. In this case the ball is located in the space of compressed air dispensed through the hole.

At the moment, a lot of engineering solutions for slip displacement sensors designing are based on the application of sensitive electroconductive rubber [11, 12].

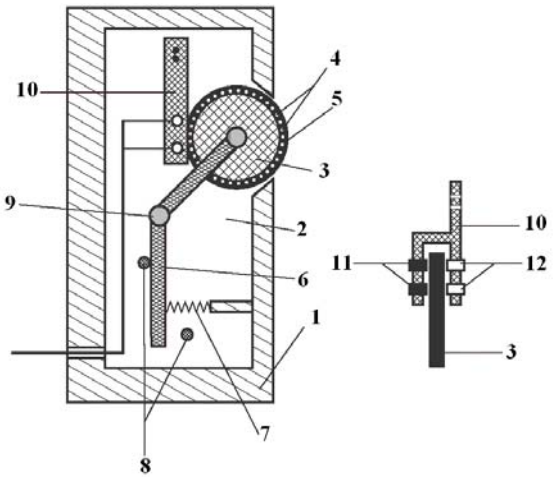


Fig. 2.1. Slip displacement sensor based on the rotated cylindrical roller and optical measurement system: 1 – gripper's finger; 2 – a cavity for SDS installation; 3 – roller (sensitive element); 4 – information components in the form of cylindrical holes in the roller body 3; 5 – elastic working surface; 6 – swivel lever; 7 – spring; 8 – limiters; 9 – hinge; 10 – bracket; 11 – light source; 12 – light receiver.

2.2.4. Advances in the Development of Self-Clamping Grippers of Intelligent Robots

The slip displacement signals, responsible for creation of the required compressive force adequate to the object mass, provide the conditions for the correction of the gripper trajectory-planning algorithm, which identifies an object mass as a variable parameter [15]. The object mass identification is carried out in response to the final value of the compressive force, recorded at the slippage signal disappearance. It is of extreme importance to employ the slip sensors with uprated response when the object mass changes in the functioning process.

In those cases, when the main task of the sensing system is the compression of the object without its deformation or damage, it is expedient in future research to project the advanced grippers of a self-clamping design (Fig. 2.2), excluding the gripper drive for the compressive force growth (at slipping) up to the required value.

In such gripper (Fig. 2.2, Fig. 2.3), the rollers have two degrees of freedom and during object's slippage they have compound behavior (rotation and translation motions). This gripper (Fig. 2.2, Fig. 2.3) has adaptive property since the object self-clamping is being

accomplished with a force adequate to the object mass up to the moment of the slippage disappearance [11, 16].

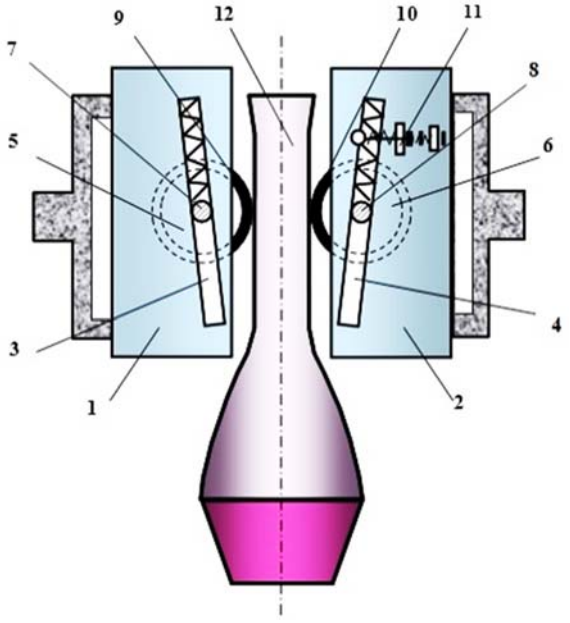


Fig. 2.2. Adjusting of self-clamping force of robot's gripper in automatic mode.

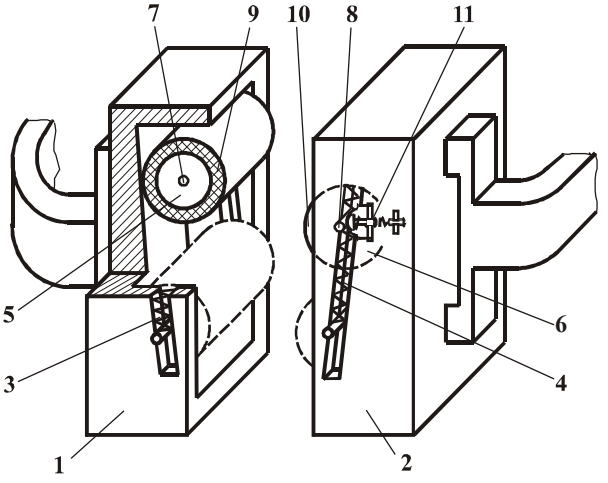


Fig. 2.3. Self-clamping gripper of intelligent robot: 1, 2 - finger; 3, 4 - directing groove; 5, 6 - roller; 7, 8 – roller axle; 9, 10 - elastic; 11 - contact sensor.

Another engineering solution [4, 17] for designing self-clamping gripper is presented in Fig. 2.4, where cams are used as clamping elements. On the left finger (Fig. 2.4) of robot gripper made two partial horizontal cross-sections through the bottom of the guide groove

and through the middle axis of cam roller. The surface of each cam roller, which is in direct contact with the object of manipulation (Fig. 2.4) covered by elastic rough material, such as rubber. Adaptive gripper of intelligent robot is equipped with the first sensor of compressive force (Fig. 2.4), which records the movement of the cam rolling element with increasing compressive force of adaptive gripper fingers. First clamping force sensor can be made, for example, as potentiometric, inductive or capacitive.

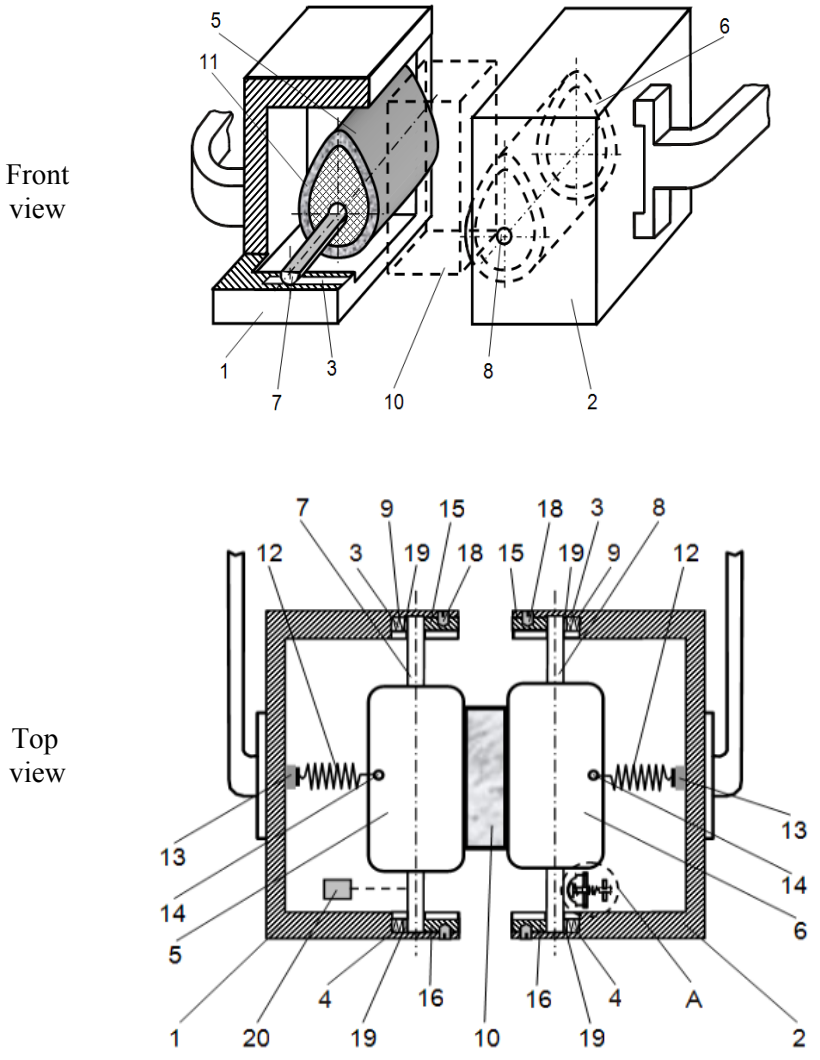


Fig. 2.4. Self-clamping gripper of intelligent robot with cam rollers: 1, 2 - finger; 3, 4 - directing groove; 5, 6 – cam's roller; 7, 8 – cam's roller axis; 9 – cylinder return spring,which is connected with cam's roller axis; 10 – manipulated object; 11 - elastic rough material; 12 - cylinder return spring, which is connected with surface of cam's roller; 13 - fixator; 14 - hinged connection;

15, 16 – pad of blocking fixator; 17 – a hole in the pad of blocking fixator; 18 – screw; 19 – segmental fixator; 20 – first sensor; A – second tactile sensor of clamping force.

When compressed fingers (1, 2), the axes of gripper (7, 8) go deep into the appropriate slots (3, 4) until compression (Fig. 2.4) of cylindrical springs (9), which is achieved with the creation of minimum (pre-set) value of compressive force $F_{\min}$ by fingers of the gripper. As with the at least one cam's roller axle interacts with a second sensor of compressive force (Figs. 2.4, 2.5), the electroconductive contacts are closed, which is a command signal for vertical motion of adaptive gripper, in particular, for lifting an object (10) by intelligent robot. The output of the first sensor of clamping force is formed by compressive force signal value that corresponds to the established compressive force between fingers of gripper and corresponds to the mass of the manipulated object.

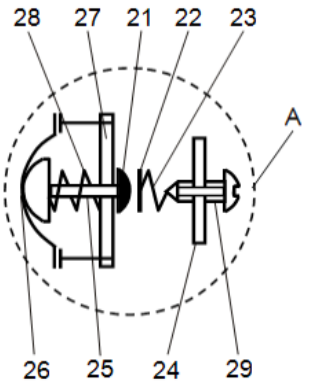


Fig. 2.5. Second tactile sensor of self-clamping gripper of intelligent robot:
21, 22 – electroconductive contact; 23, 28 – cylindrical spring; 24, 27 – bar, which fixed on the finger; 25 – rod; 26 – elastic cover; 29 – adjusting screw.

The process of required clamping force creating, according to unknown mass of manipulated object, illustrated by Fig. 2.6, where:

- Fig. 2.6 (a) presents the position of adaptive gripper with manipulation object and position of cam's rollers in situation, when intelligent robot creates the minimum value of clamping force $F_{\min}$, where h_1 is a distance between gripper fingers and base surface;

- Fig. 2.6 (b) presents the positions of cam's rollers and adaptive gripper with manipulation object in situation, when intelligent robot creates the required value of clamping force F_{req}, where h_2 is a distance between gripper fingers and base surface (30) and

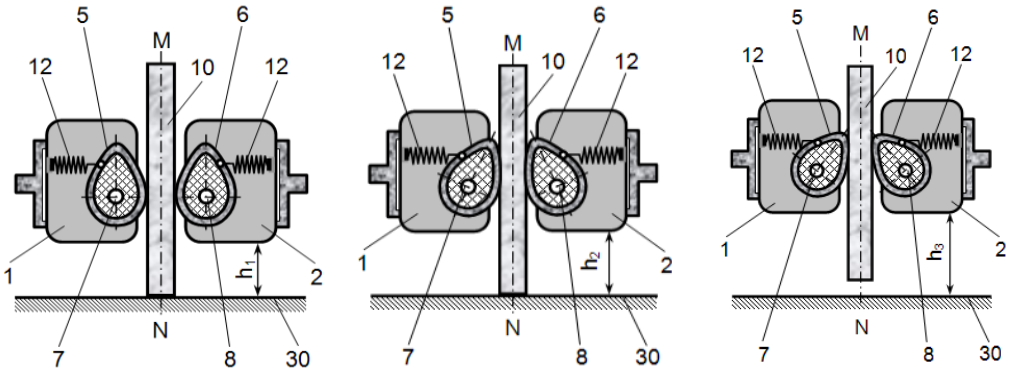
$$\Delta_1 = \left(h_2 - h_1\right) \tag{2.2}$$

is the value of vertical displacement of robot gripper during process of required clamping force creating;

- Fig. 2.6 (c) presents the position of adaptive gripper in situation when intelligent robot moves the manipulation object in vertical direction, where

$$\Delta_2 = \left(h_3 - h_2\right) \tag{2.3}$$

is the value of vertical displacement for the manipulation object.



(a) (b) (c)

Fig. 2.6. The sequence of self-clamping gripper's positions during creating minimal (a) and required (b) values of clamping force and during vertical lifting (c) of manipulated object: 30 - base surface.

2.3. Slip Displacement Sensors with Magnetic Sensitive Components

2.3.1. Mathematic Models of SDS with Magnetic Sensitive Roller

This study presents a number of sensors with magnetic sensitive components for data acquisition in real time [11, 18]. The slip displacement sensor with registration of magnetic field changes, based on the magnetic cylindrical roller and Hall-sensor, presented in Fig. 2.7.

Slip displacement sensor (Fig. 2.7) consists of magnetic system based on constant magnet (roller) and magnetic sensitive element (Hall sensor). The permanent magnet is designed as a magnetized closed toroidal magnetic core. The output voltage U_{HS} of Hall sensor depends on rotation angle α of magnetic roller.

The mathematical model of magnetic system of slip displacement sensor (Fig. 2.7), which formed based on the results of experimental investigations, can be presented in the following way:

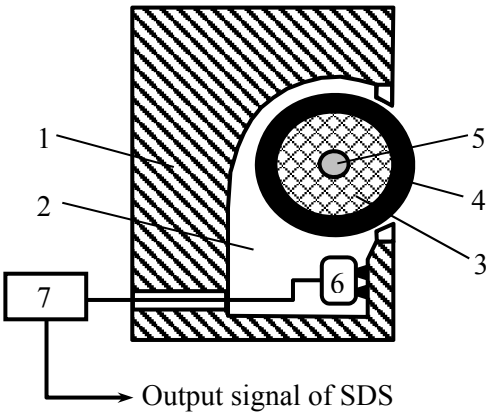


Fig. 2.7. Slip displacement sensor based on the measurement of magnetic field changes: 1 – gripper's finger; 2 – cavity for SDS installation; 3 – magnetic roller; 4 – elastic layer; 5 – axis; 6 – Hall sensor; 7 – measurement device.

$$U_{HS}(\alpha) = A_0 + A_1 \sin(\alpha + \phi_0). \tag{2.4}$$

The following parameters of mathematical model (2.1) are obtained after optimization by least square method (LSM): A_0=2.53; A_1=1.16; $\phi_0 = -0.2$. So, finally mathematical model (2.1) can be presented as follows:

$$U_{HS}(\alpha) = 2.53 + 1.16 \sin(\alpha - 0.2). \tag{2.5}$$

Comparative experimental (marked by circles) and modelling (denoted by a solid line) results for magnetic SDS, which is under consideration (Fig. 2.7), are presented in Fig. 2.8.

These results (Fig. 2.8) are very important for the discussion of next problems:

a) Uneven sensitivity of magnetic slip displacement sensor, because (Fig. 2.8)

$$\Delta U_1(\Delta\alpha_1) >> \Delta U_2(\Delta\alpha_2) \tag{2.6}$$

for

$$\Delta\alpha_1 = \Delta\alpha_2, \tag{2.7}$$

as well as the methods for increasing sensitivity in the zone $\Delta U_2(\Delta\alpha_2)$;

b) The impact of an accuracy of synthesized mathematical model of magnetic SDS to its sensitivity and to value of maximal trial movement of intelligent robot;

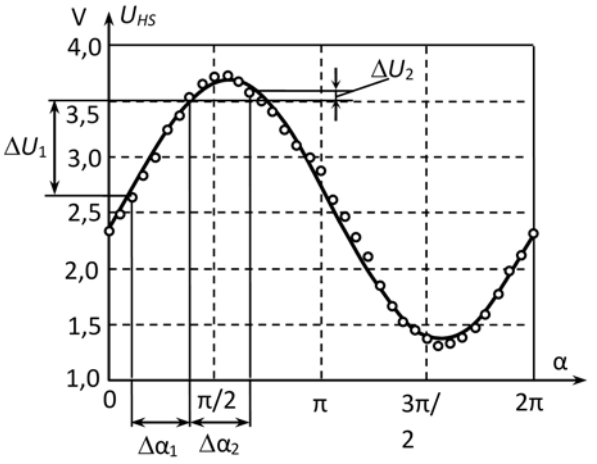


Fig. 2.8. Experimental and modelling results for magnetic SDS, presented in Fig. 2.7.

c) The implementation of developed magnetic SDS in modern robotic systems for automation of transport operations with manipulated objects, which parameters can be changed in real time of intelligent robot functioning.

Let's consider another structure and mathematical model of the developed SDS with magnetic sensitive element [18] which can detect the rod's angular deviation appearing at the object's slippage (Fig. 2.9).

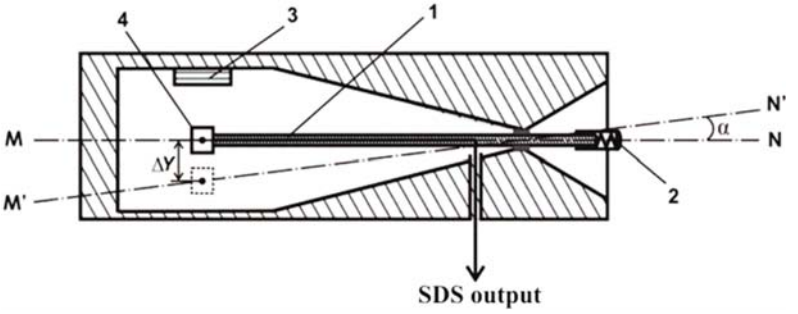


Fig. 2.9. SDS based on magnetic sensitive element's deviation:
1 – bar; 2 - head; 3 - permanent magnet; 4 – Hall sensor.

Mathematical model $U = f(\alpha)$ can be used to determine the sensitivity of the SDS and to find the minimal possible amplitudes of robot's trial motions.

To construct the mathematical models, consider the magnetic system comprising a prismatic magnet with dimensions $c \times d \times l$, which is set to ferromagnetic plane with infinite permeability $\mu = \infty$ (Fig. 2.10), where: c - width, d - length, and l - height of magnet, $(d >> l)$. The point $P(X_P, Y_P)$ is the observation point, which is located on the vertical axis and can change its position relative to the horizontal axis Ox or vertical axis Oy. Hall sensor with a linear static characteristic is located at the observation point P.

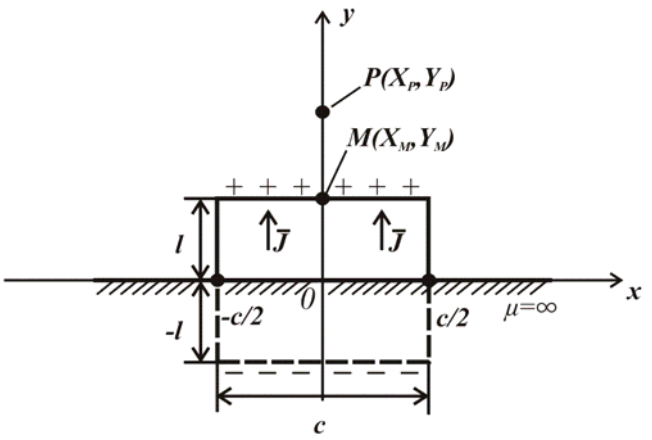


Fig. 2.10. Model of magnetic sensitive element

Let's form the mathematic model for the determination of the magnetic induction B and the output voltage $U_{out}(P)$ of the Hall sensor in relation to an arbitrary position of the observation point P under the surface of the magnet. The value of magnetic induction (outside the magnet volume) is

$$\overline{B} = \mu_0 \overline{H}, \tag{2.8}$$

where μ_0 is the magnetic constant and $\overline{H}$ is the vector of magnetic field strength.

In the middle of the magnet - the value of magnetic induction is determined by the dependence

$$\overline{B} = \mu_0 \left(\overline{J} + \overline{H} \right), \tag{2.9}$$

where J is the magnetization value:

$$J = J_0 + \chi H, \tag{2.10}$$

where χ is the magnetic susceptibility and J_0 is a residual magnetization value.

The permanent magnet can be represented [19–21] as a simulation model of the surface magnetic charges that are evenly distributed across the magnet pole faces with the surface density J_T.

Thus, a *y*-component of the magnetic field strength H_y of magnetic charges can be calculated as [3, 4, 7]:

$$H_y = -\frac{J_T}{2\pi}\left[\begin{array}{l}\left(arctg\dfrac{X_P + c/2}{l - Y_p} - arctg\dfrac{X_P - c/2}{l - Y_p}\right) - \\ -\left(arctg\dfrac{X_P + c/2}{-l - Y_p} - arctg\dfrac{X_P - c/2}{-l - Y_p}\right)\end{array}\right] \tag{2.11}$$

and *y*-component of magnetic induction B_y can be presented as:

$$B_y = -\frac{J_T \mu_0}{2\pi}\left[\begin{array}{l}\left(arctg\dfrac{X_P + c/2}{l - Y_p} - arctg\dfrac{X_P - c/2}{l - Y_p}\right) - \\ -\left(arctg\dfrac{X_P + c/2}{-l - Y_p} - arctg\dfrac{X_P - c/2}{-l - Y_p}\right)\end{array}\right]. \tag{2.12}$$

To determine the parameter J_T, it is necessary to measure the induction of the center of pole faces $B_y\big|_{x=0,y=l+}$. Value $(y = l+)$ indicates that the measurement of B_{mes} is conducted outside the volume of magnet. The value of magnetic induction at a point with the same coordinates on the inside of the pole faces can be considered equal to the value of induction from the outside pole faces, by virtue of the continuity of the magnetic flux and lines of magnetic induction, namely:

$$B_y\big|_{x=0,y=l+} = B_y\big|_{x=0,y=l-}. \tag{2.13}$$

So, we can write:

$$B_y\Big|_{\substack{x=0\\y=l+}} = B_{mes} = \mu_0 J_T + \mu_0 H_y\Big|_{\substack{x=0\\y=l-}}, \tag{2.14}$$

where B_{mes} is the value of magnetic induction measured at the geometric center of the top pole faces of the prismatic magnet.

On the basis of (2.11) we can obtain:

$$B_{mes} = \mu_0 J_T \left[1 - \lim_{Y_P \to l-} \frac{2}{2\pi} \left(arctg \frac{c/2}{l - Y_p} + arctg \frac{c/2}{l + Y_p} \right) \right] = \\ = \mu_0 J_T \left(\frac{1}{2} + \frac{1}{\pi} arctg \frac{c}{4l} \right), \tag{2.15}$$

$$J_T = \frac{2\pi B_{mes}}{\mu_0 \left(\pi + 2arctg \frac{c}{4l} \right)}. \tag{2.16}$$

For *y*-component of magnetic induction $B_y(P)$ at the observation point P the following expression was obtained:

$$B_y(P) = -\frac{B_{mes}}{\left(\pi + 2arctg \frac{c}{4l} \right)} \left[\left(arctg \frac{X_P + c/2}{l - Y_p} - arctg \frac{X_P - c/2}{l - Y_p} \right) - \right. \\ \left. - \left(arctg \frac{X_P + c/2}{-l - Y_p} - arctg \frac{X_P - c/2}{-l - Y_p} \right) \right]. \tag{2.17}$$

2.3.2. The Simulation Results for Magnetic SDS Based on Sensitive Rod's Deviation

For the analysis of the existing mathematical model (2.17), let's calculate the value of magnetic induction [3, 4, 7] on the surface of the magnet (Barium Ferrite) with parameters of c = 0.02 m, d = 0.08 m, l = 0.014 m and a value of magnetic induction $B_{mes} = 40$ mT (value measured at the geometric center of the upper limit of the magnet).

The simulation results for magnetic induction are represented as $B_y = f_i(X_P), i = 1,2,3$ above the magnet for various values of the height Y_P of the observation point $P(X_P, Y_P)$, where indicated:

f_1 - for $B_y\Big|_{\substack{x \in [-20;20] mm \\ y = l + 1 mm}}$, f_2 - for $B_y\Big|_{\substack{x \in [-20;20] mm \\ y = l + 5 mm}}$ and f_3 – for $B_y\Big|_{\substack{x \in [-20;20] mm \\ y = l + 20 mm}}$.

As can be seen from the Fig. 2.11, magnetic induction $B_y = f_1(X_P)$ above the surface of the magnet is practically constant for the coordinate $X_P \in [-5;5]$ mm, which is half of the corresponding size of magnet. If the distance from the observation point P and magnet increasing ($f_2(X_P), f_3(X_P)$ in Fig. 2.11), the curve shape changes become more gentle, with a pronounced peak above the geometric center of the top pole faces of the prismatic magnet (at the point $X_P = 0$). For the Hall sensor (Fig. 2.8) in the general case the

dependence of output voltage $U_{out}(P)$ on the magnitude of the magnetic induction B_y is defined as:

$$U_{out}(P) = U_C + kB_y(P), \tag{2.18}$$

where k is the correction factor, that depends on the type of Hall sensor and U_C is constant component of the Hall sensor output voltage.

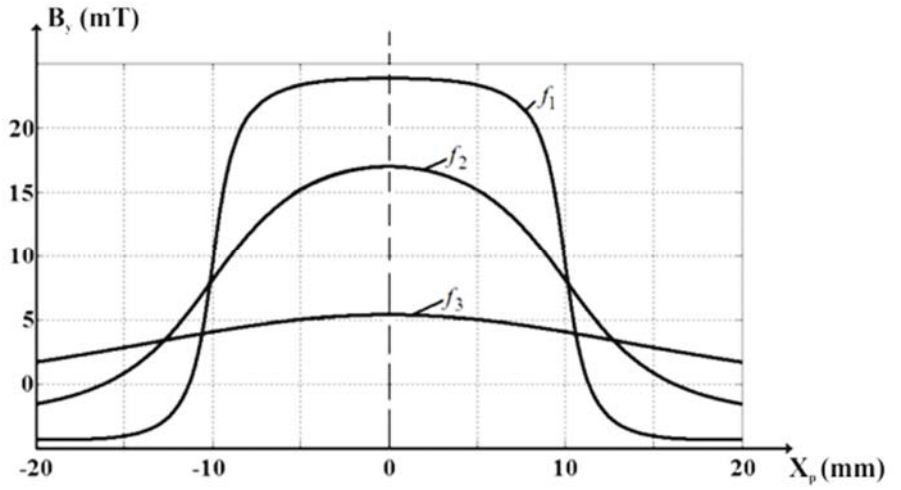


Fig. 2.11. Simulation results for $B_y(P)$ based on the mathematical model (2.17)

Authors present the mathematical model of the Hall sensor output voltage $U_{out}(Y_P)$ at its vertical displacement above the geometric center of the top pole faces of magnet ($X_P = 0$):

$$U_{out}(Y_P) = 2{,}5 + 7{,}4 \cdot 10^{-3}\left(arctg\frac{0{,}01}{0{,}014 - Y_P} + arctg\frac{0{,}01}{0{,}014 + Y_P}\right). \tag{2.19}$$

The comparative results for the dependences $U_{out}(Y_p)$, $U_E(Y_p)$ and $U_R(Y_p)$ are presented in Fig. 2.12, where $U_{out}(Y_p)$ was calculated using mathematic model (2.19), $U_E(Y_p)$ are the experimental results according to [22] and $U_R(Y_p)$ is nonlinear regressive model according to [23]. Comparative analysis (Fig. 2.12) of the developed mathematical model $U_{out}(Y_p)$ and experimental results $U_E(Y_p)$ confirms the correctness and adequacy of the synthesized models (2.11), (2.17) - (2.19).

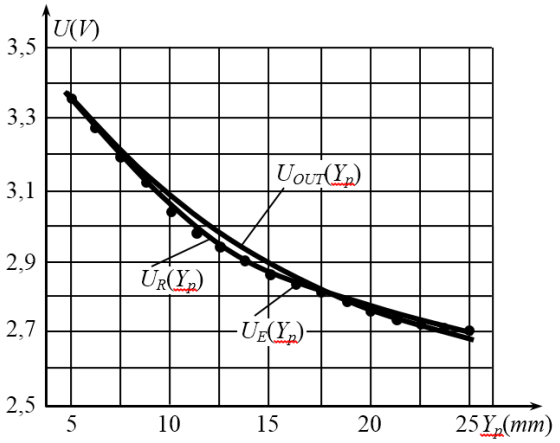


Fig. 2.12. Modeling and experimental results.

2.4. Slip Displacement Sensors with Capacitive Sensing Components

2.4.1. The Analysis of Modern Capacitive Slip Displacement Sensors

The modified method is founded on the measurement of the sensitive element angular deviation occurring at the object slippage (Fig. 2.13). The mathematical models of such slip displacement sensors with a measurement of changeable capacity are presented in [24–28].

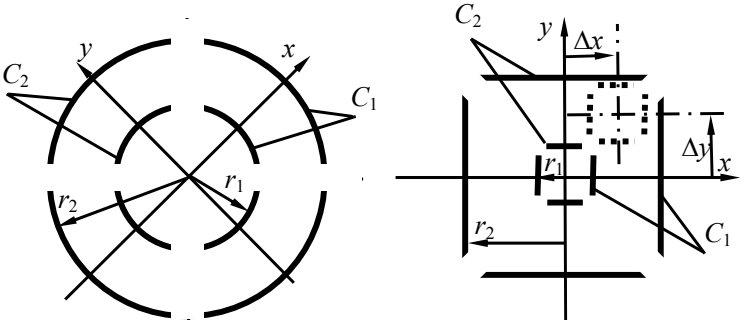


Fig. 2.13. Two different models of slip displacement sensor based on the measurement of changeable capacity C_1 and C_2: (a) cylinder model; (b) rectangular model.

The design structure of capacitive slip displacement sensor (with deviating rod as sensitive element [8]) is presented in Fig. 2.14.

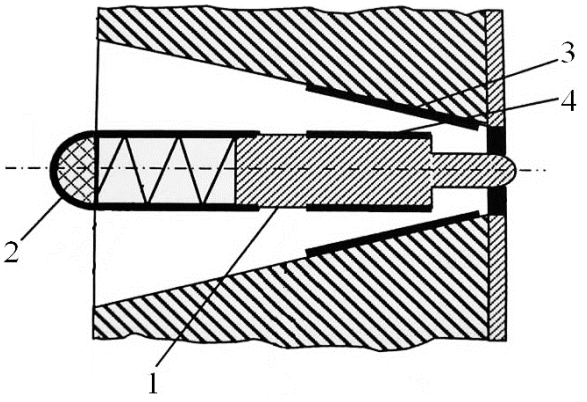


Fig. 2.14. Capacitive slip displacement sensor based on rod's deviation:
1- deviating rod (sensitive element); 2 – elastic tip; 3, 4 – capacitor plates.

The number of capacitor's sections may be various. The example of capacitive SDS with four capacitor's sections is represented in Fig. 2.15, where each capacitor's section (1, 2, 3 and 4) has separate value of corresponding capacity C_1, C_2, C_3, C_4 after the deviation of rod 5 during the slippage of manipulated object.

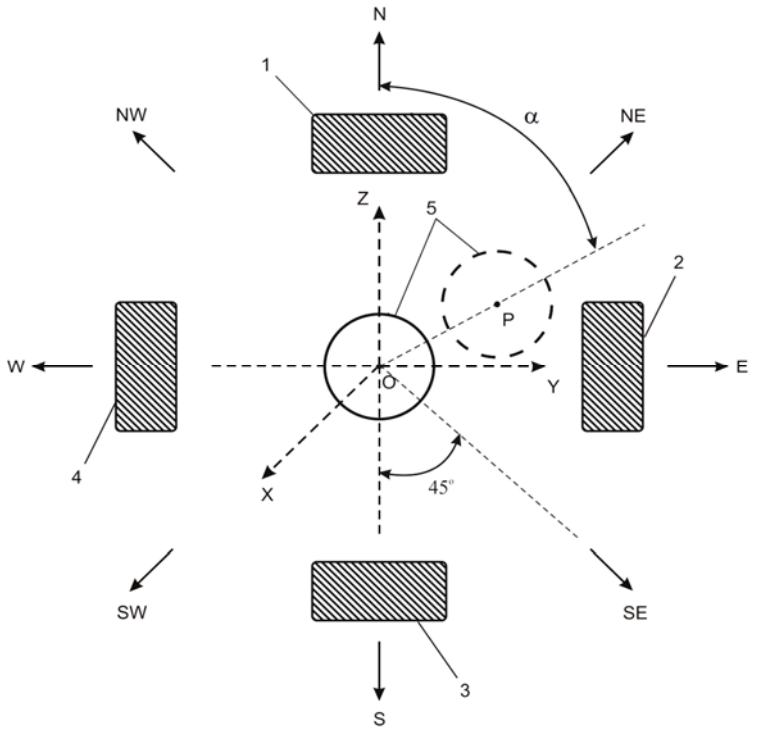


Fig. 2.15. Four-components capacitive slip displacement sensor.

2.4.2. Fuzzy-logic Approach for the Identification of the Slip Displacement Direction

During the process of manipulated object gripping the slip displacement direction is the same as a gravity force direction. The identification of slippage direction is very important for situations when robot gripper contacts any obstacle [3, 29]. The obstacles can appear randomly in the working area of intelligent robot. As result of the collision between the robot gripper and the obstacle, object's slippage may appear if current value of clamping force is not enough for reliable fixation of the manipulated object between fingers of robot gripper. In such cases the slippage directions will be different comparing with gravity force direction. Such slippage direction depends on the robot's gripper trajectory and coordinates of the obstacles in the intelligent robot's working area.

Let's consider proposed fuzzy approach for identification of slippage direction based on multi-components capacitive slip displacement sensor [29], presented in Fig. 2.15.

The following notations are used in Fig. 2.15: 1, 2, 3 and 4 – capacitors C_1, C_2, C_3, C_4; 5 – deviating rod; point O – initial position of the rod 5 before object's slippage; point P - final position of the rod 5 after object's slippage; *N* (North), *NE* (North-East), *E* (East), *SE* (South-East), *S* (South), *SW* (South-West), *W* (West), *NW* (North-West) – object's slippage directions; *OXYZ* – coordinates' system for robot's working space; *OX* – directions for creating clamping force between gripper fingers; α - slippage direction (in degrees).

If manipulated object has contact with any obstacle and if object's slippage appears, then the rod 5 (Fig. 2.15) deviates from the initial position and all values of capacities C_1, C_2, C_3, C_4 changes. Depends on the slippage direction some capacities of the slip displacement sensors may be increased and rest of them – may be decreased or stable.

It is propose to develop the fuzzy system using fuzzy inference engine of Mamdany-type [30, 31] for identification of slippage direction:

$$\alpha = f_{FS}\left(C_1, C_2, C_3, C_4\right). \tag{2.20}$$

The proposed fuzzy system $f_{FS}\left(\cdot\right)$ has for input C_1, C_2, C_3, C_4 and one output α signals.

Five linguistic terms are used for each input signal of developed fuzzy system as corresponding fuzzy sets: **L** – low; **LM** – low/middle; **M** – middle; **MH** – middle/high; **H** - high.

Nine linguistic terms are used for output signal of developed fuzzy system as corresponding fuzzy sets: **VL** – very low; **L** – low; **HL** – higher than low; **LM** – less than middle; **M** – middle; **HM** – higher than middle; **LH** – less than high; **H** – high; **VH** – very high.

Authors use triangular shape of membership functions for all linguistic terms of input and output signals. The output signals U_i, $U_i \in [U_{i\min}, U_{i\max}], (i = 1...4)$ of the four-component measurement system of capacitive SDS (Fig. 2.15), which corresponds to the values of capacities C_i, $C_i \in [C_{i\min}, C_{i\max}], (i = 1...4)$, are, at the same time, the input signals for proposed fuzzy system which provides the identification of slippage direction.

Triangular [32] linguistic terms (Fig. 2.16) of input signals U_i, $U_i \in [U_{i\min}, U_{i\max}], (i = 1...4)$ for proposed fuzzy system can be presented as following fuzzy numbers:

$$L = \left(l_1^{in}, l_0^{in}, l_2^{in}\right) = \left(U_{\min}, U_{\min}, U_{\min} + \left(U_{\max} - U_{\min}\right)/4\right); \tag{2.21}$$

$$LM = \left(lm_1^{in}, lm_0^{in}, lm_2^{in}\right) = \left(U_{\min}, U_{\min} + \left(U_{\max} - U_{\min}\right)/4, U_{\min} + \left(U_{\max} - U_{\min}\right)/2\right); \tag{2.22}$$

$$M = \left(m_1^{in}, m_0^{in}, m_2^{in}\right) = \begin{pmatrix} U_{\min} + \left(U_{\max} - U_{\min}\right)/4, \\ U_{\min} + \left(U_{\max} - U_{\min}\right)/2, \\ U_{\min} + \frac{3}{4}\left(U_{\max} - U_{\min}\right) \end{pmatrix}; \tag{2.23}$$

$$MH = \left(mh_1^{in}, mh_0^{in}, mh_2^{in}\right) = \left(U_{\min} + \left(U_{\max} - U_{\min}\right)/2, U_{\min} + \frac{3}{4}\left(U_{\max} - U_{\min}\right), U_{\max}\right); \tag{2.24}$$

$$H = \left(h_1^{in}, h_0^{in}, h_2^{in}\right) = \left(U_{\min} + \frac{3}{4}\left(U_{\max} - U_{\min}\right), U_{\max}, U_{\max}\right). \tag{2.25}$$

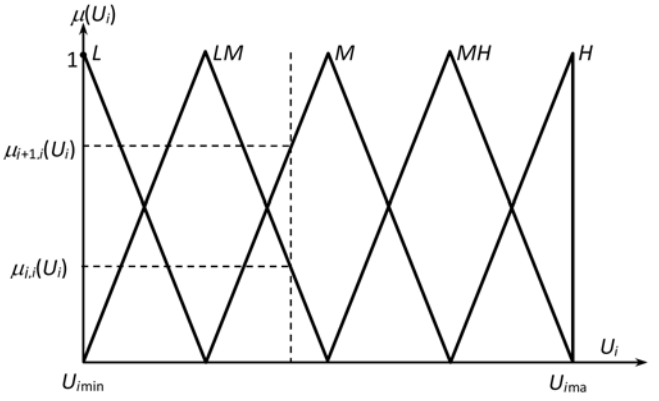


Fig. 2.16. Fuzzy sets (linguistic terms) for input signals of proposed fuzzy system.

For example, direct mathematical model for triangular fuzzy sets **M** can be presented in the following way by its triangular membership function $\mu_M(U_i)$:

$$\mu_M(U_i)=\begin{cases}0,\text{for } (U_{\min}>U_i)\cup(U_i>U_{\max});\\ \dfrac{U_i-U_{\min}+(U_{\max}-U_{\min})/4}{(U_{\max}-U_{\min})/4},\\ \quad \text{for } (U_{\min}+(U_{\max}-U_{\min})/4)\le U_i<(U_{\min}+(U_{\max}-U_{\min})/2);\\ 1,\text{for } U_i=(U_{\min}+(U_{\max}-U_{\min})/2);\\ \dfrac{U_{\min}+\frac{3}{4}(U_{\max}-U_{\min})-U_i}{(U_{\max}-U_{\min})/4},\\ \quad \text{for } (U_{\min}+(U_{\max}-U_{\min})/2)<U_i\le\left(U_{\min}+\frac{3}{4}(U_{\max}-U_{\min})\right).\end{cases} \tag{2.26}$$

Thus, any input signal $U_i,(i=1...4)$ simultaneously activates only two neighboring membership functions (linguistic terms) with satisfying well-known condition

$$\mu_{j,i}(U_i)+\mu_{j+1,i}(U_i)=1\ . \tag{2.27}$$

At the same time, the triangular linguistic terms (Fig. 2.17) for output signal α can be presented as following triangular fuzzy numbers:

$$VL=(vl_1^{out},vl_0^{out},vl_2^{out})=(0,0,45); \tag{2.28}$$

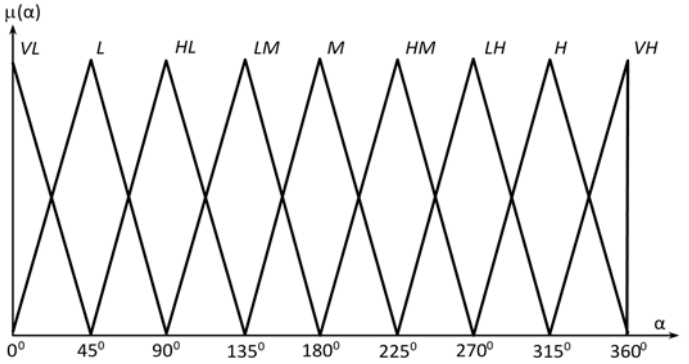


Fig. 2.17. Fuzzy sets (linguistic terms) for output signal α of proposed fuzzy system

$$\alpha=f_{FS}(C_1,C_2,C_3,C_4)$$

$$L=(l_1^{out},l_0^{out},l_2^{out})=(0,45,90); \tag{2.29}$$

$$HL = \left(hl_1^{out}, hl_0^{out}, hl_2^{out}\right) = (45,90,135); \tag{2.30}$$

$$LM = \left(lm_1^{out}, lm_0^{out}, lm_2^{out}\right) = (90,135,180); \tag{2.31}$$

$$M = \left(m_1^{out}, m_0^{out}, m_2^{out}\right) = (135,180,225); \tag{2.32}$$

$$HM = \left(hm_1^{out}, hm_0^{out}, hm_2^{out}\right) = (180,225,270); \tag{2.33}$$

$$LH = \left(lh_1^{out}, lh_0^{out}, lh_2^{out}\right) = (225,270,315); \tag{2.34}$$

$$H = \left(h_1^{out}, h_0^{out}, h_2^{out}\right) = (270,315,360); \tag{2.35}$$

$$VH = \left(vh_1^{out}, vh_0^{out}, vh_2^{out}\right) = (315,360,360). \tag{2.36}$$

For example, direct mathematical models for triangular fuzzy sets **L** and **LH** can be presented in the following way by their triangular membership functions $\mu_L(\alpha)$ and $\mu_{LH}(\alpha)$, respectively:

$$\mu_L(\alpha) = \begin{cases} 0, & \text{for } (0 > \alpha) \cup (\alpha > 45) \\ \dfrac{\alpha}{45}, & \text{for } 0 \le \alpha < 45 \\ 1, & \text{for } \alpha = 45 \\ \dfrac{90-\alpha}{45}, & \text{for } 45 < \alpha \le 90 \end{cases} \tag{2.37}$$

$$\mu_{LH}(\alpha) = \begin{cases} 0, & \text{for } (225 > \alpha) \cup (\alpha > 315) \\ \dfrac{\alpha-225}{45}, & \text{for } 225 \le \alpha < 270 \\ 1, & \text{for } \alpha = 270 \\ \dfrac{315-\alpha}{45}, & \text{for } 270 < \alpha \le 315 \end{cases} \tag{2.38}$$

The part of fuzzy rules' base (8 fuzzy rules) of proposed fuzzy system is presented in Table 2.1. Each fuzzy rule in rules base has the following structure:

IF (Condition - Antecedent)**, THEN** (Result – Consequent), (2.39)

which deals with the determination of the dependence (2.20) between slip displacement direction α for slipping manipulated object and values of signals $U_i, (i = 1...4)$, which are output signals of corresponding capacitor sensitive components and simultaneously – input signals for designed fuzzy system.

Table 2.1. Selected eight rules of fuzzy Rule Base.

№ of rule	**Conditions (antecedents)**				**RESULT (consequent)**
	U_1	U_2	U_3	U_4	α_k, degrees
1	H	M	L	M	VL
2	MH	MH	LM	LM	L
3	M	H	M	L	HL
4	LM	MH	MH	LM	LM
5	L	M	H	M	M
6	LM	LM	MH	MH	HM
7	M	L	M	H	LH
8	MH	LM	LM	MH	H

Let's consider in more details all notations, which are used in Table 2.1, in particular: $U_i, (i = 1...4)$ is output of i-th sensor component of multi-component measurement system; α_k is output signal (consequent) of k-th fuzzy rule of fuzzy inference engine, $k = 1...N$ (where N - is general quantity of rules in fuzzy rule base, maximal quantity $N = N_{\max}$ of the fuzzy rules is $N_{\max} = 625$ for our fuzzy system, when we use 5 linguistic terms $(j = 1...5)$ in antecedents of fuzzy rules for 4 input signals $U_i, (i = 1...4)$. The antecedents of fuzzy rule base are formed using linguistic terms, presented in Fig. 2.16, and consequents – based on linguistic terms, presented in Fig. 2.17.

The structure (2.39) of the fuzzy rules (Table 2.1) can be presented in the following way (for example, for 3rd fuzzy rule, $(k = 3)$):

«**IF** (U_1 = «M» **AND** U_2 = «H» **AND** U_3 = «M» **AND** U_4 = «L»),
THEN (α_3 = «HL»)».

The structure of considered Mamdani-type [30, 31] fuzzy system for the identification of slippage direction consists of Rule Base (Table 2.1) and several very important components, which provide efficient stages of fuzzy information processing for the realization of fuzzy inference engine, in particular:

- A fuzzification of input signals based on the linguistic terms, presented in Fig. 2.16;

- An aggregation of each fuzzy rule in Rule Base and a determination of corresponding consequent's membership level as output's contribution of each rule;

- An activation of fuzzy rules with identical consequents;

- An accumulation of aggregated consequents, based on the linguistic terms, presented in Fig. 2.17, for all activated rules and formation of resulting fuzzy set with membership

function $\mu_{res}(\alpha)$, which corresponds to fuzzy output signal $\tilde{\alpha}$ of the developed fuzzy system;

- A defuzzification of the resulting fuzzy set $\mu_{res}(\alpha)$ for the its transformation to crisp number α^*, which corresponds to the direction of manipulated object's slippage.

One of the most common method for defuzzification of the resulting fuzzy set $\mu_{res}(\alpha)$ is gravity center method, which can be realized in the following way:

$$\alpha^* = \frac{\sum_{j=1}^{m}\left(\alpha_{\min} + (j-1)\Delta\alpha\right)\mu_{res}\left(\alpha_{\min} + (j-1)\Delta\alpha\right)}{\sum_{j=1}^{m}\mu_{res}\left(\alpha_{\min} + (j-1)\Delta\alpha\right)}, \tag{2.40}$$

where $\alpha_{\min}$ is the left limit value of parameter α in the base of the resulting fuzzy set $\mu_{res}(\alpha)$; $\Delta\alpha$ is a step of discretization; $j, (j = 1...m)$ - a number of iteration.

Maximal number of iterations m can be calculated as

$$m = \frac{\alpha_{\max} - \alpha_{\min}}{\Delta\alpha}, \tag{2.41}$$

where $\alpha_{\max}$ is the right limit value of parameter α in the base of the resulting fuzzy set $\mu_{res}(\alpha)$, $\alpha \in [\alpha_{\min}, \alpha_{\max}]$.

In the case, if the base of the resulting fuzzy set $\mu_{res}(\alpha)$ exists at full range of parameter α (Fig. 2.17), then $\alpha_{\min} = 0; \alpha_{\max} = 360$.

It is possible to increase the efficiency of developed fuzzy system by its optimization and learning using evolutionary optimisation algorithms [31, 33-39], in particular, genetic algorithms, neuro networks, biogeography-based optimisation approach and others, as well as a set of experimental or modelling results.

According to Mamdany-fuzzy-information-processing algorithm, the output α^* (for fixed Rule Base) of developed fuzzy system depends on the input signals, weight coefficients $w_s, (s = 1...N)$ of corresponding fuzzy rules and all parameters of linguistic terms:

$$\alpha^* = F\left(\bar{W}, \bar{X}_1^{in}, \bar{X}_0^{in}, \bar{X}_2^{in}, \bar{X}_1^{out}, \bar{X}_0^{out}, \bar{X}_2^{out}\right), \tag{2.42}$$

where $\bar{W} = (w_1, w_2, ..., w_{N-1}, w_N)$ is the vector of weight coefficients;

$\bar{X}_1^{in}, \bar{X}_0^{in}, \bar{X}_2^{in}$ are vectors of adjusting parameters of linguistic terms (Fig. 2.16) for fuzzification of input signals (in antecedents of fuzzy rules);

$\bar{X}_1^{out}, \bar{X}_0^{out}, \bar{X}_2^{out}$ are vectors of adjusting parameters of linguistic terms (Fig. 2.17) for evaluation of output signals (in consequents of fuzzy rules);

For solving the multi-parametrical optimisation problem it is necessary to minimize such criterion as

$$J = \sum_{k=1}^{n} \left(\alpha_{ex}(k) - \alpha^*(k) \right)^2 = \\ = \sum_{k=1}^{n} \left(\alpha_{ex}(k) - F\left(\bar{W}, \bar{X}_1^{in}, \bar{X}_0^{in}, \bar{X}_2^{in}, \bar{X}_1^{out}, \bar{X}_0^{out}, \bar{X}_2^{out}, k \right) \right)^2 \Rightarrow \underset{W,X}{Min} \quad , \tag{2.43}$$

where $\alpha_{ex}(k)$ is the experimental output value of fuzzy system for *k*-th set of input data

$$\bar{\alpha}_{ex} = \left(\alpha_{ex}(1), \alpha_{ex}(2), ..., \alpha_{ex}(n-1), \alpha_{ex}(n) \right), k = 1...n \ ;$$

$$\bar{X}_1^{in} = \left(l_1^{in}, lm_1^{in}, m_1^{in}, mh_1^{in}, h_1^{in} \right);$$

$$\bar{X}_0^{in} = \left(l_0^{in}, lm_0^{in}, m_0^{in}, mh_0^{in}, h_0^{in} \right);$$

$$\bar{X}_2^{in} = \left(l_2^{in}, lm_2^{in}, m_2^{in}, mh_2^{in}, h_2^{in} \right);$$

$$\bar{X}_1^{out} = \left(vl_1^{out}, l_1^{out}, hl_1^{out}, lm_1^{out}, m_1^{out}, hm_1^{out}, lh_1^{out}, h_1^{out}, vh_1^{out} \right);$$

$$\bar{X}_0^{out} = \left(vl_0^{out}, l_0^{out}, hl_0^{out}, lm_0^{out}, m_0^{out}, hm_0^{out}, lh_0^{out}, h_0^{out}, vh_0^{out} \right);$$

$$\bar{X}_2^{out} = \left(vl_2^{out}, l_2^{out}, hl_2^{out}, lm_2^{out}, m_2^{out}, hm_2^{out}, lh_2^{out}, h_2^{out}, vh_2^{out} \right);$$

2.5. Computerized System for Intelligent Robot's Control Based on Tactile and Slip Displacement Sensors

In many cases, it is very important and necessary to fix the fingers of robot gripper in desired position on the manipulated object. The current position can be shifted according to desired position during slip displacement process for manipulated object. Authors developed (Fig. 2.18) the computerized information-control system, which allows (in automatic mode) fixing the fingers in desired (determinate) position on the manipulated object after identification of value of object mass (by intelligent robot) based on signals of tactile and slip displacement sensors.

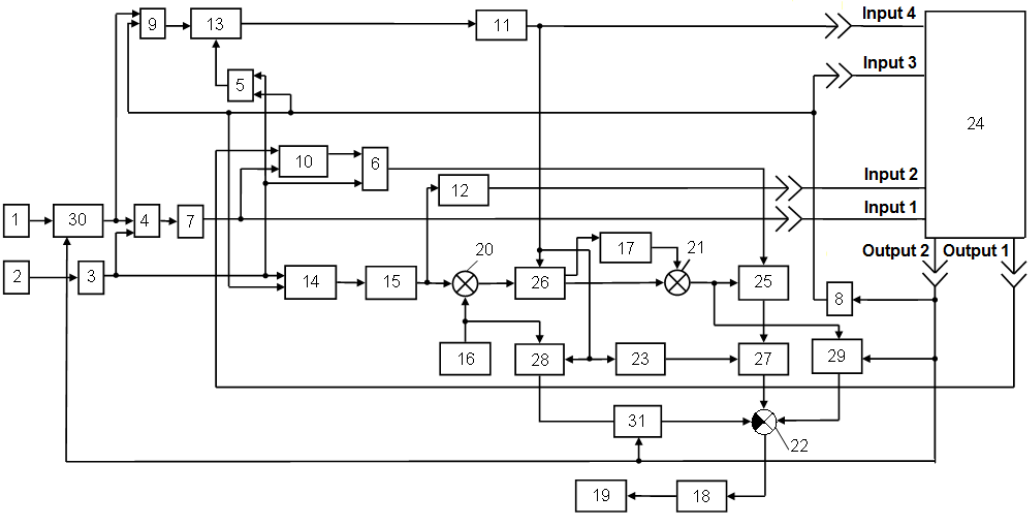


Fig. 2.18. The structure of computerized information-control system of intelligent robot.

The structure and main components of the developed computerized information-control system of intelligent robot as multi-sensor system [28, 40] are represented in Fig. 2.10, where components are marked as: 1- tactile sensor; 2 - slip displacement sensor; 3, 18 – amplifier; 4, 5, 6- element OR; 7, 8 – block of delay; 9, 10 - RS-triggers; 11, 12 - threshold element; 13 – integrator; 14 –register; 15 - digital-analog converter; 16 - reference voltage source; 17 - voltage divider; 19 – a drive of fingers; 20, 21, 22 – adders; 23 - NOT element; 24 - computer control unit; 25, 26, 27, 28, 29, 30, 31 - control switch.

As a tactile sensor (Fig. 2.18) it is possible to use any sensor [28], which is fixed in one of the gripper fingers of intelligent robot and can form a single discrete output signal corresponding to the presence (logic "1") or not (logic "0") contact of manipulated object with robot fingers. The output of slip displacement sensor should be in the form of logic "1" (if displacement is appeared) or logic "0" (if displacement is absent) [40].

Different types [3, 4, 7, 8, 9, 41, 42, 43, 44, 45] of tactile and slip displacement sensors can be used as sensitive components of proposed computerized information-control system of intelligent robot. The computer control unit (Fig. 2.10) can be realized based on the digital components [46], in particular, very perspective is its realization on the FPGA microprocessor base with possibility for reconfiguration [47, 48]. In this case, the multi-functional peculiarities of computerized information-control system may be expanded for various types of intelligent robots with different planned missions in different working environments.

2.6. Conclusions

The methods of the slip displacement signal detection considered in the present chapter furnish an explanation of the main detection principles and allow robot sensing systems obtain wide capabilities.

Authors developed wide variety of SDS schemes and mathematical models with capacitive, magnetic and light-reflecting sensitive elements with improved characteristics (accuracy, time response, sensitivity).

The results of the research are applicable in the automatic adjustment of clamping force of robot's gripper and correction of robot motion algorithms in real time.

The methods introduced by authors may be also used in random functioning conditions, in settling the problems of the automatic assembly, sorting, patterns and images recognition in the working zones of robots.

Proposed fuzzy system can be used for identification of slip displacement signal and object slippage's direction as well as for correction of trajectory of robot's gripper movement.

Advanced sensors, their models and structure of computerized information system can be used for synthesis of intelligent robot control systems [7, 37, 49, 50] with new features and for solving orientation and control tasks when intelligent robot contacts with obstacles.

References

[1]. Y. P. Kondratenko, Robotics, Automation and Information Systems: Future Perspectives and Correlation with Culture, Sport and Life Science, in Decision Making and Knowledge Decision Support Systems, Lecture Notes in Economics and Mathematical Systems, Vol. 675, A. M. Gil-Lafuente, C. Zopounidis, Eds., *Springer International Publishing Switzerland*, 2015, pp. 43–56.

[2]. H. Richter, D. Simon, Robust Tracking Control of a Prosthesis Test Robot, *ASME Journal of Dynamic System, Measurement, and Control*, Vol. 136, No. 3, paper 031011, February 2014.

[3]. Kondratenko Y. P., Kondratenko V. Y., Advanced Trends in Design of Slip Displacement Sensors for Intelligent Robots, in Advances in Intelligent Robotics and Collaboration Automation, Yuriy P. Kondratenko, Richard Duro (Eds.), Series on Automation, Control and Robotics, *River Publishing*, Denmark, 2015, pp. 167-191.

[4]. Kondratenko Y. P., Kondratenko V. Y., Klymenko L. P., Kondratenko G. V., Shvets E. A., Sensing of Intelligent Robots Based on Applications of Tactile Slip Displacement Sensors. *Sensors & Transducers*, Vol. 187, Issue 4, April 2015, pp. 29-38.

[5]. M. Ueda, K. Iwata, H. Shingu, Tactile sensors for an industrial robot to detect a slip, in *Proceedings of the 2nd Int. Symp. on Industrial Robots,* Chicago, USA, 1972, pp. 63-76.

[6]. M. Ueda, K. Iwata, Adaptive grasping operation of an industrial robot, in *Proceedings of the 3rd Int. Symp. Ind. Robots,* Zurich, 1973, pp. 301-310.

[7]. Y. P. Kondratenko, L. P. Klymenko, V. Y. Kondratenko, G. V. Kondratenko, E. A. Shvets, Slip displacement sensors for intelligent robots: Solutions and models, in *Proceedings of the*

7th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems (*IDAACS'13),* Vol. 2, Berlin, Germany, 12-14 September 2013, pp. 861–866.

[8]. Y. P. Kondratenko, Measurements methods for slip displacement signal registration, in *Proceedings of the Intern. Symposium on Measurement Technology and Intelligent Instruments,* Chongqing-Wuhan, China, Published by SPIE, USA, 1993, pp. 1451-1461.

[9]. Y. P. Kondratenko, X. Y. Huag, Slip displacement sensors of robotic assembly system, in *Proceedings of the 10th Intern. Conference on Assembly Automation,* 23-25 October, Kanazava, Japan, IFS Publications, Kempston, United Kingdom, 1989, pp. 429-436.

[10]. Y. P. Kondratenko, E. A. Shvets, O. S. Shyshkin, Modern Sensor Systems of Intelligent Robots Based on the Slip Displacement Signal Detection, in *Annals of DAAAM for 2007 & in Proceedings of the 18th International DAAAM Symposium*, Vienna, DAAAM International, Austria, 2007, pp. 381-382.

[11]. Y. P. Kondratenko, V. Y. Kondratenko, E. A. Shvets, O. S. Shyshkin, Adaptive Gripper Devices for Robotic Systems, in *Proceeding of the Intern. Scientific-and-Technological Congress Mechatronics and Robotics* (M&R'07), October 2-5, 2007, Saint-Petersburg, Russia, 2008, pp. 99-105.

[12]. F. Xu, Y. Ge, Y. Yu, J. Ding, T. Ju, S. Li, The design of novel flexible tactile sensor based on pressure-conductive rubber, *Sensors and Transducers*, Vol. 124, Issue 1, January 2011, pp. 19-29.

[13]. Q. Sheng, G. Y. Xu, G. Liu, Design of PZT Micro-displacement acquisition system, *Sensors and Transducers*, Vol. 182, Issue 11, November 2014, pp. 119-124.

[14]. Y. P. Kondratenko, O. S. Shyshkin, V. Y. Kondratenko, Device for recording the slip signal, *Patent No. 14566*, Ukraine, 2006. (in Ukrainian).

[15]. Y. P. Kondratenko, A. V. Kuzmichev, Y. Z. Yang, Robot control system using slip displacement signal for algorithm correction, in: *Selected papers from the 3rd IFAC/IFIP/IMACS Symposium on ROBOT CONTROL (SYROCO'91)*, Vienna, Austria, 1991, pp. 463-469.

[16]. Y. P. Kondratenko, E. A. Shvets, Adaptive gripper of intelligent robot, *Patent No. 14569,* Ukraine, 2006. (in Ukrainian).

[17]. Y. P. Kondratenko, E. A. Shvets, Adaptive gripper of intelligent robot, *Patent No. 93878,* Ukraine, 2014. (in Ukrainian).

[18]. Y. P. Kondratenko, O. S. Shyshkin, V. Y. Kondratenko, Device for detection of slip displacement signal, *Patent No. 79155,* Ukraine, 2007. (in Ukrainian).

[19]. Y. M. Zaporozhets, Qualitative analysis of the characteristics of direct permanent magnets in magnetic systems with a gap, *Technical Electrodynamics*, 3, 1980, pp. 19-24. (in Russian).

[20]. Y. M. Zaporozhets, Y. P. Kondratenko, O. S. Shyshkin, Mathematical model of slip displacement sensor with registration of transversal constituents of magnetic field of sensing element, *Technical Electrodynamics*, 4, 2012, pp. 67-72. (in Ukrainian).

[21]. Y. M. Zaporozhets, Y. P. Kondratenko, O. S. Shyshkin, Three-dimensional mathematical model for calculating the magnetic induction in magnetic-sensitive system of slip displacement sensor, *Technical Electrodynamics*, № 5, 2008, pp. 76-79. (in Ukrainian).

[22]. Y. P. Kondratenko, O. S. Shyshkin, Experimental studies of the magnetic slip displacement sensors for adaptive robotic systems, *Trans. of the Odessa Polytechnic University*, Special Issue, Odessa, 2005, pp. 47-51. (in Russian).

[23]. Y. P. Kondratenko, O. S. Shyshkin, Synthesis of regression models of magnetic systems of slip displacement sensors, *Radioelectronic and Computer Systems*, 6, 25, Kharkov, 2007, pp. 210-215. (in Russian).

[24]. E. P. Reidemeister, L. K. Johnson, Capacitive acceleration sensor for vehicle applications, *Sensors and Actuators*, SP-1066, 1995, pp. 29-34.

[25]. Y. P. Kondratenko, I. L. Nazarova, Mathematical model of capacity sensor with conical configuration of sensitive element, *Trans. of the Donetsk National Technical University*, No. 11, 186, DNTU, Donetsk, pp. 186-191. (in Ukrainian).

[26]. M. I. Tiwana, A. Shashank, S. J. Redmond, N. H. Lovell. Characterization of a capacitive tactile shear sensor for application in robotic and upper limb prostheses, *Sensors and Actuators,* A-165, 2011, pp. 164-172.

[27]. Y. P. Kondratenko, I. L. Nazarova, Intelligent slip displacement sensor with microprocessor system of signal processing, *Technical News*, No. 1, Issue 31, No. 2, Issue 32, 2010, pp. 74-77. (in Ukrainian).

[28]. J. G. Rosha, S. Lanceros-Mendez, Three dimensional capacitive force sensors for tactile applications, in Sensors, Focus on Tactile, Force and Stress Sensors, Jose Gerardo Rocha and Senentxu Lanceros- Mendez (Eds.), December 2008, *I-Tech, Vienna, Austria*, pp. 143-162.

[29]. Y. P. Kondratenko, G. V. Kondratenko, V. Y. Kondratenko, I. L. Nazarova, Method of adaptation of compressive force of intelligent robot and identification of direction of working object sliding, *Patent No. 68117*, Ukraine, 2012. (in Ukrainian).

[30]. E. H. Mamdani, Application of fuzzy algorithm for control of a simple dynamic plant, in *Proceedings of the Institution of Electrical Engineers*, Vol. 121, No. 12, 1974, pp. 1585-1588. O. Cordón, A historical review of evolutionary learning methods for Mamdani-type fuzzy rule-based systems: Designing interpretable genetic fuzzy systems, *International Journal of Approximate Reasoning,* 52, No. 6, 2011, pp. 894-913.

[31]. A. Kaufmann, M. Gupta, Introduction to Fuzzy Arithmetic: Theory and Applications, *Van Nostrand Reinhold Company*, New York, 1985.

[32]. D. Simon, Evolutionary Optimization Algorithms: Biologically Inspired and Population-Based Approaches to Computer Intelligence, *John Wiley & Sons*, 2013.

[33]. D. Du, D. Simon, Complex System Optimization Using Biogeography-Based Optimization, *Mathematical Problems in Engineering*, Vol. 2013, 2013, Article ID 456232, 17 pages.

[34]. P. Lozovyy, G. Thomas, D. Simon. Biogeography-Based Optimization for Robot Controller Tuning, in Computational Modeling and Simulation of Intellect, B. Igelnik (Ed.), *IGI Global*, 2011, pp. 162-181.

[35]. D. Simon, M. Omran, M. Clerc. Linearized biogeography-based optimization with re-initialization and local search, *Information Sciences*, Vol. 267, May 2014, pp. 140–157.

[36]. G. Thomas, T. Wilmot, S. Szatmary, D. Simon, W. Smith. Evolutionary Optimization of Artificial Neural Networks for Prosthetic Knee Control, in Efficiency and Scalability Methods for Computational Intellect, B. Igelnik, J. Zurada (Eds.), *IGI Global,* Hershey, PA, Chapter 7, 2013, pp. 142–161.

[37]. A. P. Rotshtein, Intelligent technology for identification: fuzzy sets, genetic algorithms, neuro networks, *Universum-Vinnytsa*, Vinnytsa, Ukraine, 1999.

[38]. Turchenko, O. Osolinsky, V. Kochan, A. Sachenko, R. Tkachenko, V. Svyatnyy, M. Komar, Approach to Neural-based Identification of Multisensor Conversion Characteristic, in *Proceedings of the IEEE International Workshop on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS'09),* Rende, Italy, 21-23 September 2009, pp. 27-31.

[39]. Y. P. Kondratenko, V. Y. Kondratenko, Information-control system of adaptive robot, *Patent No. 42905*, Ukraine, 2009. (in Ukrainian).

[40]. Modern sensors, transducers and sensor networks. Book series: Advances in Sensors: Reviews, Vol. 1, Sergey Y. Yurish (Ed.), *International Frequency Sensor Association (IFSA) Publishing,* 2012.

[41]. M. T. Restivo, F. G. De Almeida, M. De Fátima Chouzal, Displacement Measurement. Book series: Measurement of Physical and Chemical Quantities, Vol. 2, *IFSA Publishing, S. L.,* 2014.

[42]. Sensors and Biosensors, MEMS Technologies and its Applications. Book series: Advances in Sensors: Reviews, Vol. 2, Sergey Y. Yurish (Ed.), *International Frequency Sensor Association (IFSA) Publishing,* 2013.
[43]. Sensors Web Portal (http://www.sensorsportal.com).
[44]. Sensors & Signals, Sergey Y. Yurish and Amin D. Malayeri (Eds.), *IFSA Publishing, S. L.,* 2014.
[45]. J. Drozd, A. Drozd, S. Antoshchuk, V. Kharchenko, Natural development of the resources in design and testing of the computer systems and their components, in *Proceedings of the 7th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS'13)*, Vol. 1, Berlin, Germany, 12-14 September 2013, pp. 233 – 237.
[46]. A. V. Palagin, V. N. Opanasenko, Reconfigurable-computing technology, *Cybernetics and Systems Analysis*, Vol. 43, Issue 5, 2007, pp. 675–686.
[47]. F. Chen, Q. Wu, G. Tao, B. Jiang, A Reconfiguration Control Scheme for a Quadrotor Helicopter via Combined Multiple Models, *International Journal of Advanced Robotic Systems*, 11, 122, 2014, pp. 1-10.
[48]. Y. P. Kondratenko, Y. M. Zaporozhets, G. V. Kondratenko, O. S. Shyshkin, Device for identification and analysis of tactile signals for information-control system of the adaptive robot, *Patent No. 40710,* Ukraine, 2009. (in Ukrainian).
[49]. A. Kargin, Introduction to Intelligent Machines, Book 1: Intelligent Regulators, *Nord-Press, DonNU*, Donetsk, 2010 (in Russian).

Chapter 3
Oscillating Wave Sensors Based on Symmetrical Metal-cladding Waveguide

Xianping Wang, Cheng Yin and Zhuangqi Cao

3.1. Symmetrical Metal-cladding Waveguide

3.1.1. Free-space Coupling Technology

The configuration of the symmetrical metal-cladding waveguide (SMCW) is illustrated in Fig. 3.1. It consists of a dielectric slab, which acts as the guiding layer, and two metal films, which function as the coupling layer and the substrate, respectively. From bottom to up, the dielectric coefficient and refractive index (RI) of the bottom metal film, the guiding layer, the upper metal film and the air are denoted as ε_j and n_j ($j = 0,\ 1,\ 2,\ 3$), respectively. Dielectric coefficient of metal film is always a complex number since the metal can absorb light energy, and ε_2 equals to ε_0 in the common case where the upper metal film and the bottom metal film belong to the same type. The thickness (s) of the upper metal film is about several dozens of nanometer and that of the bottom metal film is thick enough to prevent light leakage. As a polarized laser beam is incident upon the upper metal film with a resonant angle, the intensity of the reflected light decreases sharply due to the energy transferring from the incident light into the guided mode. The reflection coefficient of the SMCW can be written as

$$r_{3210} = \frac{r_{32} + r_{32}r_{21}r_{10}\exp(2ik_1d) + [r_{21} + r_{10}\exp(2ik_1d)]\exp(2ik_2s)}{1 + r_{21}r_{10}\exp(2ik_1d) + r_{32}[r_{21} + r_{10}\exp(2ik_1d)]\exp(2ik_2s)}, \quad (3.1)$$

with

Xianping Wang
College of Physics and Communication Electronics, Jiangxi Normal Univeristy, Key Laboratory of Optoelectronic and Telecommunication of Jiangxi Province, Nanchang 330022, China

$$r_{ij} = \begin{cases} \dfrac{k_i/\varepsilon_i - k_j/\varepsilon_j}{k_i/\varepsilon_i + k_j/\varepsilon_j} & \text{for TM polarization} \\ \dfrac{k_i - k_j}{k_i + k_j} & \text{for TE polarization} \end{cases}, \tag{3.2}$$

where r_{ij} is the Fresnel reflection coefficient, d is the thickness of the guiding layer, k_j are the normal components of the wave vectors in each medium and expressed as

$$k_j = \sqrt{k^2 \varepsilon_j - \beta^2}\,, \tag{3.3}$$

$$\beta = k n_3 \sin\theta\,, \tag{3.4}$$

where $k = 2\pi/\lambda$ is the wavenumber in vacuum, λ is the wavelength, β is the propagation constant of the guided mode, and θ is the incident angle.

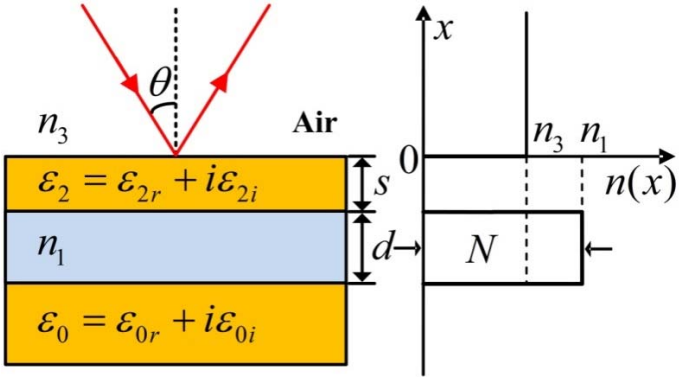


Fig. 3.1. Configuration of the SMCW.

A natural question is raised: why the light can be coupled into the guiding layer directly from the free-space without any couple component ? As well known, one important issue in the guided-wave optics is how to couple the light into the optical waveguide efficiently. The frequently used coupling techniques include the end-face coupling [1], the prism coupling [2], the grating coupling [3], and the tapered film coupling [4]. To understand the underlying physics of the above free-space coupling technology [5], a full discussion of the prism-waveguide coupling system is necessary.

The prism-waveguide coupling system is shown in Fig. 3.2. It is in essence a stratified structure with four layers consisting of the substrate, the guiding layer, the cover and the prism and being denoted by the subscripts $j = 0,\ 1,\ 2,\ 3$, respectively. Without loss of any generality, we assume

$$n_1 > n_0 > n_2\,. \tag{3.5}$$

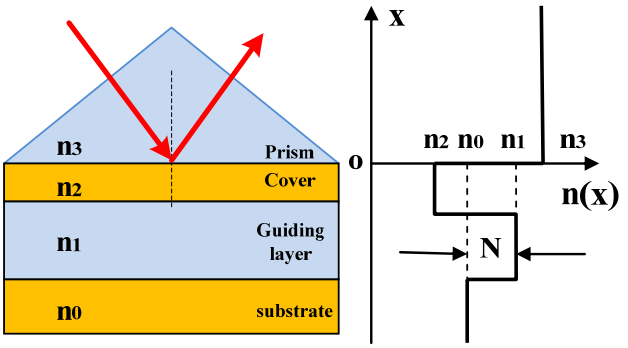


Fig. 3.2. The prism-waveguide coupling system.

For the sake of well confining the light energy in the guiding layer, the effective RI ($N = \beta/k$) is restricted in the range

$$n_1 > N > n_0 . \tag{3.6}$$

If there is no prism and the light beam is directly incident upon the cover layer with an incident angle θ_2, then the effective RI has the form

$$N = n_2 \sin\theta_2 . \tag{3.7}$$

From Eqs. (3.5), (3.6) and (3.7), we get $\sin\theta_2 > 1$. It implies that it is impossible for the prism-waveguide coupling system to couple the light energy directly from the free space (here the cover layer is assumed to be the air) into the guiding layer.

In order to satisfy the constraint in Eq. (3.6), one has to place over the cover layer with a higher RI prism, i.e.

$$n_3 > n_1 . \tag{3.8}$$

Then the light coupling is possible since the incident angle θ_3 in some range can make the effective RI

$$N = n_3 \sin\theta_3 , \tag{3.9}$$

be in the region $[n_0, n_1]$. In this case, the light fields propagate as oscillating waves in the prism and the guiding layer but as evanescent waves in the cover layer and the substrate. This coupling phenomenon is the so-called optical tunneling effect, whose quantum analogue is the scene of a microscopic particle tunneling through a potential barrier.

Let's reconsider the SMCW, in which the real part of the dielectric coefficient is negative, therefore the effective RI of the guided mode is bounded by

$$0 < N < n_1 , \tag{3.10}$$

which is larger than the corresponding range of the prism-waveguide coupling system. When the light beam is directly incident from the free space ($n_3 = 1$), the guided modes with the effective RI that obeys $N < n_3$ can be excited. In other words, the free-space coupling technology is valid in the SMCW. Its underlying physics can be understood as following: when the light beam incidents from the air to the upper metal film, it will generate an evanescent wave in the upper metal film; since the upper metal film is extremely thin, the tail of the evanescent wave can reach the metal-guiding layer interface, and then is reflected resulting in another evanescent wave with an opposite propagation direction; the interference effect between these two evanescent waves can eradiate the light energy into the guiding layer.

Fig. 3.3 shows the calculated electromagnetic energy distribution of a guided mode in the SMCW [6]. The used algorithm is the characteristic matrix method [7] and the maximum of the Poynting vector in the air is normalized to unit. It verifies that the light fields propagate as oscillating waves in the air and the guiding layer but as evanescent waves in the upper metal film and the substrate. Moreover, the amplitude of Poynting vector in the guiding layer is enhanced 68 times than that in the air.

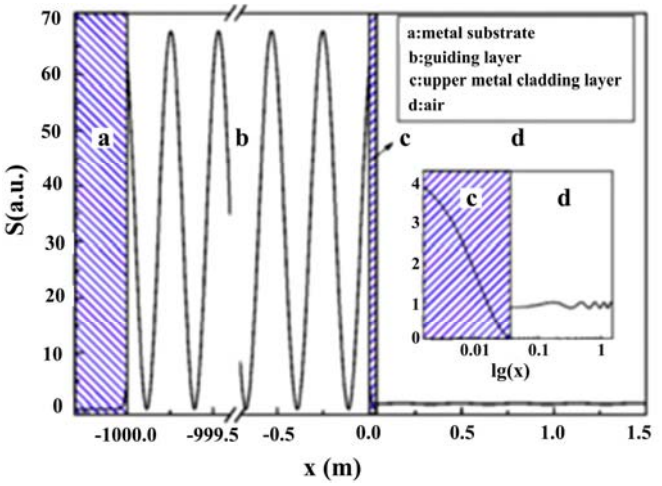


Fig. 3.3. Plot of the Poynting vector of a guided mode in the SMCW, the maximum of the Poynting vector in the air is normalized to unit (inset: close-up view for the Poynting vector in the upper metal film and air).

The experimental demonstration of the free-space coupling technology is presented. A BKT glass flat with a thickness 106 μm is sandwiched between two gold films to form the SMCW, where the thicknesses of the upper gold film and the bottom gold film are 46 nm and 300 nm, respectively. The dielectric coefficient of the glass and the gold are $\varepsilon_1 = 2.25$ and $\varepsilon_2 = -15 + i1.5$, respectively. A collimated light beam from a diode laser of 780 nm is incident upon the upper gold film, and the angular scan is carried out by a

computer-controlled $\theta/2\theta$ goniometer. A series of dips in the attenuated total reflection (ATR) spectrum due to the resonant transfer of light energy into the guided modes can be clearly observed (see solid curve in Fig. 3.4.). The theoretical (dashed curve) result is also given. It indicates that more than 50 % light energy has been fed into the guiding layer.

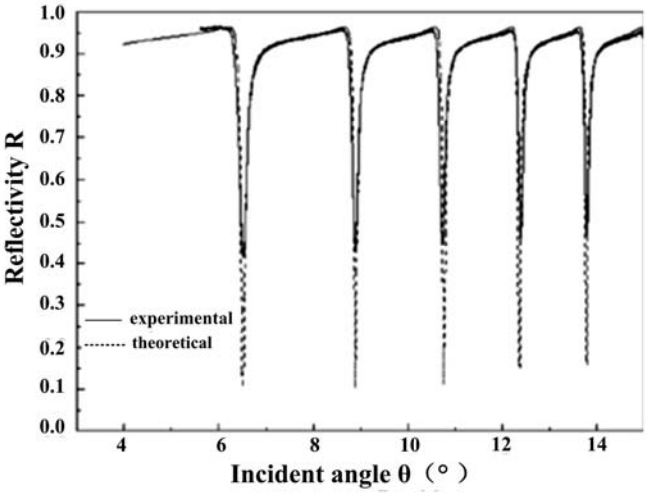


Fig. 3.4. ATR spectrum obtained from the SMCW.

3.1.2. Ultrahigh Order Mode

By ignoring the imaginary part of the dielectric coefficient of the metal film, the dispersion equation of the SMCW can be written as for TE mode:

$$k_1 d = m\pi + \arctan(\frac{\alpha_2}{k_1}) + \arctan(\frac{\alpha_0}{k_1}), \tag{3.11}$$

and for TM mode:

$$k_1 d = m\pi + \arctan(\frac{\varepsilon_1}{\varepsilon_{2r}}\frac{\alpha_2}{k_1}) + \arctan(\frac{\varepsilon_1}{\varepsilon_{0r}}\frac{\alpha_0}{k_1}), \tag{3.12}$$

where

$$\begin{cases} \alpha_0 = \left(\beta^2 - k^2\varepsilon_{0r}\right)^{1/2} \\ k_1 = \left(k^2\varepsilon_1 - \beta^2\right)^{1/2} \\ \alpha_2 = \left(\beta^2 - k^2\varepsilon_{2r}\right)^{1/2} \end{cases}. \tag{3.13}$$

If the thickness of the guiding layer is extended into a millimeter-scale, the SMCW can accommodate thousands of guided modes. For example, using the parameters: $\varepsilon_0 = \varepsilon_2 = -28 + i1.8$, $\varepsilon_1 = 2.278$, $S = 30\ nm$, $d = 0.38\ mm$ and $\lambda = 859.8\ nm$, we can get that the highest mode order (m) is 1333. When the light beam incidents with a relatively large incident angle, it is difficult to differentiate the adjacent guide modes because the mode density in the SMCW is extremely large and the corresponding ATR dips overlap with each other. On the other hand, a series of discrete guided modes can be excited at certain extremely small incident angles. These modes are called as the ultrahigh-order modes [8] in our terminology since the order of these modes are extremely large and their effective RI is small. From the perspective of ray-optics, within the same longitudinal propagation distance, the ultrahigh-order mode is rebounded at the interfaces with more times and the ray propagation length is longer since the incident angles of the ultrahigh-order mode at the top and bottom interfaces of the guiding layer are extremely small. As a result, the ultrahigh-order modes hold many peculiar optical properties different from the low-order modes.

1) The ATR dips of the ultrahigh-order modes are discrete.

On the basic of the dispersion Eqs. (3.11) and (3.12) of the SMCW, we can obtain an approximate formula

$$\Delta m \propto \sin 2\theta_3 \cdot \Delta\theta_3 . \tag{3.14}$$

By setting $\Delta m = 1$, it is clear that the incident angle θ_3 becomes smaller, a bigger $\Delta\theta_3$ will be obtained. It indicates that the ATR dips of the ultrahigh-order modes are discrete from each other. Such property is suitable to design a comb filter for the optical communication applications.

2) The polarization independence

In Eqs. (3.11) and (3.12), the maximums of the absolute value of the second term and the third term on the right side are $\pi/2$. For the ultrahigh-order modes, the deletion of these two terms will not introduce a significant error. Consequently, both dispersion equations for the TE and the TM modes can be simply approximated as

$$k_1 d = m\pi . \tag{3.15}$$

It indicates that the ultrahigh-order modes are independent on the polarization of the incident light.

3) Responding to a tiny change with a high sensitivity

Because the ultrahigh-order modes have a longer propagation time in the guiding layer, any tiny change of λ, n_1 and d will cause a dramatic variation of N. The theoretical sensitivity s is defined as the derivative of N with respect to a certain characteristic parameter, that is

$$s = \frac{dN}{d\xi}, \tag{3.16}$$

where ξ represents λ, n_1 or d. By the total differential of Eq. (3.15), we obtain

$$\frac{\partial N}{\partial n_1} = \frac{n_1}{N}, \tag{3.17}$$

$$\frac{\partial N}{\partial \lambda} = \frac{n_1^2 - N^2}{N\lambda}, \tag{3.18}$$

$$\frac{\partial N}{\partial d} = \frac{n_1^2 - N^2}{Nd}. \tag{3.19}$$

From the above three equations, it is clear that the sensitivity is inverse proportion to the effective RI (N). As a result, the sensitivity of the ultrahigh-order modes ($N \to 0$) is large and this property is extremely useful for designing the optical sensors and modulators.

4) The slow light effect

According to Eq. (3.18), a tiny change of wavelength can generate a dramatic variation of N. It indicates that the ultrahigh-order modes have the strong dispersion property and the slow light effect. Using Eq. (3.15), we can obtain the group velocity of the ultrahigh-order mode

$$v_g = \frac{d\omega}{d\beta} = \frac{N}{n_1} \cdot \frac{c}{(n_1 + \omega dn_1 / d\omega)}, \tag{3.20}$$

where the right side of the equation is composed of two sections: one originates from the first-order dispersion of the material and the other one is N/n_1, which is called as the slow light factor and is proportional to the effective RI N of the ultrahigh-order modes. The deduced results offer us a new physical mechanism for realizing the wideband slow light, and more details can see [9, 10].

3.2. Goos-Hänchen Shift

3.2.1. Theoretical Description of the Goos-Hänchen Shift

As well known, the reflection of a finite-waist light beam at an interface between two homogeneous media deviates slightly from the Snell's law. There is a later discrepancy between the reflected point and the incident one (see Fig. 3.5). Such a non-specular reflection is termed as the Goos-Hänchen (GH) shift [11] and the magnitude is given by [12]

$$\Delta z = \frac{2}{k_1}\frac{\partial \phi}{\partial \theta}, \tag{3.21}$$

where ϕ is the total reflection phase shift between the reflected point B and the incident point A.

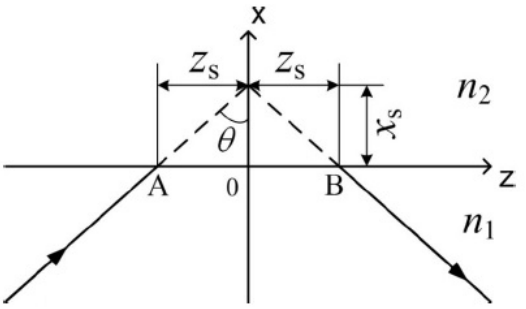


Fig. 3.5. The GH shift at the total reflection between two homogeneous media.

The reason for the emergence of GH shift is that the actual incident light is not an ideal monochromatic plane wave but a light beam with a certain spatial spectral width. Consequently, the incident light can be decomposed into a series of monochromatic plane waves and the wave vector of each one plane wave possess a slight different tangential component when compared to the other ones. In the total reflection, the reflection phase shift of each one plane wave is slight different from that of the other ones and the actual reflected light beam is composed of a series of such reflected plane waves. As a result, there is a lateral shift between the maximal intensity position of the reflected light and that of the incident light.

Let's consider the interface between two homogeneous media shown in Fig. 3.5. When the incident angle is larger than the critical angle, namely $\theta > \theta_c = \arcsin(n_2/n_1)$, where n_1 and n_2 are the RI of the two media and $n_1 > n_2$, the incident light will be totally reflected and its reflection coefficient is $r = \exp(i\varphi) = \exp(-i2\phi)$, where the reflection phase shift for the TM and the TE polarizations is

$$\phi_{\text{TE}} = \arctan\left[\frac{\sqrt{n_1^2 \sin^2\theta - n_2^2}}{n_1 \cos\theta}\right], \tag{3.22}$$

$$\phi_{\text{TM}} = \arctan\left[\left(\frac{n_1}{n_2}\right)^2 \frac{\sqrt{n_1^2 \sin^2\theta - n_2^2}}{n_1 \cos\theta}\right]. \tag{3.23}$$

To determine the magnitude of GH shift $2z_s \cos\theta$, let's consider a simple wave packet consisting of two plane waves with slight different incident angles. Supposing that the

z component of the wave vector is $\beta \pm \Delta\beta$, the complex amplitude of the incident wave packet at interface can be expressed as

$$A(z) = \left[\exp(i\Delta\beta z) + \exp(-i\Delta\beta z)\right]\exp(i\beta z) = 2\cos(\Delta\beta z)\exp(i\beta z). \quad (3.24)$$

$\Delta\beta$ is a small quantity, and the total reflection phase shift can be expanded by the differential formula

$$\phi(\beta \pm \Delta\beta) = \phi(\beta) \pm \frac{d\phi}{d\beta}\Delta\beta. \quad (3.25)$$

Therefore, the complex amplitude of the reflected wave packet at the interface is

$$\begin{aligned} B(z) &= \left\{\exp\left[i(\Delta\beta z - 2\Delta\phi)\right] + \exp\left[-i(\Delta\beta z - 2\Delta\phi)\right]\right\}\exp(i\beta z - 2\phi) \\ &= 2\cos\left[\Delta\beta(z - 2z_s)\right]\exp\left[i(\beta z - 2\phi)\right] \end{aligned}, \quad (3.26)$$

where

$$z_s = \frac{d\phi}{d\beta}, \quad (3.27)$$

is the simple form for the lateral shift of the wave packet. The vertical distance between the actual reflected light beam and the theoretical reflected light beam predicted by the geometrical optics is

$$S = 2z_s\cos\theta = \frac{2d\phi}{d\beta}\cos\theta = \frac{2}{k_0 n_1}\frac{d\phi}{d\beta}. \quad (3.28)$$

Substituting Eqs. (3.22) and (3.23) into Eq. (3.28), the GH shift at the total reflection between two homogeneous media are obtained as

$$S_{\mathrm{TE}} = \frac{2\sin\theta}{k_0\sqrt{n_1^2\sin^2\theta - n_2^2}}, \quad (3.29)$$

$$S_{\mathrm{TM}} = \frac{S_{\mathrm{TE}}}{\left[(n_1/n_2)^2 + 1\right]\sin^2\theta - 1}. \quad (3.30)$$

Other theoretical descriptions of the GH shift can be found in [13, 14].

3.2.2. Enhancement of the GH Shift

Usually, the GH shift is too small (on the order of wavelength) to encumber the quantification observations. Several configurations [15-19] have been proposed to enhance the GH shift. Moreover, theoretical and experimental results show that the magnitude of GH shift can be not only positive but also negative [18]. In this subsection, three of them for enhancing the GH shift are described in detail.

3.2.2.1. Surface Plasmon Resonance

According to the stationary-phase approach (see Eq. (3.28)), it is clear that a more abrupt change of the phase shift of the reflection coefficient can give rise to a larger GH shift. Yin et. al. [16] described a surface plasmon resonance configuration (see Fig. 3.6), with which a GH shift of greater than 50 wavelengths was observed. The dielectric coefficient of the air, the thin metal film and the prism are denoted as ε_0, ε_1 and ε_2, respectively. As a TM-polarized light beam incidents upon the interface between the prism and the metal film, and the resonance condition is satisfied, the surface plasmon wave will be excited. This excitation dramatically alters the phase shift of the reflection coefficient. In the experiment, the position of the reflected beam is detected by a position sensitive detector (PSD). By periodically modulating the incident polarization, the difference of the lateral displacement between TE and TM light is measured. Since the TE polarized incidence cannot excite any surface plasmon resonance, there is no enhanced GH shift. Therefore, it can serve as a reference beam and the measured relative displacement between TM and TE polarization indicates an absolute beam displacement (see the solid reflected light beam in Fig. 3.6) for a TM wave at the surface plasmon resonance region.

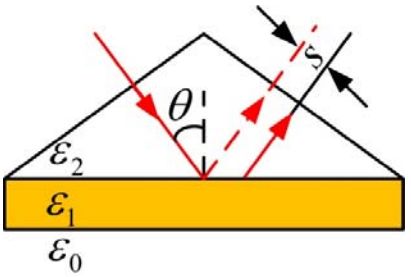


Fig. 3.6. Kretchmann configuration for the surface plasma excitation.

As scanning the incident angle θ, the minimal reflection intensity $R_{\min}$ in the ATR curve is closely related to the thickness of the thin metal film. The obtained experimental results [16] are as follows:

1) There is a critical thickness for the thin metal film. At the resonant condition, as the thickness of thin metal film approaches to h_c, $R_{\min}$ decreases and the GH shift increases significantly. A GH shift as high as 50 wavelengths was observed;

2) The GH shift is positive if the thickness of the thin metal film is smaller than h_c. Additionally, the GH shift becomes negative for the thin metal films being thicker than h_c.

3.2.2.2. Prism-waveguide Coupling System

Let's reconsider the prism-waveguide coupling system shown in Fig. 3.2. The thickness of the guiding layer and the cover layer are d_1 and d_2, respectively. Through the analysis described in [17], when the resonant condition is satisfied, we can get

$$\beta = \mathrm{Re}\left(\beta^0\right) + \mathrm{Re}\left(\Delta\beta^L\right), \tag{3.31}$$

where β^0 is the eigenpropagation constant of a guided mode in the absence of the prism. In this case, the cover layer and the substrate are both considered to be semi-infinite. While $\Delta\beta^L$, which originates from the existence of the coupling prism, represents the difference of the propagation constants between the three-layer waveguide and the prism-waveguide coupling system. The minimum of the reflectivity can be expressed as

$$R_{\min} = \left|r_{32}\right|^2 \left\{1 - \frac{4\,\mathrm{Im}\left(\beta^0\right)\mathrm{Im}\left(\Delta\beta^L\right)}{\left[\mathrm{Im}\left(\beta^0\right) + \mathrm{Im}\left(\Delta\beta^L\right)\right]^2}\right\}, \tag{3.32}$$

where r_{32} is the Fresnel reflection coefficient, $\mathrm{Im}\left(\beta^0\right)$ and $\mathrm{Im}\left(\Delta\beta^L\right)$ are the intrinsic and the radiative damplings, respectively. According to the stationary phase approach [12], the GH shift of the prism-waveguide coupling system is approximately written as

$$S = -\frac{2\,\mathrm{Im}\left(\Delta\beta^L\right)}{\mathrm{Im}\left(\beta^0\right)^2 - \mathrm{Im}\left(\Delta\beta^L\right)^2}\cos\theta, \tag{3.33}$$

From the above formula, we can obtain two important conclusions:

1) When the intrinsic damping is well-matched with the radiative damping, namely,

$$\mathrm{Im}\left(\beta^0\right) \approx \mathrm{Im}\left(\Delta\beta^L\right), \tag{3.34}$$

the corresponding GH shift will be extremely enhanced.

2) The sign of the GH shift is determined by the intrinsic and the radiative dampings. When the intrinsic damping is larger than the radiative damping ($\mathrm{Im}\left(\beta^0\right) > \mathrm{Im}\left(\Delta\beta^L\right)$), a negative GH shift ($S < 0$) can be obtained. The positive GH shift ($S > 0$) corresponds to the reverse case ($\mathrm{Im}\left(\beta^0\right) < \mathrm{Im}\left(\Delta\beta^L\right)$). Since the radiative damping mainly depends on the thickness of the cover layer, there is also a critical thickness in the prism-waveguide coupling system. The dependence of the sign of GH shift on the comparison between the thickness of the cover layer and the critical thickness [18] is similar to that of the surface plasmon resonance.

3.2.2.3. Symmetrical Metal Cladding Waveguide

If the Eq. (3.34) is satisfied in the SMCW, the GH shift of the reflected light beam will be also greatly enhanced. The used SMCW [19] in the experiment is a glass slab with two silver films coated on the upper and the bottom sides. Their thicknesses are 3.04 mm, 33.1 nm and 300 nm, respectively. Fig. 3.7 (a) illustrates the schematic arrangement for measuring the enhanced GH shift. A TE polarized Gaussian beam emitted from a tunable laser is introduced onto the SMCW. Before it reaches the upper silver layer, the light beam has passed through several devices: two apertures with a diameter of 2 mm and a distance of 0.5 m which are used to further confine the divergence of the incident light; a beam splitter which is used to reject part of the light beam into a wavemeter for the wavelength monitoring. The intensity of the reflected beam is detected by a photodiode (PD). The angular scanning is carried out by a computer controlled $\theta/2\theta$ goniometer. The incident angle is selected to be located at the maximum reflectivity of one certain reflectivity dip, where the GH shift is not remarkable. Since the magnitude of the GH shift is strongly dependent on the energy coupling between the incident light and the guided mode, it is reasonable to take this position of the reflected beam as the reference of the GH shift. After removing the PD out of its position without changing any position of the incident beam and the structure, we put on the PSD and let the reflected beam impinge onto the PSD at the center perpendicularly. Then we change the wavelength of the incident light from 859.401 to 859.415 nm, and get a tremendous GH shift greater than 1700 wavelengths.

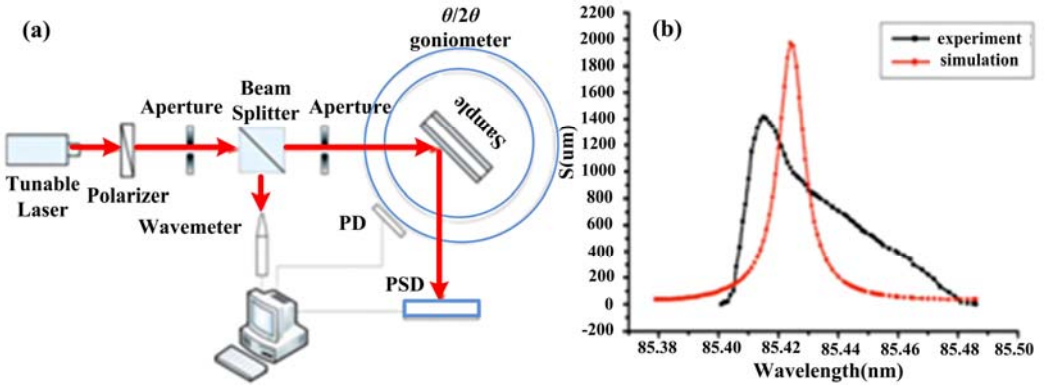


Fig. 3.7. (a) Experimental setup for measuring the enhanced GH shift, (b) Theoretical and experimental GH shift (S) with respect to the light wavelength.

Why the GH shift is largely enhanced in the SMCW can be well explained by using the theory of energy flux conversion ? The incident energy upon the SMCW is divided into two portions, one of them is reflected directly from the upper silver film; the rest one is coupled into the guiding layer in the form of an oscillating wave field and then is leaked continuously back into the incidence medium again during its lateral propagation process. The superposition between those two energy portions results in an actual reflected light laterally shifted from the ideal position. The GH shift of the reflected beam changes with

the variation of the wavelength is shown in Fig. 3.7 (b). The simulation in the Gaussian beam model is also outlined for comparison. The differences between the experimental observation and theoretical simulation most probably result from a slight non-parallelism of the guiding layers, the instability of the laser, the defect of the silver films, as well as errors in the measurements of the structure parameters. The result shows that the GH shift as large as 1.5 mm is obtained with the laser wavelength changing only by 14 pm.

3.3. Analysis on the Sensitivity

3.3.1. Definition of the Sensitivity

The sensitivity is defined as the ratio of the change in sensor output signal to the change in the quantity to be measured. Physically, the sensitivity is determined by the strength of light-matter interaction. Therefore, it is an important parameter to evaluate the sensor performance. The detailed expressions are closely related to the modulation methods. Here, three of them are listed as follows:

1) Angle modulation method

The variation of the quantity to be measured will lead to a collective movement of the ATR dip and the corresponding resonance angle. Therefore, the resonance angle can be used as an output signal to determine the quantity to be measured. In such angle modulation method, the sensitivity is written as

$$s = \frac{\partial}{\partial \xi}\left(\frac{\partial R}{\partial \theta}\right) = \left(\frac{\partial^2 R}{\partial \theta^2}\right)_{\theta=\theta_r} \cdot \left(\frac{\partial \theta}{\partial \xi}\right), \tag{3.35}$$

where θ_r is the position of the resonant angle, ξ denotes the quantity to measured (e.g. the RI, the absorption and the thickness). The first term describes the second derivative of the reflectivity with respect to the incident angle and depends on the shape of the reflectivity curve. The second term is connected with the sensing efficiency, whose physical meaning will be fully detailed in the next subsection. It is clear that if the reflectivity curve has a sharper slope or a smaller full width of half wave (FWHM), its sensitivity will be higher.

2) Intensity modulation method

In this method, the incident angle is fixed at the middle point of the falling or the rising edge of the resonant dip, where a good linearity and a high sensitivity can be achieved. A change in the quantity to be measured will lead to a movement of the reflectivity curve, which gives rise to a change in the reflected intensity. The sensitivity formula is expressed as

$$s = \frac{\partial R}{\partial \xi} = \left(\frac{\partial R}{\partial \theta}\right)_{\theta=\theta_s} \cdot \left(\frac{\partial \theta}{\partial \xi}\right), \tag{3.36}$$

where θ_s is the fixed incident angle and ξ is the quantity to be measured. The sensitivity also consists of two terms: the first one can be understood as the sharpness of the falling or the rising edge of the resonant dip and the second one is also connected with the sensing efficiency. From Eq. (3.36), it is obvious that the sensor's sensitivity can be improved by minimizing the FWHM of the resonance dip.

3) GH shift modulation method

The resonant condition and thus the magnitude of the GH shift is intimately connected to the structure parameters of the waveguide, and this tunable property of the GH shift can be employed to carry out the sensing function. As shown in Fig. 3.8, a tiny change in the quantity to be measured leads to a variation in the effective RI, which will give rise to a dramatical change in the GH shift. The theoretical sensitivity can be written as

$$s = \frac{\partial L}{\partial \xi} = \left(\frac{\partial L}{\partial \theta}\right)_{\theta=\theta_s} \cdot \left(\frac{\partial \theta}{\partial \xi}\right), \tag{3.37}$$

where L is the magnitude of the GH shift, θ_s is the fixed incident angle and ξ is the quantity to be measured. The sensitivity also consists of two terms: the first one can be understood as the sharpness of the GH shift curve and the second one is also connected with the sensing efficiency.

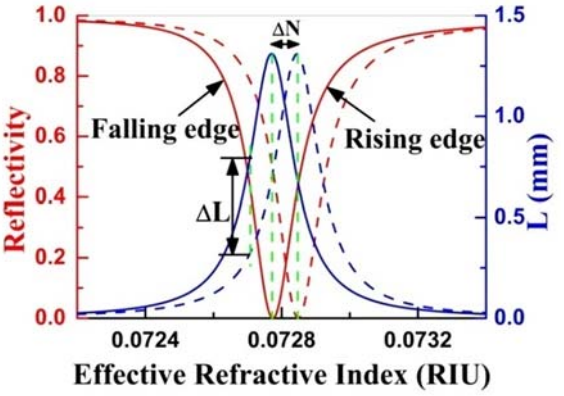


Fig. 3.8. The reflectivity (red) and the GH shift (blue) with respect to the effective RI.

3.3.2. Physical Meaning of the Sensing Efficiency

Herein, we discuss the physical meaning of $\partial\theta/\partial\xi$ in the above two sensitivity expressions. The parameter $\partial\theta/\partial\xi$ denotes the ratio of the variation of the resonant angle

to the change in the quantity to be measured. Obviously, it is equivalent to the variation in the effective RI (N) of the guided mode responding to the change in the quantity to be measured. Considering the three-layer planar waveguide shown in Fig. 3.9 and taking the TE mode as an example, we can obtain

$$\frac{dN}{dn_1}=\frac{n_1}{N}\cdot\frac{h+\dfrac{\alpha_0}{k_1^{\,2}+\alpha_0^{\,2}}+\dfrac{\alpha_2}{k_1^{\,2}+\alpha_2^{\,2}}}{h+\dfrac{1}{\alpha_0}+\dfrac{1}{\alpha_2}}=\frac{n_1}{N}\cdot\frac{P_1}{P}, \tag{3.38}$$

$$\frac{dN}{dn_2}=\frac{n_2}{N}\cdot\frac{k_1^{\,2}}{\alpha_2\left(k_1^{\,2}+\alpha_2^{\,2}\right)\left(h+\dfrac{1}{\alpha_0}+\dfrac{1}{\alpha_2}\right)}=\frac{n_2}{N}\cdot\frac{P_2}{P}, \tag{3.39}$$

$$\frac{dN}{dn_0}=\frac{n_0}{N}\cdot\frac{k_1^{\,2}}{\alpha_0\left(k_1^{\,2}+\alpha_0^{\,2}\right)\left(h+\dfrac{1}{\alpha_0}+\dfrac{1}{\alpha_2}\right)}=\frac{n_0}{N}\cdot\frac{P_0}{P}, \tag{3.40}$$

where P, P_1, P_2 and P_0 represent the power flowing in the whole waveguide structure, in the guiding layer, in the cladding layer and in the substrate, respectively. The sensing efficiency is proportional to the ratio of the power in the sensing area to the power in the whole waveguide structure. Furthermore, it is closely connected with the ratio of the RI of the measured medium to N. For the evanescent wave sensors (see surface plasmon resonance sensor [20-22], the leaky waveguide sensor [23], the reverse symmetry waveguide sensor [24, 25], etc.), N is always larger than n_0 and n_2, i.e., $n_0/N<1$ and $n_2/N<1$. But for the SMCW based oscillating wave sensor, the allowed range of the effective RI is $0<N<n_1$, which makes it possible to obtain an enormously high sensitivity on condition that the ultrahigh-order modes ($N\to 0$) at small incident angles are used as the sensing probe.

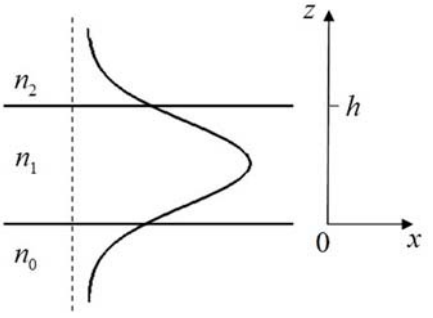


Fig. 3.9. The three-layer planar waveguide and its field distribution of the TE_0 mode.

3.4. Oscillating Wave Sensors

The common feature of the evanescent wave sensors is that the sample to be detected locates in the region where the evanescent wave of the guided mode propagates. It has shown that the sensitivity of optical sensors depend strongly on the power distribution of the guided mode. The bigger portion of power that interacts with the sample, the higher sensitivity will be. To achieve a higher sensitivity, it is essential to get as much of the optical power as possible to propagate in the sensing region. Investigation of the mode power distribution suggests one to design a configuration that contains the sample in the guiding layer of the waveguide, where the oscillating wave locates and most of the mode power is confined.

However, this idea is blocked by two barriers in the conventional waveguide. Firstly, only when the RI requirement of $n_1 > n_2, n_0$ is satisfied, a sandwiched structure can form a waveguide that the light wave can propagate stably in the guiding layer. Since the RI of aqueous solution is usually lower than the conventional waveguide materials, it is hard to find suitable materials to fabricate the substrate and the cover layer of the waveguide. Secondly, in the conventional waveguide structure, the thickness of guiding layer is usually around a micrometer scale, this limited sample room would make the sample into and out of sample room to be very difficult.

In the following subsections, we will elaborately describe our proposed oscillating wave sensors based on the SMCW, in which the above mentioned problems are no longer obstacles. As the real part of the dielectric coefficient of the metal films is negative, it suggests that not only the aqueous sample but also the gas can work as the guiding layer to be probed. So the detectable range of this proposed sensor can vary from gas with RI as low as 1 to liquid samples with RI of more than 2.0. Furthermore, the thickness of the guiding layer can be expanded to a submillimeter scale, which makes it easy to flow the aqueous sample into and out of the guiding layer.

3.4.1. Displacement Sensor

Displacement sensors have been frequently used in the ultra precision measurements of the geometrical quantities, such as positioning, vibration, and the profile measurement of the shafts and the surfaces. Among these sensors, the interferometer method is frequently employed but suffered from a relatively complicated optical arrangement and a limited resolution. In this subsection, we describe our proposed displacement sensors based on the SMCW by analyzing the fringe shift [26] and by monitoring change in the reflected intensity [27, 28] and in the GH shift [29]. Our displacement sensors can work in the real time and can provide accuracy better than those obtained by using photorefractive interferometer and speckle correlation technique.

The experimental arrangement for the displacement measurement by analyzing the fringe shift [26] is shown in Fig. 3.10. The SMCW consists of a stationary part 1 and a moving part 2. Part 1 is a glass prism, with a thin gold film of 50 nm deposited on its base, and which is rigidly attached to a heavy bar in brass. Part 2, as an object of the measurement

mounted on a translation stage, is a piece of flat glass coated with a relative thick gold film of 200 nm. The thickness of the air gap is controlled by a differential micrometer fixed on the translation stage. A light beam from a laser passes through a Glan Thomson prism P and two apertures A1 and A2, and is then reflected by a mirror M. Being spatially filtered by a spatial filter, the input beam finally focused by lens L1 at its rear focal plane, where the base of the prism is located. The reflective pattern, which is collected by another lens L2 and then recorded by a charge coupled device (CCD) placed a distance away from the focal plane of L2 along the optical axis. As a result of the energy transfer from the incident light into the guided mode, a fringe, which is attributed to the reduction in the intensity of the reflected bright spot, can be observed at the position of the resonant angle of the guided mode. If the moving part of the glass flat undergoes a slight displacement driven by the differential micrometer, the thickness of the guiding layer will then be changed with the same amount of the displacement. Substantially, the fringe will shift its angular position as well. Using this known fringe shift quantity, the displacement of the object can be determined absolutely. The measurement range of our experiment is -12.5~240 μm, the accuracy of the displacement is about 50 nm , and the relative error is less than 0.4 %.

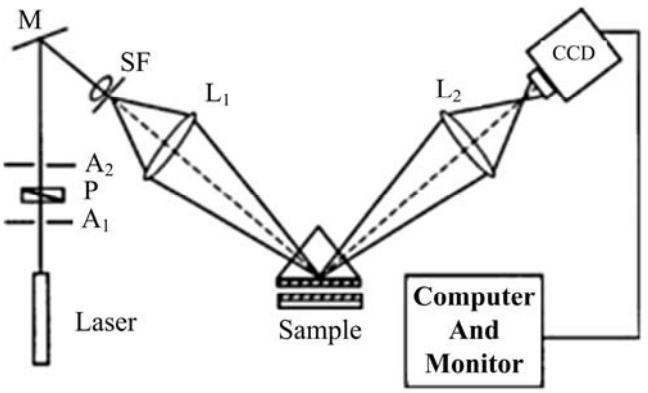


Fig. 3.10. Schematic arrangement of the displacement sensing by analyzing the fringe shift.

As shown in Fig. 3.11, the configuration of the displacement sensor by monitoring the reflected intensity [27] is composed of two parts: one is a glass prism on its base pre-coated with a thin gold film; the other is a 500- μm -thick $LiNbO_3$ slab sandwiched between two 400-nm-thick gold films and serves as a piezoelectric translator. The two components, separated by an air gap with a thickness of 100 μm, are rigidly mounted on a heavy platform. As applying a dc voltage on the pair electrodes of the piezoelectric translator, the air gap changes its thickness due to the piezoelectric effect of the $LiNbO_3$ slab. As a result, the reflection dip shifts its peak position and result in a change of the reflectivity. According to the resolution of the reflectivity variation, displacement can be evaluated from the applied voltage and the piezoelectric coefficient of the $LiNbO_3$ slab.

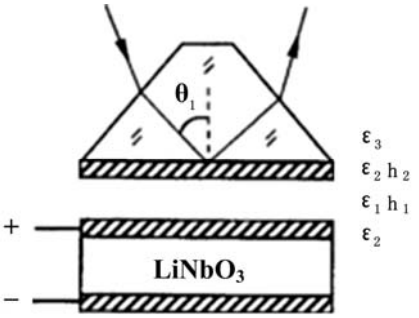


Fig. 3.11. Configuration of the displacement sensor by monitoring the reflected intensity.

The obtained reflectivity responding to displacement change is shown in Fig. 3.12. The thickness of the guiding layer (h_1) is increased and decreased in steps by increasing and reducing the voltages applied on the electrodes of the piezoelectric translator. The step-style change of voltage is 50 V. According to the piezoelectric coefficient of a Z-cut $LiNbO_3$ slab, $d_{33} = 33.45\ pm/V$, the value of the displacement resolution for the proposed configuration is determined as $S = 50 \times 33.45 \times 10^{-3} = 1.7\ nm$, which corresponds to the reflectivity change of ΔR=1% .

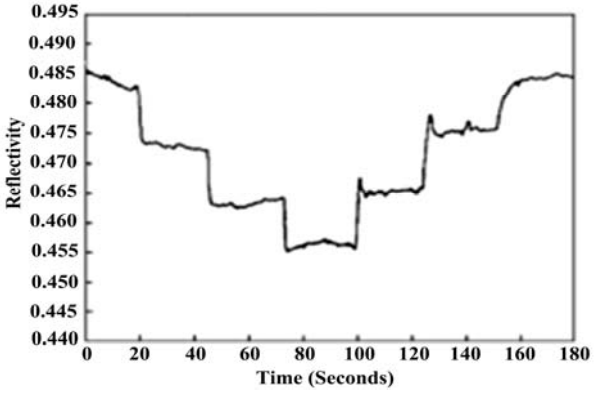


Fig. 3.12. The reflectivity responding to the displacement change.

Subsequently, an improved scheme for displacement measurement based on intensity modulation and lock-in amplification techniques is proposed [28]. The noise is greatly restrained by the lock-in amplifier. From the experimental result, a displacement resolution of 3.3 pm can be obtained.

Although a good resolution of the displacement has been achieved in the above mentioned sensors, the sensitivity is still limited by the fluctuation of the laser source. Here, an alternative approach for the displacement measurement based on the enhanced GH shift [29] is presented to obtain a higher resolution and prevent the disturbance caused by the power fluctuation in the light source.

The experimental structure is shown in Fig. 3.13. It consists of two parts; one is a glass prism with a thin gold film on its base; the other one is a z-cut $LiNbO_3$ slab sandwiched between two 300 nm thick gold film electrodes and serves as a PZT. Two components separated by an air gap with a thickness of 500 μm are rigidly mounted onto a heavy platform to prevent the relative shift. While applying a dc voltage on the pair of electrodes of the piezoelectric translator, the air gap then changes its thickness owing to the piezoelectric effect of the $LiNbO_3$ slab.

The experimental result is shown in Fig. 3.14. The voltage applied on the piezoelectric translator between each step is 10 V, and the piezoelectric coefficient of the z-cut $LiNbO_3$ is $d_{33} = 8 \times 10^{-12}\ m/V$. Thus, the thickness change per step is determined as $\Delta d = 8 \times 10^{-12} \times 10\ m = 8 \times 10^{-11}\ m$, which leads to a GH shift change of 2 μm . The experimental ripple of each step is confined to 0.5 μm . With this noise level, the sensing resolution is evaluated to be 40 pm. Since the magnitude of the GH shift is irrelevant to the incident light intensity, a power fluctuation of the laser brings no disturbance to the resolution of the displacement sensor.

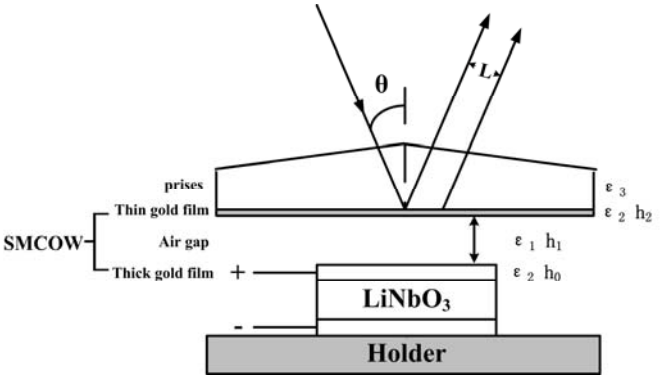


Fig. 3.13. Experimental structure of a displacement sensor based on the GH shift.

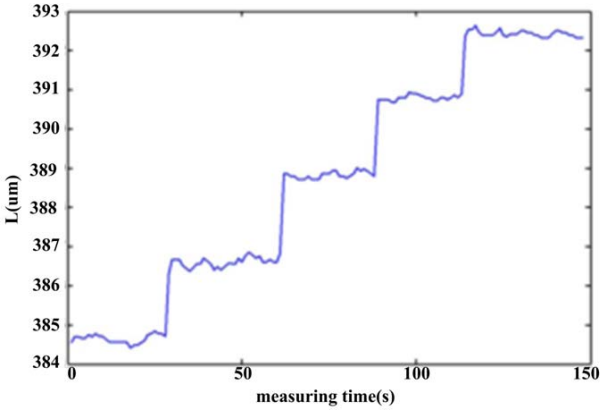


Fig. 3.14. Experimental sensitivity of the displacement sensor.

3.4.2. Angular Displacement Sensor

The Angular displacement plays an important role not only in the alignment, the assembly, and the calibration of the machine tools, but also in the fields of the earthquake measurement and the scientific experiments. There are two main ways to measure the small angle: the electronic methods and the optical methods. As one of most effective methods, capacitive transducers convert the angular displacement to the capacitance signal. By the technique of capacitance-bridge, the method reaches the angle resolution of 1.0×10^{-7} rad. Normally, the optical measurement of angular displacement is based on the principle of the optical lever [30]. The method can amplify the angular information infinitely. However, it is extremely difficult to construct an angular displacement sensor that incorporates both high resolution and compactness, and the stability of the angular displacement sensor decreases as the size is increased to increase the resolution.

A layout of the angular displacement sensor [31] based on the SMCW is illustrated in Fig. 3.15. The device is composed of two components: a prism and an oscillator. The prism is attached tightly on a platform, and the oscillator, which is suspended separately by a thin tungsten fiber, runs parallel with the hypotenuse face of the prism. The oscillator freely rotates around its suspending point O as an external torque is applied. There is an air gap with the thickness d_1 between the prism and the oscillator.

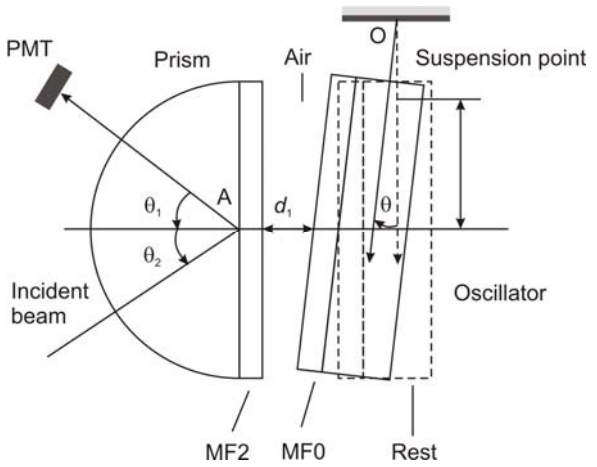


Fig. 3.15. Layout of angular displacement sensor based on the SMCW.

Under the action of the external torque, the oscillator rotates around the suspension point O with an angle of θ, and the variation of the thickness of the air gap is expressed as

$$d_1 = d_{10} + l\theta, \tag{3.41}$$

where d_{10} is the thickness of the air gap in the absence of the external torque, and l is the distance between the incident point of the laser beam and the suspending point O. In the intensity measurement scheme, a PD is employed to monitor the variation of the

optical intensity of reflected light. The minimum detectable angular displacement is 1.0×10^{-11} rad with a range of $\pm\ 2.5\times10^{-9}$ rad.

3.4.3. Wavelength Sensor

The wavelength sensing, which can monitor and stabilize the spectrum of semiconductor diode lasers, is an important task in many fields of applications. In order to obtain a high sensitivity, a probe with strong wavelength dispersion is often needed. Various sensors such as the arrayed-waveguide grating, the chirped grating, the three-layer slab waveguide, and the Fabry–Pérot etalon have been proposed. Among these methods, the guided mode of a three-layer slab waveguide structure is a preferable choice in the wavelength sensing due to its high index contrast between the guiding layer and the cladding layer, which allows for the strong optical confinement. However, the fact that a large part of the guided mode energy still disperses into the evanescent field would decreases the sensitivity.

The ultrahigh order mode excited in a SMCW exhibits strong wavelength dispersion [see Eq. (3.18)]. Experimental result [32] is shown in Fig. 3.16. The required wavelength variation is achieved by the temperature tuning of the tunable laser. The step-style change of wavelength is 0.5 pm, with the average 2.5 % change in the reflectivity ΔR. The sensitivity of about $5\times10^{10}\ m^{-1}$ is obtained with a range of 18 pm, which covers the reflectivity from 0.1 to 0.7.

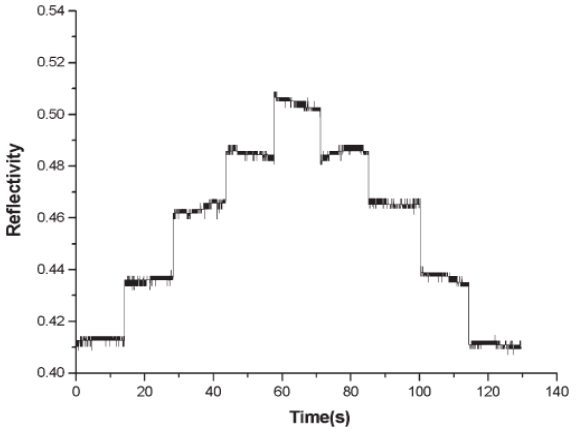


Fig. 3.16. The obtained wavelength sensitivity. The shift in the reflectivity curve was measured by changing the wavelength in steps of 0.5 pm.

The schematic diagram of the wavelength sensor with a high resolution by monitoring the variation of the GH shift [33] is shown in Fig. 3.17, which consists of two components. One is a glass prism with a gold thin film deposited on the bottom base. The other is a glass slab with another gold thin film coated on the upper surface. The mode dispersion for both the TE and the TM polarizations of the SMCW can be expressed as

$$\frac{dN}{d\lambda} = -\frac{\varepsilon_{air} - N^2}{N\lambda}. \tag{3.42}$$

From the above equation, it is easy to obtain the wave vector dispersion as

$$\frac{dk}{d\lambda} = 0, \tag{3.43}$$

for the normal component and

$$\frac{d\beta}{d\lambda} = -\frac{2\pi\varepsilon_{air}}{N\lambda^2}, \tag{3.44}$$

for the interfacial component. It is seen that k retains invariant while β changes rapidly with respect to the wavelength under the condition of $N \to 0$. The wave vector of the guided mode exhibits completely different dispersion properties in the two perpendicular directions, which indicates that the spatial position of the reflected beam will be highly sensitive to the light wavelength.

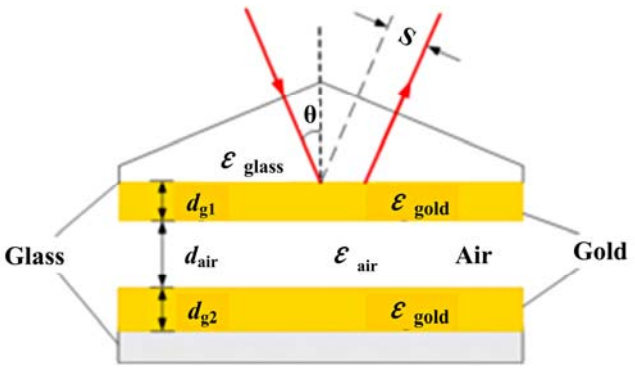


Fig. 3.17. Schematic diagram of the wavelength sensor by monitoring the GH shift.

Wavelength-shift monitoring is performed with different thicknesses of the guiding layer in the SMCW, the results of which are shown in Fig. 3.18.

Theoretical simulations based on the Gaussian beam model are also given for comparison. The sensitivity factor, which is defined as $F_S = \Delta S/\Delta\lambda$, reaches an average value of $4.5\ \mu m/pm$ in the case of $d_{air} = 0.2\ mm$. The wavelength resolution of 0.33 pm is obtained, taking into account the noise level of GH shift measurement that is about $1.5\ \mu m$. The dynamic range of the wavelength-shift monitoring is 100 pm. As d_{air} increases, the probing sensitivity becomes higher correspondingly. The value of F_S is $20\ \mu m/pm$ for $d_{air} = 0.5\ mm$ and $55\ \mu m/pm$ for $d_{air} = 1.0\ mm$ in each spectral region, which denote a wavelength resolution of 0.075 pm and 0.027 pm, respectively. The dynamic range decreases to 40 pm for $d_{air} = 0.5\ mm$ and 20 pm for $d_{air} = 1.0\ mm$.

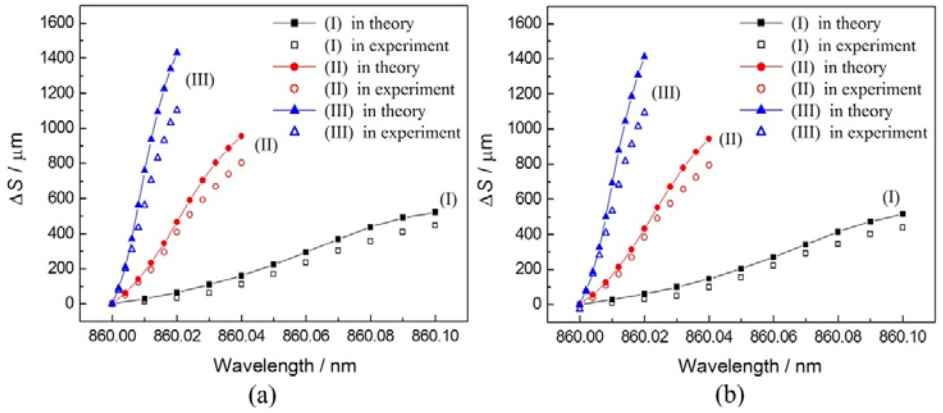


Fig. 3.18. Experimental result of the GH shift variation (ΔS) with respect to the light wavelength. (a) the TE polarization; (b) the TM polarization. The datum points are of different parameters: (I) $d_{air} = 0.2\ mm$, $\theta = 3.72^{o}$; (II) $d_{air} = 0.5\ mm$, $\theta = 3.84^{o}$; (III) $d_{air} = 1.0\ mm$, $\theta = 3.96^{o}$.

3.4.4. Aqueous Solution Concentration Sensor

The configuration of the oscillating wave sensor for the aqueous solution concentration sensing is shown in Fig. 3.19 [34]. A spacer O ring is sandwiched between two thin gold films coated on a coupling prism and a substrate glass plate, respectively, to form a sealed sample cell. The coupling prism is made of glass, with the RI of 1.5. A gold film of 30 nm is deposited on its bottom face, with the dielectric coefficient of $-11.4 + i1.5$ at the wavelength of 650 nm. The thickness of the spacer O ring is $h = 500\ \mu$m and hence the same thickness for the sample cell. Another gold film with the thickness of 300 *nm* is deposited on a glass plate substrate. Aqueous sample is pumped through the inlet and the outlet hole on the glass plate into the sample cell by a peristaltic pump. In such a manner, an oscillating wave sensor is constructed, in which the sample sealed in the sample cell serves as the guiding layer of the SMCW.

After passing through a spatial filter, a collimated TE polarized beam is focused by a lens and incident into the sensor structure, and its aperture angle is about 0.7°. The lens and the sensor structure are adjusted carefully to put the focal plane on the base of the coupling prism. At the resonant angles, energy of the incident light is coupled into the guiding layer to excite the ultrahigh-order modes. Several fringes will then be presented in the reflective pattern and captured by a CCD camera. A set of NaCl water solutions with concentration change of 250 ppm is used as samples to be probed. As the concentration of the sample pumped in the flow cell is changed, the fringe position will shift due to the minute RI change in the guiding layer. The experimental results show that a 250 ppm NaCl concentration change in the sample, corresponding to a change of 3.3×10^{-5} RIU (refractive index unit) will consequently result in a fringe shift of around 0.014°, as shown in Fig. 3.20. That means a sensitivity of 424°/RIU for TE mode. For comparison, in the

reserve symmetrical waveguide sensor presented by Horvath et al. [25], the experimental achieved sensitivities of TM and TE modes are 33.5°/RIU and 18.8°/RIU, respectively.

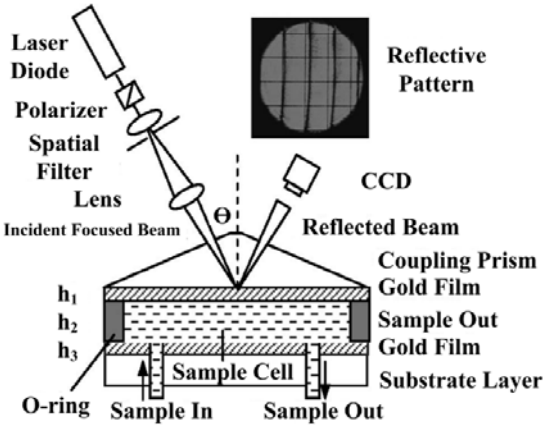


Fig. 3.19. Sensor structure and experimental layout.

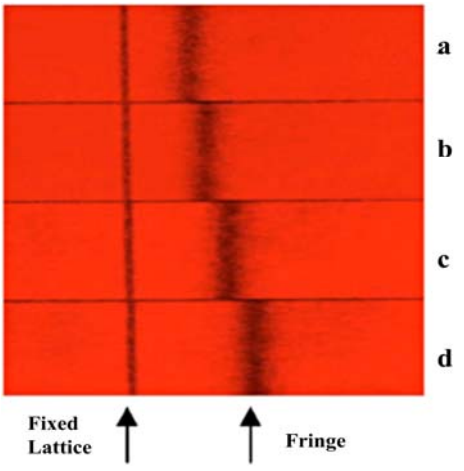


Fig. 3.20. Fringe shift captured by the CCD, the lattice is in fixed positions for reference, the samples are (a) pure water, (b) 250 ppm NaCl solution, (c) 500 ppm NaCl solution, and (d) 750 ppm NaCl solution.

A minute variation in the concentration of the aqueous solution leads to a RI change of the guiding layer, which will cause an angular shift of the resonance dip. If the incident angle is fixed at the middle point of the falling or the rising edge, the reflected intensity is another desired variable to monitor the concentration of aqueous solution [35]. Fig. 3.21 illustrates an aqueous solution concentration sensor based on the intensity modulation method. The detailed structure parameters can be found in [35].

The experimental sensor response is shown in Fig. 3.22, a 50 ppm NaCl water solution can cause a reflectivity change of about 15 %, and a 150 ppm NaCl solution sample even causes a reflectivity change of about 52 %. As the relationship between the concentration of NaCl water solution and its RI is known, it means that a 50 ppm NaCl water solution can bring an RI change of 6.6×10^{-6} over pure water, and a 150 ppm NaCl solution corresponds to a 2×10^{-5} RIU change. So the ideal resolution of the sensor is then calculated to be 8.8×10^{-8}, by assuming the detectable reflectance change to be 0.2 % [21].

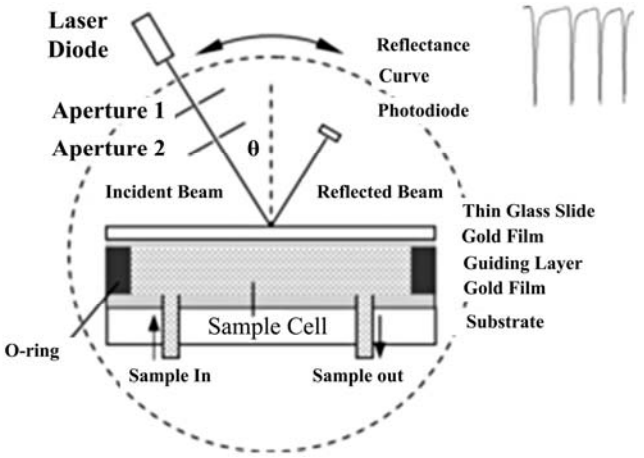


Fig. 3.21. The aqueous solution concentration sensor based on the intensity measurement.

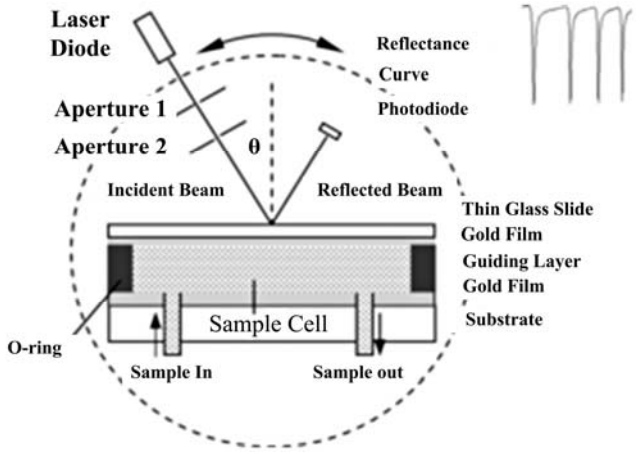


Fig. 3.22. Sensor response for NaCl water solution samples. The first intensity measurement for the pure water sample starts at B1, B2 and B3 with 1 min intervals. 1 min after a 50 ppm NaCl solution is pumped in, a second set of measurement starts at C1, C2 and C3 with 1 min intervals. Other timing spots are D1, D2 and D3 for 150 ppm NaCl water solution; E1, E2 and E3 for newly pumped in 150 ppm NaCl water solution; F1, F2 and F3 for 50 ppm NaCl solutions again and back to purewater at G1, G2 and G3.

The scheme diagram of the aqueous solution concentration sensor based on the enhanced GH shift [36] is illustrated in Fig. 3.23. A glass prism is coated with a 20 nm thick gold film to serve as the coupling layer. A 300 nm thick gold film is sputtered on a glass slab to act as the substrate. The air gap of 0.7 mm sandwiched between the two gold films works as the guiding layer, where a gasket is used to form a sealed sample cell. With the help of a peristaltic pump, sample liquids to be detected flow into the cell through the inlet and the outlet tubes embedded in the substrate glass plate.

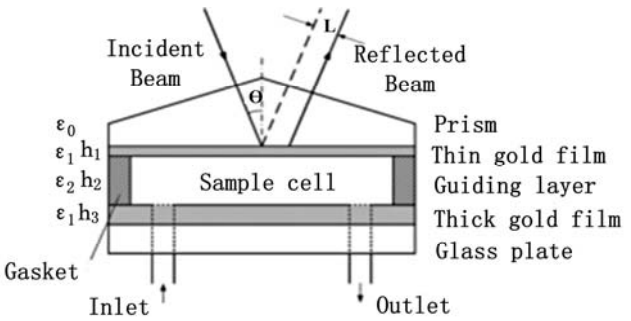


Fig. 3.23. Configuration of the aqueous solution concentration sensor based on the GH shift.

The theoretical sensitivity of the aqueous solution concentration sensor is defined by the change rate of the GH shift (L) with respect to the RI of the guiding layer (n_g) and it can be written as

$$S = \frac{dL}{dn_g} = \left(\frac{\partial L}{\partial N}\right)\left(\frac{\partial N}{\partial n_g}\right) = S_1 S_2 . \tag{3.45}$$

A high sensitivity S_1 can be obtained when the incident angle gets close to the resonance peak. According to the dispersion equation for the ultrahigh-order modes in the SMCW, S_2 can be cast in the form

$$S_2 = \frac{n_g}{N}\frac{P_g}{P_T}, \tag{3.46}$$

where P_g and P_T represent the power flowing in the guiding layer and the whole waveguide structure, respectively. For the SMCW, the energy is almost totally coupled into the guiding layer ($P_g \approx P_T$) due to the strong confinement effect. Moreover, the allowed range of the effective refractive index is $0 < N < n_g$, which makes it possible to obtain an enormously high sensitivity ($S_2 \gg 1$), when the ultrahigh-order modes ($N \to 0$) at small incident angles are used as the sensing probe.

In the experiment, the incident angle is adjusted to coincide with an ultrahigh-order mode. And this incident angle will be fixed to keep the stability of the optics. In order to attain a

high sensitivity and a good linearity, the light wavelength is tuned by using a tunable laser. As the wavelength of the incident light changes, the GH shift of the reflected beam varies correspondingly. The operation wavelength is selected at the middle area of the rising edge of the resonance peak (see Fig. 3.8) where a good linearity and a high sensitivity can be achieved. A tiny variation in the concentration of the sample solution, which corresponds to a RI change of the guiding layer, will lead to a change of the GH beam shift due to the translation of the resonance peak. The change of the GH shift is measured by using a PSD. A series of NaCl solutions with the change step of 20 ppm in concentration is used as sample sample to be probed. The experimental result is shown in Fig. 3.24. The step change of 20 ppm NaCl solution in concentration, which corresponds to a variation of 2.64×10^{-6} RIU, induces a GH shift change of at least 20 μm. By considering the noise level in the experiment, the probing sensitivity of 2.0×10^{-7} RIU is resolved since the measurement variation of the GH shift is confined within 1.5 μm for each sample.

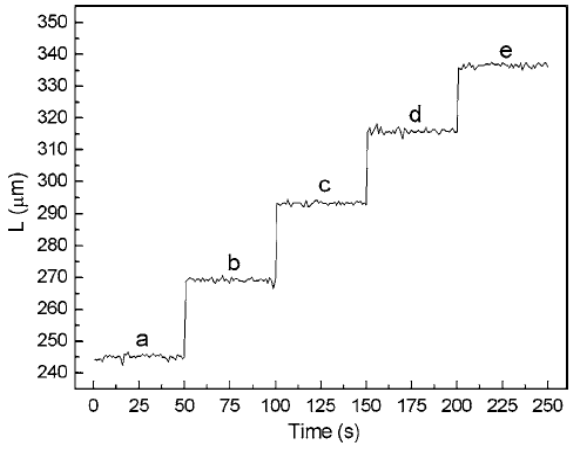


Fig. 3.24. The GH shifts with respect to solutions with different concentrations: (a) pure water, (b) 20 ppm NaCl solution, (c) 40 ppm NaCl solution, (d) 60 ppm NaCl solution, and (e) 80 ppm NaCl solution.

3.4.5. Trace Chromium (VI) Sensor

In addition to the weathering of rocks and the erosion of soils, the trace chromium (Cr) is usually found in the environment as a result of the discharge of many industrial manufacturing activities, such as the stainless-steel production, the leather tanning, the electroplating and the pigment fabrication. Cr is generally in two most stable states of Cr (III) and Cr (VI). The physiological effect of these two states on the biological systems are totally opposite: Cr (III) is essential to human health at the trace level; Whereas Cr (VI) is readily absorbed by the lungs, the digestive tracts, the mucous membranes and the skins and is toxic and carcinogenic, and is considered as a serious pollutant in the environment. The determination of Cr (VI) in the environmental and at the industrial sites is consequently important.

The minimum reflectivity of the ultrahigh-order mode is well described by Eq. (3.32) and can be cast into the form

$$R_{\min} \propto 1 - \frac{4\,\mathrm{Im}\left(\beta^0\right)\mathrm{Im}\left(\Delta\beta^L\right)}{\left[\mathrm{Im}\left(\beta^0\right)+\mathrm{Im}\left(\Delta\beta^L\right)\right]^2}. \tag{3.47}$$

If the solution is an absorbable material, the relationship between the concentration of solution and the imaginary part of the dielectric coefficient can be derived from the Beer-Lambert law as below [7]

$$\mathrm{Im}(\varepsilon) = \frac{nK_m}{\ln 10 \cdot k_0} C\,, \tag{3.48}$$

where C is the concentration of solution, K_m is the molar absorptivity, n is the real part of RI of the solution. As the concentration of Cr (VI) in aqueous solution changes, the extinction coefficient of the solution and the intrinsic damping of the SMCW change too. Combining Eq. (3.47) and Eq. (3.48), we can get the relationship between reflectivity and the concentration of the solution.

As shown in Fig. 3.25 (a), the SMCW for the trace chromium (VI) sensing [37] is composed of three parts: (i) a polished glass substrate of 0.5 mm thickness has a thin silver film coated on the top side to act as a coupling layer, (ii) two parallel glass strips of 0.5 mm thickness are placed at a distance of 4 mm to form the sample confinement walls, (iii) another polished glass substrate of 0.5 mm thickness has a relatively thick silver film deposited on the base side to prevent the light leakage. The experimental setup is shown in Fig. 3.25 (b). A TE polarized laser beam from a solid-state laser is incident upon the upper silver film at certain optical angles. An aperture with a diameter of 1 mm was inserted into the beam path to further confine the divergence angle to about 0.4 mrad. The sample solution was pumped into and out the cell room by a syringe through inlet and outlet pipes with 0.5 mm inner diameter. A computer-controlled $\theta/2\theta$ goniometer was used to carry out the angular scans. The intensity of the reflected beam was detected by a PD.

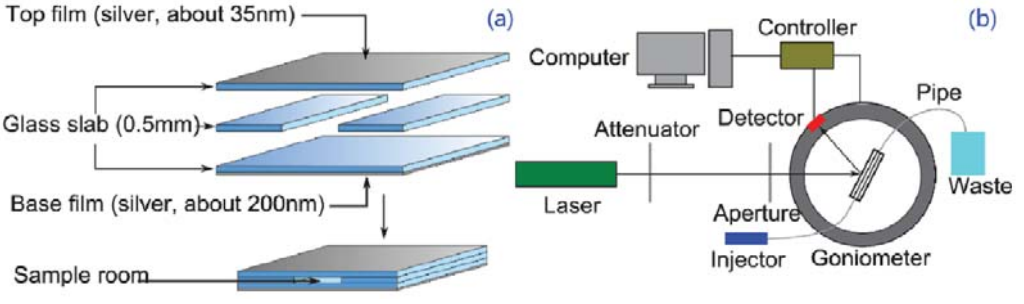


Fig. 3.25. The structure sensor (a) and experimental arrangement (b) for trace chromium (VI) sensing.

Determination of the Cr (VI) was based on the color reaction with 1,5-Diphenylcarbazide (DPC, C13H14N4O). The reaction takes place at a pH of 1.0 ± 0.3, and forms absorbing complex at the wavelength around 540 nm. Therefore, a laser of 532 nm was used as the light source in the experiment. Sulfuric acid (1:1) and phosphoric acid (1:1) were prepared by diluting 50 ml of acid in 50 ml of water. Chromogenic reagent was prepared by dissolving 2 g of DPC in 50 ml acetone, and then diluting in water with a 100 ml measuring flask. Chromogenic reagent was prepared just before use. A 400 μg/l of Cr(VI) stock solution was prepared by dissolving 0.2829 ± 0.0001 g of potassium dichromate in water with a 250 ml measuring flask. The Cr (VI) working standard solutions were prepared by appropriate dilution with water. Color reaction took place by adding 0.25 ml of Sulfuric acid (1:1), 0.25 ml of phosphoric acid (1:1) and 1 ml of Chromogenic reagent to 25 ml of each Cr (VI) working standard solution. The reaction lasted for 7 minutes. The concentration of Cr (VI) working standard solution selected in this experiment were 0, 0.24, 0.48, 0.72, 0.96, 1.2, 1.44, 1.92, 2.16, 2.64 μg/l. After reaction, the mixed solutions were injected into sample cell. And then reflectivity spectrum of each solution was measured by rotating the goniometer. Before the injection of each solution, the sample cell was washed by pumping 20 ml of de-ionized water to avoid the residue.

Fig. 3.26 (a) shows the reflectivity spectrums of the sample solutions. As the Cr (VI) concentration increases, the reflectivity at resominimum nant angle varies from 0.097 to 0.118. Fig. 3.26 (b) shows the relationship between minimum reflectivity and the concentration of Cr (VI). It produces an approximate linear response to the Cr (VI) concentrations:

$$R_{\min} = (0.007965 \pm 0.000814) \cdot C(\mathrm{Cr(VI)}) + (0.09763 \pm 0.00117), \qquad (3.49)$$

with the fitting relation coefficient, where C(Cr(VI)) is the concentration of Cr(VI) in $\mu g/l$. The detection resolution, which is defined as three times of the standard deviation of the measurement blank (a concentration of zero), is $\frac{3 \times 0.00017}{0.007965} = 0.064\ \mu g$ or 1.2 nM. It is a 16-fold improvement compared to the surface plasmon field applied to the determination of Cr (VI) [38].

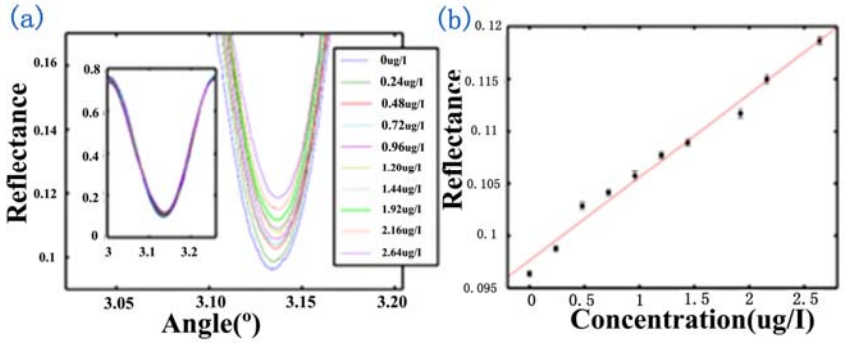


Fig. 3.26. The reflectivity spectrums (a) and the minimum reflectance (b) with respect to different concentration of Cr (VI).

3.4.6. Trace Glyphosate Sensor

To improve the outputs of the agricultural crops, the organophosphorus pesticides have been extensively researched. Among them, glyphosate (N-(phosphonomethyl) glycine) is a widely used organophosphorus pesticide because of its excellent performance on the weed control. However, glyphosate is a toxic endocrine disruptor, and its accumulation will impact the environment and pose a threat to human health. Therefore, the monitoring small concentrations of glyphosate in food and drinking water has gained an increasing importance. Analytical methods for the quantitative determination of low glyphosate concentrations in water include the electrothermal atomization atomic absorption spectrometry [39], the flame atomic absorption spectrometry [40], and etc. These spectroscopic techniques are sensitive and accurate, but suffer from the system complexity and the long testing times.

As shown in Fig. 3.27, the SMCW for sensing the trace glyphosate [41] is composed of four parts: (i) an optical prism with the vertex angle of 150^{o}, (ii) a 40 nm silver film deposited on the bottom side of the prism to act as a coupling layer, (iii) a ring-like glass gasket with the thickness of 1 mm sandwiched between the prism and the glass substrates with a rounded glass island of 980 μm thick to form a sample cell, and (iiii) a 300 nm silver film deposited on the top side of the glass island to act as another cladding layer of the waveguide. Analyte to be detected is injected into the sample cell with a 20 μm thickness to function as the guiding layer.

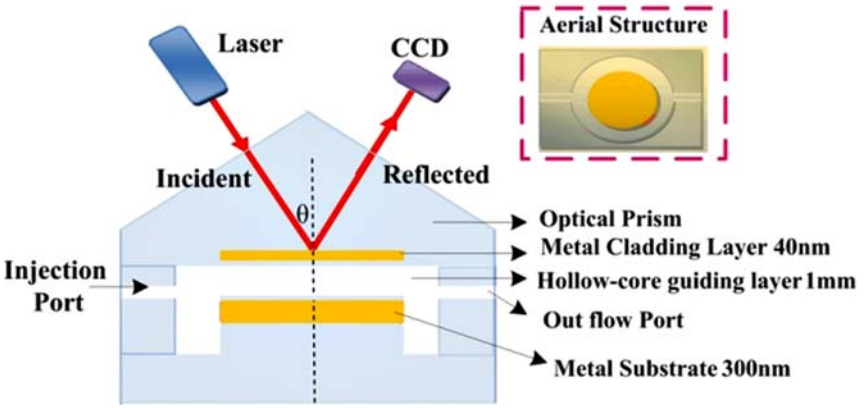


Fig. 3.27. Schematic diagram of the SMCW for sensing the trace glyphosate.

Determination of the trace glyphosate was based on the color reaction between iodine and amylum. Under a PH of 4.0 ± 0.3, the chemical reaction between the glyphosate and the sodium nitrite can oxidize the potassium iodide into idodide, which can form a chromogenic complex with amylum. Therefore, the imaginary part of the dielectric coefficient of the guiding layer is determined by the concentration of glyphosate. Similar to the above subsection, we can get the relationship between the reflectivity and the concentration of glyphosate by using Eqs. (3.47) and (3.48).

In the experiment, sulfuric acid (3 mol/L) was prepared by diluting 40 ml of acid in 140 ml of water, and a sodium nitrite solution was prepared by dissolving 1.2 g in 1 L of water. Chromogenic reagents were potassium iodide and amylum, which were prepared by dissolving 1.0 g of potassium iodide in 1 ml water and 4 mg of amylum in 1 ml water. A 1.42 mol/l glyphosate stock solution was prepared by dissolving 0.2513 ± 0.0001 g of glyphosate in 250 ml water. A 1.42 nmol/l glyphosate standard solution was prepared by diluting 1 ml of glyphosate stock solution in 100 ml water. The glyphosate standard solutions were prepared by appropriate dilution by water. The absorption peak of the formed chromogenic complex is at 532 nm, thus a laser emitting at this wavelength was used for sensing.

The ATR spectra for different glyphosate concentrations are shown in Fig. 3.28 (a). As the concentration increases from 0.00 to 4.38 nm/l, the depth of the dip at the resonant angle decreases from 0.169 to 0.046. Fig. 6(b) shows the linear relationship between the minimum reflectivity and the concentration of glyphosate from 1.42 to 4.38 nm/l, where $R_{min}= (1.81 \pm 0.014) - (0.23 \pm 0.046) * C_{(NPG)}$ (nm/l). The limit of detection is 1.42 nM $C_{(NGP)}= (m_{(NGP)}/M_{(NGP)}V) = (2.0 \times 10^{-4}$g/156 g/mol × 1 L) = 1.42 nm/l = 1.42 nM. The sample size is the small volume of the sensor cell.

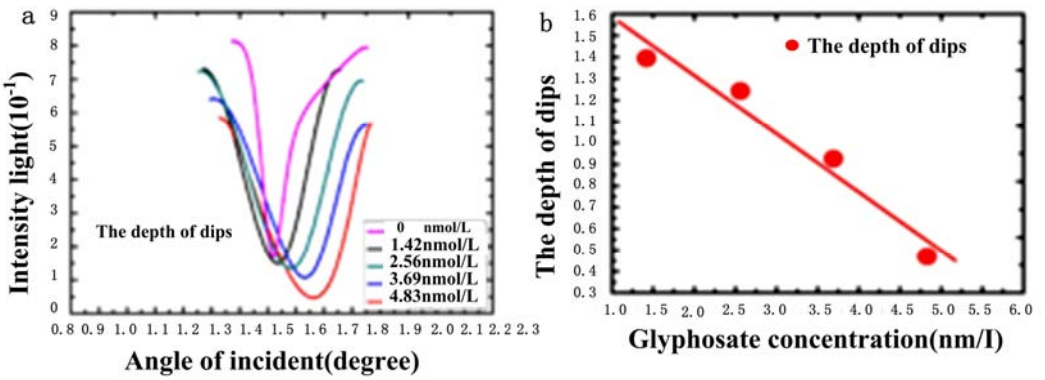


Fig. 3.28. (a) ATR spectra for different glyphosate concentrations; (b) depths of the ATR dips at resonant angles for different glyphosate concentrations.

3.5. Summary and Outlook

In the SMCW, since the real part of the dielectric coefficient is negative, the light energy can be coupled into the guiding layer by the free-space coupling technology, which makes any couple component unnecessary. When the thickness of the guiding layer is extended into a millimetre scale, the so-called ultrahigh order modes can be excited at certain small incident angles. It is found that the ultrahigh order modes are polarization independent and exhibit a high sensitivity to a tiny variation in the structural parameters. Theoretical and experimental results show that the SMCW can greatly enhance the GH shift under the resonance condition. Moreover, the enhanced GH is also highly sensitive to the structural variation. The reason for such high sensitivity is that the sample to be detected locates in

the region where the oscillating wave propagates, namely, the ratio of the power in the sensing area to the power in the whole waveguide structure is larger than that of the evanescent wave sensors. Taking the reflectivity and GH shift as the sensing probes, a series of the oscillating wave sensors based on the SMCW is selected and described in this chapter. Noted that the light energy is enhanced by two orders in the guiding layer of the SMCW, such field enhancement effect could offer potential applications in the fields of the nonlinear optics [42], the enhanced fluorescence, the enhanced Raman scattering [43], the biosensing, and the light trapping [44, 45], and so on. Further studies of the field enhancement effect of the oscillating wave in the SMCW are in progress.

Acknowledgements

This work is supported by the National Natural Science Foundation of China (Grant Nos. 61505073, 11404092, 61265001, and 11264016), and the Natural Science Foundation of Jiangsu Province (Grant No. SBK2014043338).

References

[1]. D. Marcuse, Coupled mode theory of round optical fibers, *Bell. Syst. Tech. J.*, Vol. 52, Issue 6, 1973, pp. 817-842.
[2]. P. K. Tien, R. Ulrich and R. J. Martin, Modes of propagating light waves in thin deposited semiconductor films, *Appl. Phys. Lett.,* Vol. 14, Issue 9, 1969, pp. 291-294.
[3]. M. L. Dakss, L. Kuhn, P. F. Heidrich and B. A. Scott, Grating coupler for efficient excitation of optical guided waves in thin films, *Appl. Phys. Lett.*, Vol. 16, Issue 12, 1970, pp. 523-525.
[4]. P. K. Tien and R. J. Martin, Experiments on light waves in a thin tapered film and a new light-wave coupler, *Appl. Phys. Lett.,* Vol. 18, Issue 9, 1971, pp. 398-401.
[5]. H. G. Li, Z. Q. Cao, H. F. Lu, and Q. S. Shen, Free-space coupling of a light beam into a symmetrical metal-cladding optical waveguide, *Appl. Phys. Lett.,* Vol. 83, Issue 6, 2003, pp. 2757-2759.
[6]. W. Yuan, C. Yin, P. P. Xiao, X. P. Wang, J. J. Sun, M. H. Sang, X. F. Chen, and Z. Q. Cao, Mircrosecond-scale switching time of magnetic fluids due to the optical trapping effect in waveguide structure, *Mircofluid. Nanofluid.*, Vol. 11, Issue 6, 2011, pp. 781-785.
[7]. M. Born, E. Wolf, Principles of optics: electromagnetic theory of propagation, interference and diffraction of light, A. B. Bhatia, *Cambridge University Press,* 1999.
[8]. H. F. Lu, Z. Q. Cao, H. G. Li, and Q. S. Shen, Study of ultrahigh-order modes in a symmetrical metal-cladding optical waveguide, *Appl. Phys. Lett.*, Vol. 85, Issue 20, 2004, pp. 4579-4581.
[9]. W. Yuan, C. Yin, H. G. Li, P. P. Xiao, and Z. Q. Cao, Wideband slow light assisted by ultrahigh-order modes, *J. Opt. Soc. Am. B*, Vol. 28, Issue 5, 2011, pp. 968-971.
[10]. Y. L. Zheng, W. Yuan, X. F. Chen, and Z. Q. Cao, Wideband slow-light modes for time delay of ultrashort pulses in symmetrical metal-cladding optical waveguide, *Opt. Express*, Vol. 20, Issue 9, 2012, pp. 9409-9414.
[11]. F. Goos and H. Hänchen, Ein neuer und fundamentaler versuch zur totalreflexion, *Ann. Phys.*, Vol. 436, Issue 7, 1947, pp. 333-346.
[12]. K. Artmann, Berechnung der seitenversetzung des totalreflektierten strahles, *Ann. Phys.*, Vol. 437, Issue 1, 1948, pp. 87-102.

[13]. C. F. Li and Q. Wang, Prediction of simultaneously large and opposite generalized Goos-Hänchen shifts for TE and TM light beams in an asymmetric double-prism configuration, *Phys. Rev. E*, Vol. 69, Issue 5, 2004, pp. 055601.
[14]. X. M. Liu, Q. F. Yang, Z. Qiao, T. K. Li, P. F. Zhu, and Z. Q. Cao, Physical origin of large positive and negative lateral optical beam shifts in prism-waveguide coupling system, *Opt. Commun*., Vol. 283, Issue 13, 2010, pp. 2681-2685.
[15]. H. M. Lai and S. W. Chan, Large and negative Goos-Hänchen shift near the Brewster dip on reflection from weakly absorbing medium, *Opt. Lett.*, Vol. 27, Issue 9, 2002, pp. 680-682.
[16]. X. B. Yin, L. Hesselink, Z. W. Liu, N. Fang, and X. Zhang, Large positive and negative lateral optical beam displacement due to surface plasmon resonance, *Appl. Phys. Lett*., Vol. 85, Issue 3, 2004, pp. 372-374.
[17]. X. B. Liu, Z. Q. Cao, P. F. Zhu, Q. S. Shen, and X. M. Liu, Large positive and negative lateral optical beam shift in prism-waveguide coupling system, *Phys. Rev. E*, Vol. 73, Issue 5, 2006, pp. 056617.
[18]. L. Chen, Z. Cao, F. Ou, H. Li, Q. Shen, and H. Qiao, Observation of large positive and negative lateral shifts of a reflected beam from symmetrical metal-cladding waveguides, *Opt. Lett.*, Vol. 32, Issue 11, 2007, pp. 1432-1434.
[19]. J. Hao, H. Li, C. Yin, and Z. Cao, 1.5 mm light beam shift arising from 14 pm variation of wavelength, *J. Opt. Soc. Am. B*, Vol. 27, Issue 6, 2010, pp. 1305-1308.
[20]. C. Nylander, B. Liedberg, T. Lind, Gas detection by means of surface plasmon resonance, *Sens. Actuators.*, Vol. 3, Issue 1, 1982, pp. 79-88.
[21]. J. Homola, S. S. Yee, G. Gauglitz, Surface plasmon resonance sensors: review, *Sens. Actuators. B*, Vol. 54, Issue 1, 1999, pp. 3-15.
[22]. X. D. Hoa, A. G. Kirk, M. Tabrizian, Towards integrated and sensitive surface plasmon resonance biosensors: a review of recent progress, *Biosens. Bioelectron.*, Vol. 23, Issue 2, 2007, pp. 151-160.
[23]. T. Okamoto, M. Yamamoto, I. Yamaguchi, Optical waveguide absorption sensor using a single coupling prism, *J. Opt. Soc. Am. A*, Vol. 17, Issue 10, 2000, pp. 1880-1886.
[24]. R. Horvath, H. C. Pedersen, N. Skivesen, D. Selmeczi, N. B. Larsen, Monitoring of living cell attachment of spreading using reverse symmetry waveguide sensing, *Appl. Phys. Lett.*, Vol. 86, Issue 7, 2005, pp. 071101.
[25]. R. Horvath, H. C. Pedersen, N. B. Larsen, Demonstration of reverse symmetry waveguide sensing in aqueous solutions, *Appl. Phys. Lett.*, Vol. 81, Issue 12, 2002, pp. 2166-2168.
[26]. J. Shi, Z. Q. Cao, J. Zhu, Q. S. Shen, Displacement measurement in real time using the attenuated total reflection technique, *Appl. Phys. Lett.*, Vol. 84, Issue 17, 2004, pp. 3253-3255.
[27]. F. Chen, Z. Q. Cao, Q. S. Shen, X. X. Deng, B. M. Duan, W. Yuan, M. H. Sang, S. Q. Wang, Nanoscale displacement measurement in a variable-air-gap optical waveguide, *Appl. Phys. Lett.*, Vol. 88, Issue 16, 2006, pp. 161111.
[28]. F. Chen, Z. Q. Cao, Q. S. Shen, X. X. Deng, B. M. Duan, W. Yuan, M. H. Sang, S. Q. Wang, Picometer displacement sensing using the ultrahigh-order modes in a submillimeter scale optical waveguide, *Opt. Express*, Vol. 13, Issue 25, 2005, pp. 10061-10065.
[29]. T. Y. Yu, H. G. Li, Z. Q. Cao, Y. Wang, Q. S. Shen, Y. He, Oscillating wave displacement sensor using the enhanced Goos-Hänchen effect in a symmetrical metal-cladding optical waveguide, *Opt. Lett.*, Vol. 33, Issue 9, 2008, pp. 1001-1003.
[30]. S. Z. Zhang, S. Kiyono, Y. Uda, Nanoradian angle sensor and in situ self-calibration, *Appl. Opt.*, Vol. 37, Issue 19, 1998, pp. 4154-4159.
[31]. F. Chen, Z. Q. Cao, Q. S. Shen, Y. J. Feng, Optical approach to angular displacement measurement based on attenuated total reflection, *Appl. Opt.*, Vol. 44, Issue 26, 2005, pp. 5393-5397.

[32]. L. Chen, Z. Q. Cao, Q. S. Shen, X. X. Deng, F. Ou, Y. J. Feng, Wavelength sensing with subpicometer resolution using ultrahigh order modes, *J. Lightw. Technol.*, Vol. 25, Issue 2, 2007, pp. 539-543.

[33]. Y. Wang, X. G. Jiang, Q. Li, Y. Wang, Z. Q. Cao, High-resolution monitoring of wavelength shifts utilizing strong spatial dispersion of guided modes, *Appl. Phys. Lett.*, Vol. 101, Issue 6, 2012, pp. 061106.

[34]. G. Chen, Z. Q. Cao, J. H. Gu, Q. S. Shen, Oscillating wave sensors based on ultrahigh-order modes in symmetrical metal-clad optical waveguides, *Appl. Phys. Lett.*, Vol. 89, Issue 8, 2006, pp. 081120.

[35]. J. H. Gu, G. Chen, Z. Q. Cao, Q. S. Shen, An intensity measurement refractometer based on a symmetric metal-clad waveguide structure, *J. Phys. D: Appl. Phys.*, Vol. 41, Issue 18, 2008, pp. 185105.

[36]. Y. Wang, H. G. Li, Z. Q. Cao, T. Y. Yu, Q. S. Shen, Y. He, Oscillating wave sensor based on the Goos-Hänchen effect, *Appl. Phys. Lett.*, Vol. 92, Issue 6, 2008, pp. 061117.

[37]. Y. Wang, M. Z. Huang, X. Y. Guan, Z. Q. Cao, F. Chen, X. P. Wang, Determination of trace chromium (VI) using a hollow-core metal-cladding optical waveguide sensor, *Opt. Express*, Vol. 21, Issue 25, 2013, pp. 31130-31137.

[38]. Z. Han, L. Qi, G. Shen, W. Liu, Y. Chen, Determination of Chromium(VI) by Surface Plasmon Field-Enhanced Resonance Light Scattering, *Anal. Chem.*, Vol. 79, Issue 15, 2007, pp. 5862-5868.

[39]. C. J. Miles, H. A. Moye, Extraction of glyphosate herbicide from soil and clay minerals and determination of residues in soils, *J. Agric. Food Chem.*, Vol. 36, Issue 3, 1988, pp. 486-491.

[40]. C. Hidalgo, C. Rios, M. Hidalgo, V. Salvado, J. V. Sancho, F. Hernandez, Improved coupled-column liquid chromatographic method for the determination of glyphosate and aminomethylphosphonic acid residues in environmental waters, *J. Chromatogr. A*, Vol. 1035, Issue 1, 2004, pp. 153-157.

[41]. H. L. Dai, M. H. Sang, Y. X. Wang, R. Du, W. Yuan, Z. H. Jia, Z. Q. Cao, and X. F. Chen, Determination of trace glyphosate in water with a prism coupling optical waveguide configuration, *Sens. Actuators. A*, Vol. 218, Issue 10, 2014, pp. 88-93.

[42]. J. Zhou, Z. Q. Cao, Y. L. Chen, Y. X. Chen, M. Sun, D. C. Sun, and F. M. Li, Observation of switching phenomena in a nonether polyphenylquinoxaline planar waveguide with two-wavelength, *Opt. Lett.*, Vol. 22, Issue 19, 1997, pp. 1482-1484.

[43]. T. Xu, L. M. Huang, C. Yin, Y. L. Jin, J. H. Fang, M. Z. Huang, Enhanced Raman scattering assisted by ultrahigh order modes of the double metal cladding waveguide, *Appl. Phys. Lett.*, Vol. 105, Issue 16, 2014, pp. 163703.

[44]. C. Yin, J. J. Sun, X. P. Wang, C. P. Zhu, Q. B. Han, Z. Y. Di, Z. Q. Cao, Modulated reflectivity via a symmetrical metal cladding ferrofluids core waveguide chip, *EPL*, Vol. 100, Issue 4, 2012, pp. 44001.

[45]. X. P. Wang, C. Yin, J. J. Sun, H. G. Li, M. H. Sang, W. Yuan, Z. Q. Cao, M. Z. Huang, All-optically tunable Goos-Hänchen shift owing to the microstructure transition of ferrofluid in a symmetrical metal-cladding waveguide, *Appl. Phys. Lett.*, Vol. 103, Issue 15, 2013, pp. 151113.

Chapter 4
Garnet-like Solid State Electrolyte $Li_6BaLa_2Ta_2O_{12}$ Based Potentiometric CO_2 Gas Sensor

Yongming Zhu, Venkataraman Thangadurai, Werner Weppner

Abstract

A new electrochemical cell for CO_2 sensing has been set up using Garnet-like solid electrolyte $Li_6BaLa_2Ta_2O_{12}$ and $Li_{0.36}WO_3$ as reference electrode. The electromotive force (EMF) of the cell, the reproducibility and the long-term stability of the sensor signals have been examined at temperature between 300-500 °C exposed to dry atmosphere. The EMF is dependent on the partial pressure of CO_2 and can be expressed by the Nernst equation. The sensitivities of the sensors were 73 mV/decade and 75 mV/decade at 450 and 500 °C, respectively. It was also seen that the sensor showed speedy response kinetics to CO_2. The cell is long-term stable at temperature as low as 450 °C and the sensor signals are reproducible with high accuracy.

Keywords: Garnet-like, $Li_6BaLa_2Ta_2O_{12}$, $Li_{0.36}WO_3$; CO_2, gas sensor.

4.1. Introduction

There have been increasing needs for all-solid-state CO_2 sensors which are reliable, inexpensive and compact, from the recent deepening concern about the emissions of CO_2 from the global warming issue as well as from the growing needs of CO_2 monitoring in various fields and the control of CO_2 in various technologies. Potentiometric devices combining a solid electrolyte with an auxiliary phase of carbonate are eminently suited for practical applications [1]. Lithium electrolyte based sensors are known to be promising due to their faster and less water reactive nature of the lithium ion than any other alkali metal [2, 3]. However, they have some problems such as poor long-term stability [4].

Yongming Zhu
Faculty of Applied Chemistry, Harbin Institute of Technology (Weihai), 264209 Weihai, China

There is still much need for developing new sensing materials and patterns that are consistent to work at low operating temperatures. Recently, the Li-excess garnet-related compounds $Li_5La_3A_2O_{12}$(A=Nb, Ta) have been found as a new class of fast Li-ion conductors with high bulk-conductivity σ_b around 10^{-6} S/cm at 25 °C [5].The Li concentration could increase by substitution of La^{3+} ions by divalent alkaline-earth metal ions, which would also improve the conductivity. In particular, $Li_6BaLa_2Ta_2O_{12}$ exhibits the highest total (bulk plus grain boundary) conductivity $\sigma_{total}=4.0\times10^{-5}$ Scm^{-1} at 22 °C (E_a=0.40 eV) among these compounds [6–10].

In the present work, a novel solid state potentiometric CO_2 sensor was fabricated using the garnet-like fast lithium ion conductor $Li_6BaLa_2Ta_2O_{12}$ as the solid electrolyte. Li_2CO_3 and Au paste were employed as the sensing electrode, because pure Li_2CO_3 phase has a low ionic conductivity in general working temperature range of electrochemical CO_2 sensor [4, 11]. Thus, Au paste was mixed in Li_2CO_3 to get a sufficient ionic conductivity. Lithium tungsten oxide, Li_xWO_3 (x=0.36), was used as the reference electrode, due to its good electronic conductivity and stable voltage plateau versus elemental lithium.

4.2. Experimental

4.2.1. Fabrication of Sensor Devices

Compound with the chemical formula $Li_6ALa_2Ta_2O_{12}$ (A=Sr, Ba) were found to have a garnet-like structure [5], whereas in this work the composition of $Li_6BaLa_2Ta_2O_{12}$ was prepared according to the solid state reaction process.

$$3Li_2CO_3+BaCO_3+La_2O_3+Ta_2O_5\rightarrow Li_6BaLa_2Ta_2O_{12}+4CO_2 \quad (4.1)$$

Using appropriate amounts of high purity (>99.9 %) La_2O_3 (pre-dried at 900 °C for 24 h), Li_2CO_3, $BaCO_3$, and Ta_2O_5 obtained from Fluka or Aldrich. 10 wt% excess of Li_2CO_3 was added to compensate loss due to volatilization of Li_2O during the preparation. The mixtures were heated in air and maintained at 700 °C for 6 h and then annealed at 900 °C for 24 h. Before and after the first step, the powders were ball-milled using zirconium balls for about 12 h in 2-propanol. For the second step of heat treatment, the reaction products were pressed into pellets by isostatic pressure and covered with powder of the same mother composition to reduce the loss of lithium because of volatilization. Afterwards, the raw material was cut into a shape of disk 8 mm in diam and 1 mm thick.

Li_xWO_3 (x=0.36) has the *Pm3m* cubic structure [12], whereas in this work Li_xWO_3 (x=0.36) was prepared according to the solid state reaction process. The following chemical reaction formula was used:

$$\frac{x}{2}Li_2WO_4+\frac{3-2x}{3}WO_3+\frac{x}{6}W\rightarrow Li_xWO_3 \quad (4.2)$$

Take appropriate amounts of Li_2WO_4 (Strem, 98 %), WO_3 (Alfa, 99.99+ %) and W (Alfa, 99.9 %), ball milling for 12 h used zirconium balls with 2-propanol, then dried in air

overnight. The sintered step should finish in Ar gas at temperature of 750 °C for 2 h with heating and cooling rates of 1 C/min. After that, the powder was isostatically pressed at 120 kN for 20 min, then calcined at 750 °C for 2 h in Ar. It was also cut into the shape of a disk of 8 mm in diameter and 1 mm in thickness typically.

The prepared solid electrolyte and Li_xWO_3 (x=0.36) were pressed together, and the assembly was sealed on the side surface with high temperature ceramic adhesive (940 Cotronics Corp., N.Y.). Layer of Li_2CO_3 (Merck, 99 %) and nano Au particles (Product No.C5755A, Heraeus, Inc.) (Li_2CO_3: Au = 3:2 in wt%) were screen-printed on the top side of the electrolyte, on the reference side nano Au particles were screen-printed on. Two Au wires were attached on both sides, then the assembly was heat treated at 600 °C for 2 hrs to remove the organic components. After that, ceramic adhesive was used to seal reference side. The schematic structure was depicted in Fig. 4.1.

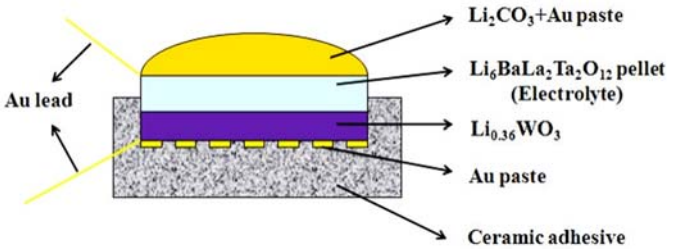


Fig. 4.1. Schematic view of the planar CO_2 sensor.

The crystal structure of the samples was characterized through X-ray diffract meter (Seifert, XRD 3000TT, using Cu K_α radiation at wavelength λ=0.1541 nm). The ionic conductivity was measured by AC impedance method using a frequency impedance analyzer (Hewlett Packard, model 4192A) in the frequency range of 5 Hz – 13 MHz and the temperature regime is between room temperature and 500 °C.

4.2.2. Evaluation of Sensing Properties

Sensor characteristics were measured using a testing chamber (Kiel Cell KC 2.0, Ionic Systems Inc.), consisting of a ceramic sensor-mounting inside a quartz-glass tube (with diameter of inside glass tube 35 mm and length of glass tube 350 mm) placed into a split tube laboratory furnace (TF 2.0, Ionic Systems Inc.), provided with a gas inlet and outlet. The sensors were heated externally through the furnace during measurement. The EMF was recorded in different CO_2 concentrations ranging from 1000 ppm (0.1 %) up to 100,000 ppm (10 %), using a Keithley 2000 digital multimeter (Keithley Instruments Inc., USA) with an input resistance of >10 GΩ and adapted data collection software. CO_2 concentrations were adjusted using different mass flow controllers (Tylan 2900 Series) varying the flows of dry synthetic air (21 vol. % O_2 in N_2), 1 % CO_2 (1 % vol. CO_2 in synthetic air) and 10 % CO_2 (10 % vol. CO_2 in synthetic air) through the quartz tube while keeping the overall flow rate constant at 70 ml/min. The sample holder needs 14.1 min to get a new CO_2 gas concentration under this flow rate.

4.3. Results and Discussion

4.3.1. Sensing Characteristics

From a microscopic point of view, the sensing reaction processes start at the triple-phase boundary (TPB). The process of O_2 and CO_2 incorporation have the gas phase into the solid electrolyte is schematically represented in Fig. 4.2.

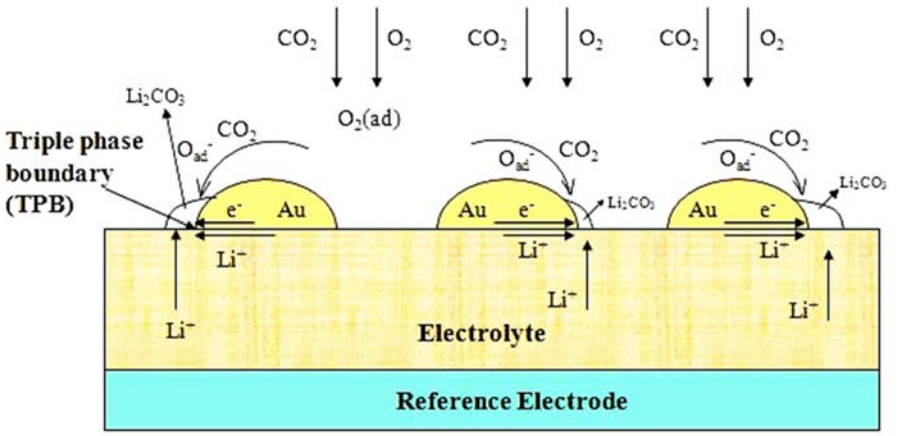


Fig. 4.2. Schematic representation of the galvanic cell reaction processed for the formation of Li_2CO_3 in a CO_2 sensor.

The process of O_2 and CO_2 incorporation may be therefore split into the following steps:

a) Oxygen and carbon dioxide diffusion in the gas phase, and oxygen dissociative adsorption of O_2 molecules at the surface of Au and at the uncovered surface area of the electrolyte (lithium ion conductor).

b) Diffusion of adsorbed oxygen along the surface of Au to the TPB.

c) Transfer of electrons to form O_{ad}^- and O^{-2} along the diffusion path of the Au surface.

d) Diffusion of lithium ions in the electrolyte and electrons in Au to the triple-phase boundary, within the contact zone to react with O^{-2} and CO_2 to form Li_2CO_3.

The sensor is composed of the following electrochemical cell:

$$Au, Li_{0.36}WO_3^{\ I} \mid Li_6BaLa_2Ta_2O_{12} \mid^{II} Li_2CO_3, Au, CO_2, O_2 \tag{4.3}$$

where, *I* and *II* show the interfaces between two component materials contacted to one another.

At interface *I* (reference electrode)

$$2Li_{0.36}WO_3 \leftrightarrow 2Li_{0.36-\delta}WO_3 + 2Li^+ + 2e^- \quad (4.4)$$

At interface *II* (sensing electrode)

$$2Li^+ + 2e^- + CO_2 + \frac{1}{2}O_2 \leftrightarrow Li_2CO_3 \quad (4.5)$$

During sensing CO_2 gas, lithium ions partially released from $Li_{0.36}WO_3$ and transferred through the solid state electrolyte $Li_6BaLa_2Ta_2O_{12}$. Meanwhile, CO_2 molecules easily diffused into the porous auxiliary layer of the sensor, leading to quick adsorption and desorption of CO_2 on the reactive sites (TPB) and then to equilibrium with lithium ions and O_2 into Li_2CO_3. When the rates of Eq. (4.4) and Eq. (4.5) are equal, the difference of the sensing and reference electrode potentials is measured as the sensing signal. And the overall cell reaction is given by Eq. (4.6).

$$2Li_{0.36}WO_3 + CO_2 + \frac{1}{2}O_2 \rightarrow Li_2CO_3 + 2Li_{0.36-\delta}WO_3 \quad (4.6)$$

4.3.2. Characterization of Electrolyte and Auxiliary Layer

Garnets are orthosilicates with the general structural formula, $A_3^{II}B_2^{III}(SiO_4)_3$, where A and B refer to eight-coordinated and six-coordinated cation sites [13], respectively. SiO_4 tetrahedral is isolated and connected to each other through ionic bonds with the interstitial B-cations. A large variety of complex oxides have also been found to crystallize in garnet-like structures with other elements replacing silicon, for example, $A_3B_5O_{12}$ (A=Ca, Mg, Y, La or rare earth; B= Al, Fe, Ga, Ge, Mn, Ni, V). Fig. 4.3 exhibits the measured and calculated powder XRD patterns of $Li_6BaLa_2Ta_2O_{12}$ sintered at 900 °C for 24 h in air, which was corresponding to the reference [14]. The powder XRD pattern for $Li_6BaLa_2Ta_2O_{12}$ was calculated using the atomic positions from single crystal data employing software *Powder Cell* program. As expected, the cubic lattice constant for $Li_6BaLa_2Ta_2O_{12}$ is 12.948(2) Å, the 8-fold coordination ionic radius for Ba^{2+} is 1.43 Å. According to JCPDF data (No.76-1497), $Li_{0.36}WO_3$ retains its crystallographic phase (cubic) and there is no new phase in the as prepared cubic structure (Fig. 4.4). The Arrhenius plot for the total (bulk + grain-boundary) electrical conductivity of $Li_6BaLa_2Ta_2O_{12}$ was shown in Fig. 4.5. The data obtained from heating and cooling cycle follow the same line with activation energy of 0.40 eV for $Li_6BaLa_2Ta_2O_{12}$, which is comparable to those values previously reported for the lithium ion conductor LISICON (0.56 eV) [15] and higher than that of Li_3N (0.25 eV) [16]. It should be mentioned that total and bulk conductivities show nearly the same magnitude for these garnet-like materials [14].This finding is a most attractive feature of the investigated garnet-like oxides compared to other ceramic lithium ion conductors.

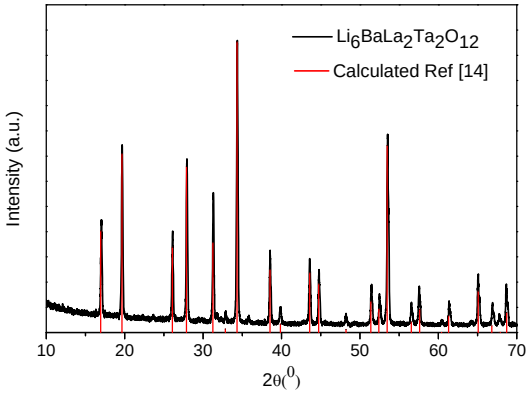


Fig. 4.3. The powder XRD pattern for $Li_6BaLa_2Ta_2O_{12}$ sintered in 900 °C for 24 h in air.

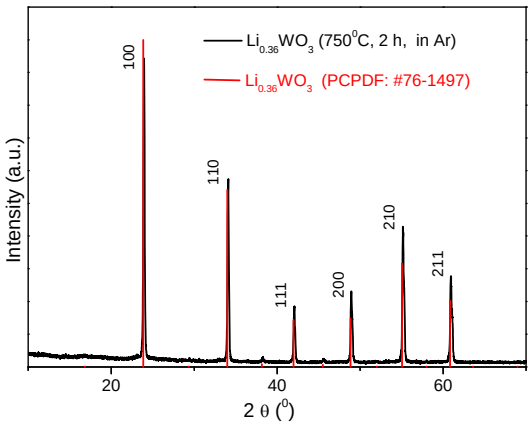


Fig. 4.4. The powder XRD pattern for $Li_{0.36}WO_3$ sintered at 750 °C for 2 h in Ar.

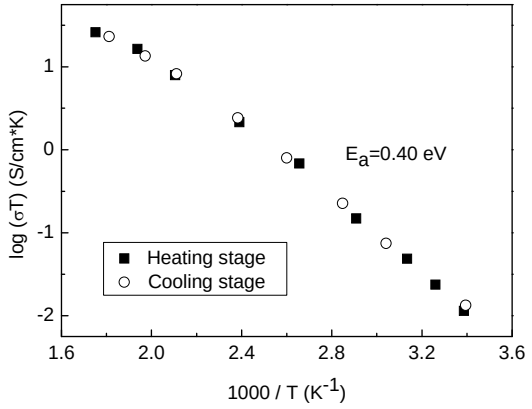


Fig. 4.5. Arrhenius plot of $Li_6BaLa_2Ta_2O_{12}$ electrolyte. (■) represents heating stage and (○) represents cooling stage.

4.3.3. CO_2 Sensing Properties

The EMF measurements were performed at 450 and 500 °C by changing the CO_2 concentration from 0.01 % to 10 % with the O_2 partial pressure remaining the same at both the reference electrode and the sensing electrode. The results are shown in Fig. 4.6 and Fig. 4.7. Upon introducing CO_2 gas into the cell, the Li activity of the sensing electrode decrease leading to an EMF decrease. Concerning the gas exchange time (14.1 min), the sensor shows reproducible response with response time (response time is defined as time taken to reach 90 % of the final signal) of less than 1 minute as shown in Fig. 4.6 and Fig. 4.7. The sensitivities of the linear fit were 73 mV/decade (calculated sensitivity is 72 mV/decade) and 75 mV/decade (calculated sensitivity is 76 mV/decade) at 450 °C and 500 °C, respectively. Fig. 4.8 shows the dependence of EMF as a function of log of CO_2 concentration (log PCO_2) at 450 and 500 °C. The sensitivities of cells agree within the experimental error with the calculated ones from Nernst's equation for a two–electron process. The sensitivity to CO_2 translated to the nominal number of electrons involved in the cell reaction, provides a value 1.96 at 450 °C and 2.02 at 500 °C (Fig. 4.8).

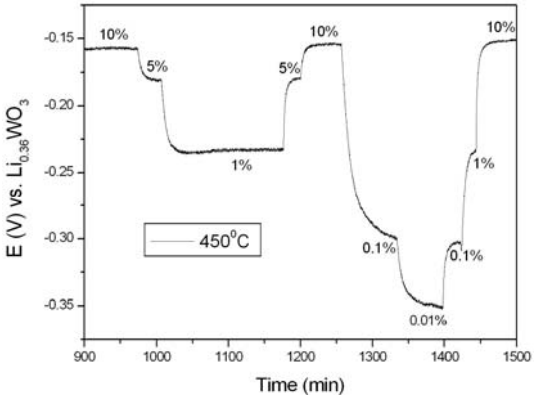


Fig. 4.6. Transient response to the cell upon changes to the CO_2 partial pressure at 450 °C.

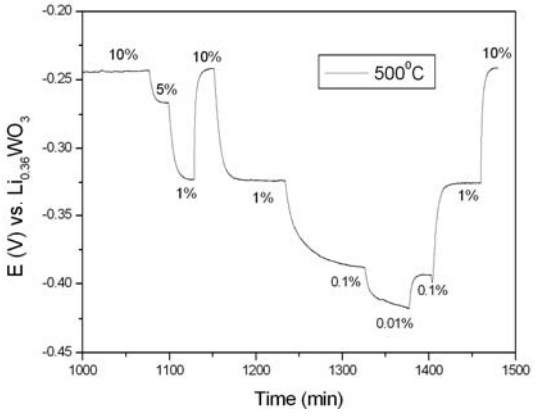


Fig. 4.7. Transient response to the cell upon changes to the CO_2 partial pressure at 500 °C.

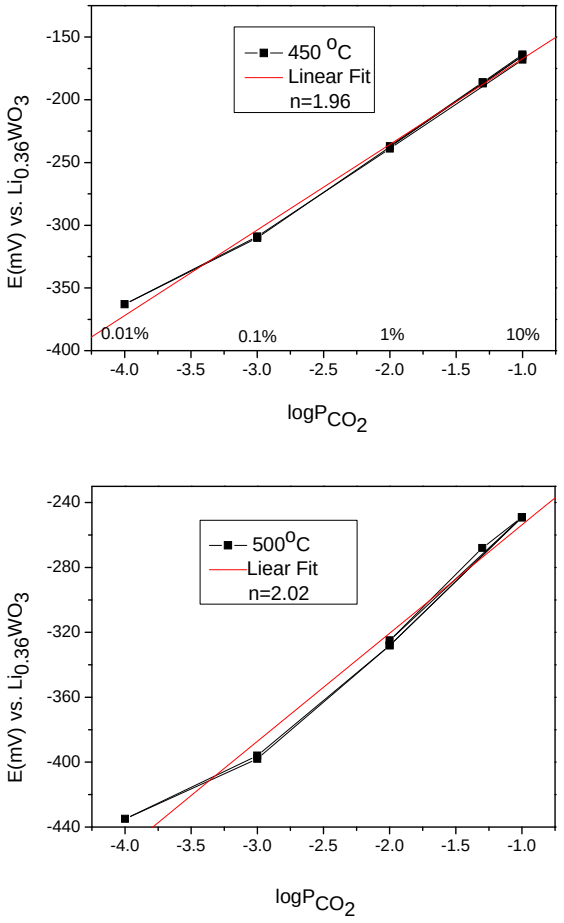


Fig. 4.8. Open circuit voltages of the cell at (top) 450 °C and (bottom) 500 °C.

It is obvious that the slope became larger with the increase of the working temperature, due to the improvement of the ionic kinetics at the effective area of the triple-phase boundary. This also indirectly indicates that the effective area of the triple-phase boundary plays an important role in increasing the sensitivity (slope) of the sensor, although the reference electrode is same. The EMF is a linear function of log CO_2 partial pressure over all the temperatures (300-500 °C) (Fig. 4.9). And the nearly no deviation from the theoretical slope indicates that there is no electronic conduction is present in the $Li_6BaLa_2Ta_2O_{12}$ electrolyte during the EMF measurement [17].

It should be mentioned that two main factors play the key roles in determining the sensitivities of this device [18]. Firstly, high electrochemical catalytic activity of the sensing electrode material to CO_2 is necessary. Secondly, the microstructure of the triple-phase boundary will affect the sensitivity. All the above-mentioned factors are entangled with each other in a complicated manner, finally determined the sensitivity to CO_2. In this work, special attention was paid on the TPB. The contact area of the electrolyte and the

sensing electrode will enlarge as the temperature increase. Larger contact area can supply more electrochemically active sites for the reaction; finally result in the increase of the sensitivity to CO_2.

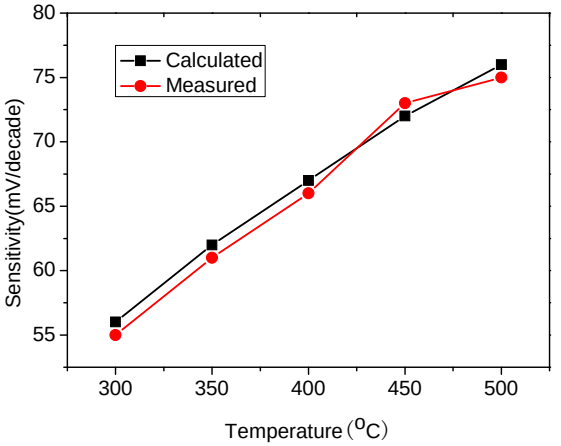


Fig. 4.9. Sensitivities of the cell between 300 °C and 500 °C.

In order to testify it, during operating the cell at elevated temperatures, impedance spectroscopy was applied. The complex impedance was measured in the frequency range from 5 Hz to 13 MHz employing a sinusoidal perturbation with 100 mV amplitude in all measurements. The frequency dispersion in the impedance plots consists of a semicircle with its value on the real axis to be the interfacial resistance. Fig. 4.10 shows the complex impedance plots at 450 °C and 500 °C. The interfacial conductivities were 6.5 x 10^{-4} S/cm at 450 °C and 8 x 10^{-4} S/cm at 500 °C. The corresponding interfacial capacitances as evaluated from the complex impedance plots were 29 nF and 27 nF at 450 °C and 500 °C, respectively. During operating the cell at variable times, impedance spectroscopy was applied. The vales at days 3-19 when the cell remain under open circuit voltage at 500 °C are shown in Fig. 4.11. The corresponding interfacial capacitances as evaluated from the complex impedance plots were 21 nF at day 3 and 9.4 nF at day 19. During aging, the contact between the sensing electrode and electrolyte was improved considerably. And the total conductivities increased with aging time. This phenomenon was also reported by Song et al. [19]. This better contact between electrode and electrolyte results in enhanced interfacial adherence and improved charge transfer kinetics. The decrease of the cell resistivity, as shown in Fig. 4.11, may be due to this reason. The sensing electrode reaction (Eq. 4.5) requires the reaction of lithium ions from the electrolyte, electrons from the electronic lead, CO_2 from the gas phase and O_2 from a dissociated oxygen molecule in the gas phase, in a single reaction step. At lower temperatures this reaction is not kinetically favorable and intermediate products with less negative Gibbs energies of formation may be formed. Only the reaction step that involved electrons from the metallic conductor contributes to the cell voltage. Purely chemical reactions at the interface occurring electrolyte-gas do not contribute. This may cause the sluggish response, especially at lower temperatures, where the formation of Li_2CO_3 is impeded by slow electrode kinetics.

The discrepancy of 1.5 mV/decade at 300-400 °C is due to this reason. The sensitivity to CO_2 increased gradually with temperature also suggesting that the *in situ* formation of Li_2CO_3 is kinetically favorable only at high temperature whereas at intermediate or low temperatures the intermediate compound lithium oxide is formed at the measuring side. These results suggest that the microstructure of an auxiliary layer for solid-electrolyte gas sensors is highly effective for improving the gas sensor properties.

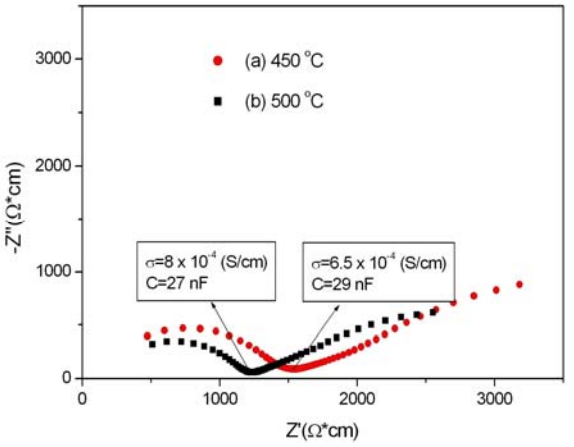


Fig. 4.10. Complex plane impedance plots for the cell at (a) 450 °C and (b) 500 °C. The frequency range is from 5 Hz to 13 MHz.

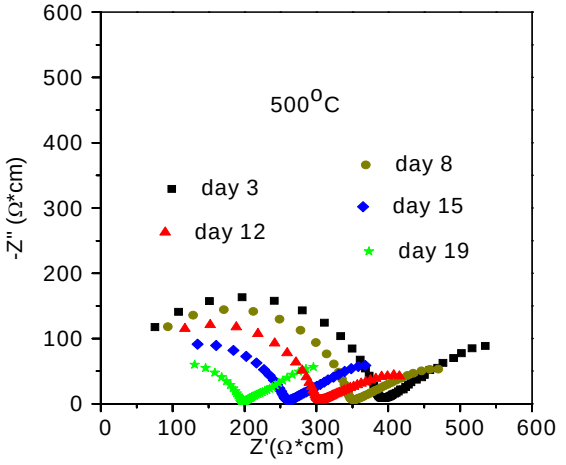


Fig. 4.11. Complex plane impedance plots of the cell for different times at 500 °C under 1 % CO_2.

Fig. 4.12 demonstrates the high stability of the response of the cell to 1 % CO_2 in dry synthetic air over elapsed periods of time. The measured cell voltage is the sum of all voltages contributing to the total redox reaction. During the aging time, EMF of the sensor

always keeps on -237 mV. The long term stability test of several cells showed, after aging of the sensors, further shift of the absolute value of the open circuit voltage between 2-8 mV per month independently on the sealing qualities of the ceramic glue. This behavior is well known for this kind of sensor which relies on *in situ* reaction in the electrode and is assumed to be caused by a loss of Li_2CO_3 at the TPB. For the sensor with sole Li_2CO_3 as an auxiliary phase, the EMF decreased with time and any steady state could not be observed in the experimental period [20]. It is known that the reforming of the auxiliary phase by adding Au results in the recovery of the sensing performances.

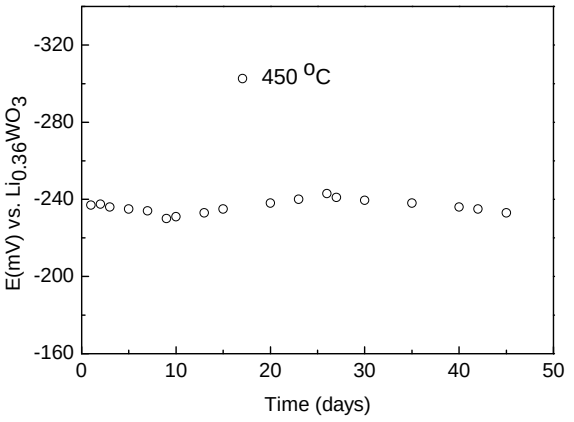


Fig. 4.12. Long term stability of the cell at 450 °C under 1 % CO_2.

4.4. Conclusions

The performance of a novel CO_2 sensor based on the garnet-like $Li_6BaLa_2Ta_2O_{12}$ solid electrolyte and $Li_{0.36}WO_3$ reference electrode operating at 300-500 °C has been investigated. The sensor is long-term stable and the signals are reproducible with high accuracy. The miniature CO_2 sensor with 450 °C working temperature shows high CO_2 sensitivity (73 mV/decade), short response and recovery time, and acceptable long term stability. With the realization of miniature structure and lower heating temperature, the power consumption was effectively decreased in comparison with the traditional tube-type sensors. The corresponding impedance analyses suggest that the microstructure an auxiliary layer for solid-electrolyte gas sensors is highly effective for improving the gas sensor properties.

Acknowledgment

The financial support of Scientific Research Foundation for the Returned Overseas Chinese Scholars, State Education Ministry (2010-149) and Natural Scientific Research Innovation Foundation in Harbin Institute of Technology (HIT.NSRIF.2011099) is gratefully acknowledged.

Reference

[1]. T. Murayama, S. Sasaki, Y. Saito, Potentiometric gas sensors for carbon dioxide using solid electrolyte, *Solid State Ionics,* 23, 1987, pp. 107-112.

[2]. W. S. Noh, L. Satyanarayana, J. S. Park, Potentiometric CO_2 sensor using Li^+ ion conducting Li_3PO_4 thin film electrolyte, *Sensors,* 5, 11, 2005, pp. 465–472.

[3]. L. Satyanarayana, G. P. Choi, W. S. Noh, W. Y. Lee, J. S. Park, Characteristics and performance of binary carbonate auxiliary phase CO_2 sensor based on Li_3PO_4 solid electrolyte, *Solid State Ionics*, 177, 39–40, 2007, pp. 3485–3490.

[4]. C. O. Park, C. Lee, S. A. Akbar, J. Hwang, The origin of oxygen dependence in a potentiometric CO_2 sensor with Li-ion conducting electrolytes, *Sens. Actuators B: Chem*., 88, 2003, pp. 53–59.

[5]. V. Thangadurai, H. Kaack, W. Weppner, Novel Fast Lithium Ion Conduction in Garnet-Type $Li_5La_3M_2O_{12}$ (M = Nb, Ta), *J. Am. Ceram. Soc.*, 86, 3, 2003, pp. 437-440.

[6]. V. Thangadurai, W. Weppner, $Li_6ALa_2Nb_2O_{12}$ (A = Ca, Sr, Ba, pp. A New Class of Fast Lithium Ion Conductors with Garnet-Like Structure, *J. Am. Ceram. Soc.*, 88, 2, 2005, pp. 411-418.

[7]. V. Thangadurai, W. Weppner, Effect of Sintering on the Ionic Conductivity of Garnet-Related Structure $Li_5La_3Nb_2O_{12}$, and In- and K-Doped $Li_5La_3Nb_2O_{12}$, *J. Solid State Chem.*, 179, 2006, pp. 974-984.

[8]. V. Thangadurai, W. Weppner, $Li_6ALa_2Ta_2O_{12}$ (A = Sr, Ba, pp. Novel Garnet-Like Oxides for Fast Lithium Ion Conduction, *Adv. Funct. Mater.*, 15, 1, 2005, pp. 107-112.

[9]. V. Thangadurai, W. Weppner, Investigations on Electrical Conductivity and Chemical Compatibility Between Fast Lithium Ion Conducting Garnet-like $Li_6BaLa_2Ta_2O_{12}$ and Lithium Battery Cathodes, *J. Power Sources*, 142, 2005, pp. 339-344.

[10]. R. Murugan, V. Thangadurai, W. Weppner, Parameter and Sintering Temperature Dependence of Bulk and Grain-Boundary Conduction of Garnet-like Solid Li-Electrolytes, *J. Electrochem. Soc.*, 155, 2008, pp. A90-A101.

[11]. F. Ménil, B. O. Daddah, P. Tardy, H. Debéda, C. Lucat, Planar LiSICON-based potentiometric CO_2 sensors: Influence of the working and reference electrodes relative size on the sensing properties, *Sens. Actuators B: Chem.*, 107, 2, 2005, pp. 695–707.

[12]. Q. Zhong, J. R. Dahn, and K. Colbow, Lithium intercalation into WO_3 and the phase diagram of Li_xWO_3, *Phys. Rev. B*, 46, 4, 1992, pp. 2554-2560.

[13]. A. F. Wells, Structural Inorganic Chemistry, 5th ed., Oxford Science Publications, *Clarendon Press,* Oxford, UK 1984.

[14]. R. Murugan, V. Thangadurai, W. Weppner, Effect of Lithium ion content on the Lithium ion Conductivity of the Garnet-like Structure $Li_{5+x}BaLa_3Ta_2O_{11.\ 5+0.\ 5x}$ (x=0-2), *Appl. Phys.,* A, 91, 2008, pp. 615-620.

[15]. U. V. Alpen, M. F. Bell, W. Wichelhaus, K. Y. Cheung, G. J. Dudley, Ionic conductivity of $Li_{14}Zn(GeO_{44}$ (Lisicon), *Electrochim. Acta,* 23, 12, 1978, pp. 1395-1397.

[16]. U. V. Alpen, A. Rabenau and G. H. Talat, Ionic conductivity in Li_3N single crystals, *Appl. Pys. Lett.,* 30, 1977, pp. 621-623.

[17]. J. W. Fergus, A review of electrolyte and electrode materials for high temperature electrochemical CO_2 and SO_2 gas sensors, *Sens. Actuators B: Chem.*, 134, 2008, pp. 1034–1041.

[18]. C. O. Park, J. W. Fergus, N. Miura, Jinsu Park, Angi Choi, Solid-state electrochemical gas sensors, *Ionics*, 15, 3, 2009, pp. 261-284.

[19]. S. W. Song, L. P. Martin, R. S. Glass, E. P. Murray, J. H. Visser, R. E. Soltis, R. F. Novak, D. J. Kubinski, Aging Studies of Sr-Doped $LaCrO_3$/YSZ/Pt Cells for an Electrochemical NO_x Sensor, *J. Electrochem. Soc.*, 153, 9, 2006, pp. H171-H180.

[20]. T. Kida, H. Kawate, K. Shimanoe, N. Miura, N. Yamazoe, Interfacial structure of NASICON-based sensor attached with Li_2CO_3–$CaCO_3$ auxiliary phase for detection of CO_2, *Solid State Ionics*, 136-137, 2000, pp. 647–653.

Chapter 5
The Characteristics of Residual Tree-ring CO_2 and H_2O Chronologies for Conifer Species

Boris G. Ageev, Aleksandr N. Gruzdev, Dmitry A. Savchuk, Yurii N. Ponomarev and Valeria A. Sapozhnikova

Abstract

The results obtained from prolonged investigations into the behavior of residual gas components in tree-ring wood of several conifer discs are presented. An integrated approach to the investigations and analysis of CO_2 and H_2O vacuum-extracted from tree-ring wood enable the behavioral features of annually accumulated (sobbed) gas components of porous wood structure, including H_2O vapor and plant cell-respired CO_2, to be revealed. The use of laser photoacoustic gas analyses provides long tree-ring chronologies of these components and suggests an association between CO_2 and H_2O chronologies and cyclic variations in the total pressure of the gas components in tree rings. The relation of CO_2 and H_2O chronologies with meteorological parameters is examined by wavelet, spectral, and cross-spectral techniques.

5.1. Introduction

Research being performed in the present period of time is of great interest from the standpoint of investigating the effect of sharp changes in atmospheric conditions on plants. Tree is a unique indicator of changes brought about by variations in the CO_2 content in the atmosphere and those found in the C isotope composition in CO_2. The aforementioned changes should influence photosynthesis in plants and translate into changes in certain wood parameters. Trees efficiently adapt themselves to changing environmental conditions, which affects both isotope composition and tree ring width chronologies [1].

Boris G. Ageev
V.E. Zuev Institute of Atmospheric Optics SB RAS, Tomsk 634021, Russia

Tree rings are generally considered to be records of past climatic conditions and environmental changes [2]. However, researchers face some problems involved in investigating the effect of atmospheric CO_2 rise on plants and in estimating internal and external CO_2 release by trunks [3]. The atmospheric CO_2 rise and climate warming would seem to enhance the global annual growth of the tree ring widths. However, this is not the case. As pointed out in [4], a direct CO_2 fertilization effect is traceable to as few as 20 % of the areas worldwide.

Extensive literature has been devoted to climate reconstruction from tree ring width chronologies (see, for example, [5–11]). However, the complexity of the chemical and physical processes at work in biological objects gives no way of describing real situations by means of mathematical models that would adequately account for the environmental impacts on the tree growth [12]. In particular, there are many questions concerning the behavior and diffusion of gas components, like CO_2 and water vapor, through stem wood.

It is known that gases fill a sufficiently large part of tree stems. Thus temperate softwood species average between 18 and 50 vol.% of gas for sapwood and heartwood, respectively. Temperate hardwood species contain 26 vol.% of gas both in sapwood and in heartwood, and tropical species with mixed sapwood and heartwood account for 18 vol.% of gas [13 and 14]. Sapling stems of tropical moist-forest species consist, on average, of 26 % of wood, 59 % of water, and 15 % of gas (per unit volume) [15]. Unfortunately, until 2004 no attempts were made to explain the presence of such a large amount of gases in tree stems [13 and 14]. The latter fact was attributed to the need to increase the tree cross-sectional area without biomass investment and thus provide mechanical support of tree trunks [13 and 14]. It was concluded that tall trees were likely to contain a large amount of gases not just because the latter are cheap fillers, but also because this might lead to an increased stability of trees [15]. In mature trees, gas may be localized in intercellular spaces, pores within cell walls, parenchyma rays, and vessel, fiber, and tracheid lumens [15]. The composition of gases extracted from tree stems was analyzed for the first time early in the 20th century: O_2 (1.24 %), CO_2 (7.21 %), CH_4 (60.9 %), and N_2 (30.65 %) [16]. The stem CO_2 concentrations were subsequently found to vary within < 1–26 %, which is much higher than those detected in the atmosphere (0.04 %) [17]. The stem CO_2 pressure would amount to 22–69 Torr [18], exceeding the atmospheric CO_2 concentration by several orders of magnitude [19 and 20]. Stem CO_2 exists both in the gaseous and in the aqueous phases (dissolved CO_2). It can be released to the atmosphere by diffusion, transported to tree branches and leaves by transpiration flow, or re-fixed in green stem parts [16 and 21]. Notably, CO_2 is transported by the transpiration flow to the place where the CO_2 gas concentration is low, being transformed from the aqueous to the gaseous phase again [22]. The use of a procedure based on the measurement of stem-respired CO_2 to absorbed O_2 ratio showed that ~35 % of plant-respired CO_2 was brought to the crown by water flow and could be re-fixed in green woody tissues/leaves [22]. The CO_2 stem diffusion (stem respiration) is known to exhibit daily and seasonal dynamics (see, for example, [23]). The main source of stem CO_2 is the respiration of living inner bark cells, cambium, and xylem, even though experiments on the C^{13} addition to the transpiration flow show that a major part of CO_2 arrives from the root system [17]. The stem gas-water ratio demonstrates daily and seasonal variations. Besides, the amount of water increases

and that of gases decreases in spring and autumn [16]. While a great deal of effort was made to perform extensive long-term investigations into stem tissue respiration, i.e. stem CO_2 efflux, it is impossible to approximate this process by the CO_2 rate alone. That is why existing quantitative models of the CO_2 balance in trees call for considerable refinement. Moreover, the stem respiration varies widely from one tree to the other. That is why it is difficult to estimate the CO_2 balance for a forest as a whole. This is due to the fact that the physiological processes involved and the physical factors controlling the CO_2 efflux and/or influencing the radial diffusion of stem CO_2 to the atmosphere are very complicated and hard to understand [16]. It is not surprising, then, that considerable attention is given to the study on the CO_2 behavior in tree stems, because carbon dioxide is the main component of the forest carbon balance, and the problem of correct measurements of the stem CO_2 efflux is still high on the agenda [22].

There is more published information about the water content than about the tree stem gas concentration [13 and 14]. The wood-water relation has been studied for at least a century now [24]. It is known that in freshly sawn trees, wood cell walls are saturated with water, and, moreover, water, water-vapor mixture, and water vapor can reside in cell lumens. The moisture content (water mass-dry wood ratio) of green wood ranges from about 30 % to more than 200 % [25]. It was found that in a porous system such as wood, there were three phases of water: solid (ice), liquid, and gaseous (vapor). Since 1981, the following wood water states have been assumed to occur: free water, freezing bound water, and non-freezing bound water. Subsequently, the existence of freezing bound water was not confirmed, because this kind of water is not present in wood. In natural tree wood, there are only two kinds of water: free water, say, in cell lumens, and non-freezing bound water in cell walls. The main stem wood-water interaction is by hydroxyl groups of wood polymers [24]. The interaction appears to be complicated by the non-linearity of the bulk flow, capillary condensation of water vapor, gas dissolution and diffusion, migration of bound water through cell walls, etc. [26].

At the present time, great attention is focused on water transport in plants and on environmental impacts upon interactions on the molecular level [27]. However, in studying the behavior of CO_2 and H_2O in tree wood, researchers leave aside the problems of the residual CO_2 and H_2O content in stems and interannual pressure variations in disc tree rings. Using our photoacoustic spectrometer with a tunable CO_2 laser, it is possible both to trace CO_2 in discs, and to characterize the H_2O distribution over tree rings as well. The water content in discs of living trees has been studied by different methods (for example, by x-ray computed tomography [28]). However, we are not aware of any attempts to describe the annual water distribution over rings in dry discs. It should be noted that the laser photoacoustic method suggested here was tested not only on dry discs, i.e. on those stored under laboratory conditions for a few years, but on a disc of a living tree as well. A laser photoacoustic analysis has been successfully used in many applications for a long time [29–37], but we were the first to apply this approach to measuring the annual CO_2 and H_2O content in disc wood.

The findings of our investigations have shown that porous wood structure is capable of annual accumulation (sorption) of stem gas components, including water vapor and plant-

respired CO_2. This generates a need for development of additional chronologies that can be used to provide a better understanding of the behavior of the stem gases. In this work, we present the main results of our investigations into vacuum-extracted CO_2 and H_2O chronologies of several coniferous tree discs. An analysis of certain features inherent in the behavior of these chronologies have revealed, among other things, 1) A high correlation between CO_2 and H_2O chronologies, 2) A statistically significant correlation between 4-year pressure variations in vacuum-extracted gas samples and CO_2 content in 300 year old larch tree rings, and 3) A tree-ring CO_2 rise treated as resulting from changes in atmospheric parameters (atmospheric CO_2 rise changes the annual concentration gradient between atmospheric and stem CO_2). Consequently, the dynamics of this process should translate into year-to-year variations in stem CO_2. Moreover, the CO_2 and H_2O chronologies under study are characterized by short- and long-period variations, with the former being independent of tree age and the latter varying with tree age.

The measured results can be used by dendrochronologists, dendroecologists, and experts dealing with carbon budget and carbon dioxide flux estimations between terrestrial ecosystems and atmosphere. It is believed that the data about interannual CO_2 variations can be employed as a marker of total gas pressure variations in disc tree rings.

5.2. Materials and Techniques

5.2.1. A description of the Discs under Study

By now more than 1700 gas samples vacuum-extracted from 30 discs of several conifer species have been investigated. A study was made on the CO_2 (and H_2O) content in gas samples vacuum-extracted from the disc tree rings of a 300 year old Siberian larch (*Larix sibirica* Ledeb), a Siberian stone pine (*Pinus sibirica* Du Tour), and some coniferous trees that grew at 56°26'N and 85°03'E in Tomsk Oblast (West Siberia, Russia). The dry discs were stored from 6 months to a few years under laboratory conditions, so the wood material can be considered to be room-dried. The larch disc (Ø ~ 66 cm, h ~ 12 cm) placed at our disposal by Tomsk Forest Museum consisted of ~ 300 tree rings, 278 of which were examined. The Siberian stone pine disc was sawn at a height of 1.5 m of the tree located near Trubachevo settlement in Tomsk region. It had ~ 132 tree rings, 100 of which were investigated. It should be noted that the performance of the method suggested here was tested using not only dry discs, but a disc of a newly sawn Scots pine disc as well.

5.2.2. The Experimental Procedure and Data Processing Techniques

Our experimental system and procedure used for investigations into the CO_2 and H_2O content in disc tree-ring wood was described elsewhere [38–43]. The measurements were performed using a computer-controlled photoacoustic (PA) spectrometer with a frequency-tunable waveguide CO_2 laser. The PA spectrum was correlated with the optical absorption spectra of the sample. Once the system was calibrated, i.e. the gas absorption with known concentrations was measured, the calibration coefficient was found, and the absorbing component concentration in the gas sample studied was obtained. The laser

emits more than 70 lines within a tuning spectral range of 9.2–10.8 μm with an output power of 0.3–3 W in the single-frequency regime. The laser line tuning and detector signal recording and preliminary processing (averaging of the data and determination of the signal ratio) were performed by means of a dedicated software. The measured results were stored in a file for associated laser lines, providing an analog of the absorption spectrum for the sample. Information about the spectrum was displayed on a computer monitor to visualize the spectrometer data. The system is pre-calibrated using a CO_2/N_2 reference mixture containing a known amount of CO_2. The ultimate absorption coefficient sensitivity of the spectrometer used was $2^{*}10^{-5}$ cm^{-1} for a laser power of 70 mW, and the calibration measurement error was no more than ± 5 %. A further increase in the sensitivity was obtained for a total PA cell detector pressure of ~ 100 Torr so air was added to attain this pressure.

Gas samples were prepared in the following way. Tree ring wood was planned off with special chisels, placed in four sealed exposure chambers pumped out for a short time to stimulate sobbed gas diffusion, and allowed to stand for 20 minutes for subsequent measurements. In a recent experiment, samples of the annual tree ring wood of approximately the same weight (~ 7 g) were placed in the exposure chambers, and the evacuation of the exposure chambers was controlled by a manometer so that the same minimum pressure in the chambers was maintained during all the measurements. Unless this condition is fulfilled, there will be additional variations in the results obtained. In the experiments, the extracted gas pressure in each of the chambers for a 20 min. exposure time was also measured by the manometer. Each series of the gas absorption measurements on the samples extracted from tree ring wood was accompanied by measurements on background air absorption. The information obtained from the difference between the gas sample absorption and that of air enables the relative content of the absorbing component (CO_2 or H_2O) in the sample to be measured for each of the rings, using a calibration curve. The measurements were performed in four lines of the tunable waveguide 10.6 μm CO_2 laser: ***P*** (20, 16, and 14) and ***R*** (20) coinciding with the CO_2 absorption lines and the ethylene absorption line ***P*** (14) and with the water-vapor absorption line ***R*** (20). In our experiments, the PA signal in the ***P*** (14) line was absent, which is why ethylene was not detected in the gas samples. The CO_2 content was averaged over the values obtained for the three laser lines, and their associated correlation coefficients were 0.85–0.9. Measurements in the ***R*** (20) line allowed for detection of the signal from the sum of the gas components (CO_2 + H_2O), and on subtraction of the CO_2 signal the data about H_2O distributions over tree rings were acquired.

To verify the fact that CO_2 in the samples studied was generated by the trees themselves and not supplied from the atmosphere, an isotope analysis of carbon CO_2 desorbed from several rings was performed using a Delta V Advantage mass spectrometer with a binary startup system.

The chemical sample preparation procedure employed standard certified methods of primary organic chemistry. To investigate the isotope composition, CO_2 and H_2O were desorbed from tree ring wood in a stream of nitrogen gas at T = 80° C. The evolving components were collected in two traps: H_2O in a cooling trap (T = 4° C) and CO_2 in a

trap filled with Ba $(OH)_2$ solution wherein it was precipitated in the form of $BaCO_3$. Carbon dioxide was produced by a reaction of $BaCO_3$ with orthophosphoric acid. Then CO_2 was frozen and injected into ampoules. A standard sample was prepared and carbon isotope composition ($\delta^{13}C$) of the resultant CO_2 was measured relative to the standard sample. Using the ($\delta^{13}C$) notation, the isotope composition was expressed in terms of deviations from internationally accepted standards Vienna Pee Dee Belemnite (VPDB) for carbon:

$$\delta^{13}C, (‰) = [(R_{\text{sample}} / R_{\text{standard}}) - 1]*1000, \quad (5.1)$$

where R is the molar heavy-to-light isotope mass ratio and $R_{sample}/ R_{standard}$ is the 13C/12C ratio in the sample and in the standard, respectively. The stable carbon isotope composition, ($\delta^{13}C$) of CO_2 in tree ring wood was measured in Laboratory of Isotopic Methods, Tomsk, Russia (Certificate No. ROSS RU 001 517930), using the DELTA V Advantage mass spectrometer. The tree ring widths were measured on ground disc surfaces to within 0.01 mm by a semi-automatic LINTAB measuring system.

To estimate the periodic and temporal variations in the tree ring CO_2 and H_2O chronologies, use was made of high-resolution spectral and cross-spectral analyses based on the maximum entropy and wavelet techniques and digital filtration of the time series. Long-term trends were eliminated from the data being analyzed. The wavelet technique has lower resolution than the spectral analysis, but it provides a gain in time localization. Fourier analysis (Fast Fourier Transform) with the use of an ORIGIN software was employed for testing periodic signals.

5.3. Results

5.3.1. An Analysis of the Desorbed CO_2 Carbon Isotope Composition ($\delta^{13}C$)

An analysis of the desorbed CO_2 carbon isotope composition (δ^{13} C) in several annual disc tree rings was performed by the mass spectrometer. Dendroisotopic analyses are known to be widely used for detecting past changes in air quality, evaluating forest responses to air pollution, and accounting for variations in local amounts of precipitation or soil water [44]. However, we are not aware of any work dealing with the CO_2 carbon isotope composition of CO_2 desorbed from disc tree rings. It is known that leaves and tree wood are characterized by a lower carbon isotope composition (from −20 ‰ to −30 ‰) [45], while the carbon isotope composition of CO_2 in air is − 8.5 ‰ [46].

By now the carbon isotope composition of 85 samples desorbed from different coniferous tree ring disc wood has been investigated. The results obtained show that the gas samples desorbed from several discs are enriched in light isotope ^{12}C up to δ^{13} C = −25.3 ‰ for spruce, varying between −25 ‰ (1894) and −36.4 ‰ (1986) for the Siberian stone pine (Tomsk Oblast, Russia) and between −26.1 ‰ (1886) and −27.7 ‰ (1995) for the spruce disc tree rings brought from the Altai Mountains, Russia. For the sake of comparison, the variations in the carbon isotope composition of CO_2 chemically extracted from disc tree ring wood of different conifer species are shown in Fig. 5.1. The gas samples of a

300 year old larch are enriched in light isotope ^{12}C from $\delta^{13}C \approx -25$ ‰ to $\delta^{13}C \approx -30$‰. It is obvious that CO_2 is formed by trees themselves due to metabolic processes at work in the trees rather than being of atmospheric origin, because according to recent evidence, the carbon isotope composition of atmospheric CO_2 is, on average, $\delta^{13}C = -8.5$‰ [46]. Interestingly, the carbon isotope composition of CO_2 for the spruce discs exhibits but slight variations with tree age.

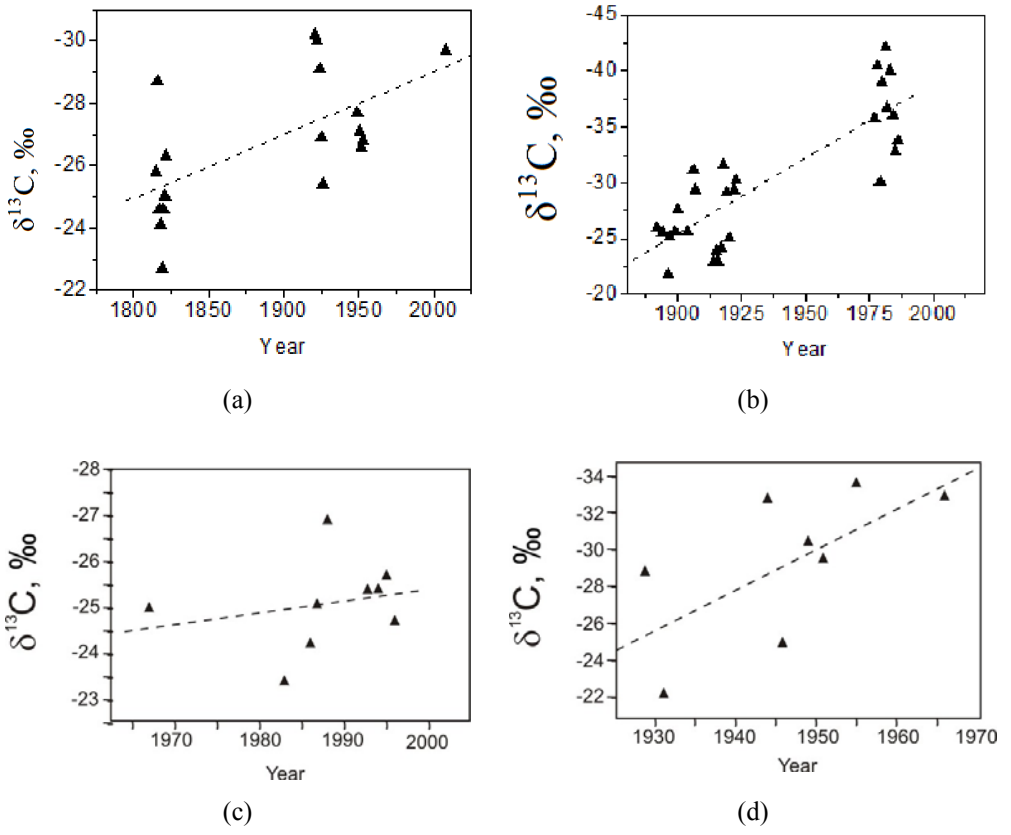


Fig. 5.1. Measured interannual variations in the carbon isotope composition of CO_2 ($\delta^{13}C$, ‰) desorbed from the tree ring gas samples of a 300 year old larch [47] (*a*); the CO_2 isotope composition variations ($\delta^{13}C$, ‰) with tree age: the Siberian stone pine disc (*b*), spruce disc 25 (*c*), and Scots pine disc 12 (*d*) [48].

5.3.2. The Total Pressure Variations in the Larch Disc Tree Rings

Concurrent with the investigation into the behavior of tree-ring CO_2, the total pressure variations in the exposure chambers were studied in the 300 year old larch disc. As a result we obtained the CO_2 and total pressure variation chronologies, which made it possible to find a correlation between chronologies and reveal a 4-year cycle of the CO_2 variations associated with a 4-year period of the total pressure variations in gas samples. Fig. 5.2(a)

and (b) demonstrates trends for the variations (polynomial approximation) in the total pressure of the gas samples in the exposure chambers and in the CO_2 content in the tree ring wood of the 300 year old larch. It is evident that the trends exhibit similar patterns: a CO_2 rise and a total pressure increase with tree age have been most noticeable since 1980. We have analyzed the measured results for the total pressure of the vacuum-extracted gas samples from the tree rings to find out whether the measurements showed cyclic variations. The results of the Fourier analysis of the variations in the tree ring pressure, using the ORIGIN software, are presented in Fig. 5.2(c): in addition to the 2-year variations, well-defined 4-year cycles can be clearly seen.

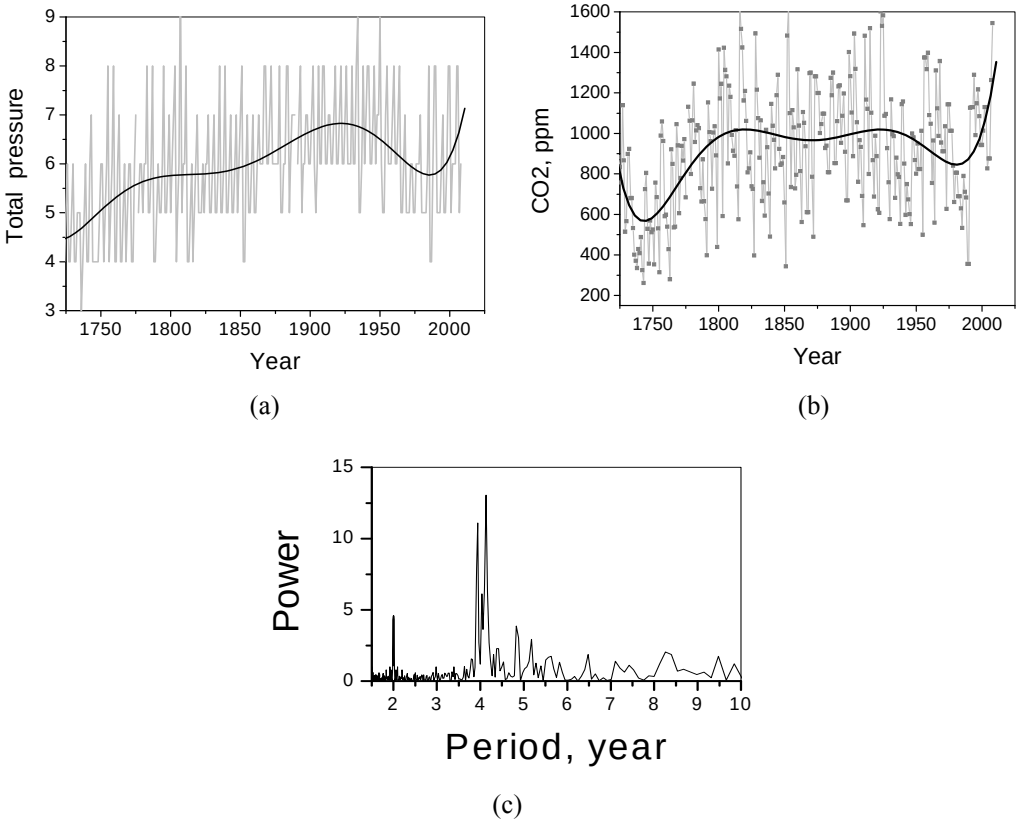


Fig. 5.2. Comparison of trends for variations in the total pressure of the gas samples in the exposure chambers (*a*) and the CO_2 content *(b*) in the 300 year old larch rings; the results of the Fourier analysis of the total pressure chronology for the wood of the 300 year old larch tree rings (*c*).

A comparison of the results of the Fourier analysis of the CO_2 and total pressure chronologies for the larch disc tree rings shows identical short-term variations (2- and 4-year cycles). The correlation function for the CO_2 and total pressure chronologies is of periodic nature, exhibiting a 4-year period for a sufficiently large shift in the time series (Fig. 5.3(a)). The autocorrelation function for the CO_2 chronologies turned out to be inherently periodic, exhibiting a 4 year cycle (Fig. 5.3(b)).

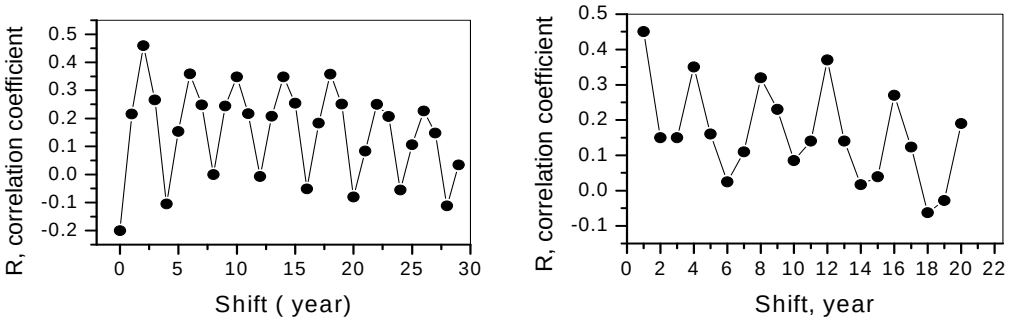


Fig. 5.3. Variations in the correlation coefficient between tree ring CO_2 and total pressure in the exposure chambers since 1722 (a) and between CO_2–CO_2 chronologies (autocorrelation), (b) for the 300 year old larch. The shift in the time series is ≤ 20 years. The trends were retained.

5.3.3. The Special Features Inherent in the Behavior of Tree Ring H_2O

The absorption by two gas components (CO_2 + H_2O) in the ***R*** (20) CO_2 laser line was recorded, because their absorption lines coincide. To derive the water vapor absorption alone, we subtracted the PA CO_2 signal from that detected in the ***R*** (20) line. Thus, two trends for the variations in annual PA signals from the two components were observed, which enabled us to compare the behavior of the CO_2 and H_2O chronologies for the disc tree rings of the 300 year old larch. The results smoothed out by an 11-year running average are demonstrated in Figs. 5.4 (a) and (b) (black) [47]. For the sake of comparison, the annual PA signal distribution for CO_2 and H_2O found in the Siberian stone pine tree rings [42] is shown in Fig. 5.4 (c).

It follows from Fig. 5.4 that the data obtained by the method proposed here show that portions of CO_2 and H_2O are stored in tree stem rings and exhibit peculiar annual distributions. The residual H_2O and CO_2 chronologies under study are characterized by distinct short- and long-term cyclic variations. Decreased PA signal intensities from H_2O in the tree rings of the Siberian stone pine (Fig. 5.4(c)) have been observed since 1980, which is attributed to the presence of a resin in the tree-ring wood.

The results of the Fourier analysis of the PA signals from H_2O in the 300 year old larch tree rings, using the ORIGIN software, confirmed the existence of the 2-and 4-year cycles (Fig. 5.5(a)) just as those found in the CO_2 and pressure chronologies. The correlation function between H_2O and total pressure variations in the exposure chambers is periodic (Fig. 5.5(b)), with a 4-year cycle being observed for a sufficiently large shift in the time series.

Thus, referring to Figs. 5.3 and 5.5, the chronologies for vacuum-extracted tree ring CO_2 and H_2O and total pressure recorded in the exposure chambers exhibit 4-year cyclic variations.

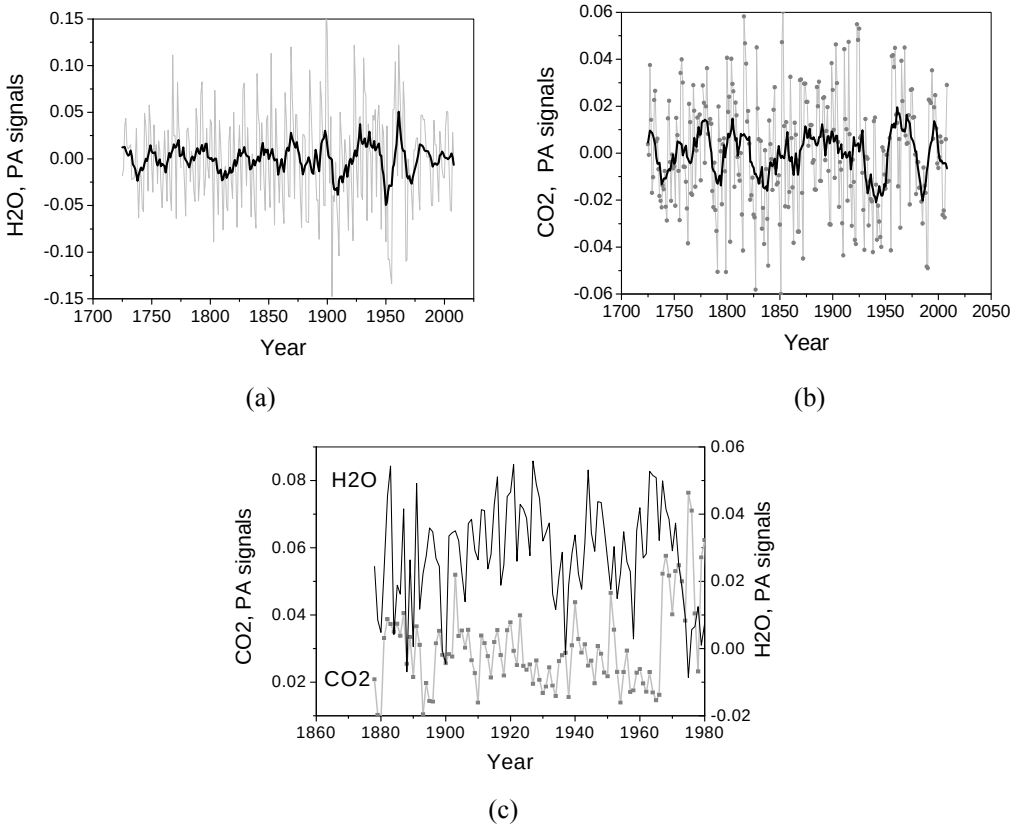


Fig. 5.4. Interannual variations in the PA signals accounting for the behavior of H_2O (a) and CO_2 (b) vacuum-desorbed from the tree rings of the 300 year old larch (without trends) and interannual variations in the PA signals accounting for the behavior of CO_2 and H_2O in the tree-rings of the Siberian stone pine (c).

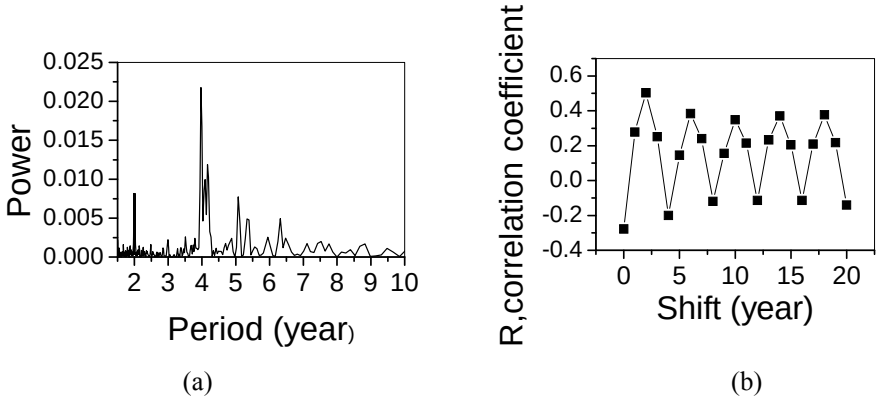


Fig. 5.5. Results of the Fourier analysis of the H_2O distribution in the annual tree rings of the 300 year old larch (*a*) and variations in the correlation coefficient between tree ring H_2O and total pressure in the exposure chambers as a function of time shift for the 300 year old larch rings since 1722 (*b*).

5.3.4. The Cyclic Components of the Chronologies

The cycles in the CO_2 and H_2O chronologies found experimentally were assumed to vary for different periods of life of the same tree species. The use of the high resolution spectral analysis made it possible to trace variations in the spectral composition with tree growth (larch). Fig. 5.6 shows the spectra of the CO_2 and H_2O content in tree ring wood for the first and second halves of the total growth period of the 300 year old larch [48].It is evident from the figure that the 2-, 3- and 4-year variations persist during the entire tree life, whereas the long-term variations are observed both in H_2O and in CO_2.

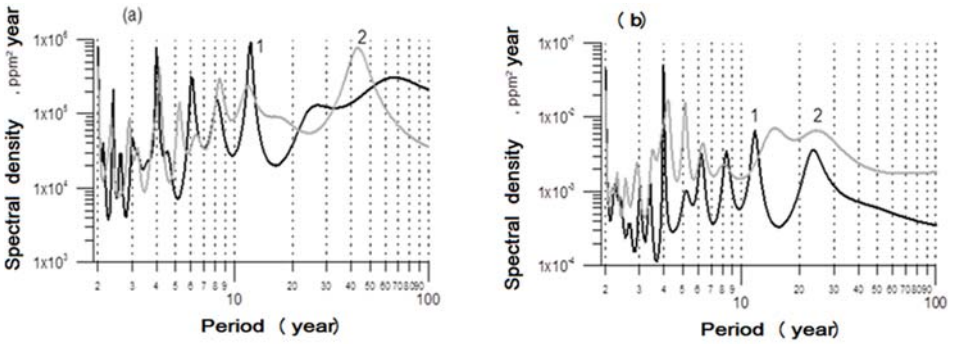


Fig. 5.6. Spectra of the CO_2 (*a*) and H_2O content (*b*) in the 300 year old larch tree ring wood in 1722–1865 (1) and 1865–2008 (2).

We have compared the features inherent in deciduous conifer trees (larch) with those observed in evergreen trees (Siberian stone pine) growing at the same place (Tomsk region). The spectra of their associated CO_2 chronologies are shown in Fig. 5.7. It is seen that the 4-year cycles exist both in the larch and in the Siberian stone pine CO_2 chronologies, but there is a difference in the long-term cycles between the two tree species [49].

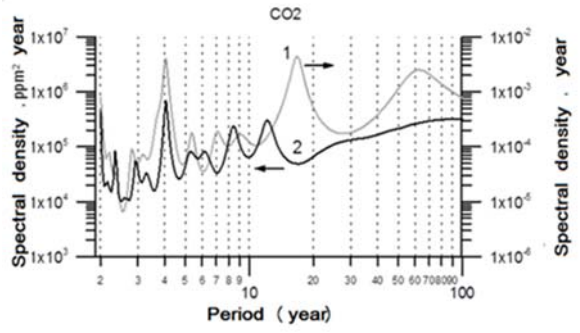


Fig. 5.7. Spectra of the CO_2 content in the Siberian stone pine tree ring wood (1) and 300 year old deciduous conifer tree (larch) (2) that grew in the same region.

As vacuum-desorbed CO_2 of tree ring wood is stem cell-respired, it needs to compare the spectra of the CO_2 chronology with those of the tree ring width chronology. A comparison between the Siberian stone pine and 300 year old larch discs shows that the 4-year cycles were observed neither in the larch nor in the Siberian stone pine tree ring width. Fig. 5.8 illustrates the spruce spectra of the CO_2 and width chronologies. It is evident from the figure that the presence of the CO_2 cycle occurs concurrently with the absence of such a cycle in the width chronology and vice versa.

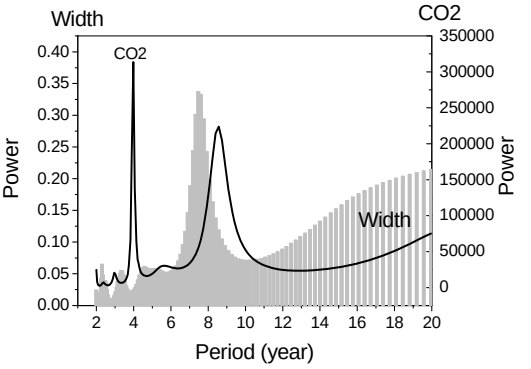


Fig. 5.8. Comparison of the spectra of the CO_2 and width chronologies for the spruce disc.

5.3.5. The Association of Chronologies and Meteorological Parameters

It is assumed that the variations in the chronologies under study are related to variations in temperature, precipitation, and cloudiness. Wavelet and high-resolution spectral and cross-spectral analyses were applied to the chronologies to find their climatic response [43 and 47]. The investigations employed the data about the monthly average temperature (1874–2009) and precipitation (1889–2009) obtained at Weather Station in Tomsk and information on the low cloud cover (1936–2000) and total cloud cover (1875–2000) [50]. The data about the temperature and cloud cover were averaged over the following periods: October–May, October–April (dormant period), May–September (vegetation period), and for twelve months from the October of the previous year to the September of the current year.

The CO_2 and H_2O content in the tree ring wood of the Siberian stone pine exhibits variations in a wide time scale range. Some of the variations are likely to be due to climatic factors. In particular, the 2- and 4-year variations in the CO_2 and H_2O content may be related to the quasi-biennial and quasi- quadrennial variations in the amount of precipitation in the dormant period, whereas the CO_2 variations on longer time scales anticorrelate with atmospheric temperature during the vegetation period ($R = -0.4$) [43].

As for the 300 year old larch, a significant correlation was found between CO_2 content and winter precipitation. The Spearman rank correlation coefficients were 0.32, 0.36, 0.37, 0.51, and 0.40 for November, December, January, February, and March, respectively, with a 95 % confidence interval for the 120-year period from 1889 to 2008.

In this case, however, the correlation was positive, which is evidence for the association of CO_2 content and spring soil moisture [47].

It was also shown in [48] that the tree ring width and CO_2 spectra exhibited weak spectral maxima with periods approaching that of the 11-year solar cycles (Fig. 5.9).

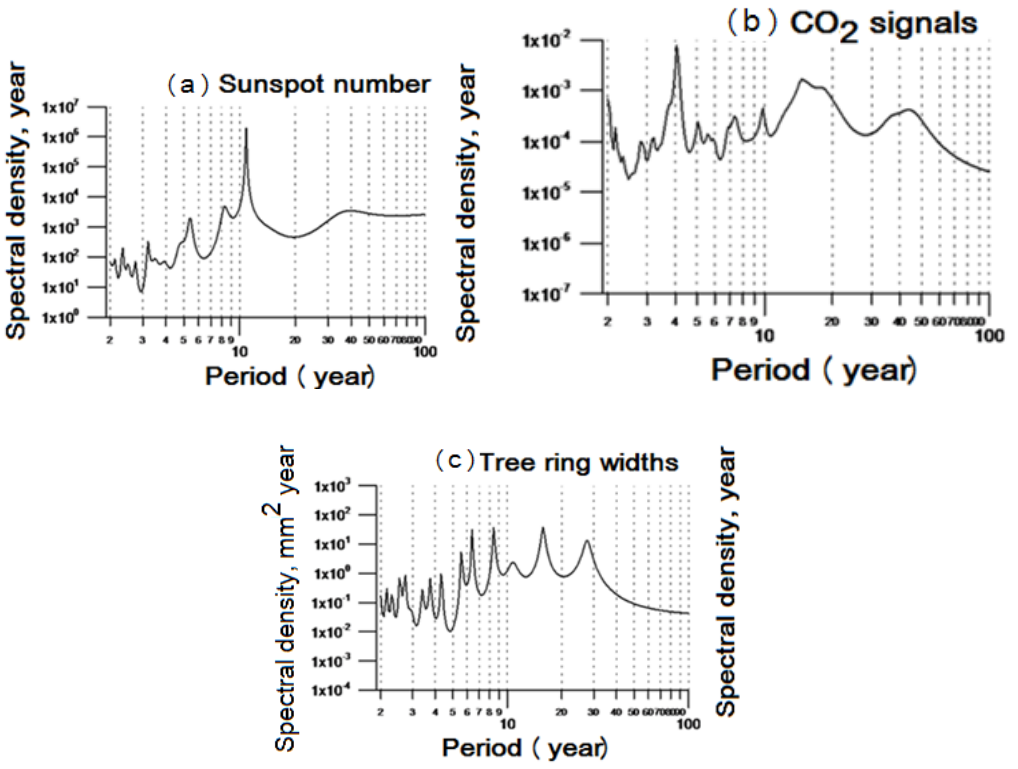


Fig. 5.9. Power spectra of the sunspot number (a), CO_2 content in the Siberian stone pine tree ring wood (b), and tree ring width (c).

To find a correlation between decadal variations in the Siberian stone pine tree ring width, tree ring CO_2 content, and solar activity, we have performed a cross-spectral analysis based on the maximum entropy technique. The power densities, coherence, and phase spectra for the foregoing parameters of the Siberian stone pine and sunspot number are shown in Fig. 5.10. All the parameters exhibit high coherence with the solar cycle. The associated width signals from the tree rings are approximately in antiphase with the solar cycle (the phase shift is about 180° for a period corresponding to the coherence peak), whereas the CO_2 signals are approximately in phase with the solar cycle (the phase shift is ~ 360°).

Simple phase relations (phase coincidence and phase opposition) between parameters of the Siberian stone pine tree rings and 11-year cycle of the solar activity lead us to suggest that the tree was affected by the level of the solar activity. A plausible mechanism

accounting for this influence is the effect of the solar cycle on the cloud amount. Fig. 5.10(c) shows that the cloud cover over Tomsk demonstrates decadal variations coherent with the 11-year solar cycle. The variations are approximately in phase with those relating to the level of the solar activity.

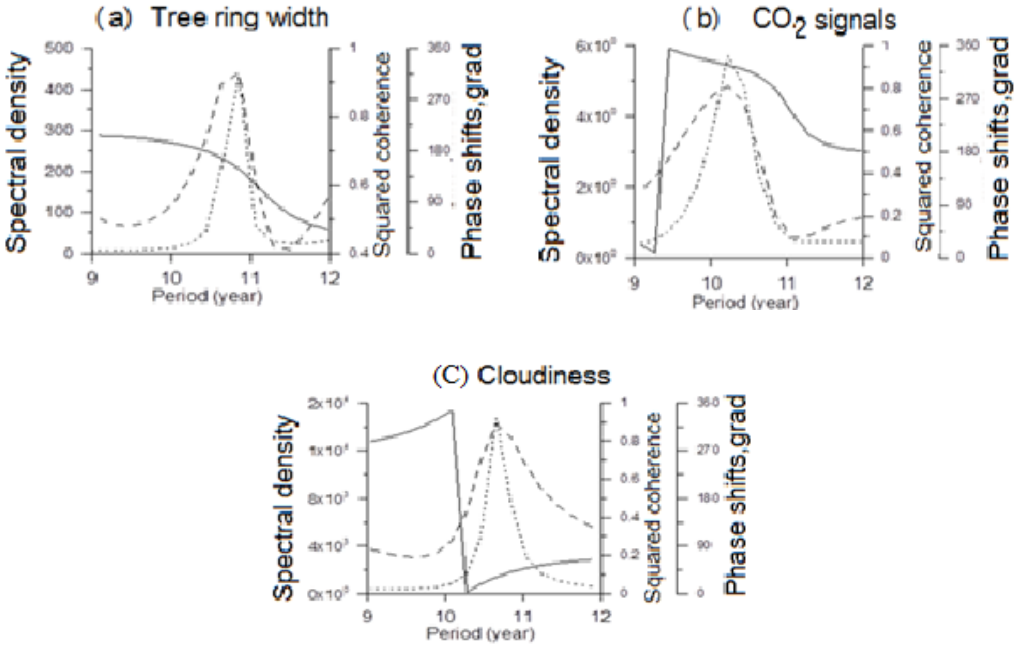


Fig. 5.10. Results of the cross-spectral analysis of the Siberian stone pine tree ring width (*a*), CO_2 content in the tree ring wood (*b*), and summer cloud amount (*c*).

5.3.6. The CO_2 Variations in Discs Brought from Different Sites

The interannual distributions of the desorbed CO_2 content found by the PA analysis in the discs of conifer species brought from different sites (lowland and mountains) were compared in [41]. The CO_2 and H_2O content was studied using the tree-ring discs of larch (*Larix sibirica* Ledeb) and Siberian stone pine (*Pinus sibirica* Du Tour). The larch disc was obtained near the settlement of Chernorud (53°00′ N, 106°43′ E), Irkutsk region, north-west coast of Middle Baikal, Russia. The Siberian stone pine disc was brought from the south-west slope of the Seminsky Range (50°04′ N, 87°45′ E) 1750 m above the sea level, Altai Republic, Russia. The measured results for the CO_2 content in the discs were compared with the data about the annual content of CO_2 in the discs of the Siberian stone pine and larch brought from the West Siberian Plain (56°26′ N, 85°03′ E), Tomsk region, Russia. The sites of all the trees differ in climatic conditions, absolute altitude, soil composition, relief, etc. For example, the Seminsky Range is the highest mountains among all Altai passes (1750 m in height) with a sharply continental climate and an annual average air temperature of −1.1 °C (the respective average temperatures in July and

January are +16.2 and − 22.1 °C). The south of the West Siberian Plain, on the contrary, is characterized by flat terrain free of sharp height drops (up to 200 m above the sea level). The sharply continental climate in the region is formed in part due to the warming effect of the Ob and Tom rivers. The average annual air temperature is − 0.6° C and the respective July and January average temperatures are + 18.1 and − 19.2° C. Priol'khonie, where Chernorud is situated, is the driest area near Lake Baikal. The radiation index of dryness corresponds to dry steppes. In certain years, there is no precipitation in that region early in the vegetation period (May–June), which affects the tree growth.

The measured results for the CO_2 content in the disc tree rings brought from the regions of Lake Baikal and Seminsky Range are demonstrated in Fig. 5.11, (a) and (b). The figure shows the presence of three- and four-year cycles in the CO_2 distribution and negative CO_2 trends. In addition, there is a distinct tendency for a decrease in the tree ring CO_2 content with tree age. For comparison, Fig. 5.11 (c) and (d) presents similar data for the Siberian stone pine and larch brought from the Tomsk region (CO_2 trend is given by a polynomial approximation). As shown in the figure, there is an appreciable CO_2 rise with increase in tree age.

The increase in the CO_2 content in the trees brought from Tomsk region can be attributed to variations in the atmospheric conditions: the atmospheric CO_2 rise and variations in the carbon isotope ratios in CO_2 [39]. The negative trends in the CO_2 tree ring distributions observed in other regions are likely to result from a shorter vegetation season in mountain areas and poor habitat quality.

5.4. Conclusions

Disc CO_2 and H_2O vacuum-extracted from coniferous tree rings were investigated using a photoacoustic laser analysis. The results obtained lead us to suggest that all coniferous discs studied exhibit peculiar tree ring distributions of CO_2 and H_2O. It turns out that large amounts of the gases are stored in annual ring wood, i.e. in tree stems. The porous wood structure is found to be capable of annual accumulation (sorption) of the stem gas components that include H_2O vapor and plant cell-respired CO_2. The tree ring carbon isotope composition of CO_2 chemically extracted from tree ring wood and analyzed by a mass spectrometer indicates that the samples are enriched in the light ^{12}C isotope, which confirms the fact that CO_2 is produced by trees themselves rather than being of atmospheric origin.

The vacuum extraction and laser photoacoustic analysis used provide long CO_2 and H_2O chronologies and enable a number of features inherent in the time series studied to be revealed. It is found that

1) Conifer species brought from the same site are characterized by short-period CO_2 and water vapor cycles with 2-, 3-, and 4-year cyclic variations independent of tree age, whereas their long-period cycles vary with tree age;

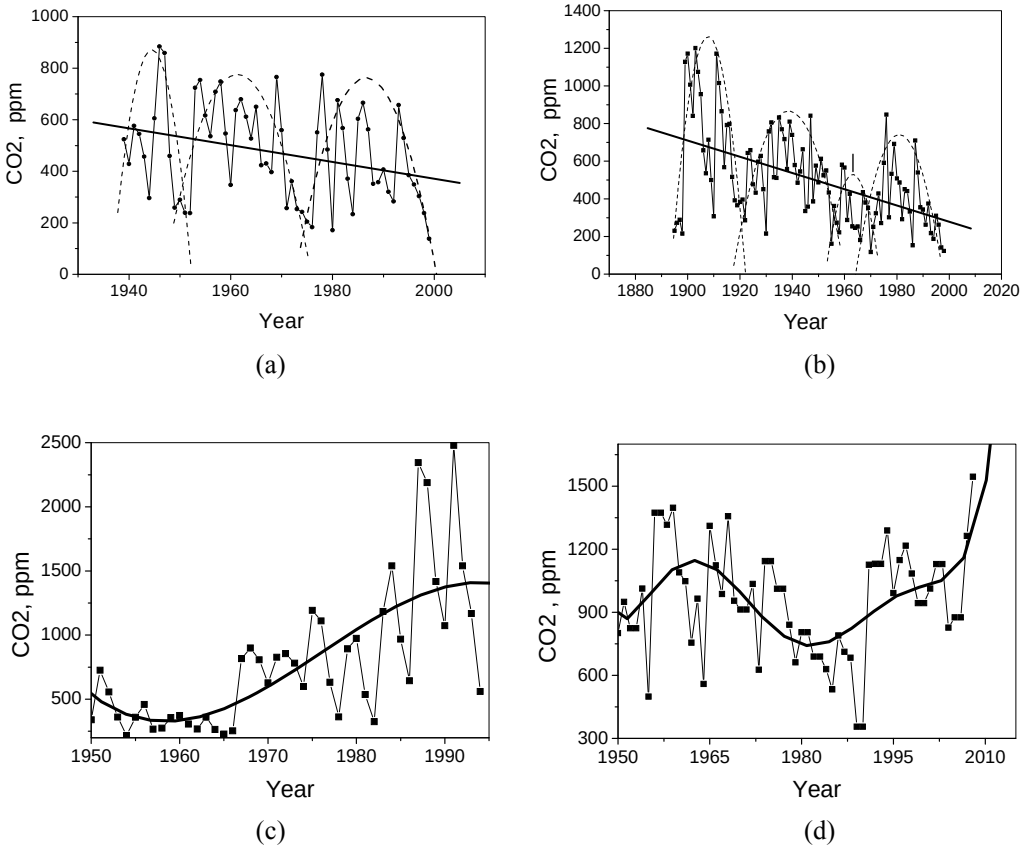


Fig. 5.11. Interannual variations in the CO_2 content in the tree-rings of the larch disc from Chernorud, Lake Baikal (a), in the Siberian stone pine from the Seminsky Range (b) and from Tomsk region: in the Siberian stone pine (c), and in the larch disc (d).

2) Measured CO_2 variations and increased total pressure in the disc tree ring wood of a 300 year old larch are associated with variations in the partial pressure of cell-respired CO_2 and transpiration water vapor. This is confirmed by the correlations between CO_2 and H_2O chronologies and total pressure variations in exposure chambers;

3) Conifers growing in the same site (Tomsk Oblast, West Siberia, Russia) exhibit increased CO_2 accumulation by tree ring wood, as the tree age is increased. The trends observed in the CO_2 content in tree-ring discs are different in sign for trees grown in different conditions;

4) It is possible that the 2- and 4-year variations in the CO_2 and H_2O content in tree rings are related to climatic factors, viz., to quasi-biennial and quasi- quadrennial variations in the amount of precipitation in the dormant state. In addition, the Siberian stone pine chronologies made it possible to establish the relation between tree ring parameters and 11-year cycle of the solar activity, which is probably due to the effect of the solar cycle

on the cloud amount. The latter is assumed to influence tree growth conditions through the modulation of solar radiation;

5). An analysis of the CO_2 chronologies has revealed another interesting feature. While the 4-year cycle in the CO_2 and H_2O content is strongly pronounced, no 4-year cycles are observed either in the larch or in the Siberian stone pine ring width chronologies. A comparison of the CO_2 and width spectra for spruce shows that the presence of the CO_2 cycle occurs concurrently with the absence of such a cycle in the width chronology and vice versa.

To sum up, the results obtained show that tree ring discs contain valuable information for estimating the CO_2 forest balance, assessing the respired CO_2 contribution to the physiological processes developing in trees, and improving the understanding of the role of gases present in tree stems. Undoubtedly, measurements of CO_2 and H_2O in tree stems will contribute to the detection of the mechanism involved in the forest adaptation for environmental and climatic changes. However, further investigations are needed to gain a deeper insight into the effect of adverse climatic conditions on the CO_2 content in tree rings. The elucidation of this problem is believed to provide a better understanding of the role of CO_2 in tree stems as a plant respiration regulator under adverse climatic conditions. The practical significance of investigations along these lines would be much greater for forest management if portable instruments for taking CO_2 cores were used. An instrument for online tree-ring CO_2 and H_2O monitoring can offer a means of investigating the forest system adaptation for climatic changes and provide unique data about forest health in ecological risk areas.

Acknowledgements

The work **i**s financed by the Siberian Branch of the Russian Academy of Sciences in the framework of a Basic Research Program (Project VIII.80.1.3).

We would like to express sincere thanks to Nikolai P. Baidin, director of Tomsk Forest Museum, for supplying a 300 year old larch disc.

References

[1]. D. McCarroll, N. J. Loader, Stable isotopes in tree rings, *Quaternary Science Reviews*, Vol. 23, Issue 7-8, 2004, pp. 771 - 801.

[2]. E. A. Vaganov, E.-D. Schulze, M. V. Skomarkova, A. Knohl, W. A. Brand, C. Roscer, Intra-annual variability of anatomical structure and δ13C values within tree rings of spruce and pine in alpine, temperate and boreal Europe, *Oecologia*, Vol. 161, Issue 4, 2009, pp. 729-745.

[3]. P. Zhao, D. Hölscher, The concentration and efflux of tree stem CO2 and the role of xylem sap flow, *Frontiers of Biology in China*, Vol. 4, Issue 1, 2009, pp. 47-54.

[4] Z. Gedalof, A. A. Berg, Tree ring evidence for limited direct CO2 fertilization of forests over the 20th century, *Global Biogeochemical Cycles*, Vol. 24, 2010, pp. 1-6 .

[5]. D. McCarroll, N. J. Loader, Stable isotopes in tree rings, *Quaternary Science Reviews*, Vol. 23, 2004, pp. 771 – 801.

[6]. O. N. Solomina, E. A. Dolgova, O. E., Maksimova, Hydroclimatic reconstructions over the last centuries in mountains of Crimea, Caucasus and Tien Shan based on tree-ring data, *Nestor-Istoriya publishing*, Moscow, S.-Petersburg, 2012. http://nestorbook.ru/uCat/field_files/5/10/693/Rekonstruktsiyagidrometeorologicheskikhusloviy.pdf

[7]. S. G. Shiyatov, E. A. Vaganov, A. V. Kirdyanov, V. B. Kruglov, V. C. Mazepa, M. M. Naurzbaev, R. M. Hantemirov, Methods of Dendrochronology, *KrasGY*, 2000.

[8]. O. V. Gerasimova, Z. Yu. Zharnikov, A. A. Knorre, V. C. Myglan, Climaticaly Induced Dynamic of Radial Growth of Siberian Stone Pine and Siberian Fir in the Mountain-Taiga Belt in Ergaki National Park, *Journal of Siberian Federal University. Biology*, Vol. 1, 2010, pp. 18-29.

[9]. K. Rozanski, S. J. Jonsen, U. Schotteren, L. G. Thompson, Reconstruction of past climates from stable isotope records of palaeo-precipitation preserved in continental archives, *Hydrological Sciences Journal*, Vol. 42, Issue 5, 1997, pp. 725-745.

[10]. M. Saurer, P. Cherubini, K. S. Treydte, W. T. Anderson, R. T. W. Siegwolf, An investigation of the common signal in tree ring stable isotope chronologies at temperate sites, *Journal of Geophysical Research*, Vol. 113, 2008, pp. 1-11.

[11]. J. Wernicke, J. Grießinger, P. Hochreuther, A. Bräuning, Variability of summer humidity during the past 800 years on the eastern Tibetan Plateau inferred from δ 18O of tree-ring cellulose, *Climate of the Past*, Vol. 11, Issue 2, 2015, pp. 327–337. www.clim-past.net/11/327/2015

[12]. E. R. Cook, N. Pederson, Dendroclimatology. Developments in Paleoenvironmental Research, *Springer Science+Business Media*, 2011. https://www.academia.edu/1906327/Uncertainty_emergence_and_statistics_in_dendrochronology

[13]. B. L. Gartner, J. A. Moore, B. A. Gardiner, Why is there so much air in sapwood? in *Proceedings of the Conference on Tree Rings and People, International Conference on the Future of Dendrochronology*, Birmensdorf, Switzerland 2001, pp. 200-201.

[14]. B. L. Gartner, J. R. Moore, B. A. Gardiner, Gas in stems: abundance and potential consequences for tree biomechanics, *Tree Physiology*, 24, 11, 2004, pp. 1239-1250.

[15]. L. Poorter, The Relationships of Wood-, Gas- and Water Fractions of Tree Stems to Performance and Life History Variation in Tropical Trees, Annals of Botany, Vol. 102, 2008, pp. 367 –375. http://aob.oxfordjournals.org/content/102/3/367.full.pdf+html?sid=2a45386c-bcff-4814-b5c8-8762ee3773a4

[16]. A. Saveyn, Dynamic interactions between CO_2 efflux, sap flow and internal CO_2 concentration in tree stems: implications towards the assessment of actual stem respiration . *Thesis in Applied Biological Sciences,* Gent, September 2007. http://buck.ugent.be/fulltxt/RUG01/001/241/146/RUG01-1241146_2010_0001_AC.pdf

[17]. J. Bloemen, M. A. McGuire, D. P. Aubrey, R. O. Teskey, K. Steppe, Transport of root-respired CO2 via the transpiration stream affects aboveground carbon assimilation and CO_2 efflux in trees, *New Phytologist*, Vol. 197, Issue 2, 2013, pp. 555–565.

[18]. P. E. Levy, P. Meir, S. J. Allen, P. G. Jarvis, The effect of aqueous transport of CO_2 in xylem sap on gas exchange in woody plants, *Tree Physiology*, Vol. 19, 1999, pp. 53 – 58.

[19]. R. O. Terskey, M. A. Mcguire, Carbon dioxide transport in xylem causes errors in estimation of rates of respiration in stems and branches of trees, *Plant, Cell and Environment*, Vol. 25, 2002, pp. 1571 – 1577.

[20]. M. A. Mcguire, R. O. Terskey, Microelectrode technique for in situ measurement of carbon dioxide concentration in xylem sap of trees, *Tree Physiology*, Vol. 22, 2002, pp. 807 – 811.

[21]. R. O. Teskey, A. Saveyn, K. Steppe, M. A. McGuire, Origin, fate and significance of CO_2 in tree stems, *New Phytologist,* Vol. 177, Issue 1, 2008, pp. 17-32 .

[22]. Angert, J. Muhr, R. N. Juarez, W. A. Muñoz, G. Kraemer, J. R. Santillan, E. Barkan, S. Mazeh, J. Q. Chambers, S. E. Trumbore, Internal respiration of Amazon tree stems greatly exceeds external CO_2 efflux, *Biogeosciences*, Vol. 9, 2012, pp. 4979–4991. www.biogeosciences.net/9/4979/2012/.
[23]. S. Etzold, R. Zweifel, N. K. Ruehr, W. Eugster, N. Buchmann, Long-term stem CO_2 concentration measurements in Norway spruce in relation to biotic and abiotic factors, *New Phytologist*, Vol. 197, Issue 4, 2013, pp. 1173–1184.
[24]. E. T. Engelund, L. G. Thygesen, S. Svensson, C. A. S. Hill, A critical discussion of the physics of wood–water interactions, *Wood Science and Technology*, 47, Issue 1, 2013, pp. 141–161.
[25]. S. V. Glass, S. L. Zelinka, Moisture relations and physical properties of wood, Wood handbook: wood as an engineering material, *General technical report FPL*, Madison, 2010. http://www.treesearch.fs.fed.us/pubs/37428
[26]. J. Khazaei, Water absorption characteristics of three wood varieties, *Cercetări Agronomice în Moldova*, Vol. XLI, Issue 2, 134, 2008, pp. 5-16. http://www.uaiasi.ro/CERCET_AGROMOLD/CA2-08-01.pdf.
[27]. U. G. Hacke, A. L. Jacobsen, R. B. Pratt, C. Maurel, B. Lachenbruch, J. Zwiazek, New research on plant-water relations examines the molecular, structural, and physiological mechanisms of plant responses to their environment, *New Phytologist,* Vol. 196, Issue 2, 2012, pp. 345-348.
[28]. Fromm, J. H., Sautter, I., Matthies, D., Kremer, J., Schumacher P., and Ganter, C., Xylem water content and wood density in spruce and oak trees detected by high-resolution computed tomography, *Plant Physiology*, Vol. 127, Issue 2, 2001, pp. 416-425.
[29]. P. L. Meyer, M. W. Sigrist, Atmospheric pollution monitoring using CO2-laser photo acoustic spectroscopy and other techniques, *Review of Scientific Instruments*, Vol. 61, Issue 7, 1990, pp. 1779-1807.
[30]. Air Monitoring by Spectroscopic Techniques, Ed. M. W. Sigrist, *Willey*, 1994.
[31]. F. J. M. Harren, G. Cotti, J. Oomens, S. te Lintel Hekkert, Photoacoustic Spectroscopy in Trace Gas Monitoring, Encyclopedia of Analytical Chemistry, Ed. R. A. Meyers, *Willey*, 2000.
[32]. R. Cernat, C. Matei, A. M. Bratu, C. Popa, D. C. A. Dutu, M. Patachia, M. Petrus, S. Banita, D. C. Dumitras, Laser photoacoustic spectroscopy method for measurements of trace gas concentration from human breath, *Pomanian Reports in Physics*, Vol. 62, Issue 3, 2010, pp. 610-616.
[33]. M. W Sigrist, R. Bartlome, D. Marinov, J. M. Rey, D. E. Vogler, H. Wächter, Trace gas monitoring with infrared laser-based detection schemes, *Applied Physics B*, Vol. 90, Issue 2, 2008, pp. 289–300.
[34]. G. R. Lima, M. S. Sthel, M. G. da Silva, D. U. S. Schramm, M. P. P. de Castro and H. Vargas, Photoacoustic spectroscopy of CO_2 laser in the detection of gaseous molecules, *IOP Publishing Journal of Physics: Conference Series,* Vol. 274, Issue 2, 2011.
[35]. M. E. Webber, T. MacDonald, M. B. Pushkarsky, C. K. N. Patel, Y. Zhao, N. Marcillac, F. M. Mitloehner, Agricultural ammonia sensor using diode lasers and photoacoustic spectroscopy, *Measurement Science and Technology,* Vol. 16, Issue 8, 2005, pp. 1547–1553.
[36]. C. Popa, Laser spectroscopy applied to analysis of active young women's breath, *Romanian Reports in Physics,* Vol. 66, Issue 4, 2014, pp. 1056–1060.
[37]. C. Popa, M. Petrus, A. M. Bratu, Ammonia and ethylene biomarkers in the respiration of the people with schizophrenia using photoacoustic spectroscopy, *Journal of Biomedical Optics,* Vol. 20, Issue 5, 2015, SPIE digital library.
[38]. B. Ageev, Yu. Ponomarev, V. Sapozhnikova, Laser photoacoustic detection of CO_2 in old disc tree-rings, *Sensors*, Vol. 10, 2010, pp. 3305-3312.
[39]. B. G. Ageev, Yu. N. Ponomarev, V. A. Sapozhnikova, Laser photoacoustic method for disc tree-ring gas analysis, *World Environment*, Vol. 2, Issue 2, 2012, pp. 4 -10.

[40]. B. G. Ageev, A. N. Gruzdev, S. N. Bondarenko, V. A. Sapozhnikova, Long-Term H_2O and CO_2 Trends in Conifer Disc Tree Rings and Meteorological Parameters, *Journal of Life Sciences,* Vol. 7, Issue 9, 2013, pp. 1002-1008.

[41]. B. G. Ageev, A. N. Gruzgev, G. V. Matyashenko, D. A. Savchuk, V. A. Sapozhnikova, Yu. N. Ponomarev, Variations and trends in CO_2 and H_2O content in disc tree rings, *Atmospheric and Oceanic Optics,* Vol. 26, Issue 3, 2013, pp. 245-251.

[42]. B. G. Ageev, A. P. Zotikova, N. L. Padalko, Yu. N. Ponomarev, D. A. Savchuk, V. A. Sapozhnikova, E. V. Chernikov, Variation of H_2O, CO_2, and CO_2 isotope composition in tree rings of Siberian stone pine, *Atmospheric and Oceanic Optics,* Vol. 24, Issue 4, 2011, pp. 390-395.

[43]. V. A. Sapozhnikova, A. N. Gruzdev, B. G. Ageev, Yu. N. Ponomarev, D. A. Savchuk, Relationship between CO_2 and H_2O Variations in Tree Rings of Siberian Stone Pine and Meteorological Parameters, *Doklady Earth Sciences,* Vol. 450, Part 2, 2013 pp. 652–657.

[44]. M. M. Savard, C. Bégin, J. Marion, L.-C. Aznar, A. Smirnoff, Chapter 9. Air Quality Changes in an Urban Region as Inferred from Tree-Ring Stable Isotopes, *Developments in Environmental Science,* Vol. 9, 2009, pp. 229–245.

[45]. D. McCarroll, N. J. Loader, Stable isotopes in tree rings, *Quaternar Science Reviews,* Vol. 23, Issue 7-8, 2004, pp. 771–801.

[46]. M. Rubino, et al., A revised 1000 year atmospheric $\boldsymbol{\delta}$ 13C-CO_2 record from Law Dome and South Pole, Antarctica, *Journal Geophysical Research: Atmospheres*, Vol. 118, 2013, pp. 8482–8499.

[47]. Boris Ageev, Yurii Ponomarev, Valeria Sapozhnikova, Dmitry Savchuk, A Laser Photoacoustic Analysis of Residual CO_2 and H_2O in Larch Stems, *Biosensors*, Vol. 5, 2015, pp. 1-12.

[48]. Boris G. Ageev, Aleksandr N. Gruzdev, Yurii N. Ponomarev, Valeria A. Sapozhnikova, Climatic response of annual tree-rings, in *Proceeding of 20th International Symposium on Atmospheric and Ocean Optics: Atmospheric Physics,* Novosibirsk, Russia, 23-27 June, 2014, in Proceedings of the SPIE, Vol. 9292, (929235-1)- (929235-8).

[49]. Boris G. Ageev, Aleksandr N. Gruzdev, Valeria A. Sapozhnikova, The special features of tree ring gas chronologies, in *Proceeding of 21th International Symposium on Atmospheric and Ocean Optics: Atmospheric Physics,* Tomsk, Russia, 22-26 June, 2015 (in press).

[50]. National Climatic Data Center. Global Synoptic Climatology Network. The Former USSR, TD9290c, 2005, p. 15.

Chapter 6
Graphene: A Unique Constructional Material for Electroanalytical Applications

Ida Tiwari, Manorama Singh

Abstract

Construction and fabrication of graphene and its hybrid forms are currently becoming more relevant material for various applications. It is a very interesting material and the chemistry of graphene is likely to play an increasingly innovative role in future developments. Functionalized graphene is more appropriate for specific applications. Because of unique characteristics of graphene, it facilitates the great interest for use in many technological fields, e.g. as enhanced materials for developing high-performance sensors, biosensors and in other hi-tech fields. Remarkable mechanical and electrical properties favor its use in industrial as well as technological fields for point of care testing.

6.1. Graphene

This 'wonderful material' is a populated allotrope of carbon which is one atom thick planar sheet of sp^2 bonded carbon atoms in honeycomb crystal lattice. It has unique characteristics of electrical, optical, mechanical, thermal and electrochemical properties. It's considerable attention for molecular sensing is in interest as it is electronically a very good low noise material [1]. It is considered as 2-D zero gap based semiconductor because no gap is present between valence band and conduction band [2].

6.2. Basic Structure and Properties of Graphene

A. K. Geim and K. S. Novaselov, 2007 [3] have succeeded in identifying, producing graphene through effective mechanical exfoliation and its characterization. In words of IUPAC, 1995: "Graphene is a single carbon layer of graphitic structure".

Ida Tiwari
Department of Chemistry, Faculty of Science, Banaras Hindu University, Varanasi(UP)-221005, India

The synthesis of graphene can be done through various ways. The first approach was chemical vapor deposition and epitaxial growth on nickel surface by the decomposition of ethylene in 1979 to synthesize the graphene [4]. Several other good methods are reported in the literature to synthesize the graphene such as micromechanical cleavage [5], thermal exfoliation [6], chemical reduction of graphene oxide [7] etc. Recently chemical synthesis from graphite is in interest [8] i.e. Hummer's method [9], Staudenmaier [10] etc.

Graphene is basically one single layer of graphite i.e. a layer of sp^2 bonded carbon atoms arranged in a hexagonal lattice. Graphene is the strongest material ever found. Due to its planar structure its thermal, electronic and acoustic properties are highly anisotropic means that photons travel more easily along the planes. Graphene with a single layer of atoms and having very high electron mobility offers incredible level of electronic conduction due to the occurrence of free π electrons for each carbon atom.

6.3. Basic Identification of Graphene and its Hybrid Materials

There are different techniques to characterize the prepared graphene and its hybrid materials. The morphology is characterized by atomic force microscopy (AFM), scanning electron microscopy (SEM) and transmission electron microscopy (TEM). AFM estimates the number of layers in graphene [11]. Raman [12], FTIR and UV-Visible spectra gives the information of inner structure of the graphene. FTIR gives the information about the functional moieties produced on the graphene samples [13]. Other techniques powdered X ray diffraction [14], electrochemical [15] and thermogravimetric analysis (TGA) [16] are also used for the characterization of the structure of graphene.

6.4. Decoration of Graphene with Different Materials e.g. Metal Nanoparticles, Organic Compounds, Conducting Polymers etc.

There are some limitations of using un-functionalized or pristine graphene i.e. dispersion in aqueous solution, low sensitivity and low detection limit [17]. Therefore, modification/ functionalization of graphene is now in interest to amplify the performance of the material in biological assays such as in biosensors. Several additional materials may be used to decorate the surface of graphene such as different metal nanoparticles (e.g. gold [18], silver [19], platinum [20], palladium [21]; conjugated dyes [22, 23], Prussian blue [24] etc.); aromatic molecules [25], carbon nanotubes [26] and conducting polymers [21] etc. Different nanoparticles and other molecules are loaded on graphene to maximize the availability of nanosized electro catalyst surface area for electron transfer and providing a better mass transport of reactants to the electro catalyst [27]. Nanoparticles provide narrow size distribution, desirable biocompatibility, reduce the over potentials and act as an 'electro catalyst' in electrochemical reactions.

Hybrids of graphene can also be synthesized by the arrangement of large aromatic molecules on the graphene sheets [24]. Porphyrin are employed to functionalize the graphene nanoribbons for electrocatalysis [28] and doping of nitrogen into graphene

increases the defects sites on the graphene surface due to which activity of the surface of the graphene increases [29]. Covalent attachment of different natural amino acids on graphene may also improve the sensitivity of graphene [14].

6.5. Application of Graphene and its Hybrid Materials in Sensors and Biosensors

This article depicts the application of graphene and its hybrid as an 'ideal sensing material' in the fabrication of different electrochemical sensors and biosensors because of its electroactive and transparent nature. It is generally used for the preparation of electrochemical sensors and biosensors due to its wide potential window and low charge resistance and good redox behavior. Graphene based electrodes are also able to load high enzyme quantity due to its increased large surface area.

6.5.1. Detection of Pesticide

The application of transition metal oxides and metal nanoparticles as catalysts is restricted due to their poor electrical conductivity and slow electron transfer; therefore they are employed with graphene. In recent years, the incorporation of metal oxides and metal nanoparticles with graphene offer synergistic effects in electrocatalytic applications i.e. MnO_2 [30], SnO_2 [31], ZrO_2 [32], CoO [33], CdS [34] etc. Wang et al, 2014 synthesized low cost highly dispersed ZnO nanoparticles on carboxylic graphene for the sensing of chlorpyrifos and carbofuran [35]. Yang et al, 2013 immobilized acetylcholinesterase (AChE) on NiO nanoparticles-carboxylic graphene –nafion (NiO NPs-CGR-NF) nanocomposite for the detection of methyl parathion (MP), chlorpyrifos and carbofuran. NiO NPs-CGR-NF nanocomposite offered an extremely hydrophilic surface for AChE adhesion [36]. Qiu et al, 2011 synthesized the TiO_2 decorated graphene for the detection of carbaryl using a facile one step solvothermal reaction where the reduction of GO and the deposition of TiO_2 on graphene occur simultaneously [37].

Several metal nanoparticles received significant attention using graphene. Song et al, 2011 have conjugated the graphene oxide with tyrosinase assembled GNPs to fabricate the enzyme based screen printed electrode [38]. Carboxylic group functionalized graphene was decorated with platinum nanoparticles by Yang et al, 2013 for the enzymatic detection of MP pesticide [39]. Liu et al, 2011 immobilized AChE on 3- carboxyphenylboronic acid/ reduced graphene oxide –gold nanocomposite modified electrode for the detection of organophosphorous (OPs) and carbamate pesticide [40]. Xue et al, 2014 electro-polymerized molecularly imprinted polymer onto N- doped graphene sheets coated gold electrode in the presence of MP as template molecules [41]. Wang et al, 2011 decorated gold nanoparticles on chemically reduced graphene oxide nanosheets in the presence of poly(diallyldimethylammonium chloride) (PDDA) for the ultrasensitive detection of organophosphate pesticide [42]. Yang et al, 2014 synthesized a nanohybrid of gold nanoparticles, polypyrrole and reduced graphene oxide sheets achieved by electrochemical deposition of reduced graphene oxide with pyrrole and the introduction of gold nanoparticles. Further silica encapsulated acetylcholinesterase was immobilized on gold nanoparticle–polypyrrole-reduced graphene oxide nanohybrid for the

amperometric detection of paraoxon ethyl. The activity of AChE was protected by encapsulation in circumambient silica matrix which provided a biocompatible environment, facilitated mass transport [43]. Zhai et al, 2014 developed an acetylcholinesterase biosensor based on graphene–gold nanocomposite and calcined layered double hydroxide (CLDH). CLDH can provide a larger effective surface area for enzyme loading. AChE loaded CLDH was immobilized onto graphene-gold nanocomoposite modified GCE [44]. Zhang et al, 2012 studied the inhibition, reaction and aging process of pesticides using graphene nanosheets/gold nanoparticles based acetylcholinesterase biosensor. They concluded that the activity of AChE can be inhibited by pesticides. OPs poisoning can be recovered with oximes while carbamates exhibits insensitive to oximes [45]. A sensitive enzymeless detection of methyl parathion was reported by Gong et al, 2011 using gold nanoparticles decorated graphene nanosheets as solid phase extraction. [46]. A graphene-Fe_3O_4 nanocomposite film was reported by Liu et al, 2013 for the determination of methyl parathion (MP). The combination of iron nanoparticles and graphene greatly catalyzed the oxidation of enzymatically generated thiocholine product and thus increasing the detection sensitivity [47]. A bienzymaatic biosensor for carbamates was reported by Oliveira et al, 2014 in a single step electrodeposition of a polyphenoloxidases-gold nanoparticles-chitosan hybrid film onto a graphene doped carbon paste electrode [48].

Sun and Wang, 2010 electrodeposited Prussian blue on GCE to enhance the electron transfer and reported an AChE-PB/GCE biosensor for the determination of organophosphate pesticide [49]. Oliviera et al, 2013 immobilized Laccase enzyme on Prussian blue film functionalized graphene doped carbon paste electrode for the quantification of carbamate pesticide applied on tomato and potato crops [50]. Zhang et al, 2012 efficiently prepared Prussian blue nanocubes/ reduced graphene oxide nanocomposite by directly mixing Fe^{3+} and $[Fe(CN)_6]^{3-}$ in the presence of GO in polyethyleneimine aqueous solution for detection of organophosphorus pesticides [51]. Zhu et al, 2012 developed poly(malachite green)/graphene nanosheets–nafion composite film-modified glassy carbon electrode for the electrochemical determination of MP [52]. Molecularly imprinted polymer was used with ionic liquid- graphene by Zhao et al, 2013 for electrochemical determination of methyl parathion [53]. Li et al, 2012 employed ionic liquid with graphene for the immobilization of AChE [54].

6.5.2. Detection of Hemoglobin

Hemoglobin (Hb) is the most important ingredient of blood and is accountable for transporting O_2 throughout the circulatory system. Several diseases and even death can be possible with the change of Hb concentration in blood. Therefore, the content of Hb in blood should be accurately determined. A chitosan-graphene (CS-GR) modified electrode was fabricated for the electroanalysis of Hb [55] by Xu et al, 2010. Ionic liquid was successfully employed with a chitosan-ferrocene /graphene by Huang et al, 2012 for the detection of hemoglobin [56]. Xu et al, 2011 studied the direct enhanced electrochemical behavior of hemoglobin entrapped in graphene-ZnO nanosphere composite film [57]. Wang et al, 2013 used facile method to synthesize graphene-ZnS nanocomposite and its application in the detection of hemoglobin [58]. Direct electron transfer and

electrocatalysis of hemoglobin immobilized on graphene-Pt nanocomposite was reported by Feng et al, 2011 [59]. The poly(diallyldimethylammonium chloride) functionalized graphene nanosheets were reported by Feng et al, 2012 and further gold nanoparticles were assembled on the surface of PDDA-G film through electrostatic interaction to construct the biocompatible interface for the detection of hemoglobin [60].

6.5.3. Detection of Heavy Metal Ions

Development of some heavy metal ion electrochemical sensors based on graphene have been reported [61-66] and their advantages are obvious in various fields due to their large detection area, unique sensing mechanism, and ease of functionalization. However, chemical binders (e.g., Nafion) have been usually employed to mix with graphene films or powders to form a kind of graphene paste with a thickness of nm or mm scale. The graphene films were used as an electrochemical sensor for detection of trace heavy metal ions e.g., Pb^{2+} [67] contained in an acetate buffer solution (pH 5.3) with a detection limit of about 7 nM for the first time. The graphene electrodes showed excellent repeatability and sensitivity in the detection of Pb^{2+} ions, and were tolerant to the presence of surface-active compounds. Because of the remarkably high electron mobility at room temperature, the electrical conductivity along the in-plane direction of graphene is high [68]. Lead oxide was employed with graphene by Ramesha et al, 2011 for the trace detection of arsenic metal [65]. Guo et al, 2013 reported ultrasensitive and selective sensing of heavy metal ions with modified graphenes [69]. A different kind of metallo- graphene platform was reported by Willemse et al, 2011 for the detection of toxic metal ions [70]. Lien et al, 2013 modified graphene with bismuth and utilized it with carbon nanotube composite for the sensing of lead ions [71]. Kong et al, 2013 reported a graphene modified gold electrode for the sensing of copper and lead ions [72]. Lead ions could also be determined by Zhang et al, 2013 graphene modified carbon nanosheets [73]. Zhang et al, 2010 decorated 1-octadecanethiol monolayers on graphene for the detection of mercury ions [74] A rapid sensing of silver ion and biothiols by nanoparticles decorated graphene quantum dots was reported by Ran et al, 2013 [75].

6.5.4. Detection of Hydrogen Peroxide

Determination of hydrogen peroxide is significant in environmental protection, clinical control and in biology. A PtNPs-graphene hybrid material was fabricated by Dey and Raj, 2010 for the detection of hydrogen peroxide [76]. Li et al, 2010 used MnO_2 and graphene to prepare the electrocatalyst for the detection of hydrogen peroxide in alkaline medium [77]. Sun et al, 2012 developed a biosensor based on immobilization of hemoglobin in thin film of chitosan containing graphene and palladium nanoparticles [78]. Pang et al, 2014 developed a sensor using graphene and gold nanorods composite for the detection of hydrogen peroxide [79]. Zhang et al, 2013 studied the electrochemical behavior of graphene/ nafion/ azure I/ gold nanoparticles composite modified electrode for the detection of hydrogen peroxide [80]. Ye et al, 2013 synthesized the MnO_2/graphene/CNT hybrid through one pot chemical reaction process for the detection of hydrogen peroxide in milk [81].

6.5.5. Detection of Glucose

Zhou et al, 2010 prepared a nafion-graphene/Gnp biocomposite material for the detection of glucose [68]. Lu et al, 2011 reported a nonenzymatic glucose sensor based on Pd-graphene nanohybrid [82]. An amperometric glucose biosensor was made by Cheng et al, 2012 through integration of glucose oxidase with palladium nanoparticles/ reduced graphene oxide nanocomposite [83]. Li et al, 2013 synthesized graphene-CuO nanocomposite on a template of graphene by one step chemical synthesis approach and used to fabricate an enzyme free amperometric glucose sensor [84]. A TiO_2-graphene composite was synthesized from a colloidal mixture of TiO_2 nanoparticles and graphene oxide nanosheets by an aerosol assisted self-assembly for the detection of glucose [85]. A novel highly efficient enzyme electrode was fabricated by Liu et al, 2010 using covalent attachments between carboxyl acid groups of graphene sheets and amines of glucose oxidase [86]. A simple solution phase approach was used by Unnikrishnan et al, 2013 to prepare exfoliated graphene oxide followed by electrochemical reduction to get reduced graphene oxide- glucose oxidase biocomposite [87]. The bionanocomposite film consisting of glucose oxidase/Pt/functional graphene sheets/chitosan for glucose sensing was reported by Wu et al, 2009 [88]. A glucose biosensor based on direct electron transfer of glucose oxidase self-assembled on the surface of the electrochemically reduced carboxyl graphene modified glassy carbon electrode has been reported by Liang et al, 2012 [89]. Graphene sheets were synthesized by thermal chemical vapor deposition under ambient pressure on copper tapes and then they were transferred into integrated Fe_3O_4-doped Polyaniline based microelectrodes for the detection of glucose [90].

6.5.6. Detection of Organic Pollutants/Pathogens

Graphene-PANI/Tyr/Nafion/GCE produced electrocatalytic oxidation of hydroquinone; the aromatic structure of hydroquinone allows it to form strong π-π interactions with graphene. In addition, the positively charged nitrogen atom of PANI could undergo electrostatic interactions with hydroquinone. Combined, these interactions generated a high load efficiency of hydroquinone on the electrode surface [16]. Yola et al, 2014 self-assembled silver nanopartilcles on to the surfaces of 2-aminoethanthiol functionalized graphene oxide sheets for the detection of quercetin and morin. These are antioxidants found in vegetables and fruits [91]. Recently our group has focused on development of E. coli biosensors based on graphene oxide. We have prepared a graphene oxide modified iron oxide-chitosan hybrid nanocomposite for sensitive detection of E. coli [92]. Further, we have developed an electrochemical sensor for the detection of pathogenic Escherichia coli specific DNA sequence based on graphene oxide- chitosan composite decorated with nickel ferrite nanoparticles [93]. Synthesis, characterization and genosensing application of graphene sheets decorated gold nanoparticles-polypyrrole based nanocomposite has also been done [94].

6.5.7. Sustainability and Uniqueness of Graphene

Graphene, a low cost material is being tried by researchers for use in the synthesis of different nanomaterials and in the fabrication of electronic devices, sensors etc. as

graphene based nanomaterials shows unique structural, chemical, electrical and mechanical properties. These materials are also efficient in loading the enzyme and other biomolecules for the clinical and biological applications especially for point of care testing. Further study is still required to understand the deep effect of the unique properties of the graphene and its hybrid nanomaterials.

References

[1]. W. Choi, I. Lahiri, R. Seelaboyina, Y. Kang, Synthesis of Graphene and Its, Applications: A Review, *Crit. Rev. Solid State Mater Sci,* Vol. 35, 2010, pp. 52-71.

[2]. A. Martin, A. Escarpa, Graphene: The cutting –edge interaction between chemistry and electrochemistry, *Trends in Analytical Chemistry,* Vol. 56, 2014, pp. 13–26.

[3]. A. Geim, K. Novoselov, The rise of graphene, *Nat. Mater.,* Vol. 6, 2007, pp. 183–191.

[4]. M. Eizenberg, J. M. Blakely, Carbon monolayer phase condensation on Ni(111), *Surf. Sci.,* Vol. 82, 1979, pp. 228–236.

[5]. W. Choi, I. Lahiri, R. Seelaboyina, Y. S. Kang, Synthesis of Graphene and Its Applications: A Review *Solid State Mater. Sci.,* Vol. 35, 2010, pp. 52–71.

[6]. R. D. Dreyer, S. Park, C. W. Bielawski, R. S. Ruoff, The chemistry of graphene oxide, *Chem. Rev. Soc.,* Vol. 39, 2010, pp. 228–240.

[7]. Z. Dursun, B. Gelmez, Simultaneous Determination of Ascorbic Acid, Dopamine and Uric Acid at Pt Nanoparticles Decorated Multiwall Carbon Nanotubes Modified GCE, *Electroanalysis,* Vol. 22, 2010, pp. 1106–1114.

[8]. V. Singh, D. Joung, L. Zhai, S. Das, S. I. Khondaker, S. Seal, Graphene based materials: Past, present and future, *Prog. Mater Sci.,* Vol. 56, 2011, pp. 1178–1271.

[9]. X. Y. Qi, K. Y. Pu, H. Li, X. Z. Zhou, S. X. Wu, Q. L. Fan, B. Liu, F. Boey, W. Huang, H. Zhang, Amphiphilic Graphene Composites, *Angew. Chem. Int. Ed.,* Vol. 49, 2010, pp. 9426-9428.

[10]. L. Staudenmaier, Verfahren zur Darstellung der Graphitsäure, *Berichte der Deutschen Chemischen Gesells Chaft,* Vol. 31, 1898, pp. 1481-1487.

[11]. K. S. Novoselov, D. Jiang, F. Schedin, T. J. Booth, V. V. Khotkevich, S. V. Morozov, A. K. Geim, Two-dimensional atomic crystals, *Proc. Natl. Acad. Sci. USA,* Vol. 102, 2005, pp. 10451–10453.

[12]. A. C. Ferrari, J. C. Meyer, V. Scardaci, C. Casiraghi, M. Lazzeri, F. Mauri, S. Piscanec, D. Jiang, K. S. Novoselov, S. Roth, A. K. Geim, Raman Spectrum of Graphene and Graphene Layers, *Phys. Rev. Lett.,* Vol. 97, 2006, pp. 187401-03.

[13]. M. Acik, G. Lee, C. Mattevi, A. Pirkle, R. M. Wallace, M. Chhowalla, K. Cho, Y. Chabal, The Role of Oxygen during Thermal Reduction of Graphene Oxide Studied by Infrared Absorption Spectroscopy, *J. Phys. Chem. C,* Vol. 115*,* 2011, pp. 19761–19781.

[14]. S. Mallakpoura, A. Abdolmalekia, S. Borandeh, Covalently functionalized graphene sheets with biocompatible natural amino acids, *Applied Surface Science,* Vol. 307, 2014, pp. 533–542.

[15]. M. Pumera, Electrochemistry of graphene, graphene oxide and other graphenoids: Review, *Electrochemistry Communications,* Vol. 36, 2013, pp. 14-18.

[16]. Y. Zhang, A. Yang, X. Zhang, H. Zhao X. Li, Z. Yuan, Highly selective and sensitive biosensor for cysteine detection based on in situ synthesis of gold nanoparticles/graphene nanocomposites, *Colloids and Surfaces A:Physicochem. Eng. Aspects,* Vol. 436, 2013, pp. 815– 822.

[17]. T. Kuila, S. Bose, P. Khanra, A. K. Mishra, N. H. Kim, J. H. Lee, Recent advances in graphene based biosensors, *Biosensors and Bioelectronics,* Vol. 26, 2011, pp. 4637– 4648.

[18]. F. Cui, X. Zhang, Electrochemical sensor for epinephrine based on a glassy carbon electrode modified with graphene/ gold nanocomposites, *Journal of Electroanalytical Chemistry,* Vol. 669, 2012, pp. 35–41.
[19]. H. W. Tien, Y. L. Huang, S. Y. Yang, J. Y. Wang, C. C. Ma, The production of graphene nanosheets decorated with silver nanoparticles for use in transparent, conductive films, *Carbon,* Vol. 49, 2011, pp. 1550-1560.
[20]. Y. Li, L. Tang, J. Li, Preparation and electrochemical performance for methanol oxidation of Pt/graphene nanocomposite, *Electrochemistry Communications,* Vol. 11, 2009, pp. 846–849.
[21]. Z. Zheng, Y. Du, Q. Feng, Z. Wang, C. Wang, Facile method to prepare Pd/graphene – polyaniline nanocomposite and used as new electrode material for electrochemical sensing, *Journal of Molecular Catalysis A: Chemical,* Vol. 353, 2012, pp. 80– 86.
[22]. H. Han, D. Pan, X. Wu, Q. Zhang, H. Zhang, Synthesis of graphene/methylene blue/gold nanoparticles composites based on simultaneous green reduction, in situ growth and self-catalysis, *J. Mater. Sci.* Vol. 49, 2014, pp. 4796–4806.
[23]. F. Y. Kong, X. R. Li, W. W. Zhao, J. J. Xu, H. Y. Chen, Graphene oxide-thionine-Au nanostructure composite: Preparation and applications in non enzymatic glucose sensing, *Electrochem Commun.,* Vol. 14, 2012, pp. 59–62.
[24]. Y. Jiang, X. Zhang, C. Shan, S. Hua, Q. Zhang, X. Bai, Li Dan, Li. Niu, Functionalization of graphene with electrodeposited Prussian blue towards amperometric sensing application, *Talanta,* Vol. 85, 2011, pp. 76–81.
[25]. Qi Su, S. Pang, V. Alijani, C. Li, X. Feng, K. Mullen, Composites of graphene with large aromatic molecules, *Adv. Mater.,* Vol. 21, 2009, pp. 3191– 3195.
[26]. A. Ghosh, K. V. Rao, R. V. Subi, J. George, Non covalent functionalization, solubilization of graphene and single walled carbon nanotubes with aromatic donor and acceptor molecules, *Chemical Physics Letters*, Vol. 488, 2010, pp. 198– 201.
[27]. Z. G. Le, Z. Liu, Y. Qian, C. Wang, A facile and efficient approach to decoration of graphene nanosheets with gold nanoparticles, *Applied Surface Science*, Vol. 258, 2012, pp. 5348– 5353.
[28]. S. Zhang, S. Tang, J. Lei, H. Dong, H. Ju, Functionalizen of graphene nanribbons with porphyrin for electrocatalysis and amperometric biosensing, *Journal of Electroanalytical Chemistry,* Vol. 656, 2011, pp. 285– 288.
[29]. D. Sheng, S. Yang, Y. Zhang, J. Yang, J. Liu, R. Li, T. K. Sham, X. Sun, S. Ye, S Knight, Nitrogen doping effects on the structure of graphene, *Applied surface Science,* Vol. 257, 2011, pp. 9193-9198.
[30]. D. Ye, H. Li, G. Liang, J. Luo, X. Zhang, S. Zhang, H. Kong, A three-dimensional hybrid of MnO2/graphene/carbon nanotubes based sensor for determination of hydrogen-peroxide in milk, *J. Electrochimica Acta,* Vol. 109, 2013, pp. 195–200.
[31]. T. Gana, Chen, Hua, C. Z. S. A disposable electrochemical sensor for the determination of indole-3-acetic acid based on poly(safranine T) reduced graphene oxide nanocomposite, *Talanta,* Vol. 85, 2011, pp. 310–316.
[32]. D. Du, J. Liu, X. Y. Zhang, X. L. Cui, Y. H. Lin, One-step electrochemical deposition of a graphene-ZrO_2 nanocomposite: Preparation, characterization and application for detection of organophosphorus agents, *J. Mater. Chem.,* Vol. 21, 2011, pp. 8032–8037.
[33]. M. Y. Wang, J. R. Huang, M. Wang, Zhang, D. E., J. Chen, Electrochemical nonenzymatic sensor based on CoO decorated reduced graphene oxide for the simultaneous determination of carbofuran and carbaryl in fruits and vegetables, J. *Food Chemistry,* Vol. 151, 2014, pp. 191–197.
[34]. K. Wang, Q. Liu, L. Dai, J. Yan, C. Ju, B. Qiu, X. Wu, A highly sensitive and rapid organophosphate biosensor based on enhancement of CdS decorated graphene nanocomposite, *Analytica Chimica Acta,* Vol. 695, 2011, pp. 84–88.

[35]. G. Wang, X. Tan, Q. Zhou, Y. Liu, M. Wang, L. Yang, Synthesis of highly dispersed zinc oxide nanoparticles on carboxylic graphene for development a sensitive acetylcholinesterase biosensor, *Sens. Actuators B Chem.,* Vol. 190, 2014, pp. 730–736.
[36]. L. Yang, G. Wang, Y. Liu, M. Wang, Development of a biosensor based on immobilization of acetylcholinesterase on NiO nanoparticles-carboxylic acid graphene-nafion modified electrode for detection of pesticide, *Talanta,* Vol. 113, 2013, pp. 135–141.
[37]. K. Wang, H. N. Li, J. Wu, C. Ju, J. J. Yan, Q. Liu, B. Qiu, TiO2 decorated graphene nanohybrids for fabricating an amperometric acteylcholinesterase biosensor, *Analyst*, Vol. 136, 2011, pp. 3349– 3354.
[38]. W. Song, D. W. Li, Y T Li, Y. T. Long, Disposable biosensor based on graphene oxide conjugated with tyrosinase assembled gold nanoparticles, *Biosensors and Bioelectron*, Vol. 26, 2011, pp. 3181-3186.
[39]. L. Yang, G. Wang, Y. Liu, An acetylcholinesterase biosensor based on platinum nanoparticles –carboxylic graphene-nafion modified electrode for detection of pesticide, *Analytical Biochemistry,* Vol. 437, 2013, pp. 144–149.
[40]. T. Liu, H. Su, X. Qu, P. Ju, L. Cui, S. Ai, Acetylcholinesterase biosensor based on 3-carboxyphenylboronic acid/reduced graphene oxide-gold nanocomposite modified electrode for amperometric detection of organophosphorous and carbamate pesticide, *Sensors and Actuators B,* Vol. 160, 2011, pp. 1255– 1261.
[41]. X. Xue, Q. Wei, D. Wu, H. Li, Y. Zhang, R. Feng, B. Du, Determination of methyl parathion Determination of methyl parathion by a molecularly imprinted sensor based on nitrogen doped graphene sheets, *Electrochimica Acta,* Vol. 116, 2014, pp. 366– 371.
[42]. Y. Wang, S. Zhang, D. Du, Y. Shao, Z. Li, J. Wang, M. H. Engelhard, J. Li, Y. Lin, Self assembly of acetylcholinesterase on a gold nanoparticles-graphene nanosheet hybrid for organophosphate pesticide detection using polyelectrolyte as a linker, *Journal of Materials Chemistry,* Vol. 21, 2011, pp. 5319–5325.
[43]. Y. Yang, A. M. Asiri, D. Du, Y. Lin, Acetylcholinesterase biosensor based on a gold nanoparticle–polypyrrole–reduced graphene oxide nanocomposite modified electrode for the amperometric detection of organophosphorus pesticides, *Analyst,* Vol. 139, 2014, pp. 3055–3060.
[44]. C. Zhai, Y. Guo, X. Sun, Y. Zheng, X. Wang, An acetylcholinesterase biosensor based on graphene-gold nanocomposite and calcined layered double hydroxide, *Enzyme Microb. Tech.,* Vol. 58, 2014, pp. 8–13.
[45]. L. Zhang, L. Long, W. Zhang, D. Du, Y. Lin, Study of Inhibition, Reactivation and Aging Processes of Pesticides Using Graphene Nanosheets/Gold Nanoparticles-Based Acetylcholinesterase Biosensor, *Electroanalysis,* Vol. 24, 2012, pp. 1745–1750.
[46]. J. Gong, X. Miao, T. Zhou, L. Zhang, An enzymeless organophosphate pesticide sensor using Au nanoparticles decorated graphene hybrid nanosheet as solid phase extraction, *Talanta,* Vol. 85, 2011, pp. 1344–1349.
[47]. Z. M. Liu, Y. F. Jing, Z. L. Wang, H. J. Zhan, Q. Shen, Highly Sensitive Electrochemical Biosensing of Methyl Parathion Pesticide Based on Acetylcholinesterase Immobilized Onto Graphene-Fe_3O_4 Nanocomposite, *Sensors Letters,* Vol. 11, 2013, pp. 531- 538.
[48]. T. M. B. F. Oliveira, M. F. Barroso, S. Morais, M. Araújo, C. Freire, P. de Lima-Neto, A. N. Correia, M. B. P. P. Oliveira, C. D. Matos, Sensitive bi-enzymatic biosensor based on polyphenoloxidases-gold nanoparticles-chitosan hybrid film-graphene doped carbon paste electrode for carbamates detection, *Bioelectrochemistry,* Vol. 98, 2014, pp. 20–29.
[49]. X. Sun, X. Wang, Acetylcholinesterase biosensor based on Prussian blue-modified electrode for detecting organophosphorous pesticide, *Biosensors and Bioelectronics*, Vol. 25, 2010, pp. 2611–2614.
[50]. T. M. B. F. Oliveira, M. F. Barroso, S. Morais, M. Araújo, C. Freire, P. L.neto, A. N. Correia, M. B. P. P. Oliveira, C. D. Matos, Laccase –Prussian blue film –graphene doped carbon paste

modified electrode for carbamate pesticides quantification, *Biosensors and Bioelectronics,* Vol 55, 2013, pp. 292–29.
[51]. L. Zhang, A. Zhang, D. Du, Y. Lin, Biosensor based on Prussian blue nanocubes/reduced graphene oxide nanocomposite for detection of organophosphorus pesticides, *Nanoscale,* Vol. 4, 2012, pp. 4674–4679.
[52]. M. R. Xu, J. Y. Zhu, H. C. Su, J. Dong, S. Y. Ai, R. Li, Electrochemical determination of methyl parathion using poly(malachite green)/graphene nanosheets–nafion composite film-modified glassy carbon electrode, *Journal of Applied Electrochemistry*, Vol. 42, 2012, pp. 509–516.
[53]. L. Zhao, F. Zhao, B. Zeng, Fabrication of microchip electrophoresis devices and effect of channel surface properties on separation efficiency, *Sensors and Actuators B,* Vol. 176, 2013, pp. 818– 824.
[54]. Y. Li, G. Han, Retracted article: Ionic liquid –functionalized graphene for fabricating an amperometric actelcholinesterase biosensor, *Analyst,* Vol. 137, 2012, pp. 3160–3165.
[55]. H. F. Xu, H. Dai, G. Chen, Direct electrochemistry and electrocatalysis of hemoglobin protein entrapped in graphene and chitosan composite, *Talanta,* Vol. *81,* 2010, pp. 334-338.
[56]. K. J. Huanga, Y. X. Miao, L. Wang, T. Gana, M. Yua, L. L. Wang, Direct electrochemistry of hemoglobin based on chitosan-ionic liquid-ferrocene/graphene composite film, *Process Biochemistry,* Vol. 47, 2012, pp. 1171–1177.
[57]. J. Xu, C. Liu, Z. Wu, Direct electrochemistry and enhanced electrocatalytic activity of hemoglobin entrapped in graphene and ZnO nanosphere composite film, *Microchim Acta,* Vol. 172, 2011, pp. 425–430.
[58]. L. Wang, W. Q. R. Su, Z. He, Facile method to synthesize graphene-ZnS nanocomposites: preparation and application in bioelectrochemistry of hemoglobin, *J Solid State Electrochem,* Vol. 17, 2013, pp. 2595–2602.
[59]. X. Feng, R. Li, C. Hu, W. Hou, Direct electron transfer and electrocatalysis of hemoglobin immobilized on graphene-Pt nanocomposite*, Journal of Electroanalytical Chemistry,* Vol. 657, 2011, pp. 28– 33.
[60]. Q. Feng, K. Liu, J. Fua, Y. Zhang, Z. Zheng, C. Wanga, Y. Dua, W. Ye, Direct electrochemistry of hemoglobin based on nanocomposite film of gold nanoparticles and poly(diallyldimethylammonium chloride) functionalized graphene, *Electrochimica Acta,* Vol. 60, 2012, pp. 304– 308.
[61]. N. L. Teradala, P. S. Narayana, A. K. Satpatib, J. Seetharamappaa, Fabrication of electrochemical sensor based on green reduction of graphene oxide for an antimigraine drug, rizatriptan benzoate, *Sensors and Actuators B,* Vol. 196, 2014, pp. 596–603.
[62]. S. Mallakpoura, A. Abdolmalekia, S. Boranden, Covalently functionalized graphene sheets with biocompatible natural amino acids, *Applied Surface Science,* Vol. 307, 2014, pp. 533– 542.
[63]. H. Zhang, Z. Li, A. Snyder, J. Xie, L. A. Stanciu, Functionalized graphene oxide for the fabrication of paraoxon biosensors, *Analytica Chimica Acta,* Vol. 827, 2014, pp. 86– 94.
[64]. T. Premkumar, K. E. Geckeler, Graphene-DNA hybrid materials: Assembly, applications, and prospects, *Progress in Polymer Science,* Vol. 37, 2012, pp. 515– 529.
[65]. G. K. Ramesha, S. Sampath, In situ formation of graphene-lead oxide composite and its use in trace arsenic detection, *Sensors and Actuators B*, Vol. 160, 2011, pp. 306– 311.
[66]. L. M. Lu, H. B. Li, F. Qu, X. B. Zhang, G. L. Shen, R. Q. Yu, In situ synthesis of palladium nanoparticles –graphene nanohybrids and their application in nonenzymatic glucose biosensors, *Biosensors and Bioelectronics,* Vol. 26, 2011, pp. 3500–3504.
[67]. Z. Wang, E. Liu, Graphene ultrathin film electrode for detection of lead ions in acetate buffer solution, *Talanta,* Vol. 103, 2013, pp. 47–55.

[68]. K. Zhou, Y. Zhu, X. Yang, C. Li, Electrocatalytic Oxidation of Glucose by the Glucose Oxidase Immobilized in Graphene-Au-Nafion Biocomposite, *Electroanalysis,* Vol. 22, 2010, pp. 259–264. .

[69]. C. Yu, Y. Guo, H. Liu, N. Yan, Z. Xu, G. Yu, Y. Fang, Y. Liu, Ultrasensitive and selective sensing of heavy metal ions with modified graphene, *Chem. Commun.,* Vol. 49*,* 2013, pp. 6492– 6494.

[70]. C. M. Willemse, K. Tlhomelang, N. Jahed, P. G. Baker, E. I. Iwuoha, Metallo-Graphene Nanocomposite Electrocatalytic Platform for the Determination of Toxic Metal Ions, *Sensors,* Vol. 11, 2011, pp. 3970–3987.

[71]. C. H. Lien, K. H. Chang, C. C. Hu, D. S. H. Wang, Optimizing Bismuth-Modified Graphene-Carbon Nanotube Composite-Coated Screen Printed Electrode for Lead-Ion Sensing through the Experimental Design Strategy, *J. Electrochem. Soc.,* Vol. 160, 2013, pp. 107–112**.**

[72]. N. Kong, J. Liu, Q. Kong, R. Wang, C. J. Barrow, W. Yang, Graphene modified gold electrode via π-π stacking interaction for analysis of Cu^{2+} and Pb^{2+}, *Sens. Actuators B Chem*., Vol. 178, 2013, pp. 426- 433.

[73]. J. T. Zhang, Z. Y. Jin, W. C. Li, W. Dong, A. H. Lu, Graphene modified carbon nanosheets for electrochemical detection of Pb(II) in water, *J. Mater. Chem. A*, Vol. 1, 2013, pp. 13139–13145.

[74]. T. Zhang, Z. Cheng, Y. Wang, Z. Li, C. Wang, Y. Li, Y. Fang, Self-Assembled 1-Octadecanethiol Monolayers on Graphene for Mercury Detection, *Nano Lett.,* Vol. 10, 2010, pp. 4738–4741.

[75]. X. Ran, H. Sun, F. Pu, J. Ren, X. Qu, Ag Nanoparticle-decorated graphene quantum dots for label-free, rapid and sensitive detection of Ag^{+} and biothiols, *Chem. Commun.,* Vol. 49, 2013, pp. 1079–1081.

[76]. S. R. Dey, C. R. Raj, Development of an Amperometric Cholesterol Biosensor Based on Graphene−Pt Nanoparticle Hybrid Material, *J. Phys. Chem. C,* Vol. 114, 2010, pp. 21427–21433.

[77]. L. Li, Z. Du, S. Liu, A novel nonenzymatic hydrogen peroxide sensor based on MnO_2/graphene oxide nanocomposite, *Talanta,* Vol. 82, 2010, pp. 1637–1641.

[78]. A. Sun, Q. Sheng, A. Zheng, Direct Electrochemistry of Hemoglobin Imobilized on a Functionalized Multi-Walled Carbon Nanotubes and Gold Nanoparticles Nanocomplex-Modified Glassy Carbon Electrode, *J. Appl Biochem Biotechnol,* Vol 166, 2012, pp. 764–773.

[79]. P. Pang, Z. Yang, S. Xiao, J. Xie, Y. Zhang, Y. Gao, Nonenzymatic amperometric determination of hydrogen peroxide by graphene and gold Nanorods nanocomposite modified electrode, *Journal of Electroanalytical Chemistry,* Vol. 727, 2014, pp. 27–33.

[80]. Y. Zhang, Y. Liu, J. He, P. Pang, Y. Gao, Q. Hu, Electrochemical behavior of graphene/nafion/azure I/Au nanoparticles composites glass carbon electrode and its application as nonenzymatic hydrogen peroxide sensor, *Electrochimica Acta,* Vol. 90, 2013, pp. 550– 555.

[81]. D. Ye, H. Li, G. Liang, J. Luo, X. Zhang, S. Zhang, H. Chen, A. Kong, A three dimensional hybrid of MnO2 /graphene/carbon nanotubes based sensor for determination of hydrogen peroxide in milk, J. *Electrochimica Acta,* Vol. 109, 2013, pp. 195– 200.

[82]. L. M. Lu, H. B. Li, F. Qu, X. B. Zhang, G. L. Shen, R. Q. Yu, Improved protein detection on an AC electrokinetic quartz crystal microbalance, *Biosensors and Bioelectronics*, Vol. 26, 2011, pp. 3500–3504.

[83]. N. Cheng, H. Wang, X. Li, X. Yang, L. Zhu, Amperometric Glucose Biosensor Based on Integration of Glucose Oxidase with Palladium Nanoparticles/Reduced Graphene Oxide Nanocomposite, *American Journal of Analytical Chemistry,* Vol. 3, 2012, pp. 312- 319.

[84]. Y. Li, F. Huang, J. Chen, T. Mo, S. Li, F. Wang, S. Feng, Y. Li, glucose sensor based on the graphene-CuO nanocomposites, *Int. J. Electrochem. Sci.,* Vol. 8, 2013, pp. 6332 – 6342.

[85]. H. D. Jang, S. K. Kim, H. Chang, K. M. Roh, J. W. Choi, J. Huang, A glucose biosensor based on TiO2-graphene composite, *Biosensors and Bioelectronics,* Vol. 38, 2012, pp. 184–188.
[86]. Y. Liu, D. Yu, C. Zeng, Z. Miao, L. Dai, Biocompatible Graphene Oxide-Based Glucose Biosensors, *Langmuir,* Vol. 26, 2010, pp. 6158–6160.
[87]. B. Unnikrishnan, S. Palanisamy, S. M. Chen, A simple electrochemical approach to fabricate a glucose biosensor based on graphene-glucose oxidase biocomposite, *Biosensors and Bioelectronics,* Vol. 39, 2013, pp. 70–75.
[88]. H. Wua, J. Wang, X. Kang, C. Wang, D. Wang, J. Liu, I. A. Aksayb, Y. Lin, Glucose biosensor based on immobilization of glucose oxidase in platinum nanoparticles/graphene/ chitosan nanocomposite film, *Talanta*, Vol. 80, 2009, pp. 403– 406.
[89]. B. Liang, L. Fang, G. Yang, Y. Hu, X. Guo, X. Ye, Direct electron transfer glucose biosensor based on glucose oxidase self assembled on electrochemistry reduced carboxyl graphene, *Biosensors and Bioelectronics*, Vol. 43, 2013, pp. 131– 136.
[90]. H. B. Nguyen, V. C. Nguyen, V. T. Nguyen, T. T. T. Ngo, N. T. Nguyen, T. T. H. Dang, D. L. Tran, P. Q. Do, X. N. Nguyen, X. P. Nguyen, H. K. Phan, N. M. Phan, Graphene patterned polyaniline-based biosensor for glucose detection, *Adv. Nat. Sci.: Nanosci. Nanotechnol.,* Vol. 3, 2012, pp. 025011.
[91]. M. L. Yola, V. K. Gupta, T. Eren, A. E. S. N. Atar, A novel electroanalytical nanosensor based on graphene oxide/ silver nanoparticles for simultaneous determination of quercetin and morin, *Electrochimica Acta,* Vol. 120, 2014, pp. 204–211.
[92]. I. Tiwari, M. Singh, C. M. Pandey, G. Sumana, Electrochemical genosensor based on graphene oxide modified iron oxide-chitosan hybrid nanocomposite for pathogen detection, *Sensor and Actuators B,* Vol. 206, 2015 276-283.
[93]. I. Tiwari, M. Singh, C. M. Pandey, G. Sumana, Electrochemical detection of pathogenic *Escherichia coli* specific DNA sequence based on graphene oxide- chitosan composite decorated with nickel ferrite nanoparticles, *RSC Advances,* 2015, Accepted.
[94]. I. Tiwari, M. Gupta, C. M. Pandey, Graphene sheets decorated gold nanoparticles-polypyrrole based nanocomposite: Synthesis, characterization and genosensing application, *Dalton Transaction,* 2015, Accepted.

Chapter 7
Gold Nanoparticle Based Colorimetric Sensors for Dopamine

Nokuthula Ngomane, Nelson Torto, Sibulelo Vilakazi and Rui Krause

7.1. Dopamine

Dopamine (3,4–dihydroxyphenylethylamine) (Fig. 7.1) is a biogenic catecholamine neurotransmitter [1]. It was first synthesized by George Barger and James Hill in 1910 at The Wellcome Laboratories in London, England by decarboxylation of 3,4–dihydroxyphenylalanine which is also known as $_{L}$–DOPA. The name dopamine (DA) was first suggested by Henry Dale in 1952 [2]. Before the mid–1950s, DA was known as just a precursor in the formation of norepinephrine. However, in the late 1950s DA began to be recognized as having important physiological effects in its own right. Further studies on DA by Carlsson *et al*. led to the realization that it is a neurotransmitter [3].

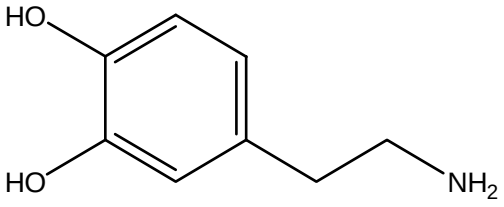


Fig. 7.1. Chemical structure of dopamine.

DA is naturally produced and widely distributed in the central nervous system (olfactory bulb, *substantia nigra*, ventral tegmental area, and the retina) of mammals. It plays a number of significant roles in the function of human metabolism, cardiovascular, central nervous, renal, as well as hormonal systems [4]. DA affects the fundamental functions of

Nokuthula Ngomane
Chemistry Department, Rhodes University, P.O. Box 94, Grahamstown, South Africa.
Ecoplanet Bamboo Laboratory, Kowie Bamboo Farm, Martindale Road, Bathurst, 6166, South Africa.

the brain such as learning, movement, emotions, the sense of smell, sensitivity to light, and the ability to experience pain. DA is also an integral to higher brain functions such as salience, euphoria, pleasure and appetitive and consummatory aspects of reward. Moreover, abnormal levels of DA in the body fluids are indications of various serious diseases and disorders such as Parkinson's disease, Huntington's disease, and Schizophrenia [5-7]. Furthermore, DA coexists with species including, ascorbic acid and uric acid in biological fluids such as urine and blood. Thus many researchers have focused on developing analytical methods that are highly selective and sensitive in order to detect DA in the presence of interfering substances [1, 8, 9].

There are various conventional analytical techniques that have been reported for detecting DA. These conventional techniques include chemiluminescent, fluorescent, spectrophotometric and electrochemical detection. Methods developed utilizing conventional techniques have challenges including the use of analytically specific instrumentations that are difficult to operate and follow time-consuming procedures. It is therefore essential to develop an analytical method that is simple yet selective and sensitive for diagnostic application. As a better alternative method, colorimetric sensors are being developed for routine analysis. Colorimetric detection relies on colour changes of materials induced by the presence of the targeted analyte and will be discussed in detail in Section 7.2.

7.2. Colorimetric Detection

Various methods for the detection of biological compounds have been reported [10-14]. The newly developed methods have been characterized by several advantages over conventional methods, particularly in sensitivity and selectivity [15-18]. Among these methods, colorimetric detection appears to be most commonly used for routine analysis because of its fascinating characteristics that include its simplicity, low cost and that detection can be achieved with the naked eye [19], thus complicated instruments are not required. Various colorimetric detection techniques have been documented, including strategies employing organic dyes (e.g. azo dyes) [20], colorimetric polymeric materials (e.g. polydiacetylene, PDA) [21] and metal NPs (e.g. silver and gold) [22]. Currently, colorimetric methods that are based on the use of metal nanoparticles are the most attractive.

7.3. Metal Nanoparticle–based Colorimetric Sensors

Metal nanoparticles have attracted great interest in science and technology mainly due to their unique properties and numerous applications. The areas of application include but are not limited to catalysis, electronics, optics [23], imaging, chemical sensors and biosensors [24]. Noble metal NPs particularly silver (AgNPs) and AuNPs are the subject of focused research owing to their optical and electronic features. Thanks to the remarkable properties of AuNPs and AgNPs that makes them ideal colorimetric probes in the development of analytical methods for detecting various analytes [25,26]. Both

AgNPs and AuNPs are extremely sensitive owing to their higher extinction coefficients compared with traditional organic chromophores [27].

However, until recently, AgNPs were mainly exploited for their antibacterial activity [28-30]. Hence, the few reports on the colorimetric detections based on AgNPs. The core challenge that has delayed the use of AgNPs as colorimetric probes is the susceptibility of the silver surface to oxidation, which makes it difficult to stabilize them for their analytical applications [24, 31, 32]. Recently, stabilization of AgNPs has been achieved through surface functionalization with appropriate ligands. Nevertheless, this review focused on AuNPs.

7.3.1. Gold Nanoparticles (AuNPs)

The AuNPs have drawn considerable attention as colorimetric probes because they are easy to synthesize and functionalize, have distinct spectral features and, long-term stability [33, 34]. They have colours ranging from red to purple/blue and violet [35]. The colour change is highly sensitive to the size, shape, stabilizing agents, refractive index of the medium, as well as the aggregation state of the AuNPs. AuNP–based colorimetric probes have been widely applied in a variety of research fields including clinical diagnostics [36, 37], detection of DNA [38], proteins [39], metal ions and small molecules [26] and so on. In order to satisfy the desired application, various procedures are followed for the synthesis of the AuNPs.

7.3.2. Methods for the Synthesis of Gold Nanoparticles

There are three main reagents for preparation of metal nanoparticles namely, a metal salt, a stabilizing or capping agent and a reducing agent. The function of the reducing agent is to reduce the metal ions to the atom that would precipitate to form nanoparticles. The stabilizing agent prevents formed nanoparticles from aggregating before their intended use. The reaction conditions also play a crucial role during the synthesis of the nanoparticles. Studies have shown that the method of synthesis of the AuNPs is important for both the mechanism of detection, their storage and may also determine their stability. As a result, depending on the intended use, AuNPs have been prepared under different conditions. For example, organic and inorganic solvents have been utilized as reaction media. Nanoparticles formed in solution are called colloidal nanoparticles. When specific shapes or sizes of AuNPs are required for a particular application, a method known as seed-grown synthesis is normally used. This method involves two or more steps (Fig. 7.2). Recently, an in situ strategy for the synthesis of NPs has been reported. The in situ method has received more attention since it involves formation of the NPs in one step. The two main chemical strategies for the synthesis of AuNPs are discussed in 7.3.2.1 and 7.3.1.2.

7.3.2.1. Seed–growth Method

As already mentioned in 7.3.1, the seed–growth approach is employed to prepare nanoparticles with desired morphology. In a general seed–growth method, gold seeds or

germ (small sized particles) are first prepared by chemical reduction of appropriate quantities of the gold salt using a suitable reducing agent in the presence of a stabilizing agent such as citrate and hexadecyltrimethyammonium bromide (CTAB) [40, 41]. The smaller particles are usually produced by using stronger reducing agents such as sodium borohydride ($NaBH_4$), phosphorus, or tetrakis(hydroxymethyl) phosphonium chloride [42-45]. The seeds are then added to a solution containing more metal salt. The reducing agent used in the second stage of the seed-growth process is a weaker one; for example, hydroxylamine and ascorbic acid [46, 47]. The weak reducing agents should reduce only the metal salt ions which are adsorbed onto the surface of the seed without forming any new nucleation or growth centres. Thus, the seeds serve as nucleation sites for the formation of particular sizes or shapes of the AuNPs [48, 49]. The whole process takes approximately 16 h or more [50]. There are various successful reports on the use of the method to produce different shapes and sizes of AuNPs.

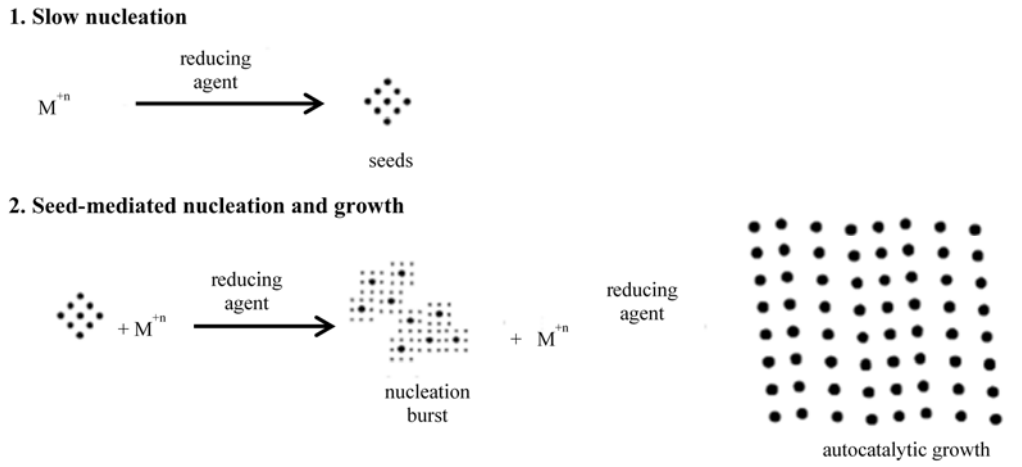


Fig. 7.2. Generalized two-step mechanism for solution-phase AuNP synthesis [51].

7.3.2.2. *In situ* Synthesis

In the *in situ* reduction method, both nucleation and growth are completed in a single step. There are two in situ reduction methods that have been applied to tetrachloauric acid ($HAuCl_4$) precursor namely photoreduction by UV irradiation and reduction by potassium (or sodium) borohydride. Reduction by UV irradiation in the presence of various stabilizing agents is a slow process and usually results in large particles; whereas the fast borohydride reduction method generally gives colloids that are stable for months with very small, spherical particles and narrower size distribution [52]. The size and shape of the nanoparticles is necessary in the development of colorimetric probes since they are the mostly affected when the analyte of interest interact with the AuNPs.

7.4. Gold Nanoparticle–based Colorimetric Detection of Dopamine

An ideal colorimetric probe must be simple, sensitive and selective towards the targeted analyte. Since AuNPs are naturally sensitive due to their high extinction coefficients, one needs to think about how to cater for selectivity. Selectivity of various analytes have been accomplished by choosing stabilizing agents that have functional groups which will allow selective binding of the target in the presence of interfering species. There are several reports for the colorimetric detection of DA employing AuNPs. Most of the approaches follow a two–step process. The first step is usually the *in situ* preparation of the AuNPs employing the citrate-mediated reduction of $HAuCl_4$. The citrate acts as both a reducing and stabilizing agent to produce wine–red solution. Basically, the citrate method involves heating of $HAuCl_4$ aqueous solution to vigorous boiling followed by the addition of sodium citrate solution under continuous stirring. The heating is stopped after boiling for few min to allow the solution to cool down to room temperature [53, 54]. In the second step, the citrate ion is exchanged for another ligand.

For example, Liu *et al.* [55] synthesized dithiobis(succinimidylpropionate)-modified AuNPs following a ligand–exchange reaction. The probe showed good selectivity in the presence of some interfering compounds with a limit of detection (LOD) of 2 nM. Recently, Feng *et al*. [56] developed a probe based on 4–amino–3–hydrazino–5–mercapto–1,2,4–triazole (AHMT) functionalized AuNPs. The selectivity of the probe relied on the hydrogen bonding between the amine and hydroxyl groups of DA with AHMT (Fig. 7.3). Employing the probe, the authors obtained an LOD of 0.07 μM.

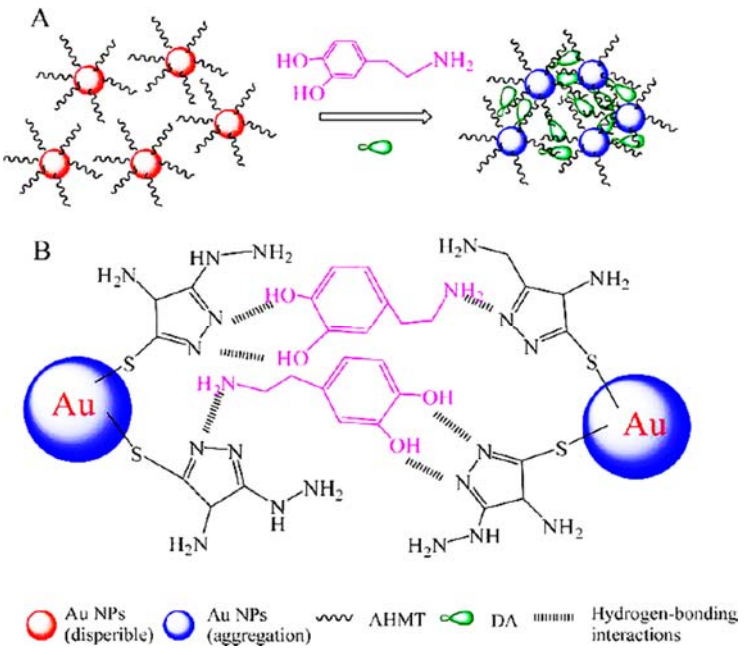


Fig.7.3. Colorimetric Detection of Dopamine Using AHMT-AuNPs as a Probe [57].

Another method utilizing AuNPs prepared via ligand–exchange reaction was reported by Lee *et al*. [58]. However, instead of following the citrate-mediated reduction, the AuNPs were prepared in a mixture of water and toluene in the presence of tetraoctylammonnium bromide (TOAB) as a stabilizing agent. In this case, $NaBH_4$ was used as a reducing agent. The TOAB–AuNPs were then mixed with a solution of 4–dimethyl–aminopyridine (DMAP). The hydrophilic DMAP molecules were replaced by the hydrophobic TOAB molecules on the surface of the AuNPs such that the DMAP–AuNPs moved into the water phase. Thus, achieving both ligand–exchange and phase transfer. The mechanism of detection by the probe is via the core etching of the DMAP–AuNPs by DA thereby resulting in a reduction in size of the prepared nanoparticles, causing colour change and an LOD of 5 nM was observed (Fig. 7.4).

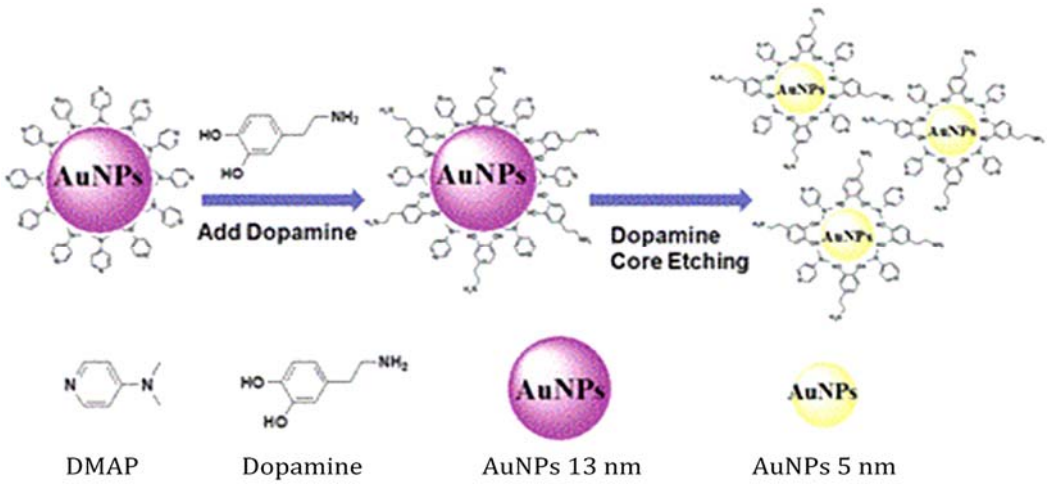


Fig. 7.4. DA detection based on the core etching effect of DMAP–AuNPs [59].

Following the citrate reduction procedure, Zheng *et al*. [60] demonstrated a strategy for detecting DA employing a DA–binding aptamer as recognition element. Kong *et al*. [7] functionalized AuNPs with 4–mercaptophenylboronic acid and disthiobis(succinimidylpropionate) for double molecular recognition of DA. In another study [61], the citrate ion on the surface of the AuNPs was exchanged for dithiobis(sulfosuccinimidylpropionate) and the resulting nanoparticles were used in conjunction with Fe_3O_4 as recognition species for DA. Liu *et al*. [62] immobilized Ag^+ on the surface of gold nanorods (AuNR). The mechanism of detection involved the reduction of the Ag^+ to Ag^0 which resulted in the change of the dielectric function, the aspect ratio of the AuNR, and colour of the solution. The structure of DA has got hydroxyl groups that can interact with other metal ions by coordination. When AuNPs are stabilized by the nitrogen group of DA, the –OH groups from two DA molecules can interact via hydrogen bonding or if a metal ion such as Cu^{2+} are introduced coordination can occur. These interactions will result in a change of the colour of the AuNPs (a schematic diagrams showing the possible interactions and colour change is shown in Fig. 7.5).

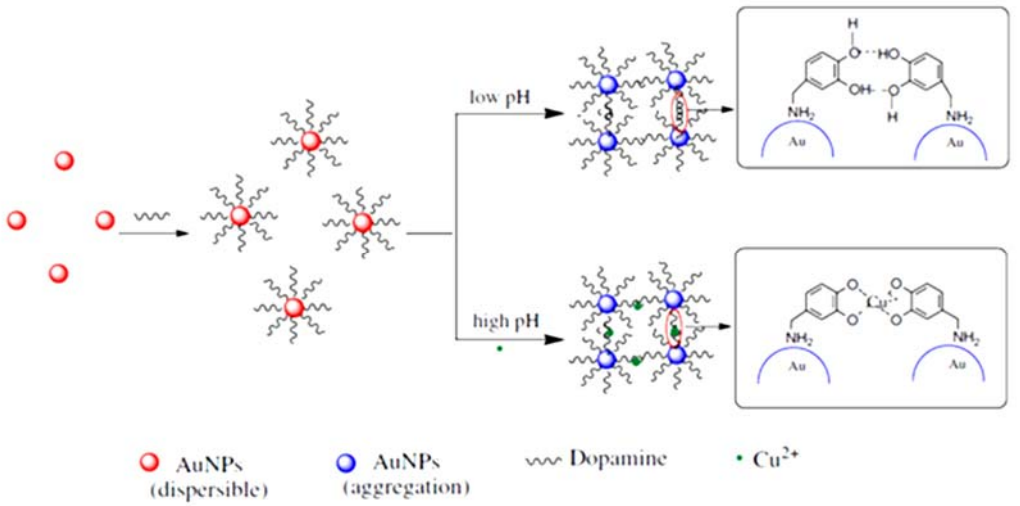


Fig. 7.5. The detection of DA based on cross-linking aggregation of AuNPs. In the absence of Cu^{2+}, AuNPs aggregation is induced by hydrogen bonding; in the presence of Cu^{2+}, DA can bind Cu^{2+} by coordination with its hydroxyl (−OH), which induced AuNPs aggregated [63].

7.5. Conclusions

All the examples of AuNP–based colorimetric sensors have shown undoubtable selectivity and sensitivity towards DA. Most of the research done on these colorimetric sensors for DA are reported from 2010 and so far there are at least 2-3 papers published annually, this means AuNPs are recognized for their potential to detect DA. However, most the reported sensors are liquid state based and are susceptible to disadvantages such as difficulty to transport to places in which the roads are not well matained. Thus, they could only be readily available to the elite. It would be interesting to see more of such studies utilizing metal nanoparticles or other chromoresphores immobilized in solid materials. The advantage of solid state sensors over liquid state is that they are easy to handle and transport.

Acknowledgements

Nokuthula Ngomane express her sincere gratitude to MINTEK (Randburg, South Africa), the Henderson Postgraduate Scholarship and NRF Scarce Skills Scholarship (South Africa) for financial support.

References

[1]. J. Huang, Y. Liu, H. Hou, T. You, Simultaneous electrochemical determination of dopamine, uric acid and ascorbic acid using palladium nanoparticle-loaded carbon nanofibers modified electrode, *Biosens Bioelectron*, Vol. 24, 2008, pp. 632-637.

[2]. C. A. Marsden. Dopamine: the rewarding years, *Br J Pharmacol*, Vol. 147, pp. S136-S144.
[3]. S. Fahn, The history of dopamine and levodopa in the treatment of Parkinson's disease*, Mov Disord,* Vol. 23, 2008, Suppl. 3, pp. 497-508.
[4]. C. R. Raj, T. Okajima, T. Ohsaka, Gold nanoparticle arrays for the voltammetric sensing of dopamine, *Journal of Electroanalytical Chemistry*, Vol. 543, 2003, pp. 127-133.
[5]. R. M. Wightman, L. J. May, A. C. Michael, Detection of dopamine dynamics in the brain, *Anal Chem*, Vol. 60, 1988, pp. 769A-779A.
[6]. R. G. W. Staal, S. Rayport, D. Sulzer, Amperometric detection of dopamine exocytosis from synaptic terminals, *CRC Press LLC*, 2007, pp. 337-352.
[7]. B. Kong, A. Zhu, Y. Luo, Y. Tian, Y. Yu, G. Shi, Sensitive and Selective Colorimetric Visualization of Cerebral Dopamine Based on Double Molecular Recognition, *Angew Chem, Int Ed*., Vol. 50, 2011, pp. 1837-1840.
[8]. Z. Zhuang, J. Li, R. Xu, D. Xiao, Electrochemical detection of dopamine in the presence of ascorbic acid using overoxidized polypyrrole/graphene modified electrodes, *Int J Electrochem Sci*, Vol. 6, 2011, pp. 2149-2161.
[9]. Y. R. Kim, S. Bong, Y. J. Kang, Y. Yang, R. K. Mahajan, J. S. Kim, H. Kim, Electrochemical detection of dopamine in the presence of ascorbic acid using graphene modified electrodes, *Biosensors and Bioelectronics*, Vol. 25, 2010, pp. 2366-2369.
[10]. D. S. Linthicum, J. Patel, N. Cairns, Antibody-based fluorescence polarization assay to screen combinatorial libraries for sweet taste compounds, *Comb Chem High Throughput Screening*, Vol. 4, 2001, pp. 431-438.
[11]. Y. Ito, S. I. Yamazaki, K. Kano, T. Ikeda, Escherichia coli and its application in a mediated amperometric glucose sensor, *Biosens Bioelectron*, Vol. 17, 2002, pp. 993-998.
[12]. E. Eteshola, D. Leckband, Development and characterization of an ELISA assay in PDMS microfluidic channels, *Sensors Actuators, B*, Vol. 72, 2001, pp. 129-133.
[13]. N. J. Goddard, K. Singh, J. P. Hulme, C. Malins, R. J. Holmes, Internally-referenced Resonant Mirror devices for dispersion compensation in chemical sensing and biosensing applications, *Sens Actuators, A*, Vol. 100, 2002, pp. 1-9.
[14]. R. Pauwels, J. Balzarini, M. Baba, R. Snoeck, D. Schols, P. Herdewijn, J. Desmyter, E. De Clercq, Rapid and automated tetrazolium-based colorimetric assay for the detection of anti-HIV compounds, *Journal of Virological Methods*, Vol. 20, 1988, pp. 309-321.
[15]. E. S. Forzani, H. Zhang, W. Chen, N. Tao, Detection of Heavy Metal Ions in Drinking Water Using a High-Resolution Differential Surface Plasmon Resonance Sensor, *Environ Sci Technol*, Vol. 39, 2005, pp. 1257-1262.
[16]. H. Li, J. Fan, F. Song, H. Zhu, J. Du, S. Sun, X. Peng, Fluorescent probes for Pd2+ detection by allylidene-hydrazone ligands with excellent selectivity and large fluorescence enhancement, *Chemistry*, Vol. 16, 2010, pp. 12349-12356.
[17]. X. Xu, J. Wang, F. Yang, K. Jiao, X. Yang, Label-Free Colorimetric Detection of Small Molecules Utilizing DNA Oligonucleotides and Silver Nanoparticles, *Small*, Vol. 5, 2009, pp. 2669-2672.
[18]. S. Kim, M. S. Eom, S. K. Kim, S. H. Seo, M. S. Han, A highly sensitive gold nanoparticle-based colorimetric probe for pyrophosphate using a competition assay approach, *Chem Commun*, Vol. 49, 2013, pp. 152-154.
[19]. Y. Song, W. Wei, X. Qu, Colorimetric Biosensing Using Smart Material, *Adv Mater*, Vol. 23, 2011, pp. 4215-4236.
[20]. D. Zhang, W. Jin, Highly selective and sensitive colorimetric probe for hydrogen sulfide by a copper (II) complex of azo-dye based on chemosensing ensemble approach, *Spectrochim Acta*, Part A, Vol. 90, 2012, pp. 35-39.
[21]. F. Zheng, Z. Wu, Y. Chen, A quantitative method for the measurement of membrane affinity by polydiacetylene-based colorimetric assay, *Anal Biochem*, Vol. 420, 2012, pp. 171-176.

[22]. C. A. Mirkin, R. L. Letsinger, R. C. Mucic, J. J. Storhoff, A DNA-based method for rationally assembling nanoparticles into macroscopic materials, *Nature*, Vol. 382, 1996, pp. 607-609.
[23]. S. Sun, S. Anders, T. Thomson, J. E. E. Baglin, M. F. Toney, H. F. Hamann, C. B. Murray, B. D. Terris, Controlled Synthesis and Assembly of FePt Nanoparticles, *J Phys Chem B*, Vol. 107, 2003, pp. 5419-5425.
[24]. Z. Chen, X. Zhang, H. Cao, Y. Huang, Chitosan-capped silver nanoparticles as a highly selective colorimetric probe for visual detection of aromatic ortho-trihydroxy phenols, *Analyst*, Vol. 138, 2013, pp. 2343-2349.
[25]. S. Song, Y. Qin, Y. He, Q. Huang, C. Fan, H. Y. Chen, Functional nanoprobes for ultrasensitive detection of biomolecules, *Chem Soc Rev*, Vol. 39, 2010, pp. 4234-4243.
[26]. D. Liu, Z. Wang, X. Jiang, Gold nanoparticles for the colorimetric and fluorescent detection of ions and small organic molecules, *Nanoscale*, Vol. 3, 2011, pp. 1421-1433.
[27]. Y. Zhang, B. Li, X. Chen, Simple and sensitive detection of dopamine in the presence of high concentration of ascorbic acid using gold nanoparticles as colorimetric probes, *Microchim Acta*, 168, 2010, pp. 107-113.
[28]. M. Rai, A. Yadav, A. Gade. Silver nanoparticles as a new generation of antimicrobials. *Biotechnology Advances*, Vol. 27, 2009, pp. 76-83.
[29]. J. S. Kim, E. Kuk, K. N. Yu, J.-H. Kim, S. J. Park, H. J. Lee, S. H. Kim, Y. K. Park, Y. H. Park, C. Y. Hwang. Antimicrobial effects of silver nanoparticles, Nanomedicine: Nanotechnology, *Biology and Medicine*, Vol. 3, 2007, pp. 95-101.
[30]. A. Kumar, P. K. Vemula, P. M. Ajayan, G. John, Silver-nanoparticle-embedded antimicrobial paints based on vegetable oil, *Nature Materials*, Vol. 7, 2008, pp. 236-241.
[31]. Y. Chen, J. Aveyard, R. Wilson, Gold and silver nanoparticles functionalized with known numbers of oligonucleotides per particle for DNA detection, *Chem Commun*, 2004, pp. 2804-2805.
[32]. H. Li, Z. Cui, C. Han, Glutathione-stabilized silver nanoparticles as colorimetric sensor for Ni2+ ion, *Sensors Actuators, B*, Vol. B143, 2009, pp. 87-92.
[33]. J. Tian, J. Xu, F. Zhu, T. Lu, C. Su, G. Ouyang, Application of nanomaterials in sample preparation, *J Chromatogr A*, Vol. 1300, 2013, pp. 2-16.
[34]. Y. Li, H. J. Schluesener, S. Xu, Gold nanoparticle-based biosensors, *Gold Bull*, Vol. 43, 2010, pp. 29-41.
[35]. D. T. Nguyen, D.-J. Kim, K.-S. Kim. Controlled synthesis and biomolecular probe application of gold nanoparticles, *Micron*., Vol. 4, 2011, pp. 207-227.
[36]. L. Wang, H. Xing, S. Zhang, Q. Ren, L. Pan, K. Zhang, W. Bu, X. Zheng, L. Zhou, W. Peng, Y. Hua, J. Shi, A Gd-doped Mg-Al-LDH/Au nanocomposite for CT/MR bimodal imaging and simultaneous drug delivery, *Biomaterials*, Vol. 34, 2013, pp. 3390-3401.
[37]. P. Baptista, E. Pereira, P. Eaton, G. Doria, A. Miranda, I. Gomes, P. Quaresma, R. Franco, Gold nanoparticles for the development of clinical diagnosis methods, *Anal Bioanal Chem*, Vol. 391, 2008, pp. 943-950.
[38]. E. Spain, R. Kojima, R. B. Kaner, G. G. Wallace, J. O'Grady, K. Lacey, T. Barry, T. E. Keyes, R. J. Forster, High sensitivity DNA detection using gold nanoparticle functionalised polyaniline nanofibers, *Biosensors and Bioelectronics*, Vol. 26, 2011, pp. 2613-2618.
[39]. E. J. Nam, E. J. Kim, A. W. Wark, S. Rho, H. Kim, H. J. Lee, Highly sensitive electrochemical detection of proteins using aptamer-coated gold nanoparticles and surface enzyme reactions, *Analyst*, Vol.137, 2012, pp. 2011-2016.
[40]. B. Nikoobakht, M. A. El-Sayed, Evidence for Bilayer Assembly of Cationic Surfactants on the Surface of Gold Nanorods, *Langmuir*, Vol. 17, 2001, pp. 6368-6374.
[41]. A. Gole, C. J. Murphy. Seed-Mediated Synthesis of Gold Nanorods: Role of the Size and Nature of the Seed, *Chem Mater*. Vol. 16, 2004, pp. 3633-3640.
[42]. J. Belloni. Metal nanocolloids, *Curr Opin Colloid Interface Sci*. Vol. 1, 1996, pp. 184-196.

[43]. J. D. Grunwaldt, C. Kiener, C. Wogerbauer, A. Baiker, Preparation of supported gold catalysts for low-temperature CO oxidation via "size-controlled" gold colloids, *J Catal*., Vol. 181, 1999, pp. 223-232.
[44]. S. Schneider, P. Halbig, H. Grau, U. Nickel. Reproducible preparation of silver sols with uniform particle size for application in surface-enhanced Raman spectroscopy, *Photochem Photobiol*, Vol. 60, 1994, pp. 605-610.
[45]. T. K. Sau, A. Pal, N. R. Jana, Z. L. Wang, T. Pal. Size controlled synthesis of gold nanoparticles using photochemically prepared seed particles, *J Nanopart Res*, Vol. 3, 2001, pp. 257-261.
[46]. M. A. Watzky, R. G. Finke. Nanocluster Size-Control and "Magic Number" Investigations, Experimental Tests of the "Living-Metal Polymer" Concept and of Mechanism-Based Size-Control Predictions Leading to the Syntheses of Iridium(0) Nanoclusters Centering about Four Sequential Magic Numbers, *Chem Mater*, Vol. 9, 1997, pp. 3083-3095.
[47]. K. R. Brown, D. G. Walter, M. J. Natan, Seeding of colloidal Au nanoparticle solutions. 2. Improved control of particle size and shape, *Chem Mater*, Vol. 12, 2000, pp. 306-313.
[48]. M. Grzelczak, J. Perez-Juste, P. Mulvaney, L. M. Liz-Marzan, Shape control in gold nanoparticle synthesis, *Chem Soc Rev*, Vol. 37, 2008, pp.1783-1791.
[49]. J. Niu, T. Zhu, Z. Liu, One-step seed-mediated growth of 30-150 nm quasispherical gold nanoparticles with 2-mercaptosuccinic acid as a new reducing agent, *Nanotechnology*, Vol. 18, 2007, pp. 325601-325607.
[50]. N. R. Jana, L. Gearheart, C. J. Murphy, Wet chemical synthesis of high aspect ratio cylindrical gold nanorods, *J Phys Chem B*, Vol. 105, 2001, pp. 4065-4067.
[51]. N. R. Jana, L. Gearheart, C. J. Murphy, Evidence for seed-mediated nucleation in the chemical reduction of gold salts to gold nanoparticles, *Chemical Materials*, Vol. 13, 2001, pp. 2313-2322.
[52]. I. Capek. Preparation of metal nanoparticles in water-in-oil (w/o) microemulsions, *Advances in Colloid and Interface Science*, Vol. 110, 2004, pp. 49-74.
[53]. Y. Zheng, Y. Wang, X. Yang, Aptamer-based colorimetric biosensing of dopamine using unmodified gold nanoparticles, *Sensors Actuators, B*, Vol. B156, 2011, pp. 95-99.
[54]. Y. Q. Dang, H.-W. Li, B. Wang, L. Li, Y. Wu, Selective Detection of Trace Cr3+ in Aqueous Solution by Using 5,5'-Dithiobis (2-Nitrobenzoic acid)-Modified Gold Nanoparticles, *ACS Applied Materials Interfaces*, Vol. 1, 2009, pp. 1533-1538.
[55]. L. Liu, S. Li, L. Liu, D. Deng, N. Xia, Simple, sensitive and selective detection of dopamine using dithiobis(succinimidylpropionate)-modified gold nanoparticles as colorimetric probes, *Analyst*, Vol. 137, 2012, pp. 3794-3799.
[56]. J. J. Feng, H. Guo, Y. F. Li, Y. H. Wang, W. Y. Chen, A. J. Wang, Single Molecular Functionalized Gold Nanoparticles for Hydrogen-Bonding Recognition and Colorimetric Detection of Dopamine with High Sensitivity and Selectivity, *ACS Appl Mater Interfaces*, Vol. 5:2013, pp. 1226-1231.
[57]. J. J. Feng, H. Guo, Y. F. Li, Y. H. Wang, W. Y. Chen, A. J. Wang, Single Molecular Functionalized Gold Nanoparticles for Hydrogen-Bonding Recognition and Colorimetric Detection of Dopamine with High Sensitivity and Selectivity, *ACS Applied Materials & Interfaces*, Vol. 5, 2013, pp. 1226-1231.
[58]. H. C. Lee, T. H. Chen, W. L. Tseng, C. H. Lin, Novel core etching technique of gold nanoparticles for colorimetric dopamine detection, *Analyst*, Vol. 137, 2012, pp. 5352-5357.
[59]. H. C. Lee, T. H. Chen, W. L. Tseng, C. H. Lin, Novel core etching technique of gold nanoparticles for colorimetric dopamine detection, *Analyst*, Vol. 137, 2012, pp. 5352-5357.
[60]. Y. Zheng, Y. Wang, X. Yang, Aptamer-based colorimetric biosensing of dopamine using unmodified gold nanoparticles, *Sensors and Actuators B: Chemical*, Vol. 156, 2011, pp. 95-99.

[61]. Z. Wang, Y. Bai, W. Wei, N. Xia, Y. Du, Magnetic Fe_3O_4-based sandwich-type biosensor using modified gold nanoparticles as colorimetric probes for the detection of dopamine, *Materials*, Vol. 6, 2013, pp. 5690-5699.

[62]. J. M. Liu, X. X. Wang, M. L. Cui, L. P. Lin, S. L. Jiang, L. Jiao, L. H. Zhang, A promising non-aggregation colorimetric sensor of AuNRs-Ag^+ for determination of dopamine, *Sensors and Actuators B*, 176, 2013, pp. 97-102.

[63]. Y. Zhang, B. Li, X, Chen, Simple and sensitive detection of dopamine in the presence of high concentration of ascorbic acid using gold nanoparticles as colorimetric probe, *Microchimica Acta*, Vol. 168, 2010, pp. 107-113.

Chapter 8
Bio Implant ECG Sensor with Continuous Arrhythmia Monitoring and Auto Diagnosis

Jong-Ha Lee

Abstract

Prolonged monitoring is more likely to result in an accurate diagnosis of atrial fibrillation patients than intermittent or short-term monitoring. In this study, we present an implantable ECG sensor to monitor atrial fibrillation patients in real time. The developed implantable sensor is composed of a micro controller unit, analog to digital converter, signal transmitter, antenna, and two electrodes. The sensor detects ECG signals from the two electrodes and transmits these signals to the external receiver that is carried by the patient. The sensor continuously transmits signals, so its battery consumption rate is extremely high. To overcome this problem, we consider using a wireless power transmission module in the sensor module. This module helps the sensor charge power wirelessly without holding the battery in the body. The size of the integrated sensor is approximately 0.12 inch × 1.18 inch × 0.19 inch. This sensor size is appropriate enough for cardiologists to insert the sensor into patients without the need for a major surgery. The data sampling rate was 300 samples/sec, and the frequency was 430 Hz for signal and power transmission. To verify the validation of the developed sensor, the small animal experiments were conducted.

8.1. Introduction

There are a numerous medical problems whose treatment requires 24/7 monitoring of vital signs of diverse body organs. Most of the time a patient is hospitalized and is kept under observation using wired equipment to measure viral signs of the patient. In remote cases patient is kept at home along with the expensive monitoring equipment and medical staff increasing the medical expenditure and reducing the human resource at hospital. A lot of work has been carried out in the research and development of wearable and implanted biomedical devices over recent years. Progress in this field has provided with benefits

Jong-Ha Lee
Department of Biomedical Engineering, Keimyung University School of Medicine, Daegu, South Korea

including lower cost, patient's free movement and, uninterrupted diagnostic data for the medical staff. Wireless biomedical devices provide enhanced mobility and efficiency with minimum disruption of monitored data [1]. Furthermore a network based on biomedical sensors can create an effective solution distributing patient's information on multiple platforms.

In this chapter, we developed an implantable ECG sensor with wireless communication as well as wireless power transmission. With the increased pace of living in contemporary society and the related reliance on tobacco, alcohol, and caffeine, the number of patients with heart conditions such as arrhythmia, is increasing. Arrhythmia, (also known as cardiac dysrhythmia), is caused by an abnormal ejection fraction, and presents an irregular heartbeat that is either faster (tachycardia) or slower (bradycardia) than usual. It can occur unexpectedly anytime and anywhere, and can lead to shortness of breath, dizziness, and fainting, and in serious cases can cause sudden cardiac arrest due to non-contraction of the ventricles, resulting in a life-threatening myocardial infarction (heart attack). An electrocardiogram (ECG) (or electrocardiograph) is used to detect cardiac abnormalities and thus predict arrhythmia. It produces a graph on which changes in electrical potential associated with the pattern of the heartbeat are recorded. Measurements using an ECG can be performed with either patch-type or insertion-type ECG sensors. A standard patch-type 12-lead ECG is the most widely used ECG sensor, where electrodes are attached to the four limbs and the anterior chest near the heart to measure and record the ECG using standard limb leads, unipolar limb leads, and chest leads. However, in the case of arrhythmia, a short-term ECG measurement is of little help because of its short duration of symptoms. It is therefore necessary to use an ECG device which is carried by the patient, and which has electrodes attached to the body's surface; the heart rhythm is then recorded using either a Holter monitor or an implantable loop recorder (ILR) surgically inserted under the skin. Unfortunately, the Holter monitor is inconvenient as it disrupts daily activities because it needs to be worn constantly, and although the insertion-type ILR is more comfortable because it need not be carried once it is implanted into the body, the disadvantages of the ILR include the need for surgical intervention for implantation, which raises safety and confidence issues [2]. In addition, at the end of the battery's life, a similar surgical procedure is necessary to either replace or remove it, which again raises safety and cost issues.

However, the drawbacks of ILR can be overcome by enabling the use of a quasi-permanent battery that is recharged via wireless power transmission, but although this is technically possible an investigation of its effects on the human body needs to be conducted before it can be used in this manner. Therefore, in this study we developed an insertion-type wireless ECG sensor and tested its performance on a human body phantom using a thermal imaging camera, and on an animal model using the implanted sensor. As a result, it was possible to identify any potential problems that could occur in the use of a wireless ECG sensor [3].

In this study, we developed a micro-sized implantable electrocardiogram (ECG) sensor that shows a high diagnostic yield. The size of the integrated sensor is approximately 0.12 inch × 1.18 inch × 0.19 inch. This sensor size is appropriate enough for cardiologists to insert the sensor into patients without the need for a major surgery. The advantages of

the use of this sensor include continuous monitoring and the capability to capture all asymptomatic and symptomatic episodes as long as the unit is worn continuously. In addition, the small size of the sensor entails minimally invasive surgery to implant it onto patients. The proposed sensor was validated with shielding experiments to mimic the action of the implant and the shielding of sensor effects. The cylindrical case for the experiment was made of three different materials: quartz, titanium, and acryl. The sensor, which transmits signals to the external receiver, was placed into the cylindrical case. To mimic the generation of ECG, the ECG simulator was attached to the sensor to generate artificial ECG signals. The experiments showed that the sensor transmits ECG signals correctly with the three different types of case materials. The quartz material shows the best results among the three materials because it sends data without any distortion.

8.2. System Design Concepts

8.2.1. Cardiac ECG Measurement and Electrodes

Monitoring heart activity through ECG signals is carried out using at least three electrodes placed on specific points on the skin to sense electrical signals generated by heart constituents. Holter monitor is one such diagnostic device which is commonly employed for active monitoring of heart activity after major heart procedures [4]. Holter monitors are proven technologically capable but these monitors are large in size and are connected to electrodes using wires limiting free movement of the patient and requires continuous placement of wired electrodes for long term monitoring. Therefore, Holter monitor is incapable in providing smooth and seamless unobtrusive continuous monitoring. Over the past few years, Holter devices have been evolved into complete wire-free miniaturized modules but they still require further improvements to ensure a totally unobtrusive monitoring architecture.

In this study, we used the Ag/AgCl electrode. An ECG traces the electrical potential differences between electrodes placed on the body's surface. However, the action potential that gives rise to the contraction and relaxation of the cardiac muscle is about 1 mV, and thus extremely difficult to measure. It is therefore necessary to amplify electrocardiographic data to make it easily perceivable to the human eye, and an operational amplifier (op-amp) is used for this purpose [4]. An op-amp amplifies an input electrical potential to the level desired by the user and produces an output potential augmented to the intended level. In this study, we fabricated an instrumentation amplifier using op-amps, and configured it to amplify the micro-fine ECG by 100-fold using a band-pass filter (BPS). The current consumption of the proposed ECG sensor is about 11 mA, and its noise generation is inversely proportional to the length of the wireless communication antenna inside the sensor.

8.2.2. Telemetry Methods

Current systems for health monitoring use various methods between sensors and the data display module. Data is normally shared between these two units using wires which

increase the redundancy of the system limiting movement of the patient. Wire based equipment provide a robust mean for communication in health monitoring systems and is low-cost but bounds normal movement of a patient in everyday routine.

Another problem which arises in wired systems is improper connection of wires due to multiple reasons which can seriously interrupt the system posing serious consequences for the patient [5]. Continuous improvisation and research is being carried out to develop smart health monitoring systems. Multiple alternative communication techniques have emerged with wireless communication being the most suitable communication method curtailing the need for wired connections between sensors and the equipment. Wireless technologies enable intra-body communication to complement monitoring systems for 24/7 health monitoring without a need for admitting patient and attaching wires. Wireless communication allows real-time monitoring of vital signs on an unwired display device in proximity of the patient and it can also be sent to a doctor via internet so that he can observe patients health without a need to call him at hospital. Wireless connectivity will also help patient to track their own health indicators using smart-phones or PDAs connected to implants or wearable sensors in real-time. This will result in better health management and prompt alerts in case of health related emergencies.

In this study, we used the medical Implant Communication Service (MICS). MICS operates in the frequency range of 402–405 MHz and is normally used in communication between body-worn monitoring systems and implants. Implantable antennas in this frequency range have been developed to transmit data from pacemakers and cardiac sensors; however, due to regulatory restrictions in hospitals, its full utilization in WBANs is limited.

8.2.3. Wireless Power for Biomedical Implants

Supplying adequate power to bio-medical implants is currently the main challenge limiting function and performance that can achieved for a near ideal bio-implant. Power consumption affects many characteristics of an implant like size, processing power, transmission range and life span etc.

Batteries have the ability to power implant for a long span by exploiting design techniques requiring extremely low power consumption. Average power consumption for a battery used in a pacemaker is about 8 uW and makes up 90 % of the total size of implant. Moreover, it requires periodic replacement after every few years thorough a costly invasive surgery. Power hungry implants like mechanical pump based cardiac and orthopedic implants require a lot of power to function and battery doesn't remain an effective power source option in these implants. Prospects and application for bio-medical implant technology is currently severely limited by the unavailability of an adequate power source. This problem can be successfully addressed using a wireless power transfer techniques which are capable of delivering uninterrupted power to ensure continuous monitoring and communication by implants.

In this chapter, we used the near-field wireless power transmission. It is assumed for inductive coupling that radiated fields are not rapidly changing. For low frequencies, displacement current does not affect the generated fields and is thus ignored. It is generally called the quasi-static approximation. Using this approximation, magnetic field was found to be concentrated in the vicinity of the source. These problem can be analytically solved using the Biot–Savart law or by finding a solution using the diffusion equation. Many techniques for coupled power links have been reported in literature. In authors used coupled self-resonant coils to power a 60 watt bulb over a distance of 2 meters with an efficiency of 40 %. This was accomplished by non-radiating magnetic induction using resonant loops. Two identical helical coils were used as coupling elements and to generate frequencies in the MHz range a standard Colpitts oscillator was used with a single copper wire loop as inductive element. This copper loop coupled inductive power to the source coil for further transmission. The authors used a light-bulb as load of the power transfer system. Experimental results showed that power transfer using non-radiative magnetic coupling can be achieved over a range of 8~9 times the radius of the coils. Authors also presented a quantitative model for explaining the power transfer with an accuracy of around 5 %.

8.2.4. Circuit Design

An ECG traces the electrical potential differences between electrodes placed on the body's surface. However, the action potential that gives rise to the contraction and relaxation of the cardiac muscle is about 1 mV, and thus extremely difficult to measure. It is therefore necessary to amplify electrocardiographic data to make it easily perceivable to the human eye, and an operational amplifier (op-amp) is used for this purpose. An op-amp amplifies an input electrical potential to the level desired by the user and produces an output potential augmented to the intended level. In this study, we fabricated an instrumentation amplifier using op-amps, and configured it to amplify the micro-fine ECG by 100-fold using a band-pass filter (BPS) as shown in Fig. 8.1. The current consumption of the proposed ECG sensor is about 11 mA, and its noise generation is inversely proportional to the length of the wireless communication antenna inside the sensor.

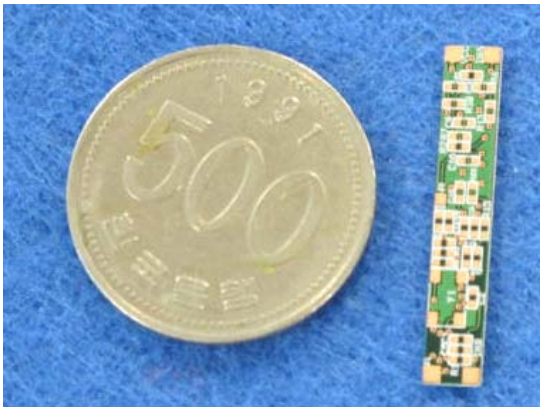


Fig. 8.1. The circuit design of implantable ECG sensor.

8.2.5. Packaging

The packaging materials selected need to be biocompatible to avoid causing inflammation or necrosis of human tissues. Additionally, they need to satisfy the strength standard depending on the insertion location, and electric waves should not be lost when passing through the material. It is also necessary to take packaging design into account, to avoid any risk of damage to tissues caused by sensor insertion and post-insertion movements. In addition, a foreign-body sensation needs to be minimized, by reducing the size of the object. Feed through needs to meet several requirements, and should therefore ensure conductivity between the internal circuit and electrodes, minimize noise, and maintain airtightness. In this study, electrode sealing methods were developed using the elasticity of polymer films.

The requirements of ECG sensor electrodes are as follows: an electrode-to-electrode distance of ≥40 mm, and an electrode width of ≥5 mm. Electrodes were fabricated using titanium, which thus satisfies the requirements for electrodes. Packaging materials are rigid and need to be coated to protect tissues. They also need to be hermetically joined using a proper joining technique to completely block any interaction between the interior of the human body and the sensor environment, and to accomplish this, either adhesives or a laser can be used. In this study, we used adhesives to produce a packaging prototype. PDMS and medical epoxy are suitable adhesives, and PDMS, parylene, polyethylene, glycol, and silicone can be used as coating materials; their biocompatibilities have been verified in numerous studies. Fig. 8.2 shows the packaged ECG sensor.

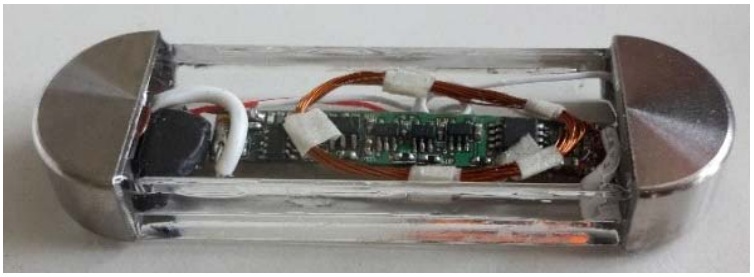


Fig. 8.2. The packaged ECG sensor.

8.3. Experimental Results

8.3.1. Self-sealing Airtightness Testing

Testing for self-sealing airtightness was performed in two steps. In the first step, the packaging was submersed in de-ionized (DI) water for 1 hour with no sensor included, and in the second test the packaging containing the sensor was submersed in DI water for 5 hour. Results of tests 1 and 2 are outlined in Table 8.1 and 8.2, respectively, and show that the materials joined with adhesives are air-tight.

Table 8.1. Self-sealing test 1 results.

No.	Before the experiment	After the experiment
1	8.941 g	8.940 g
2	8.941 g	8.941 g
3	8.941 g	8.940 g
Avg.	8.941 g	8.940 g

Table 8.2. Self-sealing test 2 results.

No.	Before the experiment	After the experiment
1	11.587 g	11.587 g
2	11.587 g	11.587 g
3	11.587 g	11.587 g
Avg.	11.587 g	11.587 g

8.3.2. Thermal Testing

Coil charging was prepared in relation to the wireless power transmission to the ECG sensor. Fig. 8.3 shows the process of sensor activation through wireless power transmission. We prepared a device that emitted an electric current that was the same as that of a real ECG, and connected it to the ECG sensor; thus establishing a wireless network-driven environment for transmitting and receiving ECG data for ECG measurements. We then prepared an infrared temperature camera to measure the temperature changes of the sensor itself, as shown in Fig. 8.3.

Fig. 8.3. The thermal testing experimental setup.

These temperature changes were measured prior to the initiation of power transmission, and measurements were then continued for 1 hour after transmission began, with the aim of determining the average temperature change.

From the result, we notice that the baseline average temperature was 23.7 °C. The temperature then harply rose by about 3.0 °C after about 10 min, and continued to rise to reach 27.2 °C after 1 hour. This temperature is thus far lower than 36.9 °C, which is the average temperature of the interior of the human body. It is considered that the packaged sensor will not undergo any considerable temperature change once inserted into the human body.

8.3.3. Insertion Experiment Using Animal Model

Prior to using the sensor within a human body, it was necessary to test the in vivo safety of the instrument to ensure both the efficient operation of the insertion-type ECG measurement system in measuring physiological functions and its efficacy in receiving external signals. A pig was therefore used as a sensor-implanted animal model because the animal's physiological characteristics are similar to those of humans. The species selected was a Hanford mini pig because the heart size is very similar to that of a human. We purchased a female pig of specific pathogen free (SPF) quality with ho history of pregnancy, 46–60 kg, and 50–57 weeks old, from Optipharm Medipig (Choongbuk, South Korea).

The insertion surgery was performed in Daegu High-Tech Medical Complex as follows. Anesthesia was induced using Zoletil (Tiletamine/Zolazepam) (2.5 mg/kg, IM) and Xylazine (2.3 mg/kg, IM) and maintained with Isoflurane (1–3 %). Lactated ringer's solution (5 ml/kg/h, IV) was administered intraoperatively. After the anesthesia, the left anterior corselet was depilated and disinfected with alcohol and povidone. An incision was made between the left 5th–7th ribs, and separated using blunt dissection to a depth of 4 mm under the skin. The sensor was placed at the site, and the skin was sutured. On completion of the wireless ECG sensor implantation as shown in Fig. 8.4, the pig's ECG data was received by wireless network, as shown in Fig. 8.5.

Fig. 8.4. Insertion experiments using animal model.

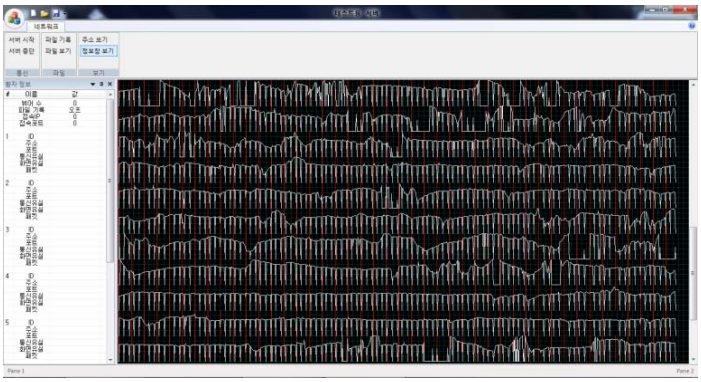


Fig. 8.5. ECG monitoring results using animal model.

8.4. Summary

In this study, we designed a quasi-permanent ECG sensor that negates the need for surgical replacement, by applying a double loop coil-shaped magnetic resonance-type wireless power transmission system. Additionally, we developed an ultra-small antenna (20 mm in width, 10 mm in length) that has a spiral-shape metal pattern, thereby minimizing the sensor size while securing a sufficient electric length. We also developed a human body phantom that has similar electrical properties to that of the human skin within the MICS band, with a 10 % error range of measurement values (specific permittivity = 43.2, conductivity = 0701 S/m), and verified the communication performance of the antenna on the developed phantom. The hermetic joining of packaging using adhesives was verified, and the biocompatibility was also experimentally verified. Finally, we performed sensor insertion surgery on a laboratory pig and succeeded in obtaining ECG data via the wireless network.

Acknowledgement

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2014R1A1A2056420) and sponsored by the global specialized technology development project of the ministry of trade, industry and energy(Registry no.: 20150364).

References

[1]. Hao, Y., Foster, R., Wireless body sensor networks for health-monitoring applications, *Phys. Meas.,* 29, 2008, pp. R27–R56.

[2]. Newman, J. D., Turner, A. P. F., Home blood glucose biosensors: A commercial perspective, *Biosens. Bioelectron.*, 20, 2005, pp. 2435–2453.

[3]. Recommended Practice for Determining the Spatial-Peak Specific Ab-sorption Rate (SAR) in the Human Body Due to Wireless Communications Devices: Measurement Techniques, I*EEE Standard* 1528/D1.2, 2003.

[4]. Pantelopoulos, A., Bourbakis, N., A survey on Wearable sensor-based systems for health monitoring and prognosis, *IEEE Trans. Syst. Man. Cybern.*, 40, 2010, pp. 1–12.
[5]. J. Malmivuo and R. Plonsey, Bioelectromagnetism, *Oxford University Press,* 1995.
[6]. Zimetbaum, P. J., Josephson, M. E., The evolving role of ambulatory arrhythmia monitoring in general clinical practice, *Ann. Intern. Med.,* 130, 1999, pp. 848–856.
[7]. Bergey, G. E., Squires, R. D., Sipple, W. C., Electrocardiogram recording with pasteless electrodes, *IEEE Trans. Biomed. Eng.*, BME-18, 1971, pp. 206–211.

Chapter 9
Fano-Resonance Plasmonic Biosensors

Xiaowei Guo

9.1. Introduction

A biosensor is an analytical device for the analysis of biomaterial samples to gain an understanding of their bio-composition, structure and function, containing a biological recognition element immobilized on a solid surface and a transduction element which converts analyte binding events into a measurable signal. There are various transducers including optical, magnetic, electrochemical, radioactive, piezoelectric, micromechanical, and mass spectrometric. In the past three decade, optical transducers development has been a fascinating and fast-paced area because they are immune to electromagnetic interference, capable of performing remote sensing, and present real time response to biomolecular interactions. Main optical sensing methods used in biosensors include fluorescence spectroscopy, interferometry, and plasmonic spectroscopy. Plasmonic methods are label free sensing methods and do not require labeling of the target molecules with different types of reagents, such as fluorescent dyes. The plasmonic biosensor which works based on surface plasmon resonance (SPR), or localized surface plasmon resonance (LSPR) can monitor a wide range of analyte surface binding interactions such as absorption of small molecules, proteins, antibody and antigen, DNA and RNA hybridization [1].

The most common method to induce SPR on the gold surface is to utilize an optical prism [2], known as the Kretschmann configuration, which leads to many commercial SPR instruments, such as Biocore series, SR series, Autolab SPR, Plasmonic, Spreeta, Nanofilm, Multiskop. In addition to the prism coupling method, metallic nanostructures offer a simple way for SPR excitation. Compared to prism-based SPR sensors, LSPR sensors based on metallic nanostructures, e.g. nanoparticles, benefit from having a small detection volume and normal light incidence [3, 4]. Recently, periodic gold nanohole arrays or nanoslit arrays have been utilized for biosensing applications [5, 6]. Using such arrays, the excitation and detection of SPRs can be realized in a collinear arrangement using a low-cost light-emitting diode and a camera. The small footprint of the plasmonic sensor array allows the biochemical characterization of the sample from a very small

Xiaowei Guo
University of Electronic Science and Technology of China

volume(~μL). They provide a feasible way to achieve chip-based, high-throughput and label-free detection for modern DNA and protein microarrays. Combining with Fano resonance [7], such plasmonic biosensors can obtain ultrahigh intensity or wavelength sensitivity and ultrasmall linewidth, and are sensitive to their identity and conformational state, representing a significant advance in the study of biomolecular interactions.

This chapter introduces the Fano resonance plasmonic biosensors. Firstly the concepts of SPR and LSPR are offered, and then the performance parameters of plasmonic biosensor are described, followed by depiction of Fano resonance principle. Subsequently the state-of-the-art Fano resonance biosensors are presented. Finally, the future scope of research and development of nanoplasmonic biosensors is provided in the conclusion section.

9.2. Basic Concepts of Plasmonic Biosensors

9.2.1. SPR and LSPR

SPPs are coherent electron oscillations that exist at the interface between any two materials where the real part of the dielectric function changes sign across the interface, typically on a metal-dielectric interface [8]. These are essentially light waves that are trapped on the surface because of their interaction with the free electrons of the metal. In this interaction, the free electrons respond collectively by oscillating in resonance with the light wave, as shown in Fig. 9.1. But the electron oscillations decay exponentially into both materials and the decay length is about wavelength order. The excitation of SPPs by light is denoted as a SPR (Fig. 9.1 (a)) for planar surfaces or LSPR (Fig. 9.1 (b)) for metallic nanoparticles or metallic nanostructures, which are of interest to a wide spectrum of scientists, ranging from physicists, chemists and materials scientists to biologists [9, 10].

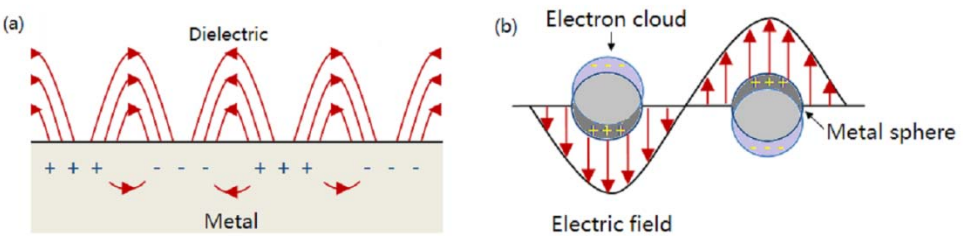


Fig. 9.1. Schematic diagrams illustrating (a) a SPR and (b) a LSPR.

In SPR case, the interaction between the surface charge density and the electromagnetic field results in the momentum of the SP mode, k_{SP}, being greater than that of a free-space photon of the same frequency, k_0. There are two main techniques by which the missing momentum can be provided. One is utilizing prism coupler to enhance the momentum of the incident light. This method is referred to as the attenuated total reflection (ATR) method [11]. The dispersive relation of SPs can be expressed as below

$$kSP= nk0sin\theta \quad (9.1)$$

with n is the refractive index in incidence medium and θ is the incidence angle.

The other is based on a periodic corrugation in the metallic surface, typically metallic grating. For two dimension grating, the momentum matching condition between the in-plane wave-vectors of the incident photons and the surface-plasmon polaritons is fulfilled when the Bragg coupling condition is met [12]

$$k_{SP}= k_x \pm mG_x \pm nG_y \quad (9.2)$$

where k_x is the wave vector along the grating plane and its value is $nk_0\sin\theta$, G_x and G_y are the reciprocal lattice vectors along x and y direction, equaling to $2\pi/\Lambda$ where Λ is the pitch. (m,n) is the grating order for reciprocal lattice vector G_x and G_y.

In LSPR case, the light scatters from a topological defect on the surface, such as nanoparticles or nanostructures [13]. Light interacts with metallic nanostructures much smaller than the incident wavelength, which leads to a plasmon that oscillates locally around the nanoparticle with a frequency. For noble spherical nanoparticles, e.g. Au or Ag, with diameters less than 30 nm, mainly dipole plasmon resonance is involved (Fig. 9.1b); however, for larger particles, quadrupole plasmon resonance from two negatively charged poles and two positively charged poles may be observed. The magnitude, peak wavelength and spectral bandwidth of the plasmon resonance associated with a nanoparticle are dependent on the particle's size, shape, and material composition, as well as its local dielectric environment, which can be obtained analytically by using the Mie scattering formalism. It is noted that SPPs in the SPR case are propagating, dispersive electromagnetic waves. Localized SPPs in the LSPR case on the other hand are non-propagating excitations of the conduction electrons of metallic nanostructures coupled to the electromagnetic field.

9.2.2. Plasmonic Biosensing Principle

For researchers in the field of chemistry and biology, one of the most attractive aspects of SPPs is the way in which these oscillations are very sensitive to the dielectric permittivity of the environment and therefore can be used as a sensing method for detection of external medium.

Fig. 9.2 reveals that the principle of affinity SPR and LSPR biosensing. Generic SPR biosensing schemes involve the immobilization of capture molecules on the metal surface. The immobilized molecules are designed to bind preferentially to the biomolecules of interest (analytes) from the sample. Therefore, although measurements of refractive index are not molecule specific, the SPR biosensor can be made highly selective to a particular biomolecule by the appropriated surface chemistry modification [14]. The binding produces a local increase in the refractive index at the metal surface. The refractive index increase gives rise to an increase in the propagation constant of SPPs which can be accurately measured by different optical means, such as intensity modulation, angular modulation, wavelength modulation, phase modulation, and even polarization modulation.

For the purpose of small footprint size and high signal-to-noise ratio, wavelength modulation is often preferred.

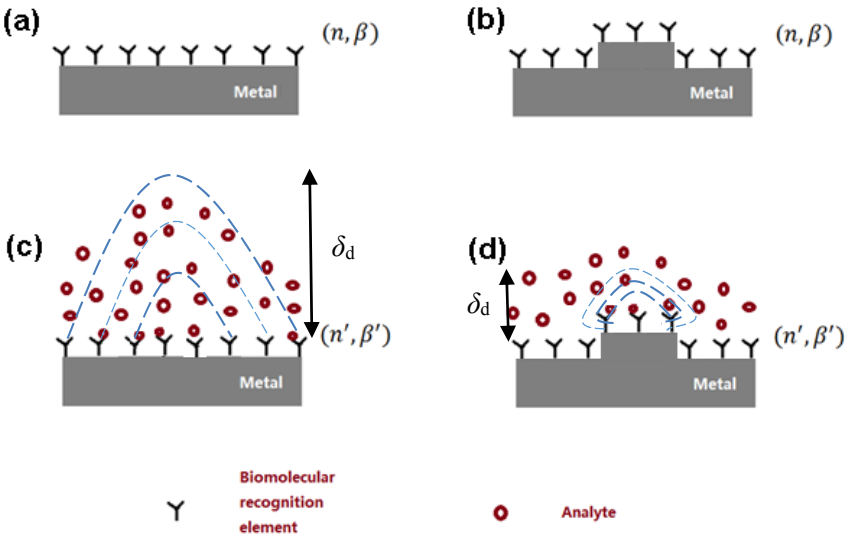


Fig. 9.2. Basic principle of affinity SPR (a, c), and LSPR (b, d) biosensing.

As shown in Fig. 9.2, for SPR on planar surfaces, the decay length δ_d is roughly of the order of the resonance wavelength. For LSPR, δ_d is often significantly smaller (around 20 nm for a 30-nm-diameter spherical gold nanoparticle at the dipolar SPR resonance). The smaller δ_d values in the case of LSPR mean that those types of sensor respond better to changes closer to the surface, as required for biosensing. This provides a unique advantage for biosensors as the size of the biomolecules is generally only about ten nanometres and the background noise can be all but eliminated by localized SPR rather than propagating SPR. As a result, the detection system can be much simpler and the signal-to-noise ratio can be high as a result of using localized SPR. On the other hand, in contrast to propagating SPPs, there is no propagation length to define in the LSPR case, which means short plasmon lifetime. In plasmonic biosensing, the propagation length value provides a lower limit on the size of the sensor element. Although very small propagation lengths negatively affect the performance of the device in sensing applications, periodic spatial arrangement, e.g. grating, can be designed to contain the lower-loss, longer-propagation SPRs, allowing dense integration of sensitive biosensors.

9.2.3. Performance Evaluation of Plasmonic Biosensors

Sensitivity and figure of merit (FOM) are two parameters that are used to analyze the performance of a plasmonic sensor. For the best performance both the parameters should be as high as possible. Sensitivity of a plasmonic sensor utilizing wavelength interrogation method depends on the amount of shift of the resonance wavelength with a change in the

refractive index of the sensing layer. For a given refractive index change if the shift in resonance wavelength increases this means an increase in the sensitivity of the sensor. Thus the sensitivity of a plasmonic sensor utilizing wavelength interrogation method is defined in nm/RIU units as [15]:

$$S_n=\Delta\lambda_{res}/\Delta n \tag{9.3}$$

However, the sensitivity increase is accompanied with a substantial broadening of the resonance peaks when the resonance wavelength increases, thus reducing the resolution and ability to discriminate small wavelength shifts. To overcome this controversy, FOM was introduced. The widely adapted metric for the intrinsic resolving power of the LSPR sensors is FOM in wavelengths units, defined as the ratio of wavelength sensitivity to a bandwidth Δw of resonance (a full width at half-maximum, FWHM) [16]:

$$\text{FOM}= S_n/\Delta w \tag{9.4}$$

9.3. The Fano-resonance

In physics, Fano resonance can be considered as a mechanism which couples narrow or discrete bound states of one channel with broad or continuum states of another [17]. Due to the superposition principle of quantum mechanics, whenever two states are coupled by different paths, interference may occur. Interference between a bound state and a continuum state produces an asymmetric line-shape, as shown in Fig. 9.3. It is named after Italian-American physicist Ugo Fano [18], who gave a theoretical explanation for the scattering line-shape of inelastic scattering of electrons from helium. Because it is a general wave phenomenon, examples can be found across many areas of physics and engineering. In contrast to a Lorentzian resonance, the Fano resonance exhibits a distinctly asymmetric shape with the following functional form [17, 19]:

$$I \propto \frac{(F\gamma+\omega-\omega_0)^2}{(\omega-\omega_0)^2+\gamma^2} \tag{9.5}$$

Where ω_0 and γ are standard parameters that denote the position and width of the resonance, respectively. F is the so-called Fano parameter, which describes the degree of asymmetry.

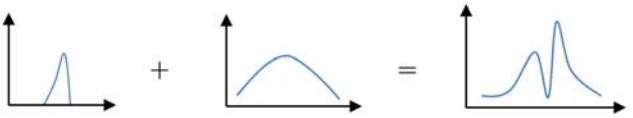
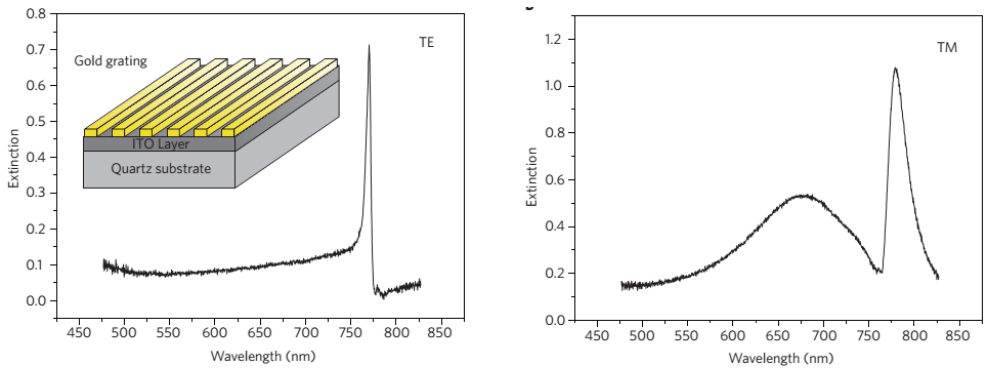


Fig. 9.3. Schematic illustration of Fano resonance.

Fano resonances have very recently become a topic of particularly high interest in nanophotonics for their potential to surpass the efficiency of conventional plasmon resonances [20, 21]. Together with the tremendous progresses that have recently been

made on low-cost fabrication techniques, characterization techniques and devices integration, the road is now open towards practical implementations of nanoscale devices for many applications. In nanophtonics, Fano resonances originate from the interference between two electromagnetic eigenmodes, referred to as 'superradiant' and 'subradiant'. [22]. The superradiant mode is a bright mode, whereas the subraiant mode is a dark mode. To obtain Fano resonance, two pathways have to be considered: the direct excitation of the bright mode and the excitation and reemission of the dark mode through its coupling to the bright mode. In periodic metallic structures, Fano resonances are abundant. For example, an array of gold nanowires placed on a single-mode slab waveguide exhibits a Fano resonance in extinction in transverse electric (TE) polarization owing to coupling between the array and the waveguide (see Fig. 9.4). The narrow waveguide mode interferes with a continuum of vacuum states. In transverse magnetic (TM) polarization, the incident light can excite a particle plasmon across the gold nanowires, which can interfere destructively with the narrow waveguide resonance. The light in the waveguide interferes destructively with the re-emitted light of the particle plasmon.



(a) (b)

Fig. 9.4. Fano resonances in a metallic photonic crystal.
Figure reproduced with permission from ref. [17].

9.4. Design Method of Fano Resonance Plasmonic Biosensor

Investigation of the physical interaction between the metal nanostructures and the electromagnetic field can be conducted by solving the Maxell wave equations, including static and dynamic analytic methods. The static analytic method includes Mie scattering theory, Discrete Dipole Approximation (DDA), Multiple Multipole Programs (MMP), Rigorous coupled wave analysis (RCWA), Method of moments (MoM), etc.; The dynamic method includes Finite-Difference Time-Domain (FDTD) method, Finite element method (FEM), and Finite integration technique (FIT), and so on.

For metallic nanoparticles, Mie theory for light scattering is a very effective method [23]. The extinction, scattering and absorption cross-sections are given by expressions of the form $\sigma = \pi a^2 Q$, where a is the radius of the particle and Q the corresponding efficiency. The scattering efficiency is defined by:

$$Q_{sca} = \frac{2}{q^2}\sum_{\lambda=1}^{\infty}(2\lambda+1)\left\{|a_\lambda|^2 + |b_\lambda|^2\right\} \tag{9.6}$$

Where a_λ and b_λ are electric and magnetic amplitudes, respectively, which depend on the size parameter $q=\omega a/c$ (c is the speed of light in a vacuum and ω the incident-light frequency and also on the dielectric permittivity ε and the magnetic permeability μ. The scattering amplitudes are defined by the Mie formulas:

$$a_\lambda = \frac{\Re_\lambda^a}{\Re_\lambda^a + i\Im_\lambda^a}, b_\lambda = \frac{\Re_\lambda^b}{\Re_\lambda^b + i\Im_\lambda^b} \tag{9.7}$$

The functions $\Re$ and $\Im$ are expressed as combinations of the spherical Bessel and Neumann functions.

For metallic nanostructures, FDTD is a major photonic analysis tool [24] in which the space is divided into a small mesh, the so-called "Yee mesh". The electric and magnetic fields in each mesh are solved according to Maxwell's equation step by step. The dynamic behaviour of the electric field can be calculated for an arbitrary material environment by this method. However, the simulation time and memory space required for solving such complex structures are considerable.

9.5. Fano-resonance Plasmonic Biosensors

As Fano resonances arise from the interference between two or more states, they possess an inherent sensitivity to changes in geometry or local environment. Its extremely sharp and asymmetric line shape promises high FOM and small perturbations can induce dramatic resonance or lineshape shifts. The most straightforward application of Fano resonant media is in the development of chemical or biological sensors. In combination with the appropriate biomarkers, Fano resonance has been used for development of versatile plasmonic biosensors, enabling a new generation of label-free chemical and bioanalysis probes adaptable to high-throughput applications [25]. Nanoparticle array [26], nanocavities [27], nanorings [28], dolmen nanostructures [29, 30], grating [31], slit-grating nanostructures [32] etc. has been used for Fano resonance platforms to conduct biosensing. For commercial purpose, the latter three platforms are compatible with next generation biosensing technology. Therefore, these platforms are described below in detail.

9.5.1. Dolmen-type Biosensor

A dolmen-type plasmonic structure is made of three metallic beams arranged as in Fig. 9.5 (a). Near-field interaction between the two beams results in parallel (electric dipole) and antiparallel (quadrupole dipole) current excitations corresponding to super- and subradiant modes, with resonance frequencies of ω_1 and ω_2, respectively. If the two parallel beams are identical, the subradiant mode is completely dark and decoupled from the normally incident light, as shown in Fig. 9.5 (b). The dipolar mode of the perpendicular beam has a larger spectral width due to radiative losses and ensures the coupling of the

non-radiative mode to the radiative continuum (see Fig. 9.5 (c)). In this system, the bright and the dark modes are supported by two separate parts of the complete system. The field overlap between the two modes results in asymmetric line shape in total spectra, as indicated by dashed circle in Fig. 9.5 (d).

The asymmetric line width and position can be tuned by changing structure parameters, e.g. the gap size *g*. Following this idea, Wu et al. [33] designed an asymmetric dolmen-type plasmonic biosensor as illustrated in Fig. 9.6. The small horizontal beam makes the metamolecule asymmetric by breaking all spatial inversion/reflection symmetries in the plane of the structure. Such symmetry breaking not only directly couples the subradiant and superradiant modes, but also couples the subradiant mode to the perpendicular polarization. The resonant frequencies of the sub- and superradiant modes are primarily determined by the length L_1 of the two parallel antennas, whereas the length L_2 and position W of the short horizontal segment determine the degree of symmetry breaking.

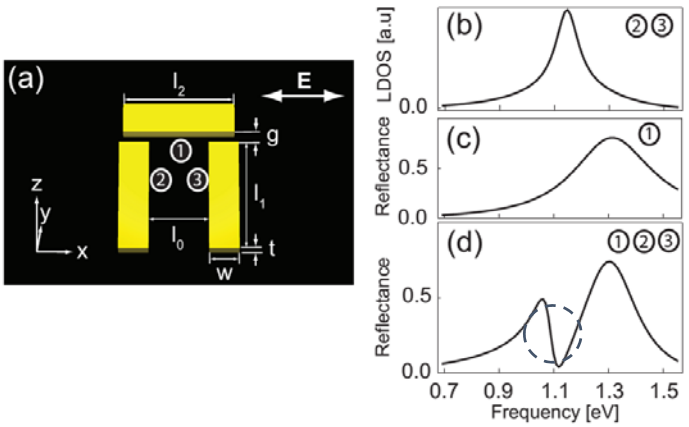
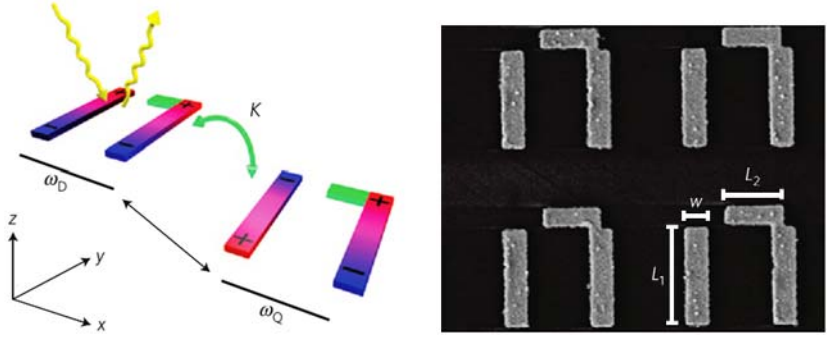


Fig. 9.5. A Fano resonance in a dolmen-type plasmonic structure [34].



(a) (b)

Fig. 9.6. An asymmetric dolmen-type biosensor. Figure reproduced with permission from [33].

Combined with infrared spectroscopy, this plasmonic biosensor provides potential to conduct fingerprinting and characterization of protein A/G monolayers [33]. Fig. 9.7 (a) shows the reflection spectra before and after functionalization with a 3-nm-thick protein A/G monolayer. The plasmonic biosensor can accurately measure the thickness of a protein monolayer that provides indirect information about their orientation on the surface and the availability of antigen binding as the presence of protein monolayer changes the dielectric environment of the infrared spectroscopy and results in frequency shifts ($\Delta\omega_Q$), thus leading to the reflection change $\Delta R(\omega_Q)$. In Fig. 9.7 (b), binding of the IgG antibody monolayer to three A/G protein-immobilized plasmonic structures is illustrated.

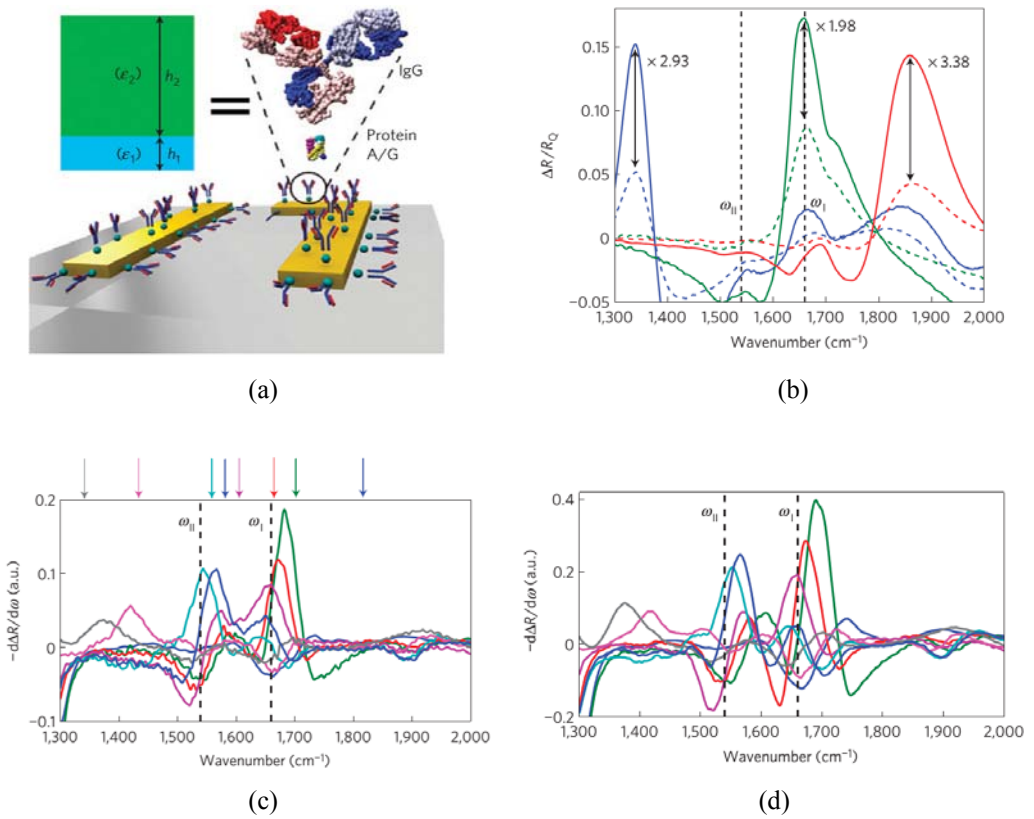


Fig. 9.7. Dolmen-type nanostructure biosensing. (a) Schematic representations of proteins' mono- and bilayers binding to the metal surface; (b) Spectra before (dashed lines) and after (solid lines) binding of IgG antibodies to three different plasmonic substrates immobilized by the protein A/G; (c, d) Spectra for the protein A/G monolayer and the protein A/G+IgG antibody bilayer. Figures reproduced with permission from ref. [33].

By measuring the differential spectra $\Delta R(\omega_Q)(\omega)/R(\omega_Q)(\omega = \omega_Q)$ of the resulting protein bilayer, the thickness h_2 of the IgG monolayer can be decided by using h_1 as a molecular yardstick. It is noted that the strongest interaction between the proteins and the plasmonic structure corresponding to the largest differential reflectivity occurs when the wavenumber of the plasmonic structure coincide with that of one of the vibrational modes

of the molecules, as indicated by green curve. By employing the first frequency derivative of the reflectivity spectrum $G(\omega_Q)(\omega)=-(d(\Delta R(\omega_Q)/d\omega)/(dR(\omega_Q)/d\omega)(\omega=\omega_Q)$, the non-specific contribution to $\Delta R(\omega_Q)(\omega)/R(\omega_Q)$ can be effectively eliminated. The biosensor also provides potential to determine the surface orientation of proteins. Furthermore, the Fano resonance can recognize chemical bonds because different chemical bonds have different spatial arrangements resulting in the change of the reflective spectra, which can be seen by comparing Fig. 9.7 (c) and (d). It is clear that the effective dipolar strength of amide I (C=O stretch) is stronger than that of amide II (C–N and C–C stretch, and N–H bend). This has been attributed to stronger alignment of amide I with the axes of the protein's alpha-helices, which are normal to the metal surface for oriented proteins.

9.5.2. Nanohole Array Sensor

It was demonstrated that for a periodic thin-film metallic grating, formed from a two-dimensional array of holes(see Fig. 9.8), the transmitted fraction of the incident light can exceed the open fraction of the array for certain wavelengths [35]. The enhancement in the transmitted zero-order beam is reported to be several orders of magnitude larger than that from pure metallic slab without holes. This phenomenon has been called extraordinary transmission (EOT) through periodic arrays of subwavelengths holes in metallic films. The common understanding of the extraordinary transmission is due to a resonant excitation of surface plasmon-polaritons by incoming radiation [36]. In addition to the resonant enhancement of the transmission the resonant suppression was observed as well. According to experimental observations, each extraordinary transmission is accompanied by resonant suppression of transmission resulting in asymmetric lineshapes, which can be perfectly fitted by the Fano formula [37]. This resonant behaviour can be understood in terms of the coupling of the incoming waves with the surface-bound states of periodic arrays. Thus, the extraordinary resonant scattering of light by modulated metal film can be described in terms of the Fano resonance, revealing the interference nature of the phenomenon.

Yanik et al. experimentally demonstrate that such EOT can be exploited for ultrasensitive label-free biosensing [31]. In Fig. 9.8 (a), at normal incidence multiple extraordinary light transmission resonances are observed with the creation of SPPs corresponding to the different grating orders, e.g. (0, ±1), (+1,0) and (-1,0). The two former are superradiant and the latter is subradiant. By altering the incidence angles, the subradiant modes can be coupled to the incidence continuum state, which leads to new detection mechanism. It is possible to achieve much high FOM than conventional prism-based SPR sensors as (+1, 0) mode has narrower line shape. It was reported the obtained figure of merit (FOM=162 in wavelength units) for the (+1, 0) subradiant dark mode is higher than the theoretically estimated upper limits (FOM=108) of the standard prism-based SPR sensors. It is noted that in this biosensor all the superradiant and subradiant modes achieved an ultrahigh sensitivity over 690 nm/RIU (Fig. 9.8 (c)). It is highlighted that the high-quality nanohole sensors enable seeing single biomolecular monolayers with naked eye. In Fig. 9.8 (d), it can be seen that after the capturing of a monolayer of the antibody IgG by protein A/G, transmitted light intensities in a nanohole array of 90 μm × 90 μm in dimensions become invisible. Because this detection scheme is not requiring dark-

environment measurements, utilization of broad-band light sources allows direct detection with human eye without any safety concerns, which is especially suitable for onsite detection. In addition, use of (-1, 0) superradiant mode provides potential to obtain high signal-to-noise ratio. In Fig. 9.8 (e), the capturing of a single monolayer of mouse IgG antibody causes red shifting of the plasmonic resonance. This resonance shift is large enough to cause spectral overlapping of the transmission minima of the nanohole array with the transmission window of the notch filter, as shown in Fig. 9.8 (f), enabling naked eye detection(Fig. 9.8 (d)).

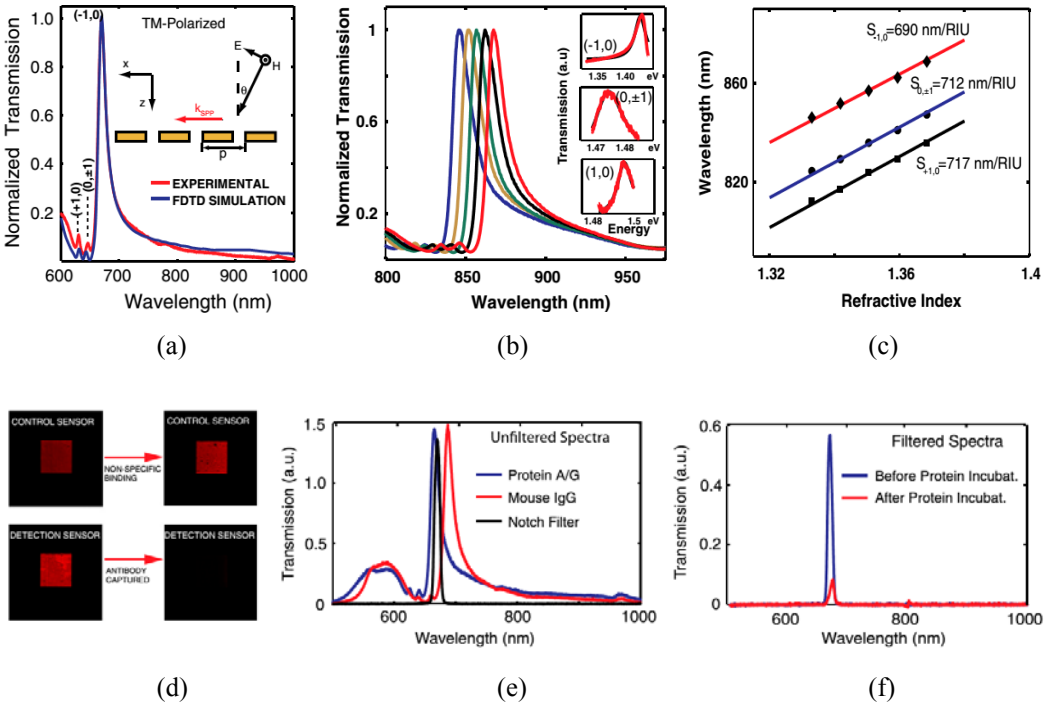


Fig. 9.8. Nanohole array plasmonic biosensing. (a) Fano resonance in metallic nanohole array. (b) Spectral measurements are obtained with the varying concentrations of NaCl. (c) Sensitivity of the devices using a linear fitting scheme. (d) Images of the transmitted light obtained from detection and control sensors. Capturing of the antibody causes a dramatic reduction of the transmitted light intensities through the detection sensors. (e) Transmission spectra before (blue curve) and after (red curve) the capturing of the antibody. (f) Transmitted light intensities in the presence of the notch filter is given before (blue curve) and after (red curve) the capturing of the antibody. Figure reproduced with permission from ref. [31].

9.5.3. Slit-groove Nanostructure Sensor

The sensor structure is similar to nanohole array sensor, but with a silt in the grating. Fig. 9.9 (a) describes a slit-embedded metallic grating sensor reported by Lee et al. [32]. The broad resonance is the LSPR in the slit and grooves. Because the LSPR is a cavity mode, the resonance wavelength increases with the increase of the depth as shown in

Fig. 9.9 (b). The discrete state is associated with Bloch wave surface plasmon ploritons (BW-SPP) on the periodic grooves that occurs when the Bragg condition is satisfied. At around the wavelength of 630 nm, there are anti-symmetrical resonances. These resonances are Fano-type resonances formed by the interference between spectrally overlapping broad resonance and a narrow discrete resonance. For the Fano-type resonance, the resonant shape is affected by the groove depth. For shallow grooves, the interference of the light in single nanoslit and the BW-SPP on the periodic grooves is weak. It forms an anti-symmetrical resonant spectrum. When the groove depth increases, the peak wavelength of the Fano resonance and LSPR modes are red-shifted. From Fig. 9.9 (b), it is obvious that choice of 20 nm-depth grooves is good for improving FOM.

Using a 90 nm wide gold nanoslit with 190 nm thickness surrounded by 600-nm-period grooves, Lee et al tested refractive index sensitivities by injecting purified water mixed with various ratios of glycerin into the devices. In water environment, the Fano resonant location was shifted to around 810 nm from 620 nm (Fig. 9.9 (c)). In addition, when the concentrations of glycerin increase, the wavelengths of Fano resonances were red-shifted. Fig. 9.9 (d) shows the sensitivities are 577 and 615 nm/RIU for resonant peak and dip, respectively. Moreover, the device was used to detect the interactions between BSA and anti-BSA. From Fig. 9.9 (e), it is observed that significant changes in wavelength shift occurred when BSA and anti-BSA were bound on the gold surface. The binding of anti-BSA resulted in a wavelength shift of almost 4nm. Finally, it is worth pointing out at the exit of the slit adding a metallic pad can further improve the wavelength sensitivity, up to 926 nm/RIU [38].

9.6. Conclusions

In contrast to conventional plasmonic biosensor, Fano resonance plasmonic biosensor has been demonstrated in theory and experiments to provide higher FOM and sensitivity. For example, dolmen type biosensor provides a FOM almost 1.5 times the theoretic limit of prism-based biosensor if the appropriate mode is chosen [33]. The refractive index sensitivity using nanohole array can reach 700 nm/RIU, better than 538 nm/RIU reported in a LSPR biosensor [39]. Similar to LSPR-based plasmonic devices, the Fano resonance biosensors have the additional advantage of easier miniaturization and multiplexing. As fabrication techniques mature and nanoparticles become more uniform and stable, and more diverse and complementary techniques are integrated, Fano resonance plasmonic biosensors have bright future. However, currently the Fano resonance biosensor is still in the early stage. The reported Fano-resonance biosensors were only used for fundamental studies of protein–protein interaction. More practical challenges remain before such biosensors can reach their full potential as analytical tools for chemical quantification, the characterization of binding kinetics, the detection of conformational changes, and molecular identification. Also, a lot of work is needed for its capability for analysis of other molecules, proteins, antibody and antigen, DNA and RNA hybridization. On the other hand, the next generation of biomedical devices will ideally be miniaturized and automated. Low-cost biosensors based on Fano resonance, designed for fast, real-time identification of biomarkers, are required for routine point-of-care clinical evaluation,

real-time diagnosis of diseases in developing countries and fast genetic mapping for personalized care.

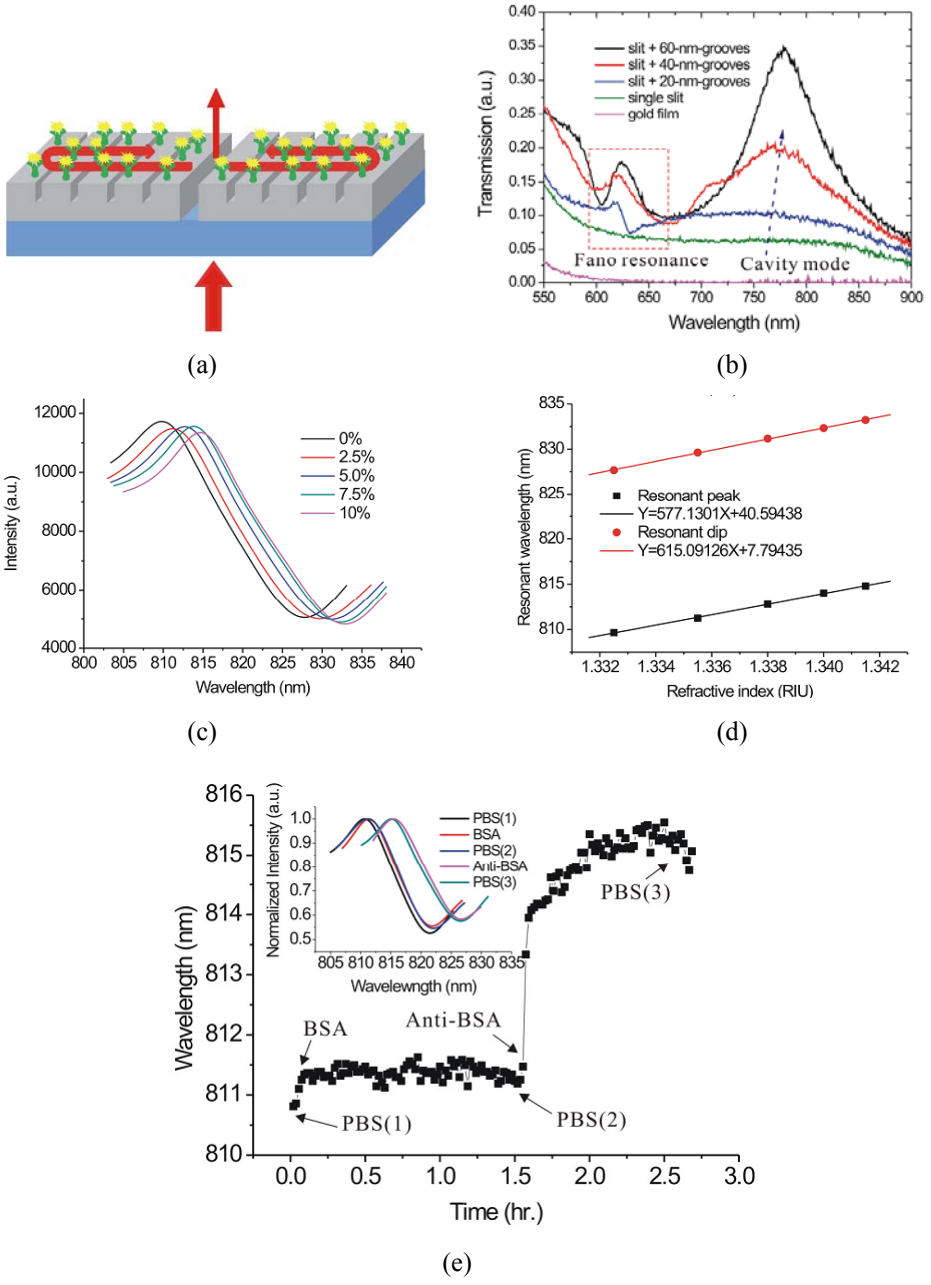


Fig. 9.9 Silt-groove nanostructure sensing. (a) Schematic illustration of the sensing device.(b) The transmission spectra of the single nanoslit with 600-nm-period grooves. (c) The Fano resonance spectra in various water/glycerin mixtures. (d) The Fano resonant peak and dip as a function of solution refractive index. (e) BSA testing using silt-groove nanostructure. Figures reproduced with permission from ref. [32].

Acknowledgements

This work was supported by the Scientific and Technique Research Project of Sichuan Province, China.

References

[1]. X. W. Guo, Surface plasmon resonance based biosensor technique: A review, *Journal of Biophotonics,* Vol. 5, Issue 7, 2012, pp. 483-501.
[2]. J. Homola, S. S. Yee, G. Gauglitz, Surface plasmon resonance sensors: review, *Sensors and Actuators B*, Vol. 54, 1999, pp. 3-15.
[3]. J. N. Anker, W. P. Hall, O. Lyandres, N. C. Shah, J. Zhao, R. P. Van Duyne, Biosensing with plasmonic nanosensors, *Nature Materials,* Vol. 8, 2008, pp. 442-453.
[4]. D. Mortazavi, A. Z. Kouzani, A. Kaynak, Nano-plasmonic Biosensors: A Review, in *Proceedings of the 2011 IEEE/ICME International Conference on Complex Medical Engineering*, Harbin, China, 22-25 May 2011, pp. 31-36.
[5]. A. G. Brolo, R. Gordon, B. Leathem, K. L. Kavanagh, Surface plasmon sensor based on the enhanced light transmission through arrays of nanoholes in gold films, *Langmuir*, Vol. 20, 2004, pp. 4813-4815.
[6]. J. C. Yang, J. Ji, J. M. Hogle, D. N. Larson, Metallic nanohole arrays on fluoropolymer substrates as small label-free real-time bioprobes, *Nano Lett,* Vol. 8, 2008, pp. 2718-2724.
[7]. U. Fano, Effects of Configuration Interaction on Intensities and Phase Shifts, *Physic Review,* Vol. 124, 2004, pp. 1866-1878.
[8]. R. H. Ritchie, Plasma Losses by Fast Electrons in Thin Films, *Physic Review,* Vol. 106, Issue 5, 1957, pp. 874-881.
[9]. W. L Barnes, A. D. Thomas, W. Ebbesen, Surface plasmon subwavelength optics, *Nature*, Vol. 424, 2003, pp. 824-830.
[10]. K. A. Willets, R. P. Van Duyne, Localized surface plasmon resonance spectroscopy and sensing, *Annu. Rev. Phys. Chem.,* Vol. 58, 2007, pp. 267-297.
[11]. E. Kretschmann, H. Raether, Radiative decay of non radiative surface plasmons excited by light, *Z. Naturforsch,* Vol. 23, 1968, pp. 2135-2136.
[12]. H. F. Ghaemi, Tineke Thio, D. E. Grupp, T. W. Ebbesen, and H. J. Lezec, Surface plasmons enhance optical transmission through subwavelength holes, *Physical Review B*, Vol. 58, 1998, pp. 6779-6781.
[13]. E. Hutter, J. Fendler, Exploitation of Localized Surface Plasmon Resonance, *Advance Material,* Vol. 16, 2004, pp. 1685-1706.
[14]. A. G. Brolo, Plasmonics for future biosensors, Nature Photonics, Vol. 6, 2012, pp. 709-713.
[15]. A. K. Sharma, R. Jha, B. D. Gupta, Fiber-Optic Sensors Based on Surface Plasmon Resonance: A Comprehensive Review, *IEEE Sensors Journal,* Vol. 7, Issue 8, 2007, pp. 1118-1129.
[16]. L. J. Sherry, S. H. Chang, G. C. Schatz, R. P. Van Duyne, B. J. Wiley, Y. N. Xia, Localized surface plasmon resonance spectroscopy of single silver nanocubes, *Nano Lett.,* Vol. 5, Issue 10, 2005, pp. 2034-2038.
[17]. B. Lukyanchuk, N. I. Zheludev, S. A. Maier, N. J. Halas, P. Nordlander, H. Giessen, C. T. Chong, The Fano resonance in plasmonic nanostructures and metamaterials, *Nature Materials,* Vol. 9, 2010, pp. 707-715.
[18]. Fano, U., Sullo spettro di assorbimento dei gas nobilipresso il limite dello spettro darco, *Nuovo Cimento,* Vol. 12, 1935, pp. 154-161.

[19]. A. E. Miroshnichenko, S. Flach, Y. S. Kivshar, Fano resonances in nanoscale structures, *Reviews of Modern Physics,* Vol. 82, 2010, pp. 2257-2298.
[20]. S. H. Fan, W. Suh, J. D. Joannopoulos, Temporal coupled-mode theory for the Fano resonance in optical resonators, *Journal of Optics Society of America A,* Vol. 20, 2003, pp. 569-572.
[21]. A. Christ, S. G. Tikhodeev, N. A. Gippius, J. Kuhl, H. Giessen, Waveguide plasmon polaritons: Strong coupling of photonic and electronic resonances in a metallic photonic crystal slab, *Physics Review Letter,* Vol. 91, 2003, pp. 183901.
[22]. J. A. Fan, K. Bao, C. H. Wu, J. M. Bao, R. Bardhan, N. J. Halas, V. N. Manoharan, G. Shvets, P. Nordlander, F. Capasso, Fano-like interference in Self-assembled plasmonic quadrumer clusters, *Nano Letter,* Vol. 10, 2010, pp. 4680-4685.
[23]. M. Gustav, Beiträge zur Optik trüber Medien, speaiell kolloidaler Metallösungen, *Annals of Physics,* Vol. 25, 1908, pp. 377-445.
[24]. A. H. Mohammadian, V. Shankar, W. F. Hall, Computation of Electromagnetic Scattering and Radiation Using a Time-Domain Finite-Volume Discretization Procedure, *Computer Physics Communications,* Vol. 68, Issue 1, 1991, pp. 175-196.
[25]. N. Liu, T. Weiss, M. Mesch, L. Langguth, U. Eigenthaler, M. Hirscher, C. Soennichsen, H. Giessen, Planar metamaterial analogue of electromagnetically induced transparency for plasmonic sensing, *Nano Letter,* Vol. 10, 2010, pp. 1103-1107.
[26]. S. Barbora, L. Petra, S. Hana, K. Pavel, R. Ivan, H. Jiri, Ambiguous Refractive Index Sensitivity of Fano Resonance on an Array of Gold Nanoparticles, *Plasmonics*, Vol. 9, Issue 4, 2014, pp. 729-735.
[27]. A. E. Cetin, H. Altug, Fano Resonant Ring/Disk Plasmonic Nanocavities on Conducting Substrates for Advanced Biosensing, *Acs Nano,* Vol. 6, Issue, 11, 2012, pp. 9989-9995.
[28]. H. L. Liu, X. J. Wu, B. Li, C. C. Xu, G. B. Zhang, L. J. Zheng, Fano resonance in two-intersecting nanorings: Multiple layers of plasmon hybridizations, *Applied Physics Letters,* Vol. 100, 2012, 153114 (1-4).
[29]. N. Liu, L. Langguth, T. Weiss, J. Kaestel, M. Fleischhauer, T. Pfau, H. Giessen, Plasmonic analogue of electromagnetically induced transparency at the drude damping limit, *Nature Material,* Vol. 8, 2009, pp. 758-762.
[30]. N. Verellen, Y. Sonnefraud, H. Sobhani, F. Hao, V. V. Moshchalkov, P. Van Dorpe, P. Nordlander, S. A. Maier, Fano resonances in individual coherent plasmonic nanocavities, *Nano Letter,* Vol. 9, 2009, pp. 1663-1667.
[31]. A. A. Yanik, A. E. Cetina, M. Huang, A. Artara, S. H. Mousavic, A. Khanikaevc, J. H. Connord, G. Shvetsc, H. Altug, Seeing protein monolayers with naked eye through plasmonic Fano resonances, *PNAS*, Vol. 108, Issue 29, 2011, pp. 11784-11789.
[32]. K. L. Lee, S. H. Wu, C. W. Lee, P. K. Wei, Sensitive biosensors using Fano resonance in single gold nanoslit with periodic grooves, *Optics Express,* Vol. 19, Issue 24, 2011, pp. 24530-24539.
[33]. C. H. Wu, A. B. Khanikaev, R. Adato, N. Arju, A. A. Yanik, H. Altug, G. Shvets, Fano-resonant asymmetric metamaterials for ultrasensitive spectroscopy and identification of molecular monolayers, *Nature Material,* Vol. 11, 2012, pp. 69-75.
[34]. G. Benjamin, Fano Resonances in Plasmonic Nanostructures: Fundamentals, Numerical Modeling and Applications, Doctoral Thesis, *Swiss Federal Institute of Technology (EPFL),* 2012.
[35]. T. W. Ebbesen, H. J. Lezec, H. F. Ghaemi, T. Thio, P. A. Wolff, Extraordinary optical transmission through sub-wavelength hole arrays, *Letters to Nature,* Vol. 391, 1998, pp. 667-669.
[36]. A. Degiron, H. J. Lezec, W. L. Barnes, T. W. Ebbesen, Effects of hole depth on enhanced light transmission through subwavelength hole arrays, *Applied Physics Letter,* Vol. 81, 2002, pp. 4327-4329.

[37]. F. J. G. de Abajo, Light scattering by particle and hole arrays, *Reviews of Modern Physics,* Vol. 79, 2007, pp. 1267-1290.

[38]. K. L Lee, J. B. Huang, J. W. Chang, S. H. Wu, P. K. Wei, Ultrasensitive Biosensors Using Enhanced Fano Resonances in Capped Gold Nanoslit Arrays, *Scientific Reports,* Vol. 5, 2014, 8547 (1-8).

[39]. E. M. Hicks, X. Y. Zhang, S. L. Zou, O. Lyandres, K. G. Spears, G. C. Schatz, R. P. Van Duyne, Plasmonic properties of film over nanowell surfaces fabricated by nanosphere lithography, *Journal of Physics Chemistry B,* Vol. 109, 2005, pp. 22351-22358.

Chapter 10
Linearization of Sensor Signal in FPGA: A Multichannel Approach for High Speed Real Time Applications

Durlav Sonowal and Manabendra Bhuyan

Abstract

Multi-channel data processing is a major requirement of many industrial applications for process monitoring and control. In this type of systems, data have to be obtained from different sensors simultaneously through multiple channels. Many sensors used for acquisition of different physical quantity have nonlinear input output characteristics. In such situation, it is required to linearize the nonlinear response of sensor signals which are simultaneously obtained from multi-channel data acquisition system for precise measurement. While various techniques have been described in literature for sensor linearization, this chapter highlighted the advantages of FPGA based sensor linearization techniques for real time multi-channel operation. In this chapter FPGA implementation of piecewise linearization (PWL), look up table (LUT) based linearization, linearization by interpolation (LI) and artificial neural network (ANN) based linearization methods have been discussed. The different aspects of these linearization techniques for real time multi-channel application and the trade-off between accuracy and implementation area in FPGA have been discussed. The comparative analysis was performed by using identical thermistors connected to 8 channels for linearization and the performance of each method was analyzed. The performance has been is estimated on the basis of logic utilization, linearization accuracy, execution time, noise immunity and speed of operation. Fixed point arithmetic has been used for data representation in the FPGA implementation.

10.1. Introduction

Linearization of sensor signals with nonlinear characteristics is a very important part of signal processing in the field of instrumentation and measurement. So far; a large number

Durlav Sonowal
Department of Electronics and Communication Engineering, Tezpur University, India.

of techniques have been proposed on sensor linearization. For ease of explanation and discussion, the various linearization techniques have been categorized as follows:
1) Analog circuit and ADC based;
2) Microcontroller based;
3) VLSI and FPGA based;
4) Software based.

1) Analog Circuit and ADC Based

The use of analog circuits with passive and active components is the oldest approach to linearize sensor characteristics which are in general widely used for resistive sensors. Additionally use of analog to digital converter (ADC) is also widely popular for linearization of sensor signal. A transducer linearization method using radiometric property of analog to digital converter (ADC) has been proposed in [2]. In this technique the maximum relative deviation of nonlinearity is found as 0.5 KPa in a range of 100 KPa. In a similar type of work in [3], a nonlinear ADC with digitally selectable quantizing characteristics was proposed for linearization of nonlinear analog signals. In [4], a programmable nonlinear ADC was proposed that uses an optimal size of ROM to realize the nonlinear characteristics. In [5] a nonlinear ADC circuit based on switched capacitor architecture was proposed that uses piecewise linear approximation of the nonlinear conversion characteristics. The maximum nonlinear deviation was found as 0.0937 l/hr. (1LSB in the range of 0-24 l/hr.). In [6], a four constant curve fit was proposed and a temperature to frequency converter circuit was developed to employ the proposed fit. The peak absolute error of linearization is found to be 0.5 K. In [7], linearization of an NTC thermistor based on 555 timer circuit with frequency and analog output was proposed. In this method, the maximum percentage nonlinearity error claimed is ± 1 % and ± 1.7 % in two different ranges. In [8], functional link artificial neural network (FLANN) based linearization is proposed using a plug-in module type of digital circuit, where the maximum full scale error is claimed as ± 2 %.

2) Microcontroller Based

Microcontrollers are widely used for implementation of various functions for small and low cost applications. Linearization of sensor signals using microcontrollers are available in many literatures. An optimal method of sensor linearization using look up table in small embedded system have been proposed in [9]. In this work a theory was also proposed for finding the minimum allowable size of a look up table without affecting the precision. However, this method is limited to LUT entries with only integer part, hence precision is also limited. A microcontroller based multiple sensor linearization using ANN has been proposed in [10]. The maximum approximation error of tanh function was predicted as less than 1 %.

3) VLSI and FPGA Based

For linearization of second and third order sensor models a MOS VLSI model has been proposed [11]. The circuit uses a simple continuous time analog signal processing cell that allows single chip fabrication. But in higher order linearization using MOS

transconductance model, there is a problem of error due to amplifier dynamics and system stability due to the feedback. The maximum frequency of operation in this method is 1 kHz. In an FPGA based sensor linearization and compensation of nonlinearity due to the environmental affect using ANN [12], the maximum full scale error is observed to be ± 1.5 %. The method in [13], proposes a low cost microcontroller for linearization of a distance sensor by look up table (LUT), piece-wise linear interpolation and extrapolation (PWLI/E), fuzzy and ANN. The nonlinearity error analysis of this work shows a maximum relative error of more than 0.3 in Fuzzy approach, 0.15 in ANN and 0.10 in PWLI.

In a previous work of the authors of this chapter, FPGA implementation of thermistor linearization have been presented, where a nonlinear thermistor characteristic for a temperature range of $0 - 100$ ℃ was linearized in FPGA [14, 15]. However the data format used in FPGA was IEEE 745 32 bit format which consumes higher logic area. Moreover the methods were suggested for real time applications of a single channel. Fixed point representation of data reduces logic area in FPGA. In this chapter FPGA implementation of linearization techniques have been represented by using fixed point format and the performance of the system was examined using real time signals for multi-channel applications.

4) Software Based

Several algorithms for linearization of sensor signals have been developed till date in software/computer based approach. A graphical approach has been proposed that uses a numerical method of iteration to obtain a linear solution. This method offers flexibility of online choice and replacement of thermistor over an extended range of operation [16]. A method for function approximation in microcomputer memory implementing the inverse characteristics of the sensor for smart sensor system (SSS) has been proposed [17]. Simulation of a two layer ANN for linearization of thermistor response using has been proposed in [18]. Some other software based sensor linearizations were- multi layer perceptron (MLP) ANN in [19, 20], sensor calibration and compensation by adaptive linear element based neural network in [21], temperature compensation of a pressure sensor using BP neural network in [22].

The above literature survey reveals different sensor linearization methods in software as well in hardware. In software or computer based linearization technique, different algorithms like interpolation, piecewise linearization, look up table, ANN etc. have been implemented using different software tools like MATLAB or programming language like C/C++. However, software modeled linearization methods are sequential with significant effects to time response for multichannel. In hardware or circuit level linearization techniques, both analog and digital methods are possible. There are applications where both digital and analog linearizations are combined in hybrid mode [23]. However, reconfiguration is difficult in analog hardware based linearization. Different types of digital implementations such as- microcontroller based implementation, DSP based implementations and FPGA based implementation have been mentioned above. Microcontroller and DSP based implementations are sequential and hence do not preserve the parallel structure which is a vital requirement for high speed multi-input-multi-output

(MIMO) systems. For example, DSP based or microcontroller based implementation of ANNs do not preserve the parallel architecture of the networks. On the other hand, FPGA devices offer a great flexibility in reconfiguration simultaneously maintaining the parallelism where required.

The use of ANN for sensor compensation and linearity correction is proposed in many literatures. ANN provides an efficient tool for mapping nonlinear input-output relations of sensor signals. Various ANN based linearization techniques are available such as - functional link artificial neural network (FLANN) based linearization [8]; ANN based sensor linearization and compensation of nonlinearity due to the environmental effect [12]; low cost microcontroller for ANN based linearization of a distance sensor [13]; simulation of thermistor linearization using a two layer ANN [18]; sensor linearization based on MLP ANN [19]; sensor calibration and compensation by adaptive linear element based neural network [21] and temperature compensation of a pressure sensor using BP neural network [22].

FPGAs are basically integrated circuits designed and configured by the user, which contains a matrix of configurable logic blocks and a hierarchy of reconfigurable interconnects that allow the blocks to be connected together as per the requirements of the user. The FPGA implementations of ANNs are discussed in many literatures [24-39]. A fully digital MLP ANN implemented with two XC3042 FPGAs and a 1 k × 8 EPROM was presented in [25, 26]. In [27], authors discussed the FPGA implementation of ANN and showed that the high speed operation in real time applications of ANN could be achieved only if the networks were implemented using parallel hardware architecture. In [28], parameterized neuron was developed in FPGA. In [29], a neural network topology was described that implements the bit-serial architecture by reducing the requirement of hardware complexity. Earlier, researchers thought that the size of an ANN importantly determines the performance of ANN, however, recently a large number of researchers believe that the network architecture is responsible for determining the performance of the network [30, 31]. In [32], a layer multiplexing technique for feedforward NN was proposed that reduces the requirement of hardware in ANN. In [33], the use of FPGAs in industrial control applications and neural network based FPGA systems were highlighted. In [34], generalized backpropagation multilayer perceptron architecture was described for online applications.

From the above literature survey it can be revealed that ANN has been implemented in FPGA for different applications – industrial control, vector control drive etc., however very little effort has been made for ANN implementation in FPGA for sensor linearization except in [16]. In [16], ANN has been used for sensor linearization and compensation of nonlinearity due to the environmental effect and the maximum full scale error is found to be ± 1.5 %. However, FPGA based linearization by piecewise segmentation, linear interpolation and look-up table is not available in the literature so far.

Although microcontroller based linearization is cost effective than FPGA for a single channel application, FPGA based linearization is promising for multichannel sensor linearization. Although microcontroller can be interfaced to multichannel ADCs the computational bandwidth cannot be optimally utilized compared to FPGA. This is

because; FPGA can be utilized in parallel hardware architecture. Therefore, FPGA based sensor linearization will be a better solution in multi-channel sensor array applications. In multi-channel sensor arrays applications different types (monotype or multi-type) of sensors are installed in different locations of an application and FPGA implementation becomes a simple task by routing the sensor outputs through an ADC to identical FPGA 'blocks' in hardware. Fig. 10.1 shows a proposed scheme for multichannel implementation of linearization in FPGA where the FPGA can be configured with parallel architecture of linearization blocks for parallel computation of sensor linearization.

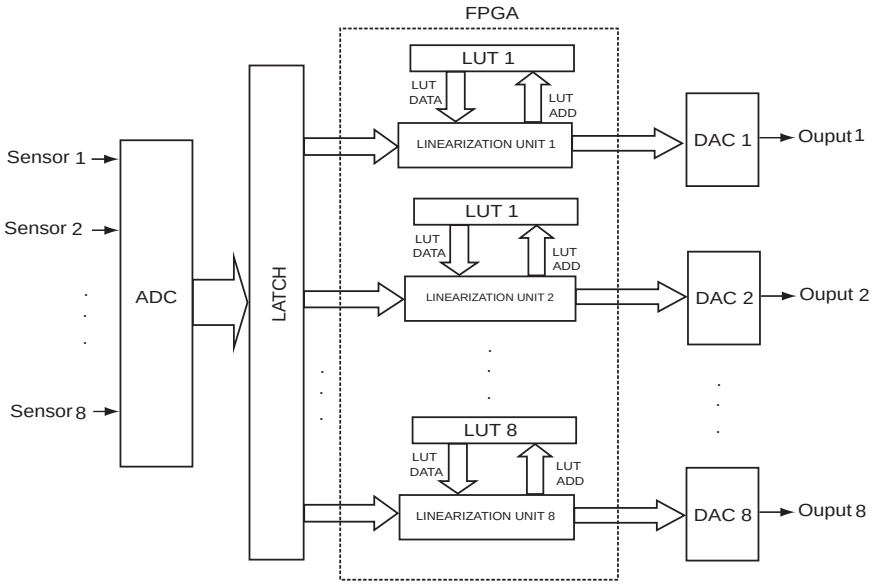


Fig. 10.1. Multiple sensor signal linearization in FPGA.

In this chapter the FPGA implementations of sensor linearization with multiple channels have been discussed using the following techniques - (1) Piecewise linearization, (2) Look up table (LUT) based linearization, (3) Linearization by interpolation and (4) Artificial Neural Network (ANN) based linearization. To realize these methods in FPGA for multi-channel operation a total of 8 channels have been used. Thermistors have been connected to all the 8 channels to acquire signals for linearization.

After the introduction in Section 10.1 the remaining part of the chapter is organized as follows; Section 10.2 presents the basic theory of linearization of sensor signal and the experimental setup. FPGA implementation of linearization methods are discussed in Section 10.3; results and discussion are presented in each subsection and concluded in Section 10.4.

10.2. Theory of Linearization and Experimental Setup

A sensor or sensor system generally converts the physical quantities into an electrical signal, preferably into voltage signal. The basic linearization theory states that if

$y = f(x)$ is a nonlinear function that describes the relationship between input variable x and its corresponding nonlinear output y, then the relationship between sensor output y and output of the linearization unit z must be an inverse function such that,

$$z = f^{-1}(y) = f^{-1}(f(x)) = x \tag{10.1}$$

A basic linearization process is illustrated in Fig. 10.2. For realization of different methods of linearization for multi-channel; 8 thermistors with nonlinear characteristics (with $R_0 = 10\,k\Omega$ and thermistor constant $\beta = 3750\,K$) has been chosen. Each of the thermistors has been connected to a voltage divider circuitwith a fixed resistor as shown in Fig. 10.3. The output voltage (V_o) of the voltage divder circuit is a nonlinear function of temperature and has to be linearized. As per Equation (10.1), the input temperature to the thermistor is x and the output of the voltage divider circuit V_o is y. To linearize the nonlinear function $y = f(x)$ the linearization unit has been configured digitally such that $g = f^{-1}$, i.e. $g(f(x)) = f^{-1}(f(x)) = x$.

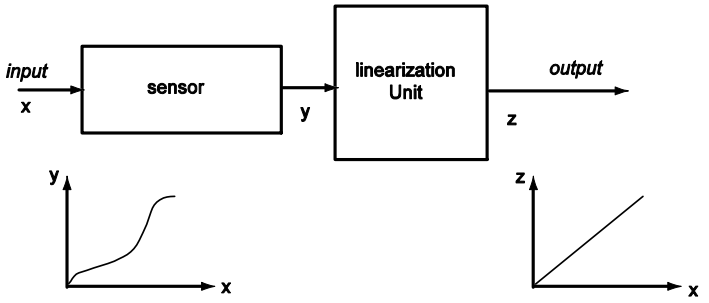


Fig. 10.2. Basic Linearization Process.

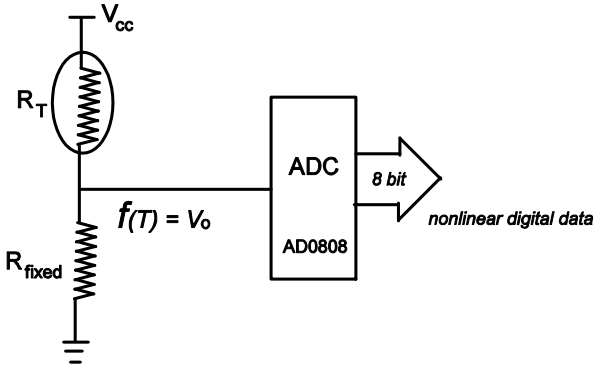


Fig. 10.3. Primary measurement circuit.

The outputs of the voltage divider circuits have been connected to all the 8 channels of ADC for analog to digital conversion. Fig. 10.3 shows the thermistor circuit connected to an 8-bit, 8 channel ADC (ADC0808) interfaced with Spartan-III XC3S400 FPGA. In the thermistor circuit a fixed resistance value of 1 kΩ (tolerance-0.005 %, temperature

coefficient of 0.2 ppm/ºC and thermal stability less than 1 sec) is used which provides output voltage in the range from $0.1569\,\text{V} - 3.2079\,\text{V}$ for a temperature range of $0 - 100$ ℃. The voltage source in the thermistor circuit is +5 V powered by a precision voltage source (ST4076) with a ripple noise specification of less than 1 mV (rms) and load regulation of $\pm$ 0.05 %.

To drive ADC, DACs and LCD finite state machines have been implemented in FPGA. DAC AD7541 provides analog version of the linearized output. Fig. 10.1 shows the block diagrams of the linearization unit. The Spartan-III XC3S400 FPGA was synthesized and configured using Xinilx ISE 11.

10.3. FPGA Implementation

As mentioned above, four different techniques of sensor linearization have been implemented and these are described in details in the following subsections.

10.3.1. Piecewise Linearization (PWL)

In piecewise linearization method, the nonlinear sensor characteristic is divided into a certain number of linear segments. Each of the segments within its linear range is fitted to a known linear equation having a slope (m) and a bias (c). The number of segments in the nonlinear characteristic is dependent on accuracy of the linearization required and the length of operating range. However, the characteristic equation for each linear segment depends on the nonlinear characteristics of the sensor for the operating range. Considering n numbers of segments in the non-linear thermistor characteristics, each segment of the characteristic equation can be expressed by a straight line equation for which the slopes (m) and intercepts (c) are different. The equations of the straight line segments can be determined either graphically or using the coordinates of the two end points of the segments. The equation of the straight line can be written as

$$y_{n-1,n} = m_n x_{n-1,n} + c_n \tag{10.2}$$

where subscript n indicates the segment number and $y_{n-1,n}$ and $x_{n-1,n}$ indicate the values in $(n-1)^{th}$ to n^{th} range. Here, initially 20 segments were considered for linearization of the nonlinear output voltage obtained from the thermistor measurement circuit and the corresponding linearized temperature values are obtained by using the following equations

$$\begin{aligned} T_{i,1} &= m_1 V_{o,1} + c_1 \\ T_{i,2} &= m_2 V_{o,2} + c_2 \\ &\dots \\ T_{i,20} &= m_{20} V_{o,20} + c_{20} \end{aligned} \tag{10.3}$$

Before configuring the linearization algorithm in FPGA the slopes (m_n) and biases (c_n) given by Equation (10.3) are calculated. Table 10.1 shows the piecewise linear equations for 20 segments that we have used in this work. The same have been repeated for $n = 30$ and $n = 40$.

Table 10.1. Piecewise Linear Equations of Thermistor Characteristics for the Temperature Range of 0-100 ℃ with 20 Segments.

Temperature Range (℃)	Output Voltage Range (Volt)	Linear Equations
[0-5]	[0.1569-0.2019]	$T_i = 123.852V_o - 19.631$
[5-10]	[0.2019-0.2567]	$T_i = 103.8084V_o - 15.6448$
[10-15]	[0.2567-0.3230]	$T_i = 87.9795V_o - 11.7344$
[15-20]	[0.3230-0.4018]	$T_i = 75.4031V_o - 7.9128$
[20-25]	[0.4018-0.4945]	$T_i = 65.3598V_o - 4.1949$
[25-30]	[0.4945-0.6021]	$T_i = 57.3075V_o - 0.5982$
[30-35]	[0.6021-0.7251]	$T_i = 50.835V_o + 2.8577$
[35-40]	[0.7251-0.8637]	$T_i = 45.6279V_o + 6.1501$
[40-45]	[0.8637-1.0178]	$T_i = 41.445V_o + 9.2532$
[45-50]	[1.0178-1.1866]	$T_i = 38.1001V_o + 12.1383$
[50-55]	[1.1866-1.3686]	$T_i = 35.4491V_o + 14.7727$
[55-60]	[1.3686-1.5619]	$T_i = 33.33806V_o + 17.12$
[60-65]	[1.5619-1.7644]	$T_i = 31.8081V_o + 19.1399$
[65-70]	[1.7644-1.9732]	$T_i = 30.6654V_o + 20.7875$
[70-75]	[1.9732-2.1856]	$T_i = 29.9016V_o + 22.0133$
[75-80]	[2.1856-2.3987]	$T_i = 29.4786V_o + 22.7628$
[80-85]	[2.3987-2.6098]	$T_i = 29.3687V_o + 22.9762$
[85-90]	[2.6098-2.8164]	$T_i = 29.5523V_o + 22.5884$
[90-95]	[2.8164-3.0163]	$T_i = 30.017V_o + 21.5284$
[95-100]	[3.0163-3.2079]	$T_i = 30.7563V_o + 19.719$

10.3.1.1. FPGA Implementation of PWL

Fig. 10.4 illustrates the piecewise linearization algorithm in FPGA. Here, the output voltages (V_o) from the multi-channel voltage divider circuits are fed to an ADC and the digital versions of V_o are compared online with previously stored voltage ranges ($V_{o,n}$). The comparator generates slopes m_n and bias c_n for the n^{th} range of input data. The corresponding slope (m_n) is multiplied with input voltage which is then added with the

corresponding bias c_n to get the linear temperature value T_i given by a straight line equation.

$$T_{i,1} = m_1 V_{o,1} + c_1$$

of the i^{th} piecewise linear segment. The linearized outputs obtained from the FPGA system are processed through the DACs to obtain the analog outputs.

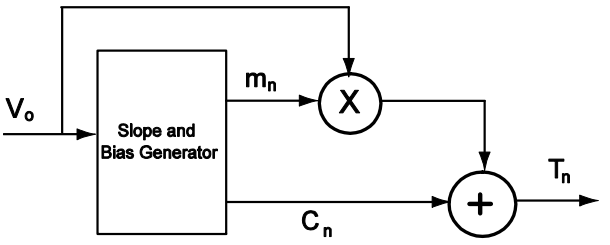


Fig. 10.4. Block Diagram of piecewise linearization.

10.3.1.2. Data Representation

In FPGA devices, arithmetic operations are performed using either floating point or fixed point arithmetic. For higher degree of accuracy and precision; floating point arithmetic such as IEEE 754 std. 32 bit and 64 bit are used. However, the floating point arithmetic is complex as compared to fixed point arithmetic and consumes higher implementation area in FPGA. Fixed point arithmetic is a trade-off between conventional integer arithmetic and floating point and therefore we have adopted fixed point arithmetic operation for implementation of linear interpolation in FPGA. The major advantage of fixed point arithmetic is that it follows the same basic mathematical rule for addition, subtraction, multiplication and division that is applicable for conventional binary numbers. For this reason fixed point arithmetic can easily be implemented in FPGA with minimum resource consumption and minimum complexity.

In this work we have used different bit lengths for input voltage (V_o), slope (m_n), bias (c_n) and output (T_i). The input data V_o is represented using $S2.10$ fixed point format, i.e. MSB is the sign bit, two bits for the integer (I) and ten fractional places (F) as shown below

SII.FFFFFFFFFF

This format is suitable to represent the voltage levels ranging from 0 volt to 3.9 volt which is sufficient for this work. The format also provides a data precision of 2^{-10} which is 9.7656×10^{-4} in this case. For illustration, let us consider a voltage level of '1.35V' for representation in $S2.10$ fixed point format. Here the number is positive, therefore $S = 0$ and the integer part '1' is represented using 2 bit binary as '01'. The fractional part '0.35', is multiplied by $1024(2^{10})$ and then converted to binary, i. e., $0.35 \times 1024 =$ $= 358.4 \sim 358$. The 10 bit binary equivalent of '358' is '0101100110'. Therefore the number '1.35' can be represented in $S2.10$ fixed point formats as '010101100110'.

The slope (m_n) and bias (c_n) are also represented in the same $S2.10$ format and output temperatures (T_i) are represented in $S8.8$ fixed point format for getting higher ranges.

10.3.1.3. Results of PWL

In this work we have implemented piecewise linearization of an array of 8 nonlinear thermistors in FPGA for three different numbers of piecewise segments, i.e. 20, 30 and 40. We have used performance measures- root mean square error (RMSE), mean absolute deviation (MAD) and standard deviation (SD) for evaluation of performance with varied number of segments. The experiment is performed with real time signals. The thermistors connected to the voltage divider circuits are placed in temperature controlled chamber. The temperature is varied from 0 – 100 ℃. The outputs of the voltage divider circuits and the outputs of DACs after linearization are recorded using LabView based data acquisition system (NI DAQ USB 6351). Outputs of the voltage divider circuits and FPGA outputs are compared with the output of a K-type thermocouple by using LabView based standard temperature measurement system.

Fig. 10.5 shows the full scale percentage error for 8 channels with 40 segments of each channel and Fig. 10.6 shows the full scale percentage error for channel 1 with $n = 20$, $n = 30$ and $n = 40$. Table 10.2 shows RMSE, MAD, SD, Maximum full scale error for all 8 channels with 40 segments. In a similar way, RMSE, MAD, SD, Maximum full scale error was calculated for linearization with 20 and 30 piecewise segments. Table 10.3 presents a comparative analysis for different number of segments for one channel. A maximum full scale (FS) error of 2.5064 % and 1.5027 % are observed with 20 and 40 piecewise linear segments respectively. The figure shows a 45.24 % reduction in RMSE due to the increase in number of linearization segments from 20 to 40. However, 4 % to 7 % increase in LUT area utilization is observed when the number of segment is increased from 20 to 40. The device uses 290 elements of 4-input LUTs with 20 segments with maximum FS error of 2.5064 %; whereas with 556 elements of 4-input LUTs for 40 segments it results a 1.5027 % maximum FS error, i.e. the maximum error is reduced by 40 % by increasing the number of linear segments from 20 to 40.

It is important to know that how much noise has been contributed by the linearization process since the overall performance of the proposed method does not depend only on the linearity error, but also on the noise contributed by the electronics such as fluctuation of electron (Johnson-Nyquist noise) due to the thermal agitation of the carrier electrons or noise due to nonlinear temperature dependent mobility in MOS transistors. The noise analysis of the sensor signals has been done by performing Fast Fourier Transform (FFT) on the sensor output signal and the linearized signal. FFT of zero level signals, viz. zero level sensor and zero level FPGA outputs are also shown in Fig. 10.7 which indicates that the FPGA contributes very less noise to the sensor output.

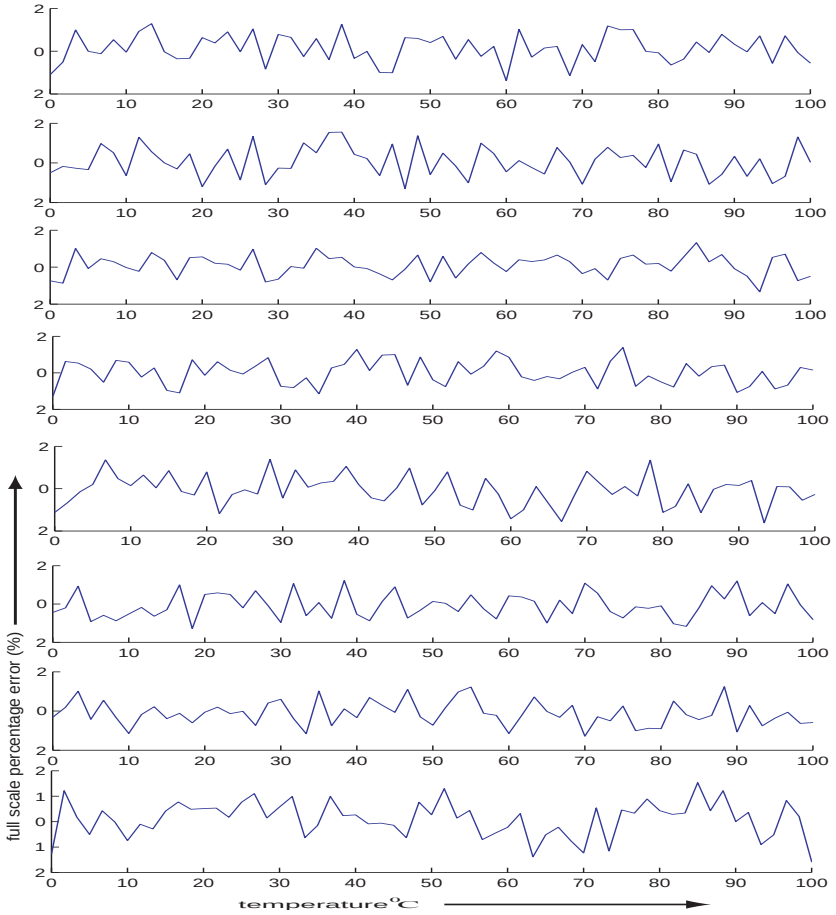


Fig. 10.5. Full scale percentage error in piecewise linearization with 40 segments for 8 channels from 0°C to 100°C.

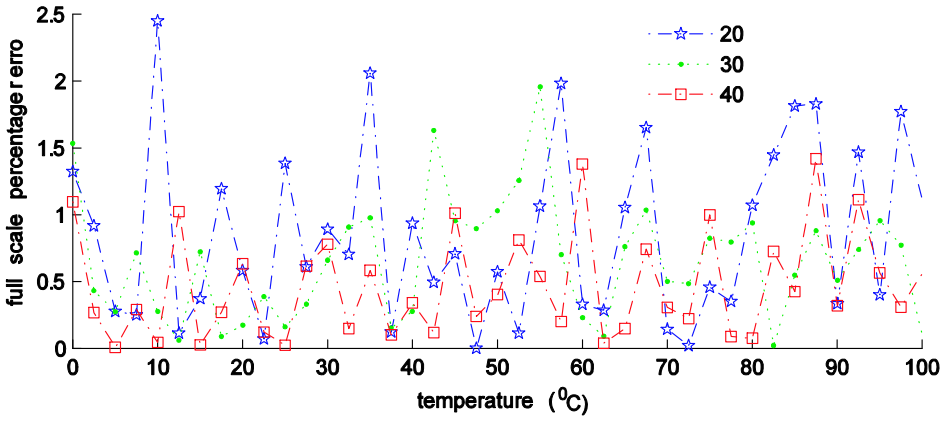


Fig. 10.6. Full scale percentage error in piecewise linearization with 20, 30, 40 segments for a single (CH1) channel from 0 °C to 100 °C.

Table 10.2. Error analysis of PWL in FPGA for 8 Channels with $n = 40$.

Channel	RMSE	MAD	SD	Maximum FS error (%)
1	0.6238	0.5030	0.6237	1.5027
2	0.6562	0.5382	0.6560	1.5896
3	0.6270	0.5157	0.6259	1.4576
4	0.6523	0.5308	0.6516	1.5262
5	0.7004	0.5638	0.7004	1.5799
6	0.5646	0.4595	0.5647	1.3904
7	0.6446	0.5168	0.6449	1.6245
8	0.6997	0.5655	0.6991	1.6950

Table 10.3. Comparative error analysis of PWL in FPGA for channel-1 with different number of piecewise segments.

Number of segments	20	30	40
RMSE	1.1391	0.8438	0.6238
MAD	0.9266	0.6903	0.503
SD	1.14	0.8445	0.6237
Maximum FS error (%)	2.5064	1.956	1.5027
No. of 4 input LUTs (out of 7168)	290 (4 %)	433 (6 %)	556 (7 %)

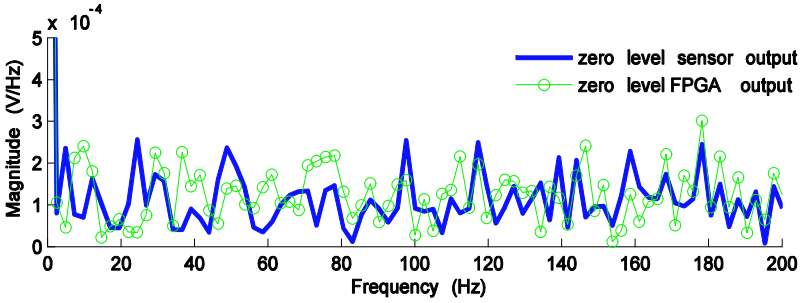


Fig. 10.7. FFT of zero level signal for sensor output and FPGA output.

The FFT of the linearized output shown in Fig. 10.8 reveals that the thermistor and the measuring circuit produces low frequency or $1/f$ noise where the noise density decreases at higher frequency. It is observed that the noise density in the sensor output and the FPGA output has very little difference. However spectral leakages are observed in the linearized output which is due to random fluctuations of the mobility carriers in the electronics. The white noise density level of the signal is found in the order of 4 mV/Hz.

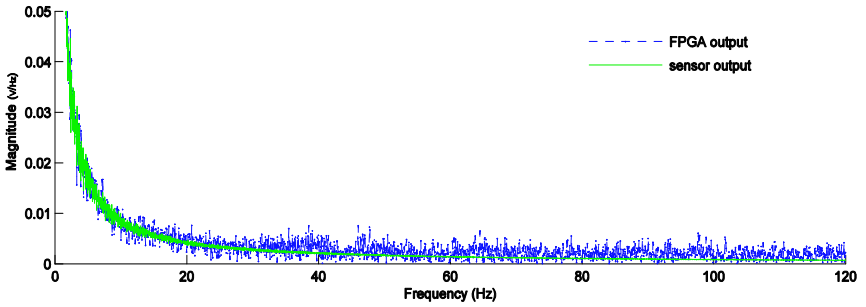


Fig. 10.8. FFT of sensor output and FPGA output in PWL.

10.3.2. Linearization by Interpolation (LI)

Interpolation is a method to estimate unknown values by interpolating known data. In piecewise linear interpolation method, a set of known data points of the nonlinear sensor characteristic is considered. Any two subsequent points of the known data form a straight line in which the unknown data points can be interpolated. The number of known data points which are equally separated from the subsequent one is dependent on the accuracy of the linearization required and the length of operating range. Let us consider two known point a($V_{o,1}$; $T_{i,1}$) and b($V_{o,2}$; $T_{i,2}$) in the nonlinear thermistor characteristics shown in Fig. 10.9. For any value of V_o within the range [$V_{o,1}$; $V_{o,2}$] the unknown value T_i can be determined or interpolated by using the straight line equation given as

$$T_i = T_{i,1} + \left(T_{i,2} - T_{i,1}\right)\frac{(V_o - V_{o,1})}{(V_{o,2} - V_{o,1})} \tag{10.4}$$

The concatenation of linear interpolants between each pair of data points of the data set ($V_{o,1}$; $T_{i,1}$); ($V_{o,2}$; $T_{i,2}$), … ($V_{o,n}$; $T_{i,n}$) form a continuous curve as shown in Fig. 10.9.

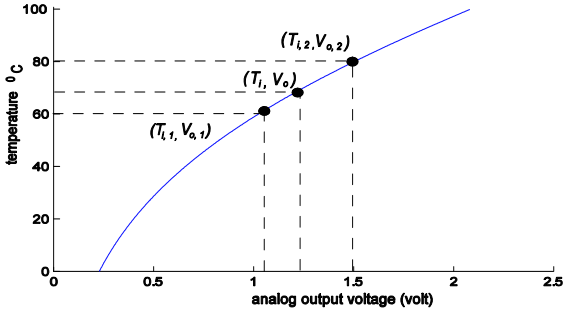


Fig. 10.9. Interpolation technique.

Initially the interpolation technique for linearization of the nonlinear thermistor is simulated in MATLAB. The data points and corresponding equations are shown in Table 10.4. In this work we have implemented the interpolation method in FPGA for 20, 30 and 40 sets of data points.

Table 10.4. Interpolation of Thermistor Characteristics for the Temperature Range of 0-100 ℃ with 20 Data Points.

Data Points	Linearized Temperature Equation
$(0.1585,0) - (0.1989,5)$	$T_i = 123.852V_o - 19.631$
$(0.1989,5) - (0.247,10)$	$T_i = 103.8084V_o - 15.6448$
$(0.247,10) - (0.3039,15)$	$T_i = 87.9795V_o - 11.7344$
$(0.3039,15) - (0.3702,20)$	$T_i = 75.4031V_o - 7.9128$
$(0.3702,20) - (0.4467,25)$	$T_i = 65.3598V_o - 4.1949$
$(0.4467,25) - (0.5339,30)$	$T_i = 57.3075V_o - 0.5982$
$(0.5339,30) - (0.6323,35)$	$T_i = 50.835V_o + 2.8577$
$(0.6323,35) - (0.7419,40)$	$T_i = 45.6279V_o + 6.1501$
$(0.7419,40) - (0.8625,45)$	$T_i = 41.445V_o + 9.2532$
$(0.8625,45) - (0.9937,50)$	$T_i = 38.1001V_o + 12.1383$
$(0.9937,50) - (1.1348,55)$	$T_i = 35.4491V_o + 14.7727$
$(1.1348,55) - (1.2846,60)$	$T_i = 33.33806V_o + 17.12$
$(1.2846,60) - (1.4418,65)$	$T_i = 31.8081V_o + 19.1399$
$(1.4418,65) - (1.6048,70)$	$T_i = 30.6654V_o + 20.7875$
$(1.6048,70) - (1.772,75)$	$T_i = 29.9016V_o + 22.0133$
$(1.772,75) - (1.9417,80)$	$T_i = 29.4786V_o + 22.7628$
$(1.9417,80) - (2.1119,85)$	$T_i = 29.3687V_o + 22.9762$
$(2.1119,85) - (2.2811,90)$	$T_i = 29.5523V_o + 22.5884$
$(2.2811,90) - (2.4477,95)$	$T_i = 30.017V_o + 21.5284$
$(2.4477,95) - (2.6102,100)$	$T_i = 30.7563V_o + 19.719$

10.3.2.1. FPGA Implementation of LI

The linear interpolation methodology is represented in Fig. 10.10. Here the range selector unit generates $T_{i,m}$, $T_{i,m+1}$, $V_{o,m}$ and $V_{o,m+1}$ for the corresponding value of V_o.

For any voltage, e.g. 0.2 volt, which is in the range of the data points (0.1989; 5) to (0.2470; 10); the corresponding temperature can be determined by interpolation as-

$$T_i = 5 + (10 - 5)\frac{(0.2 - 0.1989)}{(0.2470 - 0.1989)} = 5.118\ ℃ \tag{10.5}$$

Similar to the data representation as in PWL, we have used $S4.8$ bit format to represent input (V_o) voltage and $S8.8$ bit format to represent output (T_i) and intermediate data to implement LI in FPGA.

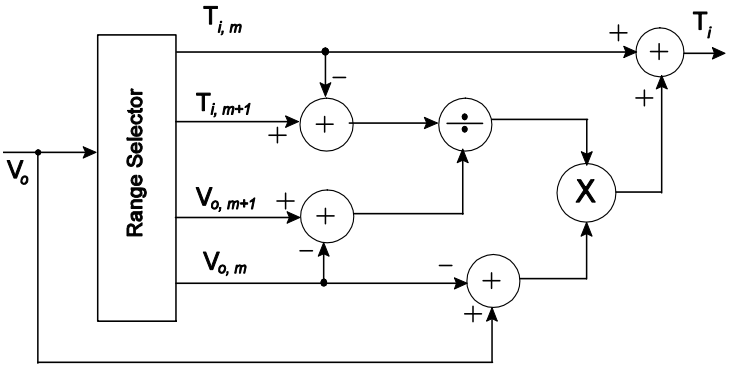


Fig. 10.10. Linearization by interpolation in FPGA.

10.3.2.2. Results of LI

The FS error in linearization by interpolation with $n = 40$ is shown in Fig. 10.11.

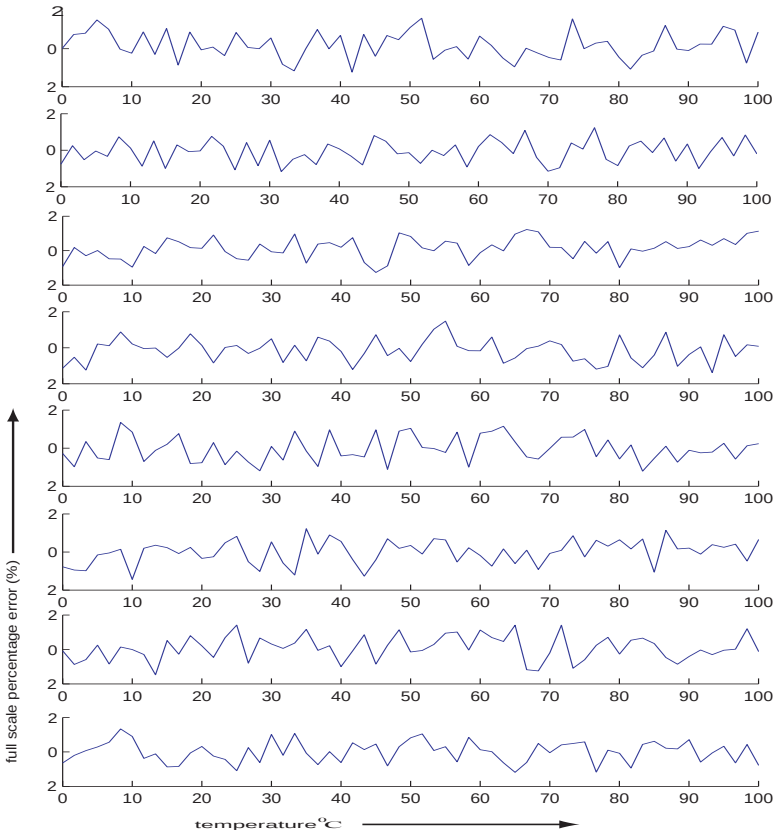


Fig. 10.11. Full scale percentage error in linearization using interpolation with 40 segments for 8 channels from 0 ℃ to 100 ℃.

Table 10.5 shows RMSE, MAD, SD, Maximum full scale error for all 8 channels with 40 segments. Table 10.6 shows the comparison of RMSE, MAD and SD respectively for different number of data points. A maximum FS error of 2.6329 % is observed with 20 data points while it is found to be 0.6237 % for 40 data points. A RMSE of 1.1793 is observed for 20 data points, while it is 0.6238 for 40 data points. Thus a 47.1 % reduction in RMSE is observed by increasing the number of data points from 20 to 40.

Table 10.5. Error analysis of LI in FPGA for 8 channels with $n = 40$.

Channel	RMSE	MAD	SD	Maximum FS error (%)
1	0.7040	0.5755	0.7040	1.6308
2	0.5780	0.4758	0.5784	1.3946
3	0.5959	0.4833	0.5963	1.4736
4	0.6321	0.5165	0.6326	1.4750
5	0.6444	0.5245	0.6441	1.5671
6	0.6696	0.5443	0.6702	1.4817
7	0.7095	0.5788	0.7084	1.5907
8	0.5788	0.4632	0.5781	1.4335

Table 10.6. Comparative error analysis of LI in FPGA for channel-1 with different Number of data points.

Number of segments	20	30	40
RMSE	1.1793	0.8262	0.6238
MAD	0.9621	0.679	0.503
SD	1.1784	0.8268	0.6237
Maximum FS error (%)	2.6329	2.0191	1.5027
No. of 4 input LUTs (out of 7168)	290 (4 %)	433 (6 %)	556 (7 %)

It is also observed that for 20 data points, FPGA device uses 290 elements of 4-input LUTs; while it is 433 and 566 for 30 and 40 data points respectively, i.e. 4 %, 6 % and 8 % for 20, 30 and 40 data points respectively. An improvement in accuracy can be observed when data points are increased from 20 to 40. Also, the LUT address size increases from 9 bits to 10 bits as the number of data points increases from 20 to 40.

The FFT analysis in Fig. 10.12 shows that the FPGA system has incorporated with higher spectral leakage however the white noise level is same as that of PWL method.

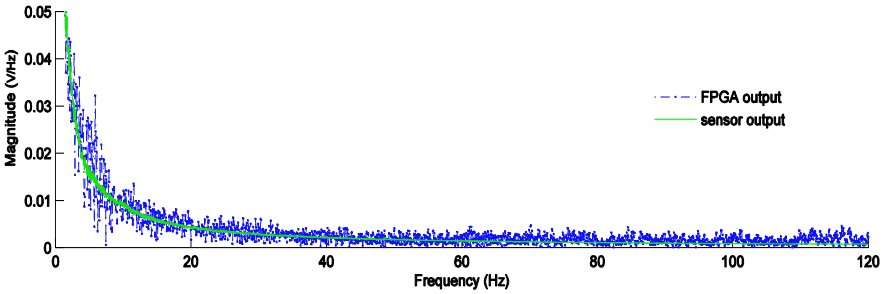


Fig. 10.12. FFT of sensor output and FPGA output in LI.

10.3.3. Look up Table (LUT) Based Linearization

A look up table (LUT) provides an interpretation between input-output values for any nonlinear system. They are used to realize a wide variety of nonlinear functions [9]. Thus a LUT can be used for mapping the input-output characteristics of the thermistor. Here, the outputs of the voltage divider circuit and the corresponding temperatures are stored in memory in a tabular form. The output of the ADC is applied as input to the linearization unit online, that addresses the corresponding linearized temperature.

10.3.3.1. FPGA Implementation of LUT Based Linearization

LUTs can be implemented in simplest controllers, but it requires much higher program memory. Typically f^{-1} is implemented in FPGA using the block RAMs of the device; however add-on memory devices like flash memory can be used to implement complex nonlinear functions with higher degree of nonlinearity and with higher precision. In this case y in Fig. 10.2 is 8 bit data from ADC and used as pointer to the memory locations. The FPGA architecture for implementation of a LUT linearization can be realized as shown in Fig. 10.13. The 8 bit ADC output which can address a maximum of 256 (2^8) data is used as pointer to the memory locations of LUT where $f^{-1}(y)$ is stored.

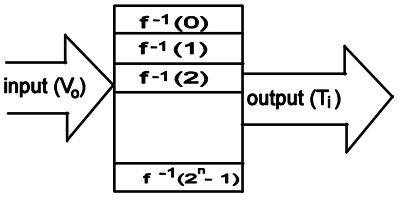


Fig. 10.13. Look up table based linearization.

10.3.3.2. Data Representation

In this method, the 8-bit ADC outputs (V_o) address the LUT data in FPGA. The LUT output data (T_i) are represented using $S8.8$ fixed point format. For a maximum

temperature range of 100 °C, the 8 bit fractional part provides a resolution of 2^{-8} i.e. 3.9×10^{-3} °C, while the microcontroller based LUT in [9] would provide only a resolution of 100×2^{-3} °C, i.e. 3 °C. In this work we have used 128 and 256 LUT data for the temperature range of $0-100$ ℃.

10.3.3.3. Results of LUT Based Linearization

The errors in linearization for different numbers of LUT data elements are shown in Fig. 10.14 and Fig. 10.15. It is observed that in case of 128 data elements maximum FS error is 1.5214 %, while it is 1.1427 % in case 256 LUT data elements. Similarly RMSE, MAD and SD are shown in Table 10.7. The results indicates a 27 % reduction in RMSE when the number of LUT data elements are increased from 128 to 256 and this increase in LUT data elements consumes 14 % to 35 % more logic area. It may be mentioned that unlike PWL and LI, in LUT based technique, linearization units were implemented for only 3 and 2 channels with 128 and 256 LUT data respectively in Spartan-III XC3S400 FPGA. However, number of channels can be increased by using high capacity FPGAs.

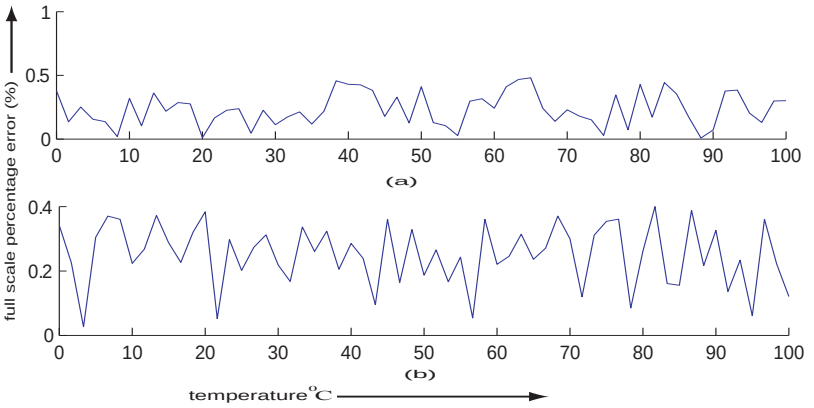


Fig. 10.14. Full scale percentage error in LUT based linearization for (a) 128, and (b) 256 LUT data from 0 ℃ to 100 ℃.

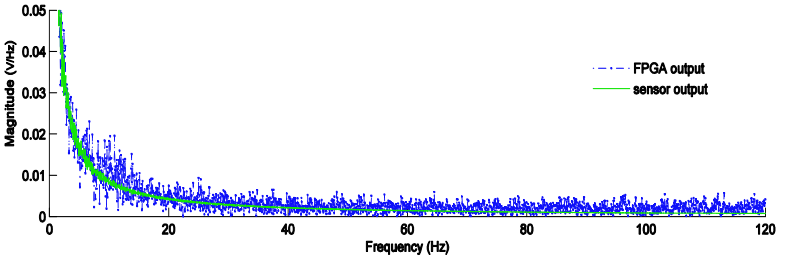


Fig. 10.15. FFT of sensor output and FPGA output in LUT.

Table 10.7. Comparative error analysis of LUT based Linearization in FPGA for channel-1 with different LUT sizes.

Number of LUT DATA	128	256
RMSE	0.6477	0.4692
MAD	0.3172	0.2311
SD	0.3794	0.2742
Maximum FS error (%)	1.5214	1.2427
No. of 4 input LUTs (out of 7168)	975 (14 %)	2500 (35 %)

The FFT of the sensor output and FPGA output of LUT based linearization are shown in Fig. 10.16. In this case the spectral density of noise is found to be higher than that of PWL and LI.

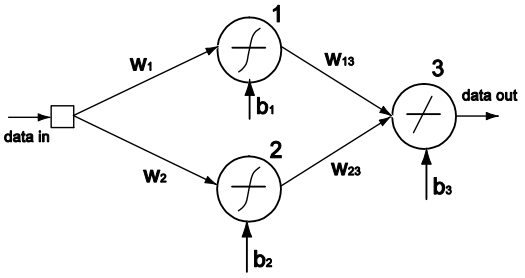


Fig. 10.16. An artificial neural network for sensor linearization.

10.3.4. ANN Based Linearization

There are different architectures of ANN for mapping nonlinear input-output functions. Based on the degree of nonlinearity of the original function, there may be different numbers of layers or different numbers of neurons in hidden layer. N. J. Medrano-Marques and B. Martin-del-Brio in [19] proposed an optimized architecture of ANN for linearization of thermistor. The network is a MLP neural network consisting only two neurons in hidden layer and one neuron in output layer as shown in Fig. 10.16. They have simulated the network for implementation in low cost processors like microcontrollers. In this work, an array of 4 such ANN structures has been implemented in FPGA for the 4 channels of sensor signals. We have taken the activation function of the hidden layer neurons as a hyper tangent activation function given by-

$$f(x) = \tanh(x) = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}}$$

where the output neuron has linear activation function. The network is initially trained in MATLAB using input-output data set given in Equation (10.3). The weight and bias of

the ANN are determined during training by simulating the network in MATLAB. The training parameters obtained are shown in Table 10.8.

Table 10.8. MatLab simulation results of ANN for Linearization channel-1.

Network Type	MLP Feed Forward Network			
Activation Function	Neuron1		Neuron2	Neuron3
	$f(x) = tanh(x)$		$f(x) = tanh(x)$	$f(x) = x$
Training Algorithm	Back Propagation			
Training data size	600			
Learning rate	2.0			
Goal	10^{-6}			
Epochs	1000			
Weights	$W1$	$W2$	$W13$	$W23$
	-0.026	1.284	- 87.348	159.44
Bias	$b1$	$b2$	$b3$	
	1.195	2.733	-85.909	

10.3.4.1. FPGA Implementation of a Neuron

The processing element of an ANN is the neuron which can be expressed by an equation of the form

$$y_i = \sum (w_i x_i + b_i)$$

where x, w and b represent input, weight and bias respectively. Fig. 10.17 shows the FPGA implementation of a neuron which includes, multiplier, adder and activation function.

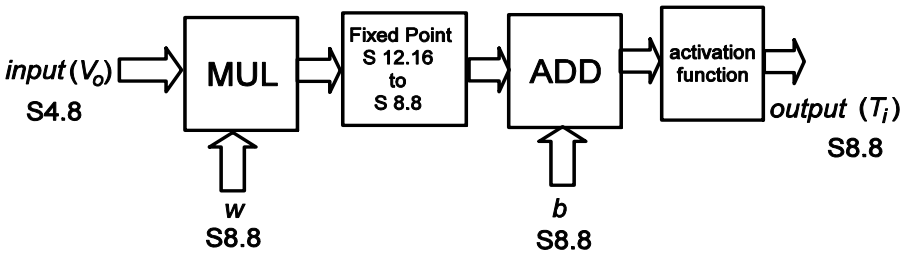


Fig. 10.17. Block Diagram of FPGA implementation of a neuron.

10.3.4.2. Implementation of Activation Function in FPGA

Efficient implementation of the nonlinear activation function on an FPGA is a difficult task faced by designers as it is not suitable for direct implementation [35, 36]. In most cases computationally simplified alternatives of nonlinear activation function are used [37, 38]. In this work we have used Piece- Wise Linear Approximation (PWL) method for implementation of the $tanh(x)$ activation function for neurons 1 and 2. The hyper tangent curve has been segmented into 10 linear segments as shown in Fig. 10.18. The piecewise linear equations for each segment are shown in Table 10.9. The Number of linear pieces can be increased to achieve the precision of the activation function at the cost of memory and computational complexity [39].

In ANN based linearization in FPGA input voltage (V_o) is represented using S4.8 bit fixed point format, while temperature (T_i), weight (w) and bias (b) are represented using S8.8 bit fixed point format.

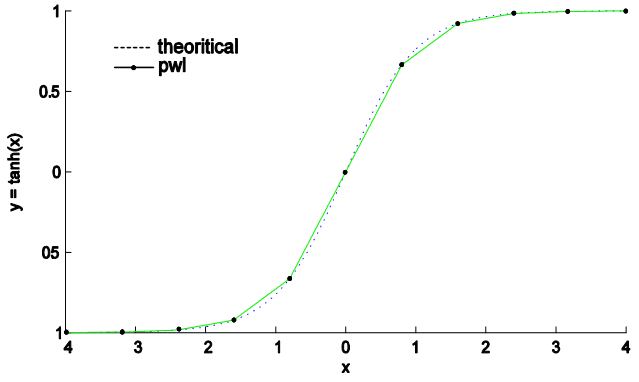


Fig. 10.18. Piecewise linear approximation of $y = tanh(x)$ function.

Table 10.9. Approximation of $y = tanh(x)$ function in FPGA.

Range	Equation
[-5,-4.1]	$y = 0.000x - 0.997$
[-4,-3.1]	$y = 0.003x - 0.985$
[-3,-2.1]	$y = 0.026x - 0.918$
[-2,-1.1]	$y = 0.177x - 0.626$
[-1,-0.1]	$y = 0.739x - 0.064$
[0,1]	$y = 0.768x + 0.044$
[1,2]	$y = 0.177x + 0.626$
[2,3]	$y = 0.026x + 0.918$
[3,4]	$y = 0.003x + 0.985$
[4,5]	$y = 0.000x + 0.997$

10.3.4.3. Results of FPGA Implementation of ANN for Linearization

ANNs have been implemented in FPGA for linearization of the nonlinear thermistors for the range 0 ℃ to 100 ℃ as discussed above. The networks implemented here are a MLP networks in which 600 data sets are used to train each of the networks with back propagation algorithm. The result in Fig. 10.19 shows that ANN can perform linearization with better accuracy as compared to the other methods discussed above. As shown in Table 10.10, a maximum FS error of 1.241 % is observed in ANN based linearization with an average logic utilization of 19 % per channel and 89 % for 4 channels. The FFT of the sensor output and FPGA output in Fig. 10.20 of ANN based linearization shows that the noise voltage fluctuation is less severe than that of LUT based linearization and similar to LI, however higher than that of PWL based linearization.

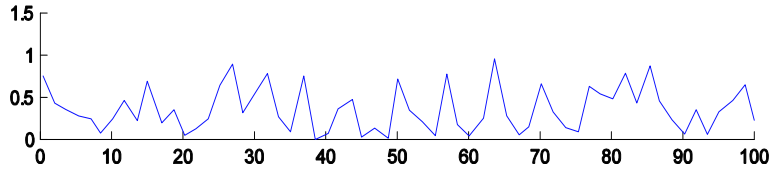


Fig. 10.19. Full scale percentage error of ANN based linearization for 4 channels from 0 ℃.

Table 10.10. Error analysis of ANN based linearization in FPGA for 4 channels.

Channel	RMSE	MAD	SD	Maximum FS error (%)
1	0.4805	0.2376	0.2857	1.1901
2	0.4549	0.2195	0.2707	1.1981
3	0.4958	0.2357	0.2811	1.1914
4	0.4751	0.224	0.2709	1.2410

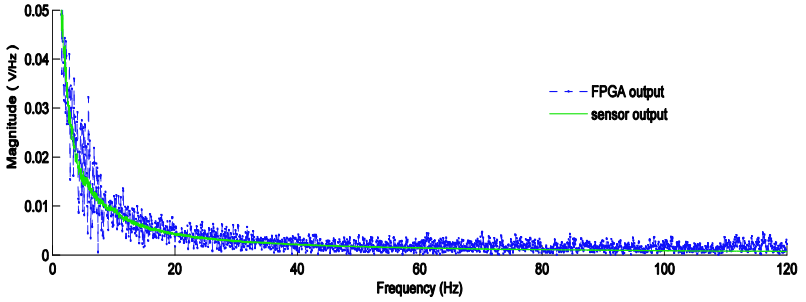


Fig. 10.20. FFT of sensor output and FPGA output in ANN based linearization.

The dynamic response of the linearization unit has also been evaluated. Since temperature variation in a thermistor itself is a slow process, therefore to test the speed of operation of the FPGA system, a fast varying analog input has been applied to the ADC and the FPGA

processed output in LUT based linearization through the DAC has been recorded as shown in Fig. 10.21. With an ADC conversion time of 100 µs per channel (for a clock frequency of 200 kHz), DAC settling time of about 150 ns and FPGA execution time of 90 ns, the maximum speed of linearization per channel was estimated as 100.24 µs (100 µs + 150 ns + 90 ns). For 8 channels this figure will be 801.29 µs (100 µs×8 + 150 ns×8 + 90 ns). This gives a maximum frequency of 9.976 kHz for a single channel while 1.248 kHz for 8 channels.

A linearity plot for the four types of linearization methods is also shown in Fig. 10.22 along with the nonlinear thermistor response.

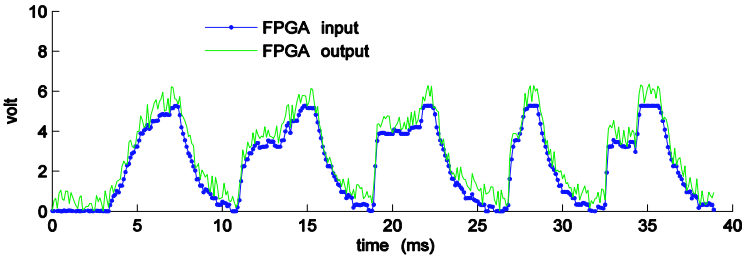


Fig. 10.21. Signals showing fast variation of an analog signal and corresponding FPGA based linearized (LUT) output.

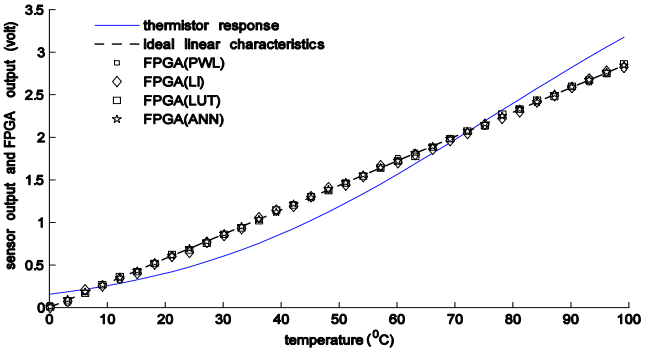


Fig. 10.22. Linearity plot for all 4 methods of linearization in FPGA.

10.4. Conclusion

In this work, implementation of four different techniques of thermistor linearization in FPGA for simultaneous 8-channel sensor linearization has been presented. The performances of the techniques have been tested using real time signals from thermistors using LabView data acquisition system. The performance of each of the linearization scheme has been estimated considering mainly accuracy, speed of operation and cost of implementation. In the result sections of each implementation, full scale percentage error, RMSE, MAD and SD were presented by comparing with different numbers of data sets.

From the comparative result of all the four methods of linearization shown in Table 10.11; it is observed that ANN is showing better accuracy, while LUTs consumes highest logic area of FPGA device. The piecewise linearization and interpolation consumes almost same logic area in FPGA, but in case of linear interpolation the execution time of the linearization block is higher than that of PWL. While in case of ANN based implementation, the logic area utilization is almost 19 % as compared to 7 % in case of PWL and LI. It can be estimated that for same amount of full scale error of 1.19 % in ANN; PWL and LI would utilize logic area of about 8 % and LUT based linearization would utilize 36 %. Thus PWL and LI are better option for this case. Further, the implementation complexity is higher in case of ANN and the accuracy of the method is dependent on several steps of implementations - e.g. accuracy of activation function. However cost of implementation is less in case of ANN as compared to LUT. The FPGA implementation using LUT is the simplest method to map the nonlinear thermistor characteristics in comparison to the other three methods. From the point of execution time PWL, LI, LUT outperforms ANN based linearization technique. On the other hand, in spite of lower performance functions, ANN can be improved considerably due to its learning capability for higher degree of nonlinearity.

Table 10.11. Comparison of four methods of linearization in FPGA.

	PWL	**LI**	**LUT**	**ANN**
RMSE	0.6238	0.6238	0.4692	0.4805
MAD	0.503	0.503	0.2311	0.2376
SD	0.6237	0.6237	0.2742	0.2857
Execution Time	120 ns	125 ns	90 ns	140 ns
FPGA LUTs used (%)	7 %	7 %	35 %	19 %
FS error (max)	1.5 %	1.5 %	1.24 %	1.19 %

The FPGA based linearization system has been tested with fast varying analog inputs to illustrate that the FPGA based linearization can handle the fast varying signals also. It is estimated that the system can linearize analog signals with a maximum frequency of 9.976 kHz for a single channel while 1.248 kHz for 8 channel operation.

The FPGA based linearization techniques are suitable for systems with multiple sensor inputs as FPGA devices are suitable for multi-channel operation with parallel architecture with high speed signal processing capability.

Acknowledgements

The authors would like to thank Tezpur University, Assam, India for partially supporting the work under start up grant scheme.

References

[1]. M. Bhuyan, Intelligent Instrumentation: Principles and Applications, *CRC Press*, 2010.

[2]. G. Iglesias and E. Iglesias, Linearization of transducer signals using an analog-to-digital converter, *IEEE Transactions on Instrumentation and Measurement,* Vol. 37, No. 1, 1988, pp. 53–57.

[3]. J. Lygouras, Nonlinear ADC with digitally selectable quantizing characteristic*, IEEE Transactions on Nuclear Science,* Vol. 35, No. 5, 1988, pp. 1088–1091.

[4]. D. Anvekar and B. S. Sonde, A programmable nonlinear ADC using optimal-sized rom, *IEEE Transactions on Instrumentation and Measurement*, Vol. 40, No. 6, 1991, pp. 1031–1035.

[5]. E. Volpi, N. Nizza, and P. Bruschi, A non linear ADC for sensor linearization, in *Proceedings of the Research in Microelectronics and Electronics Conference* (*PRIME 2007),* 2007, pp. 5–8.

[6]. S. Kaliyugavaradan, P. Sankaran, and V. Murti, A new compensation scheme for thermistors and its implementation for response linearization over a wide temperature range, *IEEE Transactions on Instrumentation and Measurement,* Vol. 42, No. 5, 1993, pp. 952–956.

[7]. Z. Nenova and T. Nenov, Linearization circuit of the thermistor connection, *IEEE Transactions on Instrumentation and Measurement,* Vol. 58, No. 2, 2009, pp. 441–449.

[8]. J. Patra, G. Panda, and R. Baliarsingh, Artificial neural network based nonlinearity estimation of pressure sensors, *IEEE Transactions on Instrumentation and Measurement,* Vol. 43, No. 6, 1994, pp. 874–881.

[9]. L. E. Bengtsson, Lookup table optimization for sensor linearization in small embedded systems, *Journal of Sensor Technology*, Vol. 2, 2012, pp. 177–184.

[10]. N. Cotton and B. Wilamowski, Compensation of nonlinearities using neural networks implemented on inexpensive microcontrollers, *IEEE Transactions on Industrial Electronics,* Vol. 58, No. 3, 2011, pp. 733–740.

[11]. N. Khachab and M. Ismail, Linearization techniques for nth-order sensor models in MOS VLSI technology, *IEEE Transactions on* Circuits and Systems, Vol. 38, No. 12, 1991, pp. 1439–1450.

[12]. J. Patra, G. Chakraborty, and P. Meher, Neural-network-based robust linearization and compensation technique for sensors under nonlinear environmental influences, *IEEE Transactions on Circuits and Systems I: Regular Papers,* Vol. 55, No. 5, 2008, pp. 1316–1327.

[13]. H. Erdem, Implementation of software-based sensor linearization algorithms on low-cost microcontrollers, *ISA Transactions*, Vol. 49, No. 4, 2010, pp. 552 – 558.

[14]. D. Sonowal and M. Bhuyan, FPGA implementation of neural network for linearization of thermistor characteristics, in *Proceedings of the International Conference on Devices, Circuits and Systems (ICDCS'12),* March 2012, pp. 422–426.

[15]. D. Sonowal and M. Bhuyan, Linearizing thermistor characteristics by piecewise linear interpolation in real time FPGA, in *Proceedings of the International Conference on Advances in Computing, Communications and Informatics (ICACCI'13),* Aug 2013, pp. 1976–1980.

[16]. D. Patranabis and D. Ghosh, A novel software-based transducer linearizer, *IEEE Transactions on Instrumentation and Measurement,* Vol. 38, No. 6, 1989, pp. 1122–1126.

[17]. J. Sturcel and M. Kamensky, Function approximation and digital linearization in sensor system, *AT & P Journal*, 2006, pp. 13–17.

[18]. M. Attari, F. Boudjema, and M. Heniche, Linearizing a thermistor characteristic in the range of zero to 100 degree c with two layer artificial neural networks, in *Proceedings of the IEEE Instrumentation and Measurement Technology Conference on Integrating Intelligent Instrumentation and Control (IMTC'95)*, 1995, pp. 119–122.

[19]. N. Medrano-Marques and M. del Brio, A general method for sensor linearization based on neural networks, in *Proceedings of the IEEE International Symposium on Circuits and Systems (ISCAS'2000),* Geneva, Vol. 2, 2000, pp. 497–500.

[20]. N. Medrano-Marques and B. Martin-del Brio, Sensor linearization with neural networks, *IEEE Transactions on Industrial Electronics,* Vol. 48, No. 6, 2001, pp. 1288–1290.

[21]. S. A. Khan, D. Shahani, and A. Agarwala, Sensor calibration and compensation using artificial neural network, *ISA Transactions*, Vol. 2, No. 3, 2003, pp. 337 – 352.

[22]. X. Wei, Sensor temperature compensation technique simulation based on BP neural network, *TELKOMNIKA*, Vol. 11, No. 6, 2013, pp. 3304–3313.

[23]. D. K. Ghara, D. Saha and K. Sengupta, Implementation of linear trace moisture sensor by nano porous thin film moisture sensor and NLAmp, *International Journal on Smart Sensing and Intelligent Systems*, Vol. 1, No. 4, December 2008, pp. 955-968.

[24]. A. Muthuramalingam, S. Himavathi, and E. Srinivasan, Neural network implementation using FPGA: Issues and application, *International Journal of Information Technology*, Vol. 4, No. 2, 2008, pp. 86–92.

[25]. N. Botros and M. Abdul-Aziz, Hardware implementation of an artificial neural network using field programmable gate arrays (FPGA's), *IEEE Transactions on Industrial Electronics,* Vol. 41, No. 6, 1994, pp. 665–667.

[26]. N. Botros and M. Abdul-Aziz, Hardware implementation of an artificial neural network, in *Proceedings of the IEEE International Conference on Neural Networks,* 1993, Vol. 3, pp. 1252–1257.

[27]. X. Yu and D. Dent, Implementing neural networks in FPGAs, in *Proceedings of the EE Colloquium on Hardware Implementation of Neural Networks and Fuzzy Logic,* 1994, pp. 1/1–1/5.

[28]. J. Zhu, G. J. Milne, and B. K. Gunther, Towards an FPGA based reconfigurable computing environment for neural network implementations, in *Proceedings of the 9th International Conference on Artificial Neural Networks (ICANN'99),* Vol. 2, 1999, pp. 661–666, (Conf. Publ. No. 470).

[29]. Y. Chen and W. du Plessis, Neural network implementation on a FPGA, in *Proceedings of the 6th IEEE African Conference in Africa (AFRICON'2002)*, Vol. 1, 2002, pp. 337–342.

[30]. J. M. Granado, M. A. Vega, R. Perez, J. Sanchez, and J. Gomez, Using FPGAs to implement artificial neural networks, in *Proceedings of the 13th IEEE International Conference on Electronics, Circuits and Systems (ICECS'06)*, 2006, pp. 934–937.

[31]. J. Liu and D. Liang, A survey of FPGA-based hardware implementation of ANNs, in *Proceedings of the International Conference on Neural Networks and Brain (ICNN B '05)*, Vol. 2, 2005, pp. 915–918.

[32]. S. Himavathi, D. Anitha, and A. Muthuramalingam, Feedforward neural network implementation in FPGA using layer multiplexing for effective resource utilization, *IEEE Transactions on Neural Networks,* Vol. 18, No. 3, 2007, pp. 880–888.

[33]. E. Monmasson, L. Idkhajine, M. Cirstea, I. Bahri, A. Tisan, and M.-w. Naouar, FPGAs in industrial control applications, *IEEE Transactions on Industrial Informatics,* , Vol. 7, No. 2, pp. 224–243, 2011.

[34]. A. Gomperts, A. Ukil, and F. Zurfluh, Development and implementation of parameterized FPGA-based general purpose neural networks for online applications, *IEEE Transactions on Industrial Informatics,* Vol. 7, No. 1, 2011, pp. 78–89.

[35]. M. Tommiska, Efficient digital implementation of the sigmoid function for reprogrammable logic, *IEE Proceedings - Computers and Digital Techniques,* Vol. 150, No. 6, 2003, pp. 403–411.

[36]. K. Basterretxea, J. Tarela, and I. Del Campo, Approximation of sigmoid function and the derivative for hardware implementation of artificial neurons, *IEE Proceedings - Circuits, Devices and Systems,* Vol. 151, No. 1, 2004, pp. 18–24.

[37]. A. Mishra, Z. Zaheeruddin, and K. Raj, Implementation of a digital neuron with nonlinear activation function using piecewise linear approximation technique, in *Proceedings of the International Conference on Microelectronics (ICM'07)*, 2007, pp. 69–72.

[38]. Tisan, S. Oniga, D. Mic, and A. Buchman, Digital implementation of the sigmoid function for FPGA circuits, *Electronics and Telecommunication, Acta Technica Napocensis*, Vol. 50, No. 2, 2009, pp. 15–20.

[39]. Z. Xie, A non-linear approximation of the sigmoid function based on FPGA, in *Proceedings of the 5th IEEE International Conference on Advanced Computational Intelligence (ICACI'12),* 2012, pp. 221–223.

Chapter 11
Low Value Capacitance Measurements for Capacitive Sensors: A Review

Prabhu Ramanathan, Sudha Ramasamy, Prateek Jain, Hardik Nagrecha, Saurav Paul, P. Arulmozhivarman, Rao Tatavarti

Abstract

The novel results in the field of measurements of very small capacitance changes are presented. The operating principles like double differential principle, RC phase delay, charging and discharging, oscillator, capacitance to phase angle conversion used for measuring very small capacitance changes are discussed. The basic and most important parameters of measurements are dealt with measurement range, sensitivity, percentage error and applications where it can be used. The aim is to provide an elaborative summary on various methods and their parameters available for capacitance measurement.

Abbreviations

ADC- Analog to Digital Converter
PMOS- p-channel Metal Oxide Semiconductor
NMOS- n-channel Metal Oxide Semiconductor
TSMC- Taiwan Semiconductor Manufacturing Company
CMOS- Complementary Metal Oxide Semiconductor
ASIC- Application Specific Integrated Circuit
PSD- Phase Sensitive Detector
FSR - Full Scale Reading
RC - Resistance Capacitance Pair
IC - Integrated Circuits
DC - Direct Current
pF - Pico Farads
PWM - Pulse Width Modulation

Prabhu Ramanathan
School of Electrical Engineering, VIT University, Vellore, India

CMRR - Common Mode Rejection Ratio
Voh - Output high Voltage
PCB - Printed Circuit Board

11.1. Introduction

Capacitance measurement is one of the most important parameters in industrial applications. From measurement point of view, change in dielectric, area, distance between plates of capacitor is used as a factor to measure various material properties. Capacitive sensors are constructed in such a way that the parameter to be measure causes a change in the capacitance of capacitive sensors. The demand for capacitive sensors in industrial applications is on the rise. Interfacing capacitive sensors is demanding and useful concept for real time applications. Interfacing is required for implementing the circuit with real time world. The raw data from the sensor is converted to the corresponding electrical signal and then interfacing is done for further processing of data. Various interfacing techniques have also been reported to measure very small capacitance change from the sensor and process it to the computer or microcontroller. Accurate measurement of very small capacitance change has always been a problem. A very small change in capacitance can easily alter the output. This very small change has to be considered for industrial and real time applications. The high degree of linearity and accuracy is always desired. High resolution measurement always requires a very precise circuit design and careful layout of the entire system. To achieve high degree of resolution, several methods such as double differential principle [1, 2], RC phase delay [3], charging and discharging [4], switched oscillator [5], capacitance to phase conversion [6] have been reported. These techniques used have few limitations but can work and provide excellent results in their respective ranges. These techniques used to measure various parameters like mass measuring system, sugar level content, angular speed sensing, liquid level sensing, pressure sensing and position sensing are just a few of these. For these applications capacitive sensors have become more popular. The principle adopted by various measurement techniques has been discussed. The papers reviewed are discussed one by one on various parameters like measurement range, sensitivity, percentage error and applications. A tabular column comparing the various parameters is provided to present a brief idea about the various designs and their limitations. Several other articles were also published in the domain of capacitive interfacing and sensing applications [7, 8]. Sergey Y. Yurish [8] in his article has elaborated the capacitance-to-digital converters and its state-of-the-art review along with direct capacitive sensors interfacing.

11.2. Methodology

11.2.1. Measuring Capacitance Using Double Differential Principle

D. M. G. Preethichandra *et al.,* in his article [1] presents a technique which is used to measure very small capacitance change in capacitive sensors. This technique uses a 24 bit ADC to provide a high resolution. With the help of double differential technique, the

capacitance measured is independent of the initial capacitance of the sensor. The technique also has an advantage of automatic cancellation of sensor-connecting parallel plate capacitances.

Fig. 11.1 shows a schematic diagram of the circuit. The excitation signal is given to sensor's capacitance and a reference capacitance. The reference capacitor is realized using a gyrator circuit. An active rectifier and low pass filter is used for each arm to minimize the error. The DC output voltage from this is fed to 24 bit Sigma-delta ADC. It converts the time varying output to an average DC value of the signal. The output of ADC is then stored in computer's memory. With the help of this ADC output, the change in capacitance is measured and is independent of the initial capacitance.

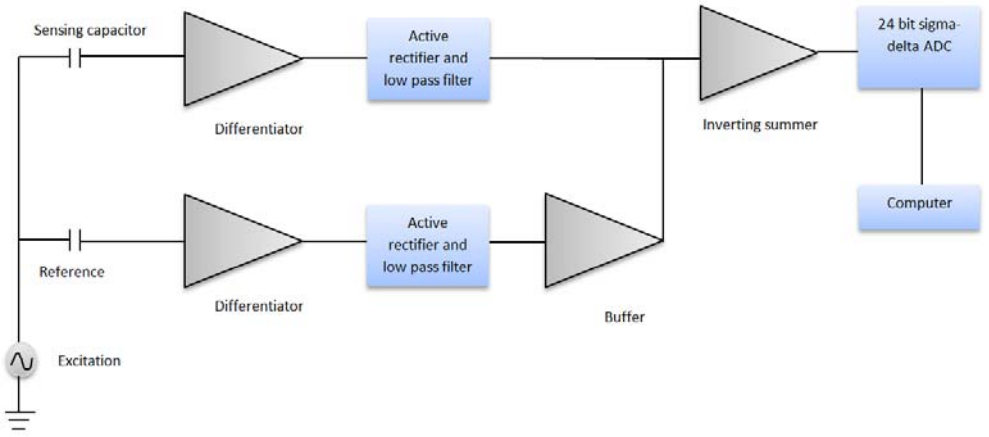


Fig. 11.1. Schematic Diagram for measuring capacitance using double differential principle.

For capacitive sensors of the range of Pico Farad, a high resolution change of output (ΔC) of order of 10-18 Farads is observed. The output observed is independent of the initial capacitance and is very precise and accurate. This circuit has a very simple design, low cost and can be used for various capacitive sensors. This type of interfacing circuit is very useful where the initial capacitance is unknown. It can be used for various real time industrial applications like position sensing, angular speed sensing, liquid level sensing and pressure measurement.

11.2.2. Capacitance Measurement with High Resolution and High Linearity

M. R. Haider et al. [2] has proposed a low power capacitance circuit for measuring small capacitance with high resolution and high degree of linearity. This capacitive measurement technique could be used for bio medical applications. The differential structure of the circuit helps in eliminating even order distortions. The circuit uses two current sense amplifiers, two PMOS-diode rectifiers and an instrumentation amplifier. Fig. 11.2 shows a schematic diagram of the circuit. The circuit has a reference capacitor and a sensing capacitor. The excitation signal is applied such that one of their plates

becomes the common terminal for both the capacitors. The other plates are connected to current sense amplifiers. The output of current sense amplifier is fed to PMOS diode rectifier to convert it into DC signal. PMOS diode is used instead of NMOS diode because of its lower flickering noise. The DC signal is then fed to an instrumentation amplifier with adjustable gain. With the help of variable gain, the sensitivity can also be increased or decreased correspondingly. The prototype of the circuit was developed by fabrication using TSMC 0.35 µm bulk CMOS process. It has been designed and simulated using Cadence Spectre simulator. The technique usually uses a reference capacitor of pF range. The circuit is capable of producing a change of 1.32 mV for a change of 1 fF. The circuit can be useful for measurement of very small capacitances in the range of fF (10-15 Farads). Due to its high Common Mode Rejection Ratio (CMRR) it is very much suited for bio medical sensor applications.

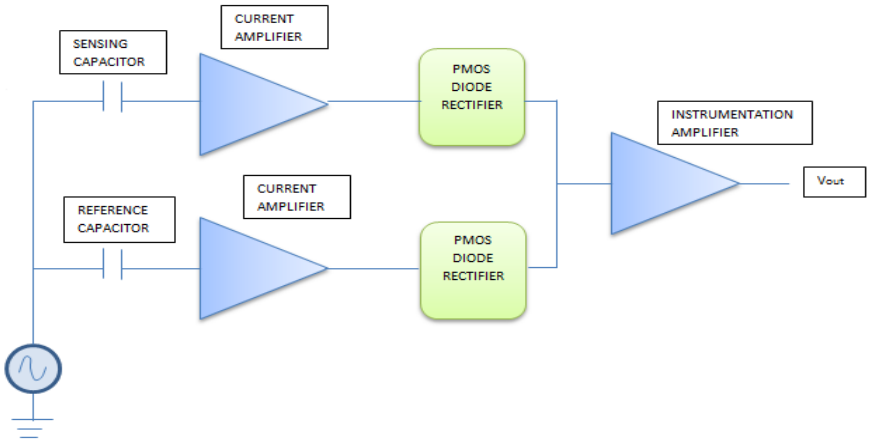


Fig. 11.2. Schematic Diagram for Sensor Read-out Circuit.

11.2.3. Measuring Capacitance Based on RC Phase Delay

Robert N. Dean et al., has presented a technique [3] using RC phase delay to measure capacitance change. Based on the phase delay through the RC network, there will be a change in state of input square wave to produce a pulse width modulated signal. Fig. 11.3 shows a schematic diagram of the circuit. When the input is low, the output remains low till the input goes high. When the input is high, the output goes high and remains high, i.e., duty cycle is proportional to unknown capacitance. The technique uses a MOSFET switch to avoid the problem of nonlinearity by discharging the unknown capacitor during each measurement cycle. The prototype circuit was implemented on a surface mounted printed circuit board using CMOS hex inverter IC and n-channel MOSFET. It uses a variable capacitor of 0.7-20 pF. The output obtained from this circuit is a pulse width modulated signal where the duty cycle is proportional to the change in capacitance. The prototype circuit provides a sensitivity of 246 mV/pF. The output PWM signal was low pass filtered, level shifted and amplified. The circuit is very useful for measurement of capacitance of the range of tens of pF or smaller. The output obtained is a linear response

with high sensitivity. The circuit uses CMOS electronics which can be easily implemented as an ASIC in chip. The measurement technique requires only a small number of components and is relatively simple.

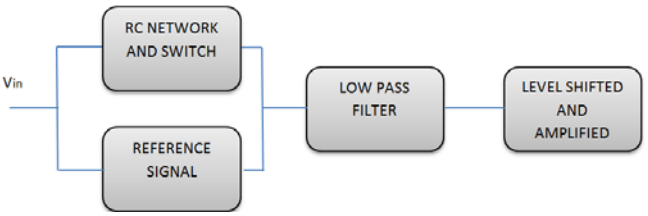


Fig. 11.3. Schematic Diagram for Capacitance measurement based on RC phase Delay.

11.2.4. Micro Controller Interface for Low value Capacitance Sensors

F. Reverter et al. has proposed a very simple and low cost microcontroller interface circuit [4] for low value capacitive sensors. The basic principle driving the measurement is that the microcontroller measures the time needed to discharge an unknown capacitance through a known resistance 'R', and this measured time 't' is utilized for determining the unknown capacitance. Fig. 11.4 shows a schematic diagram of the circuit.

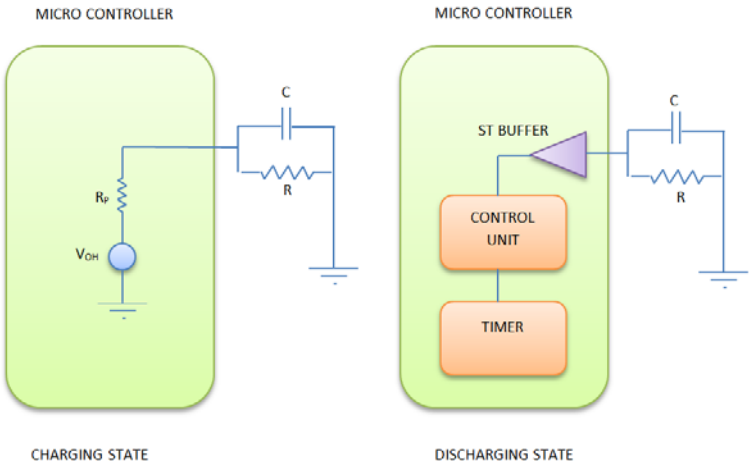


Fig. 11.4. Schematic of Measuring Capacitance with microcontroller based interfacing.

The technique used in the article to implement the same is that first a digital port pin of the microcontroller is set as output to provide a high logic, and hence charging the unknown capacitance 'C' between the port and ground to a voltage Voh (output high voltage) through the internal pull up resistance 'Rp' of the microcontroller. Secondly the same port pin is set to high impedance input mode, and the embedded timer starts timing the process as the capacitance discharges through 'R' (a resistance R>>Rp, included in parallel with 'C'). As soon as the voltage drops to the threshold voltage Vth of the

embedded Schmitt trigger buffer, the timer stops. Vth, Voh, R are considered constant and hence the measured discharge time‘t’ is utilized to calculate the value of the sensor’s unknown capacitance.

In order to minimize the effect of stray capacitance, a method of three signal calibration has been adopted. This technique requires the measurement of (a) offset measurement (b) sensor measurement and (c) reference measurement. The absolute error obtained by applying this interface circuit is below 4 % FSR for 1 pF < Cs (sensor capacitance) < 10 pF, when Cr = 10 pF (reference capacitance), and below 1.5 % for 10 pF < Cs < 100 pF, when Cr = 100 pF.

11.2.5. Measuring Capacitance Using Oscillator

Frank M. L. van der Goes et al. describes a new capacitive sensor interface based on a novel type of oscillator [5], whose frequency is insensitive to low and high frequency interfering signals by the application of a third order high pass filter. The fully integrated 0.7 µm CMOS circuit shows an inaccuracy of less than 100 aF (Attofarad) with respect to a 2 pF reference capacitor over a -30 ºC to 70 ºC temperature range.

Fig. 11.5 shows a schematic diagram of the circuit. The basic system is a first order system with a third order filter to suppress low frequency interference. It has low sensitivity to noise and the locking effects to the microcontroller have been reduced by application of dithertechniques. The two port measurement system consists of a network of sensor capacitance cursed with two parasitic capacitors. The only way to eliminate the effects of the two parasitic capacitors is to maintain a specific potential difference across the two terminals of the system and to transmit a specific current at the input. A microcontroller is used to measure the oscillator period. The entire system is continuously auto-calibrated for additive and multiplicative errors using a reference capacitor in place of sensor’s capacitance. The microcontroller is used to measure three time periods namely reference time, offset time and measured time for the specific sensor’s capacitance. These periods are measured by the microcontroller, yielding the digital numbers which in turn are used to calculate the final capacitance value.

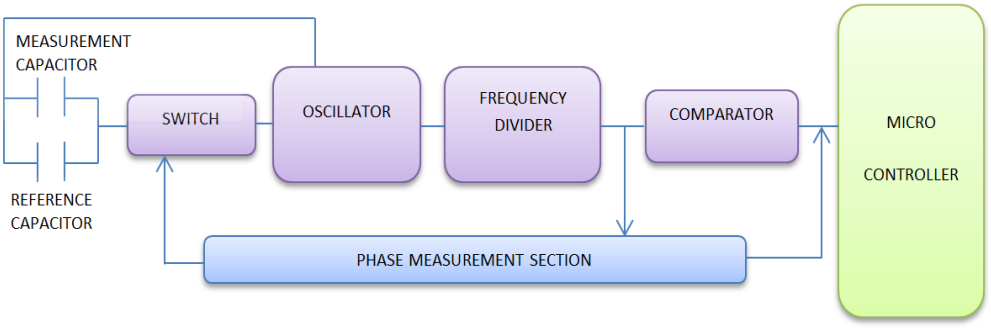


Fig. 11.5. Schematic Diagram for measuring capacitance using oscillator.

The output is independent of the gain and offset time as the calibration is done continuously and also slow variations in gain and offset time do not affect the output. The measurement range is 0 to 2 pF and the sampling time obtained is 330 ns. The result has negligible influence of parasitic capacitance of up to 300 pF. The system shows very low sensitivity to 1/f noise, thus enabling the use of low cost CMOS process.

11.2.6. Measuring Capacitance Based on Phase Angle

Ashkan Ashrafi et al., provides a technique of capacitance-to-phase angle conversion for measuring very small capacitance change [6]. There is a linear relationship between the output phase and initial capacitance change. The main features of this technique are extremely high stray immunity and very high resolution.

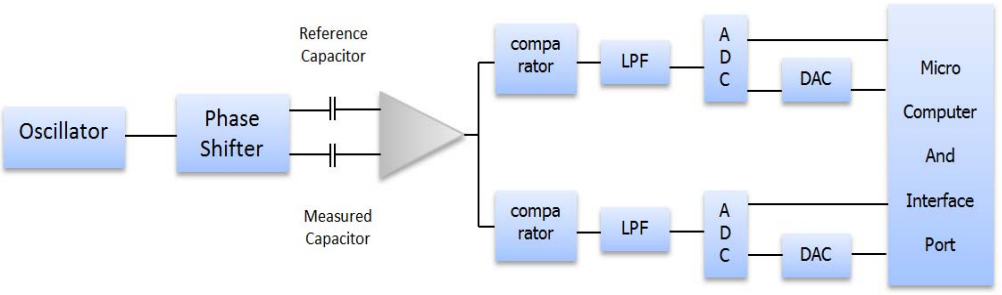


Fig. 11.6. Schematic diagram based on capacitance to phase angle converter for measuring capacitance.

A sinusoidal wave is generated using an oscillator and then this wave is used for generating two balanced signals with 180º phase difference. Phase of one of the signals is shifted using R and C. To overcome the nonlinear problem a conventional quadrature phase sensitive detector (PSD) is used. PSD provides the cotangent of the output phase. By using the switching multipliers, DC output is obtained which is then fed into ADC. The output from ADC is given to the microcomputer. The reference voltage for ADC is provided by passing it through DAC so that the variation of temperature changes on reference voltage is very small. For maximum highest possible resolution the resolution of ADC is adjusted. The prototype was developed to test the circuit. Testing results show that it can measure minimum displacement of 7.7×10^{-12} m. The nonlinearity is less than 0.5 %. A linear and high resolution output is obtained by this scheme. This scheme can be used to measure very small displacement in various applications.

11.2.7. Capacitance to Frequency Converter Suitable for Sensor Applications Using Telemetry

S. Chatzandroulis et al. describes a capacitance to frequency converter, which is suitable for capacitance type sensor interface [09]. The circuit can be used with passive telemetry systems as its output frequency is independent of the power supply (range of 3.7 to

5.5 V). The capacitance-to-frequency converter circuit has been designed taking into consideration the fluctuations in the system.

The proposed circuit consists of two modules; one is the band-gap reference voltage generator (produces a band-gap reference voltage and a regulated internal power supply) and an oscillator. The oscillator is designed around a current mode comparator that results in an output frequency independent of power supply variations and temperature drift to an extent. The oscillator circuit also includes two current sources and a capacitor which periodically charges and discharges the sensor capacitance. The switching period is initially calculated using the various parameters and is in-turn used to find the output frequency. Various relations to find the frequency are mentioned in the article. The frequency drift remains better than 800 ppm/°C at temperature of 37 °C.

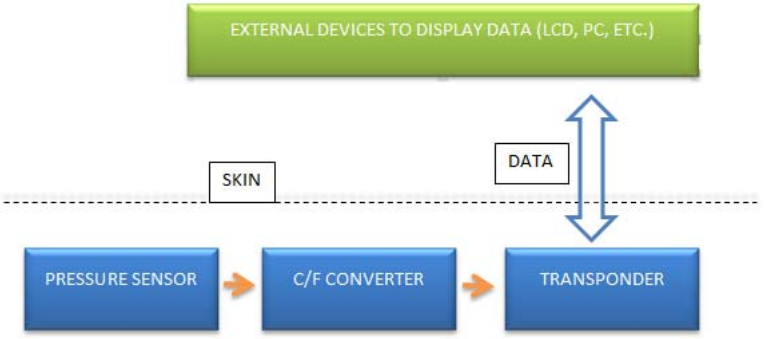


Fig. 11.7. Schematic diagram for Capacitance to Frequency Converter Suitable for Sensor Applications Using Telemetry.

11.2.8. Universal Capacitive Sensors and Transducers Interface (USTI)

Sergey Y. Yurish has presented a simple, cost-effective universal capacitance-to-digital converter based on the Universal Sensors and Transducers Interface (USTI) circuit [10]. The designed integrated circuit principle is based on three-signal measurement technique. The measurand, reference and offset will be converted into three time intervals. There are only a few components required to be connected externally, including a reference capacitance. A relationship between the aforementioned three parameters determines the unknown capacitance. The relationship is as follows [10]:

$$C_x = \frac{N_x - N_{off}}{N_{ref} - N_{off}} . C_{ref} \quad (11.1)$$

where

N_X - number of reference frequency pulses counted during the measurand,
N_{off} - number of reference frequency pulses counted during the offset cancellation,
N_{ref} - number of reference frequency pulses counted during the reference,
C_{ref} - precision reference capacitor.

The Fig. 11.8 shows the USTI in 32 pad Micro Lead Frame (MLF) packaging with dimensions of 5 mm × 5 mm × 1 mm. The converter has three popular serial interfaces - RS232, Serial Peripheral Interface Bus (SPI) and Inter-Integrated Circuit (I^2C) for communication to Personal Computer (PC) or master microcontroller. The circuit has an average relative error of ± 0.036 % [10] in a wide capacitance range from 50 pF to 100 μF. This is a significant reduction in the error when compared to some commonly used capacitive to digital converting techniques. The relative error for further frequency-to-digital conversion is constant in the whole frequency range and does not exceed ± 0.0005 %.

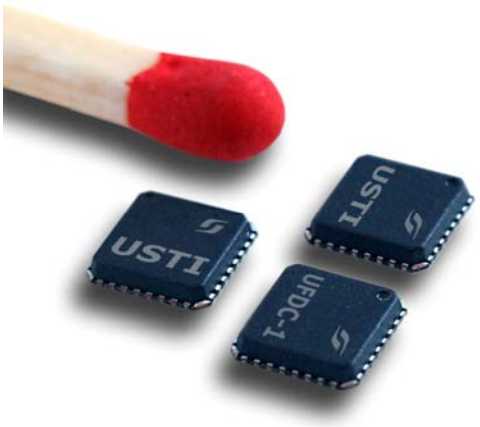


Fig. 11.8. Universal Sensors and Transducers Interfacing circuit (USTI) in miniaturized 32-pad MLF 5 mm × 5 mm × 1 mm package [10].

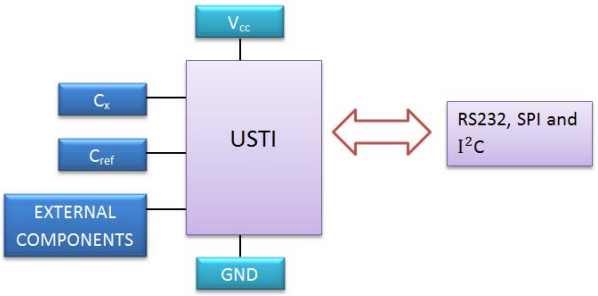


Fig. 11.9. Block diagram of capacitive sensors measurement based on the USTI.

11.2.9. An Integrated Interface Circuit with a Capacitance-to-voltage Converter as Front-end for Grounded Capacitive Sensors

Ali Heidary *et al.* presented a design for an integrated interface for grounded capacitive sensors [11]. This principle is based on a feed-forward technique to reduce the effects of parasitic cable capacitances and also provides immunity with a combination of a front-end amplifier. The standard 0.7 μm CMOS technology circuit has been designed and

implemented using as an integrated circuit. Theoretical verification and simulation has been performed and analyzed with experimental data for parasitic capacitance introduced by the connection cable.

Fig. 11.10 shows a block diagram of the designed circuit. The circuit consists of a multiplexer, a new capacitance-to-voltage converter, a voltage-to-period converter and a control unit. This circuit uses a three-signal calibration technique. The multiplexer is used to switch capacitance for input to capacitance-to-voltage converter. It is used such that the effect of parasitic capacitance can be eliminated without having the instability problem. The voltage-to-period converter is the next stage. The end user can optimize it independently of the sensor capacitance range.

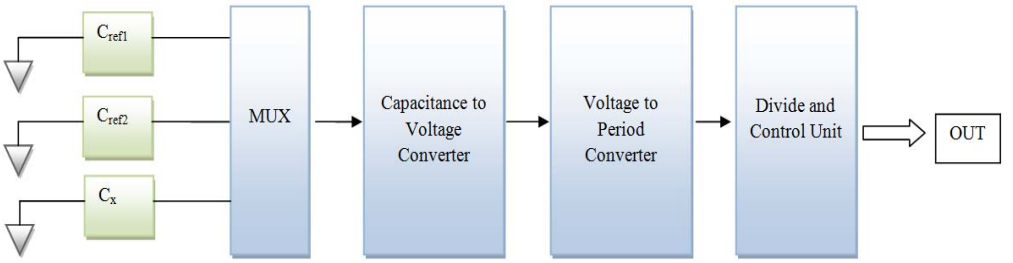


Fig. 11.10. Block diagram of capacitance-to-voltage converter for grounded capacitive sensors.

The major non-idealities like the amplifier offset, the switched charge injection and the CMOS switch ON resistance of the circuit has been discussed and analyzed. The voltage-to-period converter has been designed such that it is only sensitive to peak-to-peak voltage and hence independent to amplifier offset. The switch charge injection error can be removed by applying three-signal auto-calibration. The residual error can be reduced by adjusting the reference capacitors accordingly. The switch ON resistance error can be reduced using the called fast/slow mode.

Results show that for capacitance of 10 pF with connection cable of 30 m can be measured with error of less than 0.3 pF. The measured nonlinearity for capacitance ranging from 10 pF to 330 pF with cable length of 30 m is less than 3×10^{-4}. The circuit can be used for measurement of small capacitance in capacitance transducer applications.

11.2.10. A 16-channel Capacitance-to-period Converter for Capacitive Sensor Applications

Ioannis Ramfos *et al.* proposed an ASIC implementation of a capacitance-to-period conversion system using a relaxation oscillator [12] which has a linear characteristics and can read arrays of up to 16 different capacitive type sensors. The system multiplexes sixteen different channels to provide a ratio metric result. The ASIC is developed using 0.35 μm CMOS technology and mounted on PCB along with a supervising microcontroller which produces ratio metric measurements and minimizes the parasitic

effects. The variation in the capacitance of the connected sensor is translated linearly into equivalent output period variation on a pulse waveform, which is measured using the supervising microcontroller. As shown in the Fig. 11.11. The ASIC consists of three main components namely a multiplexer, a band-gapreference and a relaxation oscillator. The microcontroller unit coordinates the function of each block and provides with ratio metric measurements.

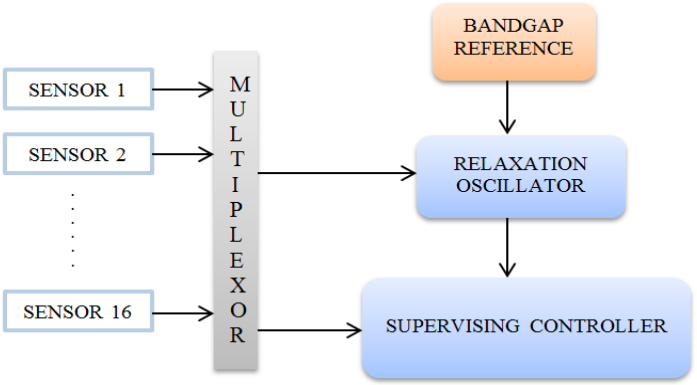


Fig. 11.11. System block diagram.

The multiplexor unit connects the capacitive sensors one at a time to the relaxation oscillator so that there are no crosstalk errors. A band-gap reference voltage is provided to the oscillator unit so as to provide a stable reference voltage which is insensitive to temperature and power supply variations. The final unit, the relaxation oscillator, consists of two parts: (a) a comparator operating in current mode and (b) a set of current sources that charge and discharge the sensor's capacitance. This unit provides output period variations in form of a pulse waveform which is linearly related to the sensor capacitance.

To minimize various errors such as interconnection parasitic, low frequency voltage variations, and circuit temperature fluctuations, ratio metric measurements using reference capacitance is performed by the supervising microcontroller unit, followed by a process of averaging via controlling software to further minimize the errors. To determine the unknown capacitance values, two sets of parasitic capacitances in parallel to the unknown capacitance were used, 343.2 pF and 15.8 pF and the sensitivity obtained were 0.89 μs/pF and 0.69 μs/pF respectively.

11.2.11. A CMOS Integrated Capacitance-to-Frequency Converter with Digital Compensation Circuit Designed for Sensor Interface Applications

Cheng-Ta Chiang *et al.* has presented a low cost CMOS integrated capacitance-to-frequency converter with digital compensation circuit for sensor interfacing applications [13]. The unique characteristic of the circuit proposed is that it produces a digitized output signal without realizing an analog to digital convertor, thereby reducing the hardware cost

along with increasing the time response. A notable feature of the circuit is that, due to pulse stream output the data can be transmitted over wide range of transmission media like radio, optical, Personal Satellite Network (PSN), Infra Red (IR), ultrasonic etc., hence making the circuit suitable for remote sensing applications as well. The circuit works on the principle of switched-capacitor technique that was developed for integration of digital and analog signals on a single silicon chip.

The circuit incorporates three stages to implement capacitance to frequency convertor namely capacitance-to-voltage converter (CVC), followed by a voltage-to-frequency converter (VFC), and the digital compensation circuit as shown in Fig. 11.12. The stage-I consists of a feedback capacitance 'CF'and reference capacitance 'CR' which are utilized to derive the voltage at the end of the 1st stage. The output voltage of stage-I (Capacitance to voltage converter) acts as the input to stage-II (the voltage to frequency convertor which consists of the switched -capacitor integrator, a comparator, and a D flip-flop). After application of the switched capacitance technique the circuit applies charge redistribution method. Finally the circuit uses a unique and firstly proposed digital compensation circuit which eliminates the quantization error and provides a dynamic output frequency range. The digital compensation circuit includes a detector, control, and frequency circuits.

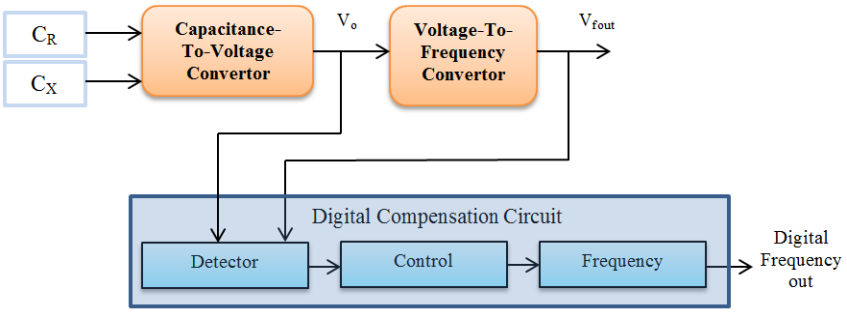


Fig. 11.12. The schematic of the capacitance-to-frequency converter.

The output frequency range achieved by the circuit is from 0.5 to 500 kHz under the variable capacitance of the detected sensor ranges from 4 to 24 pF. The area of this chip is 940 × 1080 μm^2 and the power consumption is 6.4 mW.

11.3. Summary

In this chapter, we have discussed various techniques for measuring very small change in capacitance. The various techniques used aims at providing a high resolution and linear output. The circuits been discussed are based on the principle to measure change in capacitance. The principles used are double differential principle, RC phase delay, charging and discharging, oscillator and capacitance to phase angle conversion. Following is a brief description of various papers that involves use of discrete electronic devices rather than fabrication of micro-chips. A simple interfacing circuit using double

differential principle is that this can measure a change of 10^{-18} F. The main feature of the design is that the change in capacitance is independent of the initial sensor's capacitance. A low power capacitance circuit is to measure capacitance change using differential structure. It uses TSMC 0.35 μm bulk CMOS process for fabrication. It has very high resolution and high degree of linearity. It is useful for biomedical sensor applications and used a RC phase delay concept to measure capacitance change. It uses MOSFET switch to avoid nonlinearity. The output of this is stable and linear with very high resolution. The design is simple and requires a small number of components. Interfacing with a microcontroller measures charging and discharging time of unknown capacitor through known capacitance. It's simple and low cost technique. It is mainly used for interfacing circuits. Used novel oscillator concept to measure capacitor change, it gives very high resolution, high accuracy and high stability. It has very less sensitivity to noise. The output has high stability and high accuracy. It has negligible influence of parasitic capacitance and a capacitance to phase angle conversion for measurement. It has a high resolution and linear output. The system has an extremely high stray immunity. It is used to measure very small displacement.

Table 11.1 shows comparing the various parameters of the design for capacitance measurement is provided. It provides the details regarding methodology, measurement range, sensitivity, error and application of the design. The application of these Capacitance Measurement techniques could be extended to the MEMS based sensors as well [14].

Table 11.1. Comparison summary of various design techniques and their applications.

Methodology	Minimum measurement value	Sensitivity	Error	Applications
Double Differential Principle	aF	2.5 ppm	0.92 pF for 1 μF capacitor	Measure very small capacitance change where initial can be unknown mainly in industries and interfacing circuits.
Differential Principle	fF	1.32 mV/fF	-	Bio-medical sensor application
RC Phase Delay	pF	246 mV/fF	2.6%	Micro sensors
Charging and Dis-charging using microcontroller	pF	-	Below 4% FSR for 1 pF to 10 pF	Interfacing circuits
Oscillator	pF	-	100 aF for 2 pF capacitor	Interfacing circuits
Capacitance to phase angle conversion	pF	32 ppm	Less than 0.5%	Very small displacement measurements.

Table 11.1. (Continued).

Methodology	Minimum measurement value	Sensitivity	Error	Applications
A CMOS Integrated Capacitance-to-Frequency Converter with Digital Compensation Circuit Designed for Sensor Interface Applications	4-24 pF	NA	NA	Sensor interface
A 16-channel capacitance-to-period converter for capacitive sensor applications	NA	NA	NA	Sensor interface
Capacitance to Frequency Converter Suitable for Sensor Applications Using Telemetry	NA	NA	NA	Suitable design for passive telemetry systems
Universal Capacitive Sensors and Transducers Interface (USTI)	50 pF to 100 μF	NA	±0.036 %	Varied applications. Suitable for wide range C/F converters

References

[1]. D. M. G. Preethichandra and Katsunori Shida, A Simple Interface Circuit to Measure Very Small Capacitance Changes in Capacitive Sensors, *IEEE Trans. Instrum. Meas.*, Vol. 50, Dec. 2001, pp. 1583-1586.

[2]. M. R. Haider, M. R. Mahfouz, S. K. Islam, S. A. Eliza, W. Qu and E. Pritchard, A Low-power Capacitance Measurement Circuit with High Resolution and High Degree of Linearity, *IEEE Circuits and Systems*, Sep. 2008, pp. 261-264.

[3]. Robert N. Dean and AditiRane, An Improved Capacitance Measurement Technique Based of RC Phase Delay, in *Proceedings of the IEEE Instrumentation and Measurement Technology Conference (I2MTC)*, 2010, pp. 367-370.

[4]. F. Reverter, M. Gasulla, and R. Pallàs-Areny, A low-cost microcontroller interface for low-value capacitive sensors, in *Proceedings of the 21st IEEE Instrumentation and Measurement Technology Conference (IMTC'04),* Como, Italy, Vol. 3, May 2004, pp. 1771–1775.

[5]. Frank M. L. van der Goes and Gerard C. M. Meijer, A Novel Low-Cost Capacitive-Sensor Interface, *IEEE Trans. Instrum. Meas.,* Vol. 45, Apr. 1996, pp. 536–540.

[6]. Ashkan Ashrafi and Hossein Golnabi, A high precision method for measuring very small capacitance changes, *Rev. Sci. Instrum.,* Vol. 70, No. 8, August 1999, pp. 3483-3487.

[7]. Prabhu Ramanathan, Marimuthu. R, R. Sarjila, Sudha Ramasamy and P. Arulmozhivarman, Performance Comparisons of Interface Circuits for Measuring Capacitances, *Journal of Theoretical and Applied Information Technology*, Vol 39. No. 1, May – 2012, pp. 001 - 005.

[8]. Yurish S. Y., Digital Sensors and Sensor Systems: Practical Design, *IFSA Publishing*, 2011, http://www.sensorsportal.com/HTML/BOOKSTORE/Digital_Sensors.htm

[9]. S. Chatzandroulis and D. Tsoukalas, Capacitance to Frequency Converter Suitable for Sensor Applications using Telemetry, *Analog Integrated Circuits and Signal Processing,* 27, 2001, pp. 31–38.

[10]. Sergey Y. Yurish, Universal Capacitive Sensors and Transducers Interface, in *Proceedings of the XXIII Eurosensors Conference*, Procedia Chemistry, 1, 2009, pp. 441–444.

[11]. Ali Heidary and Gerard C M Meijer, An integrated interface circuit with a capacitance-to-voltage converter as front-end for grounded capacitive sensors, *Meas. Sci. Technol.*, 20, 2009, 015202 (7 p).

[12]. Ioannis Ramfos, Stavros Chatzandroulis, A 16-channel capacitance-to-period converter for capacitive sensor applications, *Analog Integr. Circ Sig. Process*, 71, 2012, pp. 383–389.

[13]. Cheng-Ta Chiang, Chi-Shen Wang, and Yu-Chung Huang, A CMOS Integrated Capacitance-to-Frequency Converter with Digital Compensation Circuit Designed for Sensor Interface Applications, in *Proceedings of the IEEE SENSORS Conference*, 2007, pp. 954 - 957.

[14]. Kochuthomman Joseph Mampilly, Arjun Ashok, Sudha Ramasamy and Prabhu Ramanathan, Review of MEMS Based Application in Medical Industries, *Sensors & Transducers,* Vol. 139, No. 4, April 2012, pp. 25-44.

Chapter 12
Design and Validation of Unimorph Piezoelectric Energy Harvesters

Ashok Batra and Almuatasim Alomari

12.1. Introduction

Energy harvesting is the process of driving and capturing energy from external sources and storing it to power autonomous wireless devices, e.g. wireless sensor network nodes, microelectromechanical systems (MEMS), and biomedical sensors. Piezoelectric energy harvesting devices are attracting interest to minimize the consumption and use of batteries which are considered expensive, harmful to the environment, and have limited life. While a number of piezoelectric materials exist, thin film piezoelectric polyvinylidene fluoride (PVDF) polymer has become one of the most common materials for the creation of piezoelectric harvester devices [1]. This is due to the fact that the piezoelectric coefficient of PVDF is ten times larger than any other polymer; it is also a material of high flexibility, high toughness and can be readily formed in a thin film profile. This combination of properties make it an excellent choice for many applications including dynamic strain gauges and vibration sensors for alarms, switches, speakers, and contact microphones [2]. In a continuous effort to increase the performance of piezoelectric energy harvesting devices researchers in both academic and industrial communities in field of renewable energy have proposed many techniques and procedures to increase the power generation capability and enhance the bandwidth of vibration energy harvesters. Tailoring the type of piezoelectric material and dimension of its geometrical shape have been the studied by a number of researchers [3-14]. In addition a number of theoretical models have been proposed to simplify the dynamic system of the complete energy harvesting system [15-22].

In the present work a unique unimorph cantilever system is investigated both via modeling and experiments. The system consists of a piezoelectric energy harvester attached to one end of a structural cantilever beam to form as a dynamic magnifier system. The dynamic system is of interest in this work since it is a combined mechanical and electrical system

Ashok Batra
Department of Physics, Chemistry and Mathematics, Alabama A&M University, Normal, AL 35762 USA

(electro-mechanical system) and a lumped parameter model (LPM) is expressed in this work to simplify the modeling approach for the dynamic magnifier energy harvester [23]. Generally, a single unimorph piezoelectric cantilever is attached to a base excitation and is characterized as a single degree of freedom (SDOF) piezoelectric energy harvester system (PEHS) is described by a LPM. Researchers reported [15] that the SDOF is valid only for harvesting energy in a region close to its resonance frequency. Therefore, the connection of a unimorph cantilever to an additional mechanical beam, as proposed here, requires two independent coordinates or two degrees of freedom (2DOF). Finally, the theoretical results of 2DOF piezoelectric energy harvester system [21] can be considered as a solution of an energy harvester with dynamic magnifier.

Different from single UCB, multiple cantilevers or cantilever arrays integrated in on energy harvesting device as seen in this work can easily achieved continuous wide bandwidth if the geometric parameters of the harvester are appropriately selected [24]. Different energy harvesters with cantilever array were also implemented compatibly with current standard MEMS fabrication techniques [25-28]. A unimorph cantilever beam (UCB) configuration in simple form includes one piezoelectric layer and one shim layer which is commonly showing a single resonance frequency at each natural mode shape. A multifrequency energy harvester devices are easily achieved by integrating multiple cantilevers or cantilever arrays in one cantilever beam [29, 30].

The present work is organized as follows: theoretical predictions of output voltage and power of a 2DOF system under an external uniform force in covered in section 12.2. The electrical output parameters and mode shapes of a 5 commercial piezoelectric unimorph cantilever arrays with same thickness and different length integrated in aluminum beam. Modeling is compared with the experimental data. In section 12.3, we describe experimental results and modeling of a 5 commercial piezoelectric unimorph arrays with same thickness and different length intergrated in amluminum cantilever beam.

12.2. Parametric Study and Effect of Beam Material on a Unimorm Energy Harvester

12.2.1. Mathematic Concept of Mechanical Energy Conversion

The bending stiffness of the materials in the composite section in Fig. 12.1 are given as follows:

Substrate layer
$$EI_s = E_s w\left(\frac{1}{12}h_s^3 + h_s\left(\frac{h_p}{2} + h_e + \frac{h_s}{2}\right)^2\right) \tag{12.1}$$

Electrode layer
$$EI_e = E_e w\left(\frac{1}{12}h_e^3 + h_e\left(\frac{h_p}{2} + \frac{h_e}{2}\right)^2\right) \tag{12.2}$$

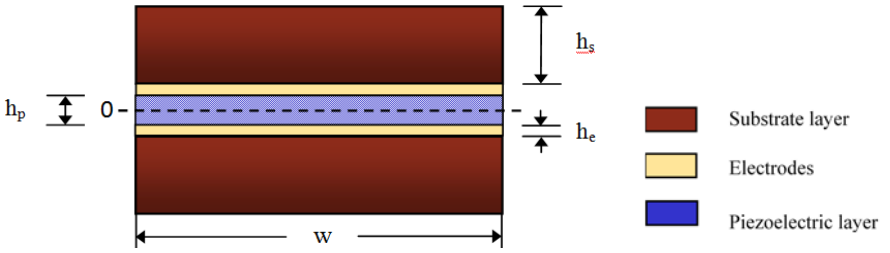


Fig. 12.1. Cross-sectional views of the composite sections of the unimorph beam.

Piezoelectriclayer $$EI_p = E_p \frac{wh_p^3}{12} \tag{12.3}$$

The total bending stiffness is the sum of the bending stiffness of each material section, this gives

$$EI = IE_p + 2IE_s + 2IE_e \tag{12.4}$$

The mass density of the unimorph beam is given by:

$$m = w\left(\rho_p h_p + 2\rho_s h_s + 2\rho_e h_e\right) \tag{12.5}$$

The effective mass of the unimorph beam with tip mass (m_t) is given by:

$$m_{eff} = \frac{33}{140} mL + m_t \tag{12.6}$$

The natural frequency of beam can be obtain from:

$$\omega_n = \beta_n^2 \sqrt{\frac{3EI}{mL^4}} \tag{12.7}$$

Electromechanical coupling coefficient can be obtained by:

$$\theta = \int_0^L J_p \varphi''(x) dx \tag{12.8}$$

where

$$J_p = \frac{e_{31}}{h_p}\left(\int_0^w \int_{\frac{t_p}{2}}^{\frac{t_p}{2}+h_e} z dy dz + \int_0^w \int_{\frac{t_p}{2}+h_e}^{\frac{t_p}{2}+h_e+h_s} z dy dz \right)$$

The eigenfunction corresponding undamped free vibration problem is denoted by

$$\varphi(x) = C_r\left[\cos\beta_n x - \cosh\beta_n x + \varsigma_r(\sin\beta_n x - \sinh\beta_n x)\right]$$

β_n the eigenvalues of the system at different mode shapes are obtained from:

$$1 + \cos\beta_n \cosh\beta_n + \frac{m_t}{mL}\beta_n(\cos\beta_n \sinh\beta_n - \sin\beta_n \cosh\beta_n) -$$
$$-\frac{\beta_n^3 I_t}{mL^3}(\cosh\beta_n \sin\beta_n + \sinh\beta_n \cos\beta_n) + ... + \frac{\beta_n^4 m_t I_t}{m^2 L^4}(1 - \cos\beta_n \cosh\beta_n)$$

$$I_t = \frac{1}{6} m_t l^2$$

$$\varsigma_r = \frac{\sin\beta_n L - \sinh\beta_n L + \beta_n L\frac{M_t}{mL}(\cos\beta_n L - \cosh\beta_n L)}{\cos\beta_n L + \cosh\beta_n L - \beta_n L\frac{M_t}{mL}(\sin\beta_n L - \sinh\beta_n L)}$$

In this case:

$$\beta_n = 1.5548,\ 4.2071,\ 7.2619,\ 10.34502$$

The output voltage and of lumped-parameter system (SDOF) is given by [9] as:

$$|V| = \frac{m_{eff} R_{eq} d_{31} \omega_n \Omega \omega^2 u_o e^{i\omega t}}{\sqrt{([1-(1+2\xi r)\Omega^2]^2 + [(1+K_e^2)r\Omega + 2\xi\Omega - r\Omega^3]}} \tag{12.9}$$

The governing equations of dynamic system of harvester described in Fig. 12.2 can be expressed as follows [31]:

$$m'\ddot{u}' + c'\dot{u}' + k'u' + \theta\ V = -m'\ddot{u} - m'\ddot{u}_o \tag{12.10}$$

$$(m + m')\ddot{u} + c\dot{u} + ku + m'\ddot{u}' + (m + m')\ddot{u}_o = 0 \tag{12.11}$$

$$-\theta\dot{u} + c_p\dot{V} + \frac{V}{R_L} = 0 \tag{12.12}$$

The output voltage and power of 2DOF can be described as [24]:

$$V = \left| \frac{\dfrac{a^2 + i2\xi\ a\ \Omega}{a^2 - (1+\mu)\Omega^2 + i2\xi\ a\Omega}\dfrac{m'\omega^2 u_o e^{i\omega t}}{\theta}}{\left(1 - \Omega^2 + i2\xi'\Omega - \dfrac{\mu\Omega^4}{a^2 - (1+\mu)\Omega^2 + i2\xi\ a\Omega}\right)\dfrac{ir\Omega + 1}{irK_e^2\Omega} + 1} \right| \tag{12.13}$$

$$P = \frac{V^2}{R_L} \tag{12.14}$$

where,

$$a = \frac{\omega_1}{\omega_2}\ ,\ \Omega = \frac{\omega}{\omega_2}\ ,\ r = \omega_2 C_p R_L\ ,\ C_p = \frac{\varepsilon L w}{h_p}\ ,\ K_e = \frac{\theta}{\sqrt{C_p k'}}\ ,$$

m, m' are the mass of each system, c, c' are the damping coefficient of the both systems.

k, k' are the spring stiffness of both systems. θ is the electromechanical coupling coefficient of the system. C_p is the equivalent capacitance of piezoelectric model. R_L is load resistance. V is the voltage across R_L. $u' = u_2 - u_1$ and $u = u_1 - u_o$ are the relative displacements of both systems.

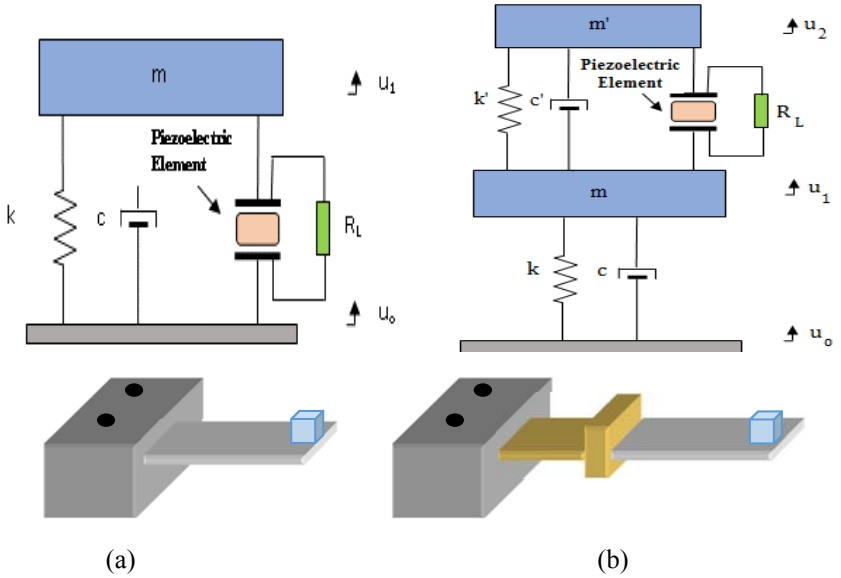


Fig. 12.2. Simplified (a) single degree of freedom (SDOF) and (b) two degree of freedom (2DOF) mechanical model of the proposed energy harvesting device.

12.2.2. Experimental and Modelling of Unimorph Energy Harvester with Tip Mass

Energy harvesting experiments were carried out by initially attaching a commercial unimorph cantilever beam (UCB; PVDF based) directly to shaker as single device with a single degree of freedom (SDOF). The UCB was then connected to carbon fiber reinforced plastic (CFRP) to form a two degree of freedom (2DOF) harvester. A schematic of how the commercial unimorph cantilever beam (UCB-PVDF based) was connected at the front of a beam is shown in Fig. 12.3. The dimensions, electrical, and mechanical properties of cantilever beams are given in Tables 12.1 and 12.2. The parametric study conducted was performed by the experimental set-up shown in Fig. 12.3.

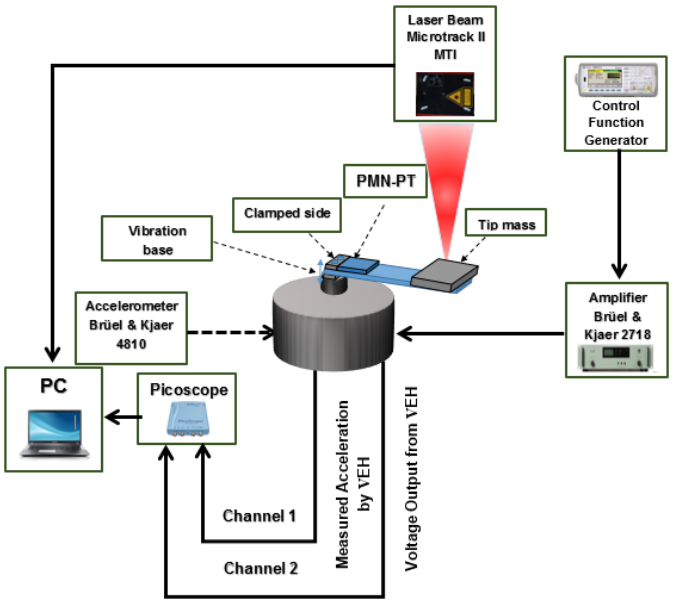


Fig. 12.3. Schematic diagram of an experimental set-up.

Table 12.1. Properties of cantilever beams investigated.

Unimorph cantilever beam	PVDF layer	Polyester layer	Electrode
Length (m)	0.041	0.041	0.041
Width (m)	0.016	0.016	0.016
Thickness (μm)	52	300	10
Young's modulus (GPa)	3	5	83
Density (kg/m^3)	1780	1820	10500
Dielectric constant	12	-	-
Piezo strain constant (10^{-12}C/m)	23	-	-
Capacitance (nF)	2.7	-	-
Electromechanical Coefficient θ (N/V)	4.5e^{-5}	-	-

Table 12.2. The properties of piezoelectric unimorph cantilever [CFRP = carbon fiber reinforced composite].

Type of beam	CFRP
Length (m)	0.066
Width (m)	0.016
Thickness (m)	0.0004
Young's modulus (GPa)	77
Density (kg/m^3)	402

In this section, we present the modeling and experimental results of the proposed single UCB and the suggested design of piezoelectric energy harvester with dynamic magnifier.

Generally, the model assumes that the piezoelectric unimorph cantilever coupled to a mechanical cantilever beam can be represented by an oscillatory system which can be simply described as a spring-mass-damper system for each part as in Fig. 12.2. The system as whole can then be described as coupled second-order differential equations as mentioned in Section 12.2. The model was divided into two parts. Firstly, as a single piezoelectric unimorph cantilever beam (UCB) as illustrated in Fig. 12.2 (a) and secondly as a UCB that is connected to mechanical cantilever beam as illustrated in Fig. 12.2 (b). The model was applied to the designed harvester to determine the output parameters and bandwidth of suggested design. In order to validate the model, both output voltage and output power versus frequency were compared with experimentally determined data. The critical damping coefficient depends on the mode of cantilever beam, each mode has different damping coefficient and each beam has different damping coefficient as well. The following formulas are used to calculate ξ from the experimental data in Fig. 12.4. Table 12.3 shows a summary of damping ratio of experimental results at first resonance frequency of each beam in both cases as SDOF and 2DOF.

$$\delta = \frac{1}{n}\ln\left(\frac{B_1}{B_{n+1}}\right) \tag{12.15}$$

$$\xi = \frac{\delta}{\sqrt{(2\pi)^2 + \delta^2}} \tag{12.16}$$

where δ is the logarithmic decrement.

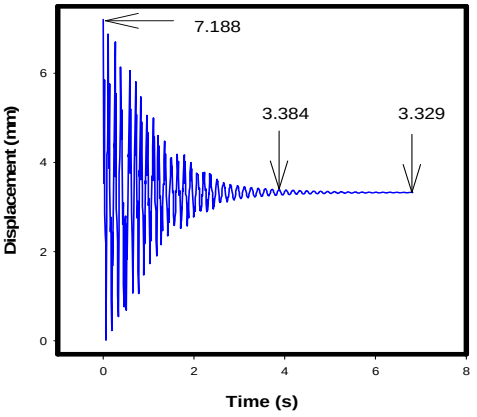


Fig. 12.4. Unimorph Beam response from tip accelerometer for first mode.

Table 12.3. Experimental damping coefficients of studied systems.

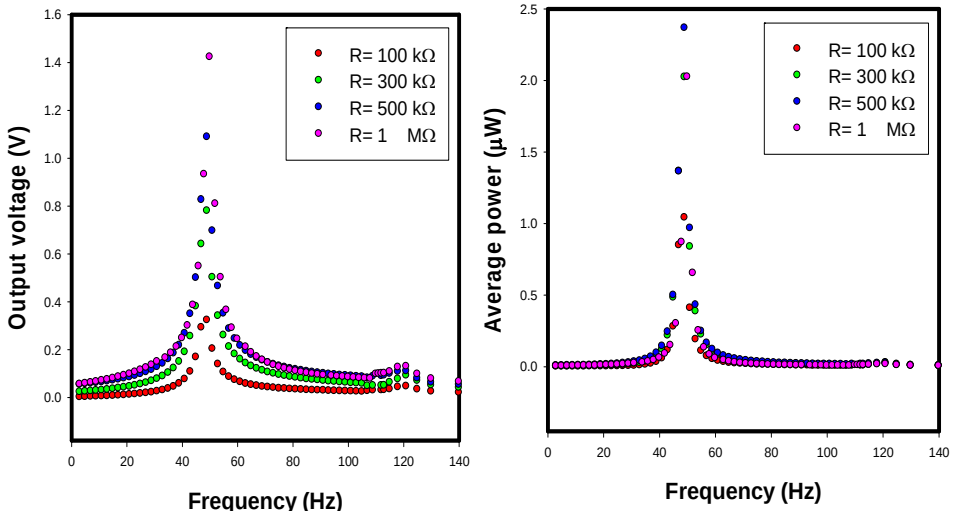
Beam type	Damping ratio (ξ) SDOF	Damping ratio (ξ) 2DOF
Unimorph	0.02	0.01
CFRP	-	0.05

Table 12.4 shows a summary of both experimental and theoretical results at the resonance frequencies of the system. The last column in each part represents the percentage error of theoretical model to experimental data. It was found that, the system exhibited a single resonance peak for single UCB which represents a single degree of freedom (SDOF) system, as in Fig. 12.2 (a) and two resonance peaks when the UCB was connected to the mechanical cantilever beams, as in Fig. 12.2 (b), which can be represented as two degree of freedom (2DOF) system. Different mode responses have been revealed and are shown in Figs. 12.5 and 12.6 when the UCB was a single and when it was attached to the CFRP mechanical cantilever beam of Table 12.1. These modes are clearly revealed since we employed two resonators which are represented by UCB and mechanical cantilever beam. However, our goal of adding mechanical cantilever beam to UCB is focused on increasing output voltage and output power at a resonance frequency that is ideally close to ambient environmental frequency. In addition the use of a 2DOF provides a route to achieve multiple resonance peaks and therefore enhance the frequency range of the harvesting device. Fig. 12.5 and Fig. 12.6 show the output voltage and average power results of a single UCB and UCB with magnifier beam, respectively. It is shown in Figs. 12.5 (b) and 12.6 (b) that the optimal resistance appear at load resistance R_L =500 kΩ. It is shown in Table 12.3 that the output voltage and average power of the SDOF at optimal resistance of UCB are 1.08 V and 2.37 μW respectively at a first resonance frequency of approximately 49 Hz, and 0.11 V and 0.03 μW respectively at a second resonance frequency of approximately 121 Hz as shown in Fig. 12.5. Also, It is shown in Table 12.3 that the output voltage and average power of the 2DOF of UCB with magnifier beam are 1.18 V and 2.78 μW respectively at a first resonance frequency of approximately 54 Hz, and 0.42 V and 0.36 μW respectively at a second resonance frequency of approximately 120 Hz as shown in Fig. 12.6. However Figs. 12.5 (b) and 12.6 (b) highlights the narrowband harvesting capability of such a system since the power generated by the harvester rapidly decreases as the vibration frequency moves above from resonance conditions.

Table 12.4. Resonance frequency, voltage and power output parameters at optimal resistance R_L=500 kΩ.

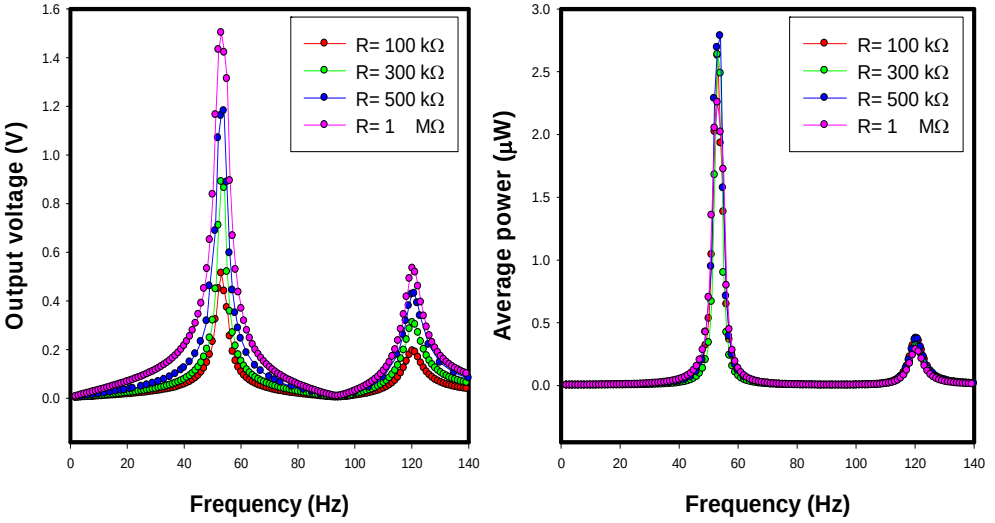
Cantilever beam type System	Resonance frequency (Hz)			Output voltage (V)			Output power (μW)		
	Exp.	Model	Error %	Exp.	Model	Error %	Exp.	Model	Error %
UCB (SDOF)	49 121	49.2 -	0.4 -	1.08 0.11	1.04 -	3.7 -	2.37 0.03	2.14 -	9.7 -
CFRP (2DOF)	54 120	48 124	11.1 3.3	1.18 0.42	1.19 0.40	0.8 6	2.78 0.36	2.82 0.33	1.4 8

Experimental and simulation results are well matched for first resonance frequency as shown in Fig. 12.7 using a model of a SDOF system for a single UCB, Also experimental and simulation results are well matched for first and second resonance frequency as shown in Fig. 12.8 using a model of a 2DOF system for a UCB with magnifier beam.



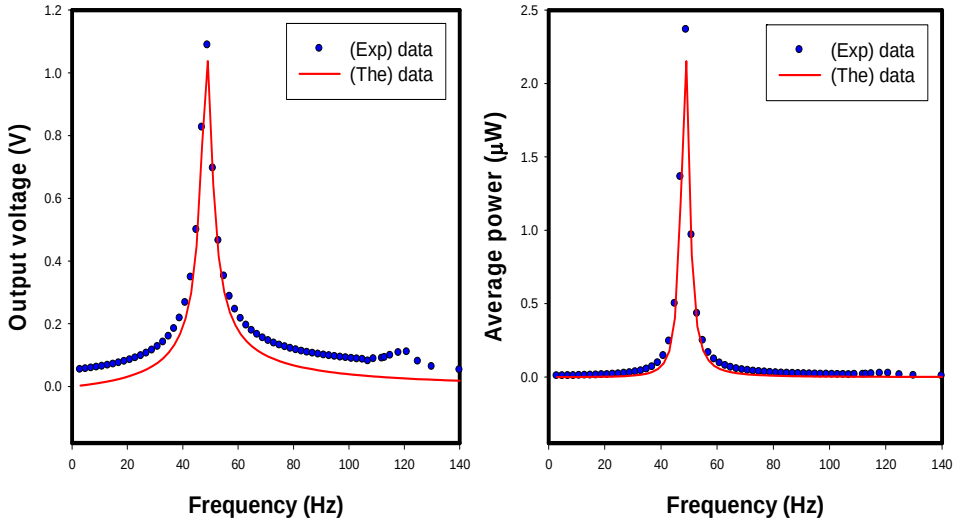
(a) (b)

Fig. 12.5. Measured of (a) output voltage and (b) average power magnitudes versus frequency for a single UCB (SDOF).



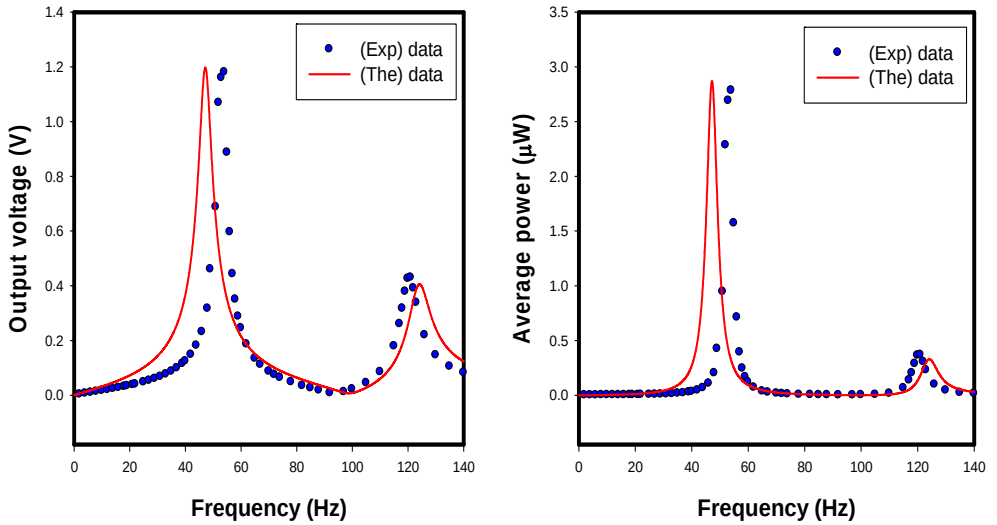
(a) (b)

Fig. 12.6. Measured of (a) output voltage and (b) average power magnitudes versus frequency for a UCB with magnifier beam (2DOF).



(a) (b)

Fig. 12.7. Experimental and model data of output RMS voltage versus frequency of UCB.



(a) (b)

Fig. 12.8. Experimental and model data of output RMS voltage versus frequency of CFRP.

Figs. 12.9 and 12.10 show the variation in voltage and power as the load resistance is changed, again with good agreement between experiment and model. The voltage increases with load resistance (R_L) since the piezoelectric UCB acts as a capacitor during vibration and discharges to a lesser extent as the load resistance, and the associated RC time constant, increases. The optimum power is at the condition $2\pi.f.\ R_L.C=1$, where C is

the capacitance of the UCB element [32]; for example at 49 Hz the optimum load resistance is 1.2 MΩ based on a device capacitance of 2.7 nF which is why the power level with load resistance flattens out in at 1 MΩ in Figs. 12.9 (b) and 12.10 (b).

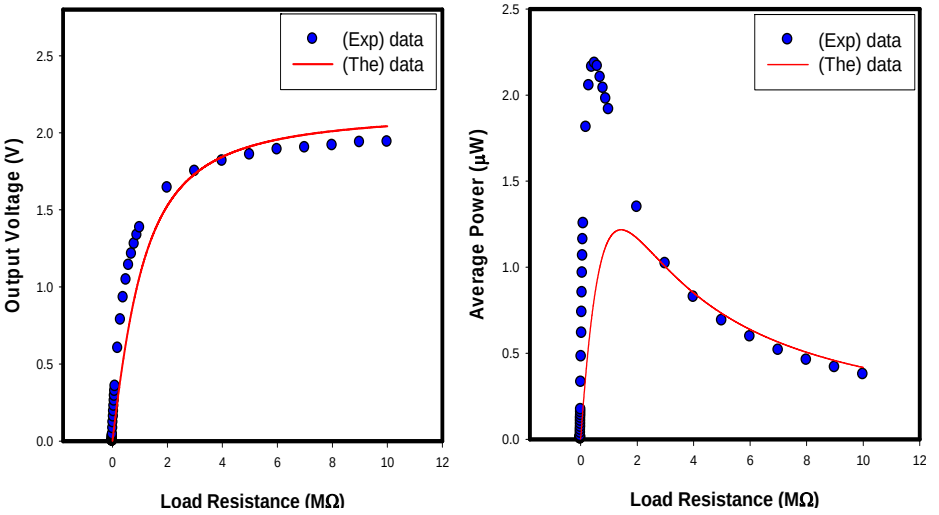


Fig. 12.9. Simulated and measured (a) voltage, and (b) power magnitudes versus load resistance for single UCB (SDOF) at 49 Hz.

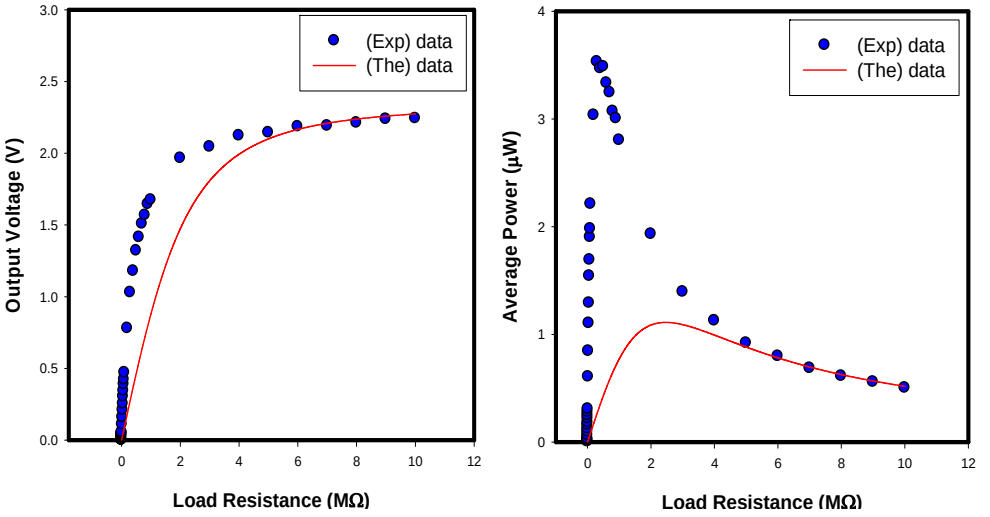
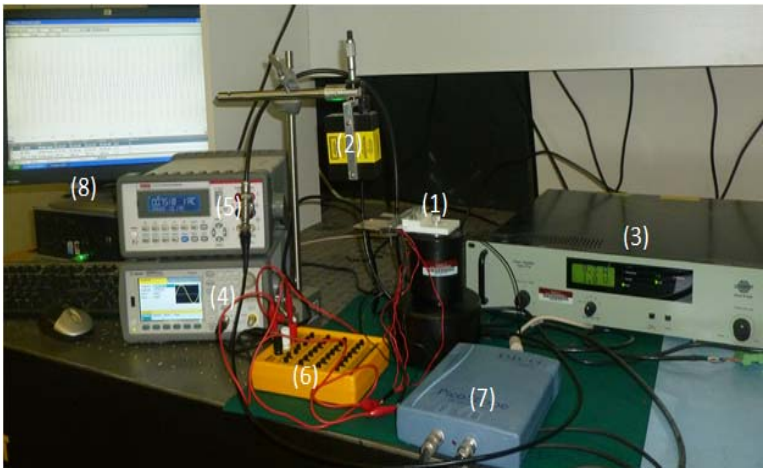


Fig. 12.10. Simulated and measured (a) voltage, and (b) power magnitudes versus load resistance for single UCB with CFRP (2DOF) at 49 Hz.

To conclude, this section described a compact bi-modal 2DOF Piezoelectric Energy Harvester (PEH) that is composed of a piezoelectric element attached to the one end of a mechanical cantilever. A theoretical 2DOF model is utilized for predicting the dynamic response of the proposed piezoelectric energy harvester. Both simulation and experiment indicate improved performance in voltage output of CFRP cantilever. The resonance frequency varies with type of cantilever materials used. Since the 2DOF system provides additional modes of resonance it provides a simple route to the formation of a broader band vibration energy harvesting system. Thus, depending upon the applicable resonance frequencies available, the geometry and materials of the cantilever can be selected in the fabrication of an optimum PEH.

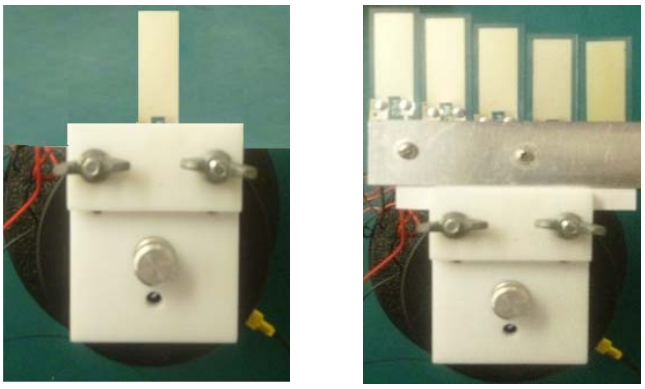
12.3. A Unimorph Energy Harvester with Cantilever Arrays

In this section, we present experimental results of our design of energy harvester with cantilever arrays. A parametric study was undertaken using the experimental setup as described in the previous section and shown in Fig. 12.11. Energy harvesting measurements were carried out initially by attaching a commercial polyvinylidene fluoride (PVDF) unimorph cantilever beam at the front of Teflon base, the configuration shown in Fig. 12.12 (a). The second configuration involved attaching multiple polyvinylidene fluoride (PVDF) unimorph cantilever beams or arrays at the front of aluminum cantilever beam, as shown in Fig. 12.12 (b). Both devices were then connected to a shaker system. The dimensions, electrical, mechanical properties of cantilever beams are shown in Table 12.5.



(1) Shaker with an accelerometer (Bruel&Kjaer 4810) and the cantilever arrays,
(2) Laser vibrometer (Microtrack II) MTI, (3) Fixed gain amplifier (Bruel&Kjaer 2718),
(4) Control function generator (GF8046 ELENCO), (5) Digital multimeter (Keithley 2110),
(6) Variable resistances box, (7) Picoscope, and (8) Data acquisition system.

Fig. 12.11. Experimental setup used for the frequency response measurements of a unimorph cantilever arrays (photos by A. Alomari, 2015).



(a) (b)

Fig. 12.12. Close views of the proposed system tested under base excitation (a) single UCB, (b) multiple UCB's (photos by A. Alomari, 2015).

Table 12.5. Properties of cantilever beams investigated.

Type of beam	Aluminum beam	Piezoelectric unimorph cantilever beam	
		PVDF layer	Polyester layer
Length (mm)	L_m=100	L_p=41, 39, 37, 35 ,33	L_s=41, 39, 37, 35 ,33
Width (mm)	w_m=24	w_p=16	w_s=16
Thickness (mm)	h_m=0.5	h_p=0.25	h_s =0.25
Young's modulus (GPa)	E_m=69	E_p=4	E_s=8.3
Density (kg/m^3)	ρ_m=2700	ρ_p=1780	ρ_s=1820
Dielectric constant	-	ε_{33}^{T}=12	-
Piezo strain coefficient (10^{-12} C/N)	-	d_{31}=23×10^{-12}	-
Capacitance (nF)		C_p=2.7	

Fig. 12.13 shows the frequency response functions (FRF) of output voltage and average output power for single UCB at various load resistance. It can be shown from Fig. 12.13 that there is only one peak which represents the resonance frequency of UCB. The energy harvesting bandwidth at resonance frequency of the harvesting beam is between (47, 55) Hz. Fig. 12.14 shows the frequency response functions (FRF) of output voltage and average output power for multiple of UCB's at various load resistance. It can be shown from Fig. 12.14 that there are five peaks which represent the resonance frequency of UCB's. The energy harvesting bandwidth at resonance frequency of the harvesting beams in this case was between (22,88) Hz. The maximum power for mode 1 excitation is around 0.29 μW at 22 Hz. For mode 2 excitation, the maximum power output is around 0.63 μW at 30 Hz. The maximum power output for mode 3 excitation is 1.54 μW at 45 Hz. For mode 4 excitation, the maximum power output is around 4.7 μW at 65 Hz. The maximum power output for mode 5 excitation is 5.22 μW at 88 Hz. The resonance frequencies, output voltage, and average power of the first fifth modes at optimum resistance for single UCB and multiple UCB's from the graphs in Fig. 12.13 and Fig. 12.14 are summarized in Table 12.6 at the end of this letter.

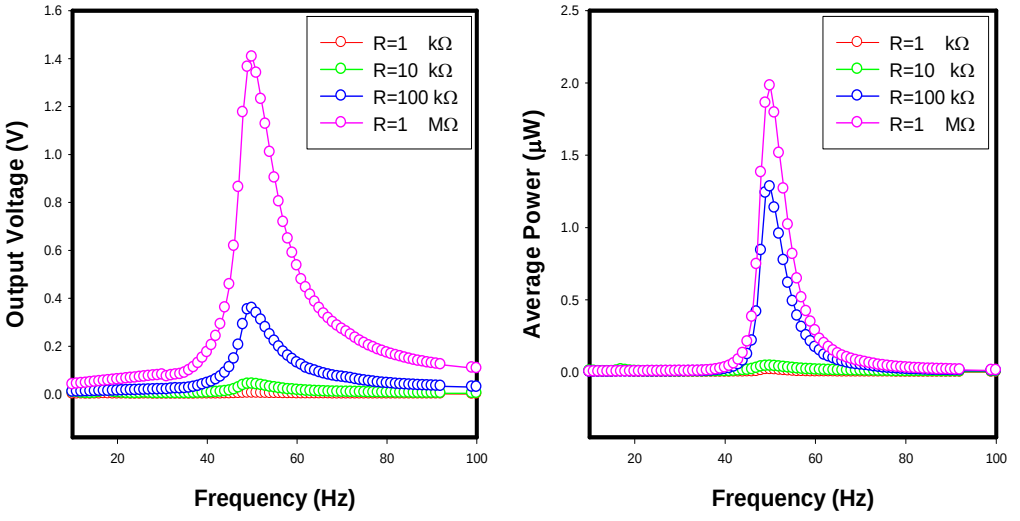


Fig. 12.13. Experimental data of FRF of single UCB at various load resistance of output (a), voltage (b) power.

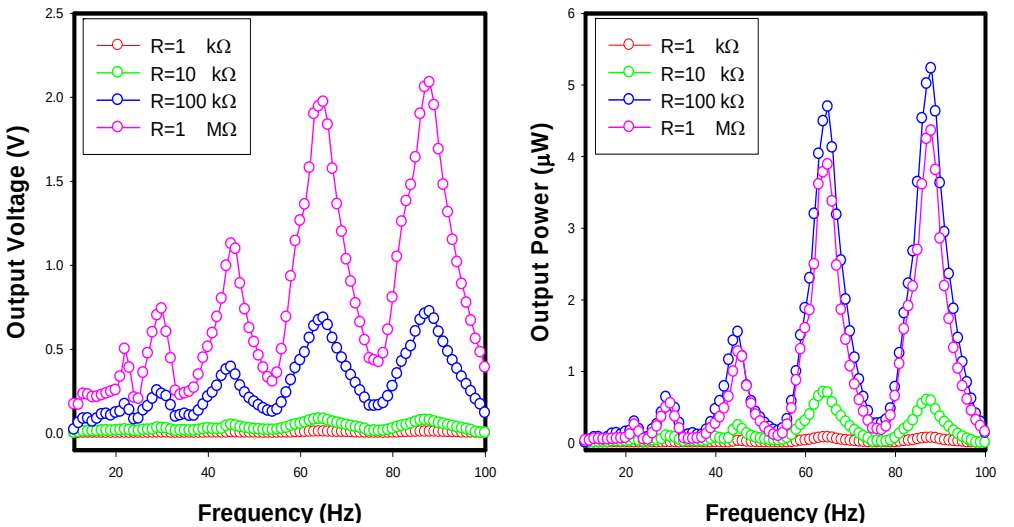


Fig. 12.14. Experimental data of FRF of multiple UCB's at various load resistance of output (a), voltage (b) power.

Fig. 12.15 shows the output voltage and average power for single UCB at various range of load resistances, R_L, from 100 Ω to 10 MΩ. It can be seen from Fig. 12.15 (a) that the output voltage increases with increasing R_L and the resonance frequency of a Piezoelectric energy harvester in UCB depends on the external load resistance R_L. Fig. 12.16 shows the

output voltage and average power for multiple UCB's at various range of load resistances, R_L, from 100 Ω to 10 MΩ obtained for excitation at the first fifth mode shapes.

Table 12.6. Resonance frequency, voltage and power output parameters of multiple UCB's**.**

Cantilever beam type	Mode shape	Resonance frequency (Hz)	Output voltage (V)	Output power (μW)
Single UCB	1st Mode	50	1.41	1.97
Multiple UCB's	1st Mode	22	0.17	0.29
	2nd Mode	30	0.25	0.63
	3rd Mode	45	0.39	1.54
	4th Mode	65	0.69	4.69
	5th Mode	88	0.72	5.22

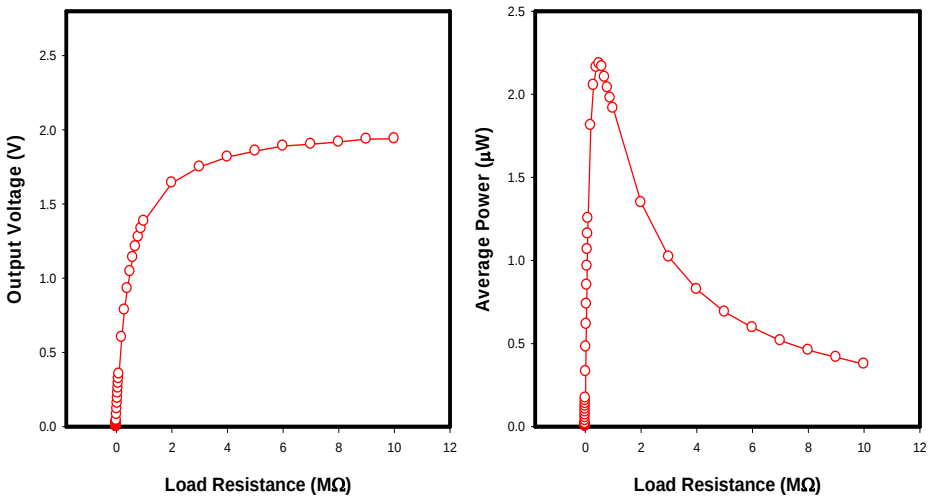
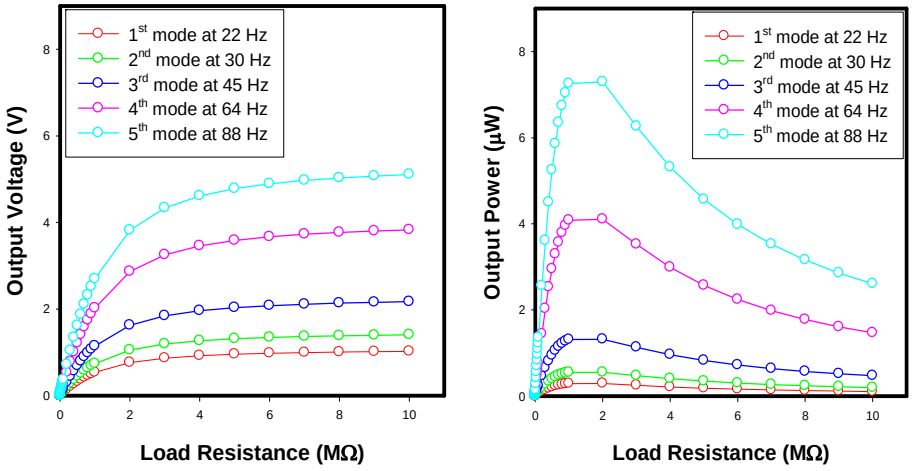


Fig. 12.15. Experimental data of (a) output voltage, and (b) output power versus load resistance of single UCB at resonance frequency.

12.3.1. Modeling Results

A 3 dimensional UCB's with aluminum beam are used for the simulation in COMSOL. The model is designed in COMSOL as shown below in Fig. 12.17 (a). The model consists of 5 UCB's with different lengths attached at the front of aluminum beam. The lengths of shim layer and piezomaterial are made equal. Using solid mechanics module, one end of the model is fixed and the other end is made to move freely. Meshing a geometry is done using size parameters for free tetrahedral, with a fine mesh near the clamped end as shown in Fig. 12.17 (b). The complete mesh consists of 22954 domain elements, 13676 boundary elements, and 1333 edge elements for a total number of degrees of freedom of 139605.

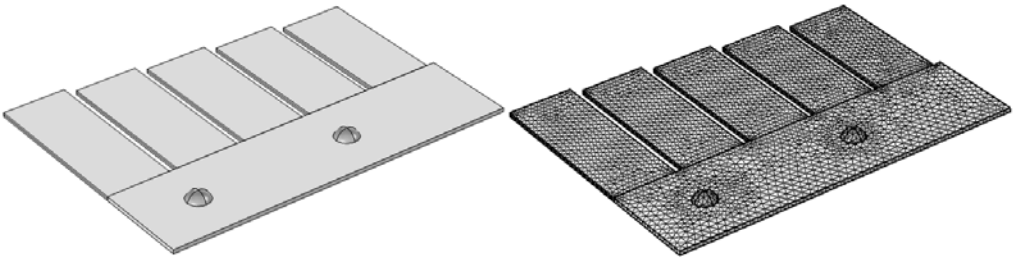


(a) (b)

Fig. 12.16. Experimental data of (a) output voltage and (b) output power versus load resistance of multiple UCB's at first fifth of mode frequency shapes.

The Eigen-frequency analysis is done with different modes of a suggested model beams as shown in Fig. 12.18 below. The resonance frequency of each vibration mode in Fig. 12.18 is summarized in Table 12.7.

In this section, we have described the effect of attaching a multiple unimorph cantilever beams with same thickness and different length of shim and piezoelectric layers at the front of aluminum beam on the output electrical parameters and bandwidth of both an experimental and modelling level. The experimental results show an increasing in maximum output power and bandwidth of multiple piezoelectric unimorph cantilever beams (UCB's comparing to single UCB. Connecting arrays of 5 commercial unimorph beams made of polyvinylidene difluoride (PVDF) in series showed increasing in output power from 2 μW to 5 μW and widening in the bandwidth from (47, 55) Hz to (22, 88) Hz. The first fifth mode shapes of five piezoelectric unimorphs cantilever arrays integrated in one aluminum beam are investigated theoretically using COMSOL multi-physics and experimentally, with a good agreement between model and experiment.



(a) (b)

Fig. 12.17. (a) Designed model, and (b) Meshing proposed system in COMSOL.

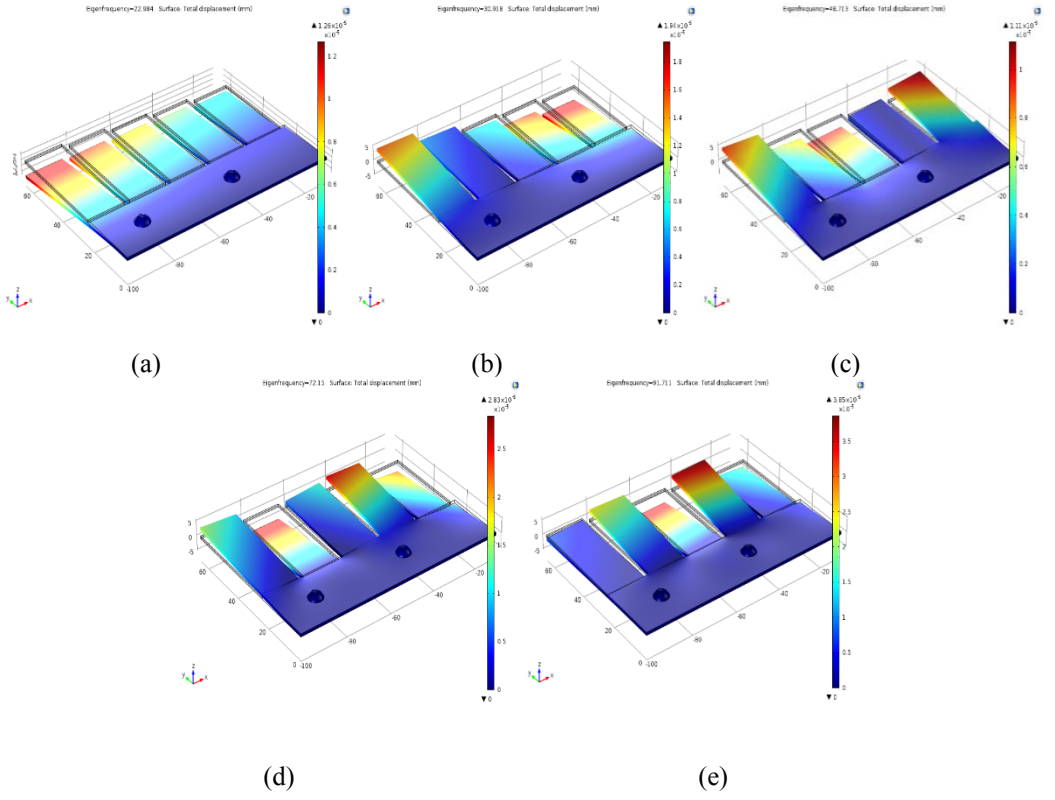


Fig. 12.18. Modeling the resonance frequency using COMSOL for multiple UCB's (a) First mode at 22.98 Hz (b) Second mode at 30.91 Hz (c) Third mode at 48.70 Hz (d) Fourth mode at 72.1 Hz (e) Fifth mode at 91.7 Hz.

Table 12.7. Experimental, and COMSOL resonance frequency results of multiple UCB's.

Cantilever beam type	Mode shape	Resonance frequency (Hz)		
		EXP.	COMSOL	Error (%)
Multiple UCB's	1st Mode	22	22.98	4.45
	2nd Mode	30	30.91	3.03
	3rd Mode	45	48.70	8.22
	4th Mode	65	72.10	10.92
	5th Mode	88	91.70	4.20

Acknowledgements

The authors gratefully acknowledge support for this work through the National Science Foundation grant #EPSCoR R-II-3 (EPS-1158862). Authors thank Dr. Chance M. Glenn, Dean, College of Engineering, Technology and Physician Sciences and Dr. M. D. Aggarwal, Chairman, Department of Physics, Chemistry and Physics for their keen interest in this work.

References

[1]. Lan. Li, Winnie Wong-Ng, Jeff Sharp, Polymer Composites for Energy Harvesting, Conversion and Storage, *Oxford University press,* American Chemical Society, Washington, DC, 2014.

[2]. M. Gad-el-Hak, MEMS: Design and Fabrication, *CRC Press,* London, 2005.

[3]. C-K. Mo, G-S. Ban, D. Charnegie and W. W. Clark, Energy harvesting from a vibrating piezoelectric unimorph bender, *J. Korean Society of Industrial Application,* Vol. 10, 2007, pp. 157-163.

[4]. C. Keawboonchuay and T. G. Engel Maximum power generation in a piezoelectric pulse generator, *IEEE Transaction on Plasma Science,* Vol. 31, 2003, pp. 123-128.

[5]. D. Koyama, and K. Nakamura, Array configurations for higher power generation in piezoelectric energy harvesting, *Japanese Journal of Applied Physics*, Vol. 49, 2009, pp. 1973-1976.

[6]. H-Y. Wang, X-b. Shan and T. Xie, An energy harvester combining a piezoelectric cantilever and a single degree of freedom elastic system, *J Zhejiang University-Science A (Applied Physics & Engineering),* Vol. 13, 2012, pp. 526-537.

[7]. H. S. Kim, J-H. Kim and J. Kim, A review of piezoelectric energy harvesting based on vibration, *International Journal of Precision Engineering and Manufacturing,* Vol. 12, 2011, pp. 1129-1141.

[8]. L. Mateu and F. Moll, Optimum piezoelectric bending beam structures for energy harvesting using shoe inserts, *Journal of Intelligent Material Systems and Structures,* Vol. 16, 2005, pp. 835–845.

[9]. N. E. du Toit, B. L. Wardle and S-G Kim, Design considerations for MEMS-scale piezoelectric mechanical vibration energy harvesters, *Integrated Ferroelectrics,* Vol. 71, 2005, pp. 121–160.

[10]. S-L. Kok, M. F. Abd Rahman, D. F. Weng and Y. Hwa Ho, Bandwidth widening strategies for piezoelectric based energy harvesting from ambient vibration sources, in *Proceedings of the Conference on Computer Applications and Industrial Electronics (ICCAIE 2011)*, Durian Tunggal, Malaysia, 4-7 December 2011, pp. 492-496.

[11]. S. Kim, W. W. Clark and Q-M Wang, Piezoelectric energy harvesting with a clamped circular plate: experimental study, *Journal of Intelligent Material Systems and Structures,* Vol. 16, 2005, 855–863.

[12]. S. Roundy, and P. K. Wright, A Piezoelectric vibration based generator for wireless electronics, *Journal of Intelligent Material Systems and Structures,* Vol. 13, 2004, pp. 1131–1142.

[13]. Y. J. Lai, D. G. Senesky and A. P. Pisano, Genetic algorithm optimization for MEMS cantilevered piezoelectric energy harvesters, in *Proceedings of the Conference on Power MEMS 2010, University of Leuven,* Leuven, Belgium, 2010, pp. 111-114.

[14]. Y. Ting, G. Hariyanto, B. K. Hou, S. Ricky, S. Amelia and C-K. Wang, Investigation of energy harvest and storage by using curve-shape piezoelectric unimorph, in *Proceedings of the Conference on IEEE International Symposium on Industrial Electronics,* 2009, Seoul, South Korea, 5-8 July 2009, pp. 2047-2052.

[15]. A. Erturk, and D. J. Inman, On mechanical modeling of cantilevered piezoelectric vibration energy harvesters, *Journal of Intelligent Material Systems and Structures,* Vol. 19, 2008, pp. 1311–1325.

[16]. D. F. Berdy, B. Jung1, J. F. Rhoads and D. Peroulis, Increased-bandwidth, meandering vibration energy harvester, in *Proceedings of the 16th International Conference on Solid-state Sensors, Actuators and Microsystems (TRANSDUCERS)*, Beijing, China, 5-9 June 2011, pp. 2638-2641.

[17]. G-Q. Wang, and Y-m. Lu, An improved lumped parameter model for a piezoelectric energy harvester in transverse vibration, *Hindawi Publishing Corporation Shock and Vibration*, Vol. 2014, 2014, pp. 1-12.
[18]. H. Wu, L. Tang, Y. Yang and C. K. Soh, A novel two-degrees-of-freedom piezoelectric energy harvester, *Journal of Intelligent Material Systems and Structures,* Vol. 24, 2012, pp. 357–368.
[19]. L. Dhakar, H. Liu, F. E. H. Tay and C. Lee, A new energy harvester design for high power output at low frequencies, *Sensors and Actuators A,* Vol. 199, 2013, pp. 344– 352.
[20]. L. Zhang, S. R. Oh, T. C. Wong, C. Y. Tan and K. Yao, Piezoelectric Polymer Multilayer on Flexible Substrate for Energy Harvesting, *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control,* Vol. 60, 2013, pp. 2013-2020.
[21 X. Tang and L. Zuo, Vibration energy harvesting from random force and motion excitations, *Smart Materials and Structures,* Vol. 21, 2012, 075025.
[22]. Y. Yang and L. Tang, Equivalent Circuit Modeling of Piezoelectric Energy Harvesters, *Journal of Intelligent Material Systems and Structures,* Vol. 20, 2009, pp. 2223–2235.
[23]. L. Meirovitch, Fundamentals of Vibration, *Waveland Press, Inc,* Illinois, 2000.
[24]. N. Elvien and A. Erturk, Advances in energy harvesting method, N. Elvien and A. Erturk, *Springer-Verlag,* New York, 2013.
[25]. J. Liu, H. Fang, Z. Xu, X. Mao, X. Shen, D. Chen, H. Liao, B. Cai, A MEMS-based piezoelectric power generator array for vibration energy harvesting, *Microelectronics Journal,* Vol. 39, Issue 5, 2008, pp. 802–806.
[26]. I. Sari, T. Balkan, H. Kulah, An electromagnetic micro power generator for wideband environmental vibrations, *Sensors and Actuators A*, Vol. 145–146, Special Issue, 2008, pp. 405–413.
[27]. S. Cheng, Y. Jin, Y. Rao, D. P. Arnold, A bridge voltage doubler AC/DC converter for low-voltage energy harvesting applications, in *Proceedings of the Power MEMS,* Washington DC, USA, 1-4 December 2009, pp. 25–28.
[28]. H. Yu, J. Zhou, L. Deng, and Z. Wen, A Vibration-Based MEMS Piezoelectric Energy Harvester and Power Conditioning Circuit, *Sensors*, Vol. 14, Issue 2, 2014, pp. 3323-3341.
[29]. W. Al-Ashtari, M. Hunstig, T. Hemsel, W. Sextro, Enhanced energy harvesting using multiple piezoelectric elements: Theory and experiments, *Sensors and Actuators A,* Selected Papers from the 9th International Workshop on Piezoelectric Materials and Applications in Actuators, Vol. 200, 2013, pp. 138-146.
[30]. L. Tang, Y. Yang, C. K. Soh, Advances in Energy Harvesting Methods, N. Elvin, A. Erturk, *Springer*, New York, 2013.
[31]. L. Tang and Y. Yang, A multiple-degree-of-freedom piezoelectric energy harvesting model, *Journal of Intelligent Material Systems and Structures,* Vol. 23, 2009, pp. 1631–1647.
[32]. W-Y. Chang and C-H. Hsu, Electromechanical characteristics of Polyvinylidene Fluoride for flexible electronics, *Transactions of the Canadian Society for Mechanical Engineering,* Vol. 37, 2013, pp. 325-333.

Chapter 13
Towards Tactical Military Software Defined Radio with the Assistance of Improved Data Gathering Tools

Tapio Saarelainen

13.1. Introduction

This chapter presents an idea-phase introduction of a tactical level (battalion and below) communication system to be used by a tactical end-user performing in a battlespace. The idea of the new communication system has been introduced earlier in [1]. As this chapter's contents represent an early idea stage of concept development, the system drafted and its features described have neither been operationalized nor field-tested as described in [1]. When creating any new functional system, the first step of the development process concerns outlining an idea of what a functioning system necessarily needs to comprise. This idea phase of a concept development process turns into a fully-fledged concept with operationalized features to be tested once the end-user devices discussed in this chapter have first been brought into being. Testing the devices described in an environment similar to what the book chapter outlines requires resources, i.e., policy-driven decisions allowing funding, personnel, and time. Furthermore, combat settings can be neither operationalized nor modelled in laboratories. The overall communication topology in a battlespace is challenging from the perspective of connectivity. The layout of the communication system can be introduced as depicted in Fig. 13.1.

The key issue in modern warfare continues to be communication. Without communication there are neither coordinated operations nor success. High Data Rates (HDR) are needed for the type of data necessary in Battle Management Systems (BMS) and in Command Posts (CP) where operations are commanded and controlled and orders issued onwards for tasks to be executed by lower echelons.

Tapio Saarelainen, Major, Title of Docent, IARIA FEELOW
Army Academy, Väinö Valveen katu 4, 53900 Lappeenranta, Finland, tapio.saarelainen@mil.fi
Lappeenranta University of Technology, PL 20, 53851, Lappeenranta, Finland, tapio.saarelainen@lut.fi

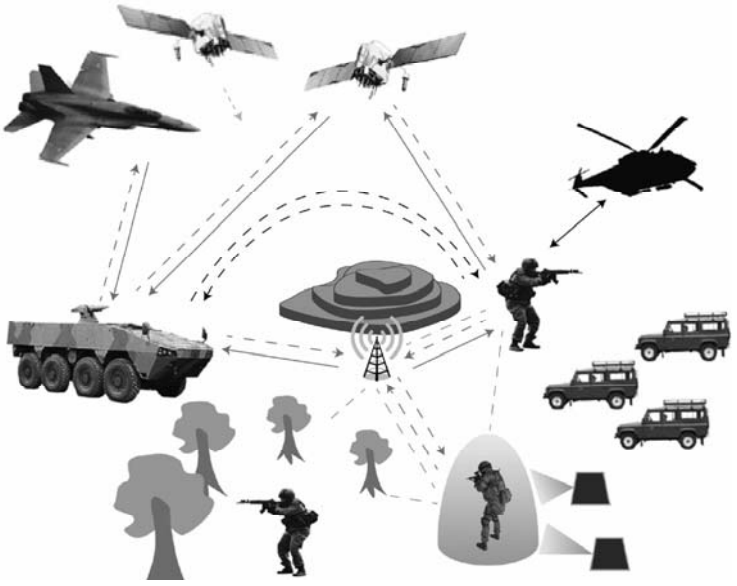


Fig. 13.1. Composition of network systems of Battle Management Systems.

Military environment is challenging also from the perspective of communications. Hostile military environment possesses challenges of several types for the need to communicate. First of all, the communication environment, a battle zone, is hostile. An adversary party tries to deny the free use of the frequency spectrum. Similarly, attempts of jamming the adversary's communication devices are typical of military actions executed in different frequencies and waveforms in the battlespace. Secondly, the soldier operating in a hostile territory using Cognitive Radios (CR) needs to establish mutual contact by using CR equipment to forming an ad hoc network. The challenge to create a functioning network in this case is exacerbated by the likely lack of accurate knowledge of the usage patterns of a radio spectrum in the hostile territory. If we compare this situation with a civilian case, where the frequencies and platforms are known in advance, the challenge in the military case has to be solved somehow in order to create a functioning communication network.

This chapter introduces the challenges of the Future Force Warrior from the perspective of a consumer of communication services. This issue is essential for both the research community and the relevant industry. Once problems are pointed out, the process of finding solutions is easier. This chapter presents one solution for how to facilitate fighters' need for constant capability to communicate in the battlespace.

This chapter examines tactical level military operating referring to commanded tasks being executed at the level of company and below. Both soldiers and commanders of any kind are dependent on radios to execute missions. A single soldier relies on radio communication in order to be commanded. This asks for a reliable communication tool and a robust ubiquitous network system that allows for precision and minimized collateral damage. A battlespace can be understood as an environment, where operations are being

executed including land, sea, air, underwater and cyber operational environments. Warriors, sensors and Unmanned Vehicles (UVs) of several types operate and communicate in the versatile, constantly changing battlespace. As indicated in [2], militaries are using sensors as part of their battlefield strategy. As mentioned in [2], the integration of data collecting capabilities is in an essential role in expanding the communication platform capabilities. The key issues involve collecting the data, analyzing these data, and forwarding the analyzed data reliably and in an intact form to the end-user, a Future Force Warrior (FFW). Networks have to be organized to cover the needs of the end-user at all levels, as explained in [3]. Issues such as Quality of Service (QoS) and Speed of Service (SoS) are seminal in tactical communications [3].

Militaries concentrate on sustaining and developing their capability to communicate in a battlespace. To optimize performance, a Future Force Warrior needs only one communication device, which can be Software Defined Radio (SDR) for the reasons listed. First, one communication device covers all the communication needs of an FFW instead of him or her needing to use several communication devices. The focus of the FFW has to be in fighting. This means that an FFW keeps his or hands around the weapon, monitors the threats in the battlespace and fights. Secondly, mobility- and action-critical matters from the perspective of an FFW, such as size, weight, and power and cost (SWAPC), are relevant. Thirdly, one power source should cover the need of communication devices instead of several sources (i.e., power source for radio and Personal Digital Assistant). Lastly, one communication device will ease the communication burden of an FFW and he or she can focus on the main functions: to monitor the environment and fight to survive. Present military communications are based on combat net radios (CNR), dominated by the enhanced position location reporting system and the single channel ground and airborne radio systems. This means that the varying levels of tactical communication can comprise several actuators. The distances between the communicating elements can vary from only a few meters to tens of kilometers and more. Tactical communication utilizes unmanned vehicles, drones and satellites acting as hubs or relay stations. The term tactical refers to the operative capabilities of a given military force. For example, a maneuver, which is tactical for the U.S. Army with its special forces, can be an operative maneuver for an army smaller in size and its operative capability.

The basic problem in communication is that the High Data Rate gives shorter range in communication (e.g., 4 × rate = ½ range). Therefore we have to solve this problem with different means than just increasing the data range with increased transmission power. Despite the system characteristics, the communication system for military use in lower echelons (i.e., companies and below) has to fulfil the requirements of operational security, coverage, connectivity and Low Probability of Detection (LPD) and Low Probability of Identification (LPI). Military operations are dependent on covert high-speed networks, which also represent functional requirement of modern infantry and special operations warfare [4].

This chapter introduces an idea-phase solution for utilizing a swarm of Unmanned Aerial Vehicles (UAVs) and SDRs. Swarms of UAVs can be recognized as parts of Unmanned

Aircraft Systems (UAS). The swarm of UAVs is seen as a platform for a communications system, in which the distances between the UAVs must be short to ensure the message throughput in a hostile communication environment. This aims at ensuring a reliable data exchange process and fulfilling the requirements of LPD and LPI.

SDRs are used by FFWs performing at the tactical level and also embedded into each UAV to ensure a reliable data exchange process. The swarms of UAVs in this system are in a central role to ensure the message throughput in a case when one or several UAVs are destroyed. Once a UAV becomes incompetent to act as a relay-station, a neighboring UAV takes over its functions and, with the assistance of Self-Organizing Networks (SON), the routing of communication can be reorganized and ensured. This chapter examines how to enhance the performance of an FFW to allow for executing the set tasks.

This chapter is organized as follows: Section 13.2 discusses Unmanned Aircraft Systems (UAS), when Section 13.3 concentrates on Service Oriented Architecture (SOA) in military operations. Section 13.4 focuses on Military Communication Environment, Section 13.5 deals with the Challenges of Future Force Warrior, Section 13.6 talks about SDR, and Section 13.7 looks into Universal Software Radio Peripheral (USRP). Section 13.8 concentrates on explaining the idea of Cognitive Radio (CR). Section 13.9 discusses the significance of a Graphic User Interface, and Section 13.10 introduces the proposed communication system. Section 13.11 introduces the composition of the set up and utilization of Sensor Element Munitions, Section 13.12 focuses on airborne sensors and Section 13.13 on a comprehensive targeting process. Section 13.14 concentrates on means to analyze collected data, Section 13.15 comprises discussion, Section 13.16 features results and Section 13.17 concludes this chapter.

13.2. Unmanned Aircraft Systems

Representing an integral part of battle command and warrior systems, Unmanned Aircraft Systems (UAS) enable the warrior to collect data and remain in contact with the higher echelons, report and receive tasks in near real-time. This contributes to the overall performance capability of warrior systems in Net centric Warfare. In light of recent development concerning the technology of UAS, most UAS tend to be designed for military purposes [5]. Military operations executed with the assistance of UAS require that the flight plan be tailored to meet the needs the mission to be accomplished. UAS solutions rely on waypoint- based control systems unless the UAS are remotely piloted. In this chapter, the use of UAS is based on the flight-plan which is generated with the assistance of Service Oriented Architecture to gain the most optimal reconnaissance results from the designated Area of Interest (AOI).

UAS are currently used primarily for military applications, but with the evolution of avionics technology, a market in civil applications keeps emerging [5]. Military operators can benefit from this evolution from the perspective of planning and optimizing the flight plan, payload and mission planning based on inventions available in the form of Commercial-Off-The-Shelf (COTS) products.

As indicated in [6], the planning of sequences to automate the UAS mission management is in an essential role. For instance, in UAS Service Abstraction Layer, the following services need to be carefully planned: flight, mission, payload and awareness services. In addition, the processing systems of gathered data have to be planned. This includes accounting for real time data processing and storage services for the accrued data. Apart from serving as a tool in tailoring flight plans, SOA can be utilized in the processes concerning mission planning and data gathering.

13.3. Service Oriented Architecture

Allocating resources at all the planning and implementation levels of an organization is enabled by means of relying on Service Oriented Architecture (SOA). This ensures an optimized utilization of the given resources rationally and effectively. Being able to utilize Service Oriented Architecture in a military operation requires comprehending SOA's capabilities and nature and being aware of the services available to allow for organizing a communication system to support a given military operation.

SOA can be used in different phases of operational planning. The use of SOA supports the use of Unmanned Aircraft Systems in mission planning, in flight route planning and in the processes of accruing data from designated areas. When planning the detailed flight routes of unmanned aerial vehicles or routes for unmanned aircraft systems, SOA can serve as a tool in optimizing the flight time, flight altitudes and the holistic process of data accruing. Also, SOA can be benefitted from in the process of data mining and data fusion when the accrued data have been successfully collected.

When needing to affect the sequencing and pace of events in the name of increasing efficiency, SOA is needed as an accelerator. This means that using SOA makes the planning process efficient by saving time as redundantly layered tasks and processes may be omitted. In a successful military operation assisted by SOA, the end result can equal an operation with minimized instances of collateral damage and fratricide. Moreover, a successful operation means efficient use of resources (ammunition, troops, vehicles, medical support and time). Thereby SOA is viewed as an aid for a military commander in the decision making process, for instance, when choosing Courses of Action (COAs).

SOA enables organizations and entities to operate complicated systems and enhance interoperability and collaboration, see [7], and foster the reusing of components and interfaces. SOA can be used in service collaboration. With the correct framework, SOA allows publishing services in a service registry and exchanging data through the Simple Object Access Protocol (SOAP) [8]. SOA offers an adjustable solution for systems integration, applications, protocols, data sources and processes to form a cohesive system that supports the execution of critical BPs [9]. SOA can be used as a collaboration tool in crises management and military environments if the challenges of real-time SOA [10] are solved.

In Network Centric Warfare (NCW) contexts, SOA has been recognized to act as an enabler of services. SOA is an architecture style that encourages loose coupling between

services to facilitate interoperability and the reuse of existing resources as described in [11]. SOA serves as a tool in enabling agility to handle the changing dynamic evolution needed in network enabled capability, see [12]. The concept of NCW can be viewed as an integration of assets to meet a mission objective, as discussed in [12].

13.4. Military Communication Environment

Enabling an organization to meet its set objectives effectively requires that communication systems be in place to watch over the battle command systems and their executors, machines and humans included. Military communication environment differs from its civilian counterpart. Any civilian communication environment tends to be non-hostile and its features thoroughly known in that the transmission distances, frequencies and waveforms used are common knowledge. In contrast, the military communication environments are part of a battlespace and abound in uncertainties in connectivity and latency may vary uncontrollably due to incessant hostile electronic warfare attempts. Communication break-down in a battlespace typically results in compromising someone's life.

A given battlespace comprehends also the communication environment, in which war is waged. Military frequencies tend to be mandatory and always commanded from higher echelons in order to control the electromagnetic spectrum most effectively from the execution perspective of own military operations.

Military troops equipped with varying end-user devices transform the battlespace of 21st century into network centric warfare with Network Centric Operations in a central role. SDRs will provide a flexible tool suited for the changing military environments in that they allow versatile communication in the battlespace [13]. In a constantly changing battlespace the commanded troops can be mobile or static. Often the communication tools, end-user devices, are handheld and lightweight. A ground level military performer, soldier, has to be able to communicate also via satellites in order to contact higher echelons, for example, when executing special operations in rocky terrain. The size of the used system, together with its weight, power and cost (SWAP-C) become vital from the perspective of the system user and provider.

For the end-user, to be able to cover multiple battlespace scenarios, the simultaneous requirement of communication requirements such as voice, video and data together with the capabilities and megabit bandwidths set design challenges. Moreover, to sustain secure communication by means of the end-user devices, new military waveforms have been designed to fulfil the requirements of the end-user, a soldier. An example of such waveforms suitable for an SDR-based system implementation is WiMAX 802.16e, which has been modified to operate in the military frequency range of the NATO UHF band of 225 – 400 MHz [14].

A future military SDR-platform should support multiple radio frequency frontends. Depending on the available and sufficient frequency bands, different frontends could be installed. The flexible use of different frontends and waveforms enable finding a suitable

system configuration for all the planned operational scenarios. The next generation SDR-based platform should enable at least the following benefits for tactical networking: First, mobility support for mobile ad hoc network (MANET). Second, sufficient communication capacity must be guaranteed at the tactical level, this is a minimum throughput of 1 Mbps to support mobile user. Third, from a perspective of life cycle management, a SDR platform must be independent from the waveforms and frequencies used. Fourth, communication flexibility has to be ensured with radio frequencies in SDR frontends and with used waveforms. Fifth, interoperability with national and coalition waveforms has to be granted [14].

In a civilian communication environment an end-user can benefit from reliable and fast communication, high throughput of messages, issues of low latency, the constant capability to communicate, adequate bandwidth, good Quality of Service (QoS), and Speed of Service (SoS). Civilian communication systems offer the possibility to benefit from constant power supply or the capability to recharge the battery of the used communication device when necessary. The communication process usually suffers from only slight if any hostile interference or jamming.

Communication systems utilized in a military environment can confront all types of interferences. These include jamming and all means electronic warfare with a constant threat of becoming annihilated by the adversary if a communication tool has been detected, pinpointed, and placed in the targeting process to be destroyed. Table 13.1 lists the differences between civilian and military communication environments.

As Table 13.1 indicates, in a military communication environment the characteristics of communication involve restrictions and constant uncertainty due to hostility in the battlespace.

Table 13.1. Differences between civilian and military communication environments.

Characteristics of Communication	Military communication environment	Civilian communication environment
Free use of spectrum	Restricted	More possibilities
Latency	Varying, sometimes high	Typically low
Energy	Limited, hard to recharge	Possibility to recharge fast
Hostility	High	Low
Jamming	Possible	Low
Adequate bandwidth	Limited, altering	Typically high
Limitations in use	Often restricted	No limitations

13.5. Challenges of the Future Force Warrior

Seminal in influencing the development of events in combat, FFW performance capability presupposes that the warrior be in the right location and execute the correct tasks in a time critical environment. The following overview lists six challenges which have been identified within the military community. The writer has encountered these challenges

while conducting research on related issues, such as a nationwide Company Attack Study performed during 2004 – 2007 in Finland. The first challenge is related to the main task of a fighter: the main task of an FFW is to fight in performing the given mission. This means he or she has to monitor the environment to stay alive and to be able to execute the commanded tasks. He or she engages the enemy with all the weaponry available. This also means that the FFW relies on connectivity and capability to communicate at all times. The constant connectivity poses the second challenge. Mission success requires the capability to transmit and receive data and commands. The constant communicating ability requires that a single warrior be capable of acting as an executor of an operation, or a military commander at some level. Connectivity remains the key. Some sort of a network must be available at all times. Location data and commands can be forwarded only by means of a functioning network and a reliable hand-held or soldier-mounted communication device.

The third challenge is linked to the usage of warrior platforms. A warrior platform consists of several subsystems and their control units. For example, the systems can be controlled via a wrist-worn user-interface presented here. All the communication controls can be easily and rapidly found from the wrist-held device which is embedded onto the arm and acts as a supporting hand when using a personal firearm. Contrary to a visor-embedded system, this wrist-held device does not hamper viewing the environment with a constant data flow. Fig. 13.2 features a wrist-held device.

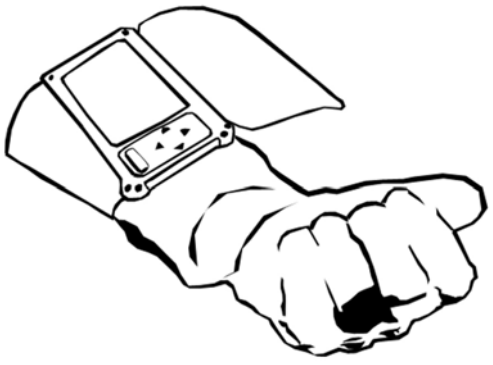


Fig. 13.2. A wrist-worn control system into which SDR can be embedded [1].

The fourth challenge involves the number of networks and data sources on which an FFW relies. If the communication network or operating unit malfunctions, FFWs get in trouble because they lack the necessary resources either because of the different frequency of waveform used or the network becoming out of coverage. This slows down a single FFW and usually harms the whole military operation.

The fifth challenge equals the access to different types of Battle Management Systems (BMSs). BMSs support the efficient utilization of military units at all levels. The access process into a BMS requires more bandwidth than using voice and text messaging when issuing commands. Fig. 13.3 features BMS.

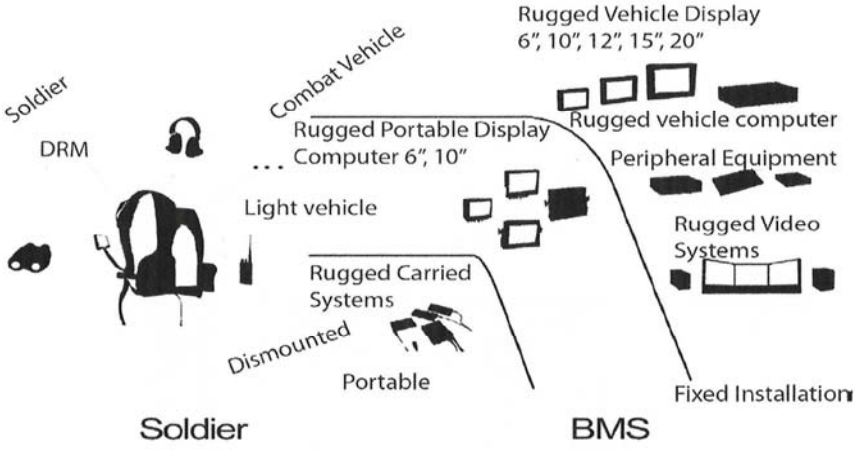


Fig. 13.3. Soldier Systems linkage into Battle Management Systems [1].

This sets more bandwidth demands for SDRs used. The data may surface in varying waveforms and frequencies. In order to benefit from BMSs, the location and identification of friend or foe are relevant. These data are needed from the battlespace to ensure the effective use of different weapon systems.

Systems related to location data and identification systems keep improving. The command and control capabilities necessary for units serving at battalion level and below are provided by means of BMS. Present combat net radios can only be seen as beneficial tools in supporting geographical based situational awareness. When several operational BMSs are integrated, it is possible to maximize the amount quantitative and qualitative data for analysis purposes. This involves applying the concept of System of Systems (SOS) and utilizing robotic platforms. The sixth challenge is the type of a Graphic User Interface (GUI) of a communication device, which can be used for various purposes. GUIs can be utilized in creating the Situational Awareness and Common Operational Picture as well as for location services and for weapon selection process. Although one of the challenges is the variety of used waveforms and the bandwidths, they fall outside of the scope of this study. This chapter discusses a possible solution for the listed challenges.

13.6. Software Defined Radio

Software Defined Radio stands for a developed communication device for an FFW in that its qualities are changeable through programming. SDR is a radio communication system in which components that have typically been implemented in hardware, for example, mixers, filters, amplifiers, modulators/demodulators, and detectors, are instead implemented by means of software on a personal computer or an embedded system. Usually SDR can be programmed to support frequencies from 100 MHz to 6 GHz with using 130 nm Complementary Metal Oxide Semiconductor (CMOS) technology [15]. Depending on the system configuration, typically supported signal bandwidths can vary between 700 kHz and 40 MHz or 200 kHz and 40 MHz, depending on the CMOS used

[15]. In other words, SDR is a wireless communications system where the traditional hardware is replaced by software modules [16]. While the concept of SDR is not new, the rapidly evolving capabilities of digital electronics enable executing many such processes which earlier used to be only theoretically possible.

Currently, most SDR related products and studies focus on analog communication and voice transmission. The SDR platform consists of Field Programmable Gate Array (FPGA) -based radio hardware and open source SDR software module [16].

The main features of SDR include: 1) Radio spectrum sensing; 2) Reconfigurable radio modules and 3) Link for digital data communication. These features form an important basis to accomplish Cognitive Radio technologies.

The mobile devices can afford the high speed and complex computation owing to the advance in computing ability of the processor, such as Personal Digital Assistant (PDA), Smart Phone, or Ultra-Mobile PC (UMPC).

Most of these mobile devices equipped with Wi-Fi, WiMAX or other wireless modules enable end-users to access services anywhere. The traditional hardware radio system comprises a variety of analogy elements such as filters, converters, modulators and demodulators. The hardware is expensive and has low compatibility with other components.

The reason why SDR becomes increasingly popular is that it allows using SDR technology for realizing many applications relatively effortlessly in the integration of different components. The most used software architecture for SDR is the Software Communications Architecture (SCA), which is considered as the standard for military domain [13]. The novelty of SCA lies in the availability of SCA-based tools to allow designers to create component-based SDR-applications as assemblies of components and logical devices. When these types of systems are being created, the communication between the components and devices must be carefully orchestrated. In this process it is possible to benefit from the use of Common Object Request Broker Architecture (CORBA) [13].

When discussing SDRs, security issues must be considered. When new Software (SW) is being loaded, the consequent threat of having unauthorized and potentially malicious SW installed on the platform becomes possible, if security precautions have not been taken, by, for example, adding a digital and verified signature in the code before the new software is being transmitted.

In order to successfully benefit from the products and performance of SDR, we have to focus only the performance produced via SDR. The added value from SDR can be seen via tactical communication requirements for the FFW operating in Battlespace which are: Situational Awareness (SA), Common Operational Picture (COP), Command and Control systems, identification friend or foe (cf. Fig. 13.4), (IFF)/Blue Force Tracking, capability to co-operate with UAVs and Unmanned Ground Vehicles (UGVs) and robots, data from sensor to shooter, Voice, Navigation, messaging, Imaging, Video, Security.

One interesting possibility is to embed Radio Frequency Identification (RFID) system into SDR by using Quadrature Amplitude Modulation (QAM). The system has been explained in [17]. The identification friend or foe (IFF) process can be embedded as part of SDR functions. Fig. 13.4 features an IFF process.

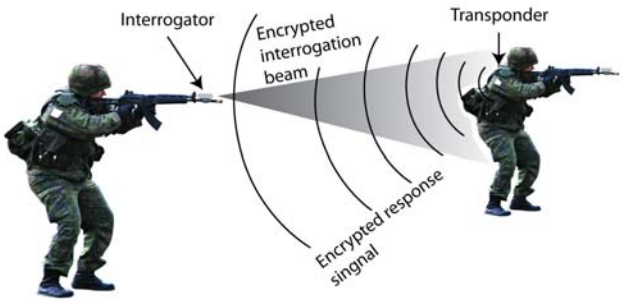


Fig. 13.4. Identification process Friend or Foe (IFF) in progress [1].

By simply downloading a new program, a SDR is able to interoperate with different wireless protocols, incorporate new services, and upgrade to new standards. One solution is depicted below above while the process is introduced in Fig. 13.5.

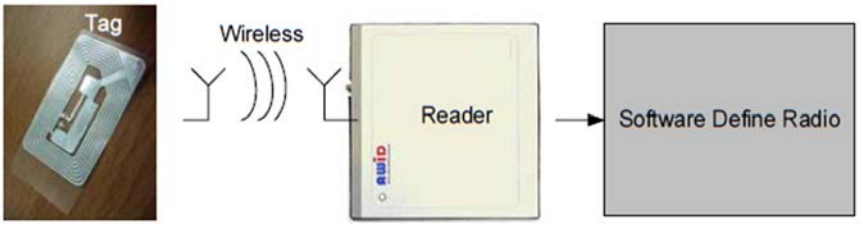


Fig. 13.5. RFID system with SDR [17].

As Fig. 13.5 indicate, by combining the Radio Frequency Identification tags and IFF – process it is possible to decrease the total mass of the gear a FFW carries. Similarly, this enables simultaneously decreasing the amount of transmission energy necessary for identification purposes.

RFIDs can be sensitive to electromagnetic interrogation signal by nature and need little energy when responding once only a slight transmission signal focuses on them. Thereby the amount of response energy transmitted towards of the interrogator can be significantly small. This means Low Probability of Detection as regards the surveillance tools spread in the battlespace. Relevant for operational security purposes, the interrogation process remains undetected and discreet.

13.7. Universal Software Peripheral Radio

In Universal Software Peripheral Radio, we examine only the functionality of USRP. In doing so, we notice it offers more performance than its predecessor SDR. The technology used in USRP is located in the hardware implemented in frontend for sending and receiving waveforms. USRP offers different frequencies, bandwidths and frequencies for specific purposes. The USRP can be fixed to respond to the end-users' requirements by selecting appropriate motherboards for controlling the frequencies and waveforms [16].

USRP can be divided into two parts based on the transmission path. These are the transmitting signal path and receiving signal path. For example, on transmit signal path, users can define the setting parameters by software on personal computer such as radio protocols, modulation types, frequency of spectrum modulation. Then the USRP receives the parameters, and FPGA executes Intermediate Frequency (IF) processing on Digital Up Converter (DUC) and Digital Down Converter (DDC). After the Intermediate Frequency process, users adjust the baseband to the frequency band selected before.

The last step on USRP motherboard is that digital to analog (DAC) converts the digital signal into analog signal. Finally, the analog signal is transmitted to the antenna through the interface side on the daughterboard, as illustrated in USRP block diagram in Fig. 13.6.

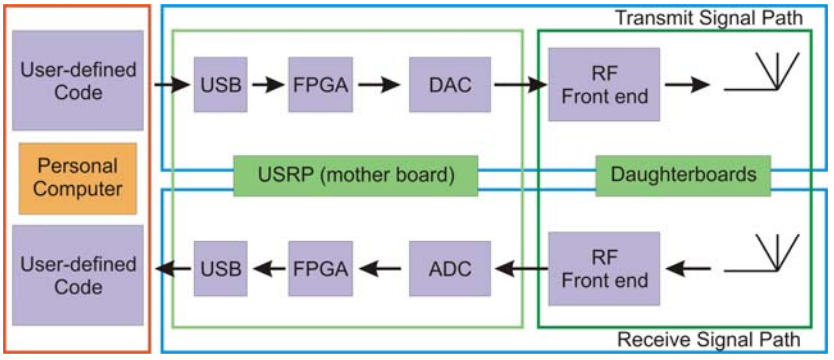


Fig. 13.6. Composition of USRP [16].

As Fig. 13.6 demonstrates, the composition of the introduced USRP system offers flexibility in using different waveforms and frequencies. The flexibility can be offered by different daughterboards which can be tailored to meet the requirements of different frequencies and waveforms. By changing and tuning the performances of daughterboards, the FFWs have improved communication devices as communication tools in a constantly altering battlespace and varying missions.

13.8. Cognitive Radio

When moving on towards the communication device suitable for an FFW, we have to take a quick glance at Cognitive Radio (CR). As widely known, Software defined Radio is a

platform for Cognitive Radio [18]. Without going into the details of the technical structure or composition of CR, we focus on the listed and wanted end-products and functionalities of the CR from an FFW perspective.

Cognitive Radio capabilities and functionalities include the following features. First, Spectrum Awareness (SAw), which means being able to detect quickly and robustly the presence of incumbent (preemptive) users to avoid causing interference. Second, Dynamic Spectrum Access (DSA), which means CRs will access the spectrum on an opportunistic basis. Third, Dynamic Spectrum Sharing (DSS), which means CRs must be aware of other CRs' coexistence. Lastly, CRs are Spectrum Agile (SAg), which means that CRs should provide seamless operation over multiple channels. Also challenges related to adaptive coding, modulation and multi-access have to be solved as indicated in [18].

There is a long way towards CR, which meets the listed requirements. From the perspective of an end-user, a consumer, an FFW, it is essential to meet the requirements listed. This also applies as regards the scientific community and the industrial community. The necessary requirements have to be identified prior to being able to produce a wanted end-product.

One solution for designing the suitable architecture and configurations for the future force radio communication device could be Software Communications architecture. SCA has been created to assist in the development process of SDR communication systems. SCA allows for waveform application software to be more easily ported across radio platforms. At the moment publicly available specifications can be found for SCA 2.2.2 and 4.0, as well as for SCA Appendices and SCA APIs. As a matter of fact, it is possible in next generation products to take full advantage of the following: First, SCA 4.0, which is to empower more freedom to do the SDR implementation. Second, Programmable SDR chip sets. Third, offering more efficient SDR development tools and use of more efficient higher level modelling methods. Fourth, adding a new approach to waveform portability and full utilization of SDR work done in commercial domain. Lastly, designers are focusing on developing of sophisticated RF front-end technologies.

SOA-technology involves assisting processes performed in military operations. As indicated in [19], the SOA has been used to design and construct the CR systems. When an FFW can benefit from the possibilities offered by a successful adoption SOA, also in communication services, the result can be improved overall performance in military operations.

Fig. 13.6 features how the data are gathered, processed, analyzed and then transmitted as commands to an FFW. If the data are correctly collected, analyzed and successfully transmitted to the performer, an FFW, the process of waging war can be improved and collateral damage minimized. Various battlespace sensors transmit data to a context-aware reasoning layer. In this layer, data are converted to context and an inference engine transmits the data to a ubiquitous main layer for analyzing purposes. The data are verified, analyzed and transmitted as information for the execution of the operation through several interfaces described in [20]. This process of data flow is depicted in Fig. 13.7.

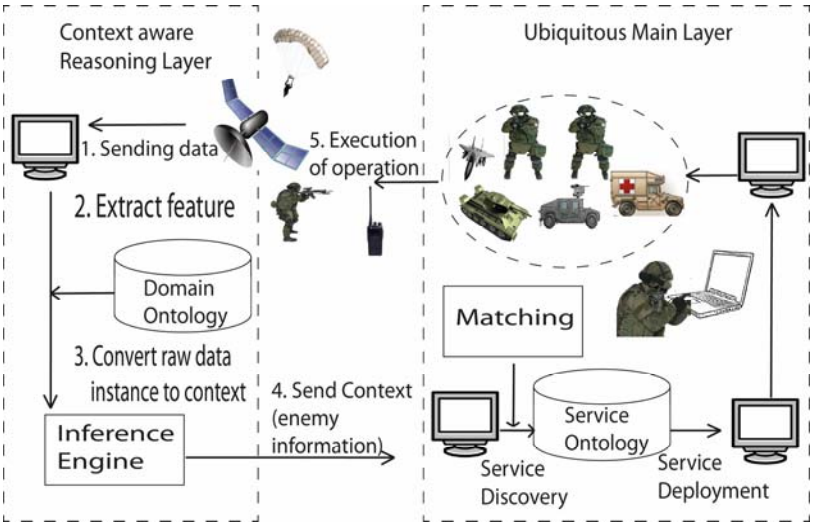


Fig. 13.7. Performance can be gained via successful data utilization [11].

Once the collected data have been analyzed, they can be forwarded to the military performers who need these data most. The transmission process has to be automated to ensure sustaining overall performance.

13.9. Graphic User Interface

Forming a picture of the incessant change occurring in a battlespace presupposes that an FFW be able to do this visualizing both by means of observing and receiving the information transmitted via electronic end-user devices. This calls for the presence of a Graphic User Interface (GUI) with its functionalities. A new type of communication device for an FFW has to fulfil the specific communication requirements set by the FFW's combat performance capability. The communication system and the GUIs have to be defined to meet these needs. Fig. 13.8 features the actuators affecting the system definition process.

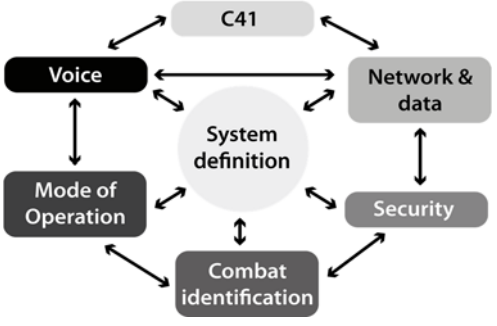


Fig. 13.8. The idea of configuration of the future communication device [1].

According to results presented in [21], the visualization of events can improve the human capability to accelerate the Military Decision Making Process (MDMP) by offering necessary information in required time and understandable form. From the perspective of a consumer, an important role is set for the type of a GUI. A functional GUI is a means to present collected data, a control panel to access networks and guide UVs, a tool for a weapon selection process, and, of course, a communication tool for the entities of higher and lower echelons. Fig. 13.9 features one possible figure caption of a functional GUI.

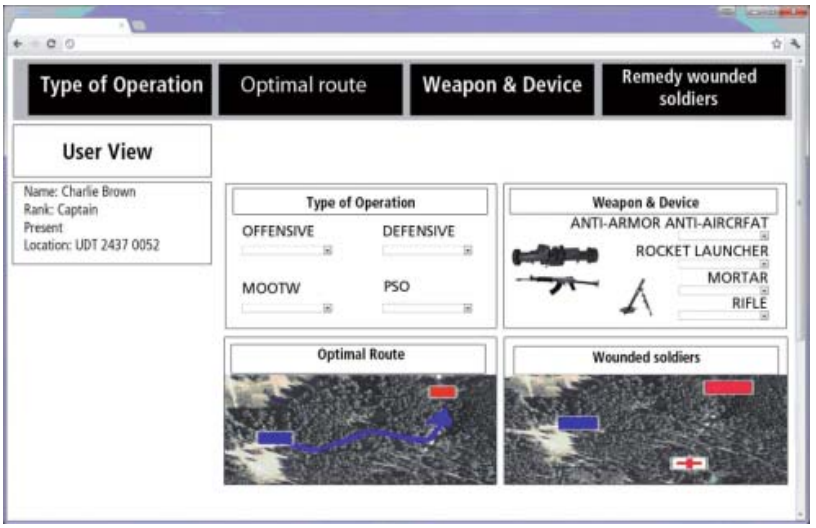


Fig. 13.9. A view of a Graphic User Interface [1].

Apart from the mentioned facts, an FFW has to be able to access BMSs of various types and different databases with a new type of communication tool. Fig. 13.10 features one type of BMS where an FFW can be constantly connected to optimize the performance while executing tasks.

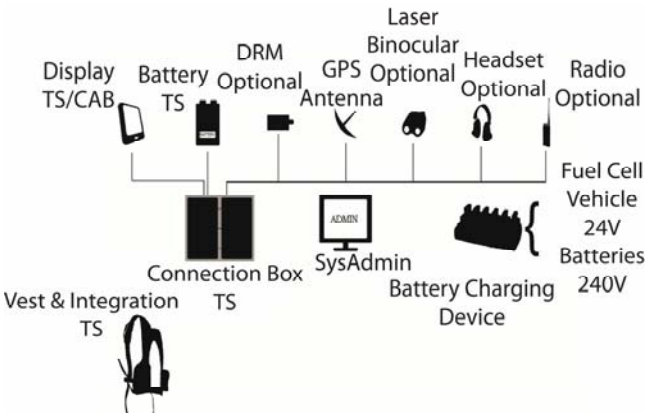


Fig. 13.10. A view of a BMS [1].

Lastly, an FFW has to be able to control the systems embedded on his Battle Dress Uniform (BDU). Fig. 13.11 features the FFW's electronic skeleton and its functions [22].

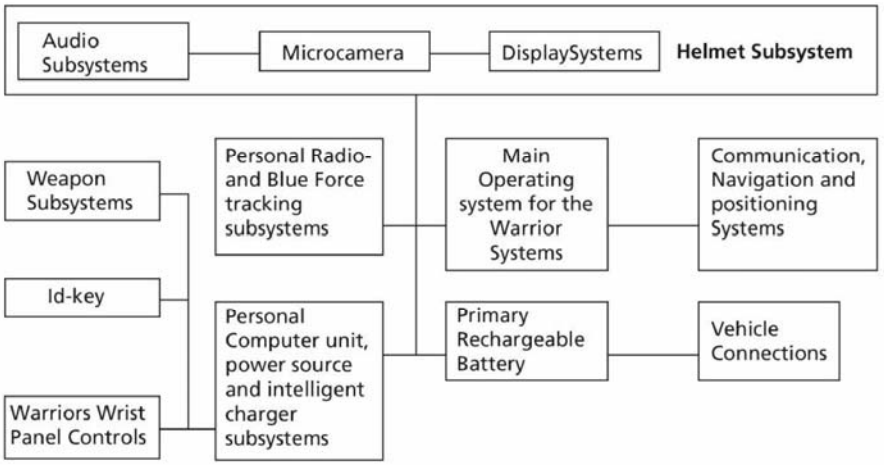


Fig. 13.11. A composition of a FFW's electronic skeleton [22].

An FFW uses the control unit embedded into his BDU as an essential tool to control and monitor the functions of own gear. In Fig. 13.12, the controlling system is wrist-worn.

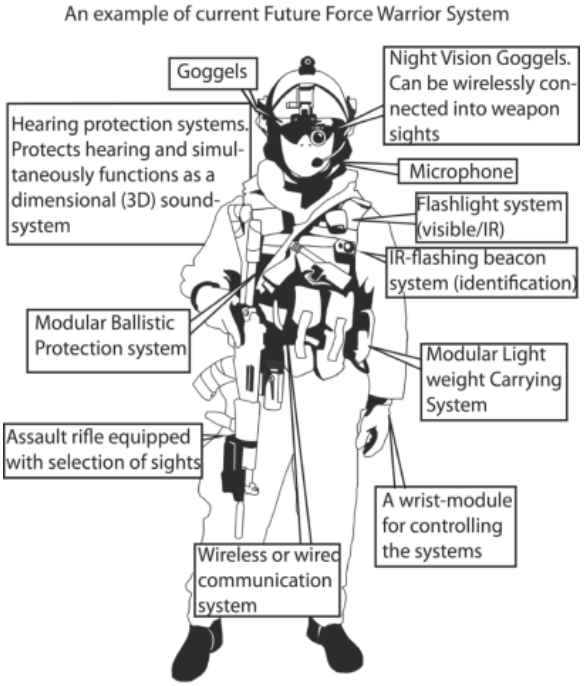


Fig. 13.12. An example of an FFW as a platform with selected gear [23].

By embedding all the control units of the FFW's electronic gear into one wrist-worn controller, the number of controlling units can be decreased. This may increase the overall performance of the FFW as the FFW can find all the control units in one location instead of needing to separately control each different embedded system listed in Fig. 13.12.

13.10. A New Communication System

As noted in [24], mobile ad-hoc networking of dismounted combatants is necessary as regards the future involving net centric operations. The amount and variety of data transmitted in the battlespace keep increasing. Issues such as bandwidth, type of waveform, frequency and security are only a few of the issues that have to be accounted for. Low Probability of Detection and Low probability of Identification remain critical in covert operations, as mentioned in [24]. Single UAVs are utilized as tools tailored for Special Forces and a system relying on Advanced Encryption System (AES) encrypted network with a range of 3 kilometers [21].

A new communication system is possible to create if we utilize the capabilities of SDR, swarms of UAVs, SDR and Self-Organizing Networks (SON) implemented in 4G networks. It has to be highlighted that in this system, SDRs are implemented in FFW gear and inside an UAV. As noted, SON aims to configure and optimize the network automatically, in a manner that the interaction of human can be reduced and the capacity of the network can be increased.

The main functionality of SON includes the following: self-configuration, self-optimization and self-healing. SON is described as a part of 3GPP LTE and it is a key feature for effective and automatic operation and maintenance (O&M) of 4G networks. Besides that, SON maximizes overall performance of network and reduces the cost of installation and need of management by simplifying operation and maintenance through self-configuration, self-optimization and self-healing. SON also reduces the power consumption and results in reduced operational expenses and produces an environmentally friendly approach. Fig. 13.13 features the SON as seen in [25].

When a swarm of UAVs is utilized, the distances needed to communicate with an FFW-worn SDR must be minimal in order to ensure the message throughput in this system.

Different types of data can be transmitted from a soldier to a higher echelon via an UAV. Security issues remain essential when dealing with UAVs utilized in Network Centric Warfare at a tactical level. This means opting for low transmission power and thus minimizing the chances of the UAV becoming detected, targeted, and destroyed.

One possible solution for increasing the security of transmission is presented in [26]. When a secure transmission protocol is offered, the accrued data can be utilized also at the lowest level of operations, namely in the use of systems an FFW is composed of, the electronic skeleton, see [22]. When these data are available, instances of fratricide may be minimized with the system described in [27]. One possible solution to benefit from the use of UAVs is explained in [28].

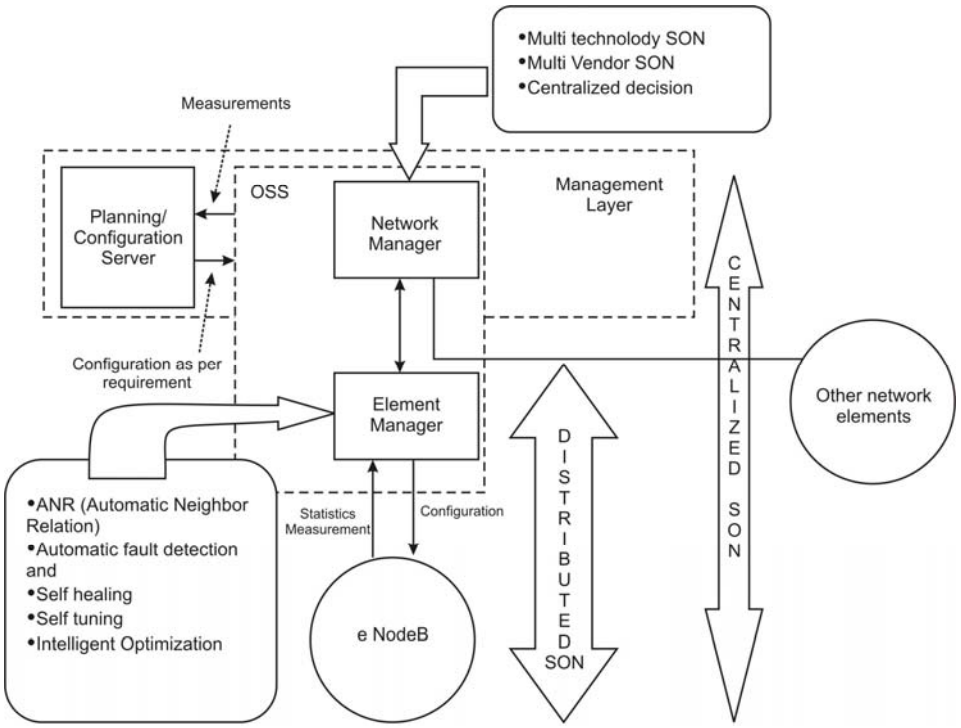


Fig. 13.13. A composition of a SON implementation in 4G as seen in [25].

When transmission distances remain short between the ground (FFW) and aerial stations composed of swarms of UAVs, the accrued and transmitted data can be better secured and requirements of LPD and LPI can achieved. When SON utilizes all the SDRs, those embedded into the UAVs and those embedded into soldier worn systems, this data exchange can be executed successfully and thus ensure that the data remain intact and coherent. Fig. 13.14 features the data-exchange process via a command post and UAVs with the assistance of embedded SDRs into the mentioned entities.

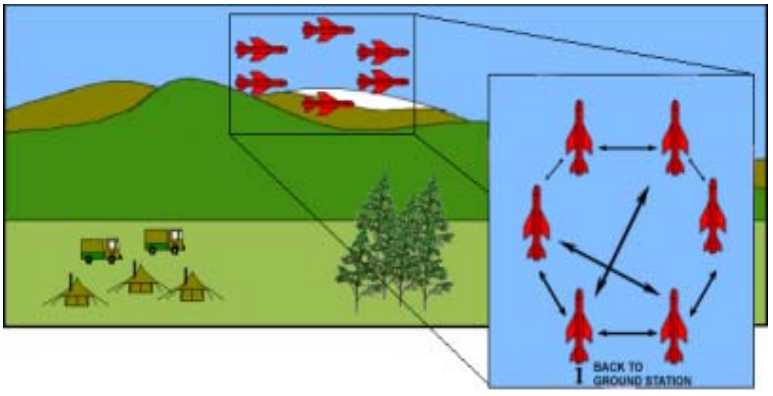


Fig. 13.14. A data exchange process with the swarm of UAVs [1].

The swarms of UAVs will forward the data automatically via the network system created by UAVs. Demands of LPD and LPI can be fulfilled, because the transmitting energy used via the transmission protocols by means of UAVs remains low.

13.11. The set-up and Utilization of Sensor Element Munitions

Situation Awareness data must be collected from the battlespace by any possible means. Rapid to deploy and effective in short range, Sensor Element Munitions (SEMs) has been devised for this purpose. The data collecting and reconnaissance carried out by means of SEMs take after the standard High Explosive ammunition used in mortars and howitzers. The main difference is in the payload, in which the explosive charge has been replaced with a parachuted Sensor Element (SE).

This SE is strong enough to withstand the forces of acceleration of a regular munition. The munition is delivered to a hostile area with similar procedures as standard High Explosive munition. The SE acting as a payload will be exhausted from the ammunition shell while airborne.

Structurally, the SE comprises a power source, an array of sensors, a transmitting unit and a relay-unit. The SE can act simultaneously in two roles: in accruing data and in a relaying role between two SEs. The SE does not receive data, but only transmits the data gathered, including GPS-data of its own position. The SE comprises sensors such as a visual sensor applicable to monitor targets both in daylight and low-light conditions as well as in the darkness. The central sensor is forward-looking infrared (FLIR) which is an important sensor for its advantages in night vision in securing military camps, reassuring soldier security, and detecting suspected terror activities in the battlespace [29].

The SE carries an image intensifier element and low-light sensors. It also features shortwave infrared (SWIR) and longwave infrared (LWIR). In addition, the SE includes detection elements for sensing acoustic, seismic and magnetic interference. Moreover, the sensor package features detection elements capable of detecting infrared and the movement of an individual and a wheeled or tracked vehicle.

Once the SE has been ejected out of the munition, it immediately starts to gather and transmit information from its area to own troops in an Ultra-Wideband using the frequency of 2.4 GHz for securing the transmission QoS. Another suitable method for transmitting the data is Worldwide Interoperability for Microwave Access (WiMAX), which is based on IEEE 802.16 standard utilizing frequencies of 4.4 -5.0 GHz.

The WiMAX standard 802.16d is applicable for slowly moving users whereas 802.16e is tailored for mobile users [30] and therefore we concentrate on Portable (Mobile) WiMAX, 802.16e, the channel sizes of which are 5 MHz, 8.75 MHz and 10 MHz. The usable WiMAX, 802.16e is based on orthogonal frequency division multiplexing (OFDM), orthogonal frequency division multiple access (OFDMA) [31]. In short, WiMAX combines OFDMA, an advanced multiple-input multiple-output (MIMO) as well as beamforming (BF) features [32]. These features together offer flexible bandwidth and fast

link adaptation, creating a highly efficient air interface exceeding the capacity of existing 3G radio access networks [32]. These systems are suitable for military surveillance applications.

In low-level operations in the battlespace, data collecting can be facilitated by SEMs. The shell of SEMs can be manufactured of either steel, composite, or heat-treated plastics. SEMs can be deployed to a target area either by a howitzer or a mortar. In what follows, we take a closer look at SEMs and examine the processes of munition deployment, data collecting and data distribution.

Firstly, the ammunition shell of SEMs can be manufactured of various materials. One of these is composite, originally tailored for ballistic protection. The benefits of this material are its strength and suitability for munition core material in that it is lighter than steel and easily forged into the desired shape and structure. When munition is lighter, the payload can be heavier, if desired. Light-weight munition can be deployed further behind enemy lines by using the same charge as in steel munition.

Moreover, composite represents a material, which can be surfaced with materials capable of absorbing radar beams, making the SEM less visible in enemy counter-artillery radars. This means that SEMs and the SEs are invisible on the screens of an adversary's counter-artillery radars while SEMs are being deployed to enemy territory by air.

Furthermore, as the SE is made of composite, when it hovers in the air above an adversary, the SE manufactured of composite is less visible compared to an SE made of traditional steel. In short, an adversary receives no early warning of the incoming munition and is unlikely to be capable of locating either the positions of an artillery weapon or the hovering elements early enough. Therefore, it is highly unlikely that any counter measures be executed for there are no indications of any oncoming actions whatsoever. Fig. 13.15 illustrates the structures of various types of SEMs with the encased SEs.

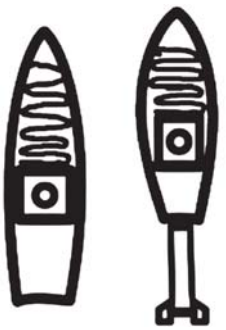


Fig. 13.15. Structures of Sensor Element Munitions: An artillery SEM (left), a mortar SEM (right). Artillery and mortar shells can be manufactured of various materials.

Secondly, the tactical use of the SEM-based reconnaissance system is as follows: 1) When reconnaissance data beyond the Event Horizon are needed, a commander issues the order

to deploy the munition to the target area, 2) Mortars or/and howitzers perform the tasks to the designated areas, 3) The SE transmits the data to own receivers, 4) Resulting improved SA is utilized in commanding troops and shooters to the designated areas or targets to maximize the performance of own troops (and gain the initiative). If more data are wanted, the described phase two can be repeated and more data can be gathered. This process is explained in Fig. 13.16.

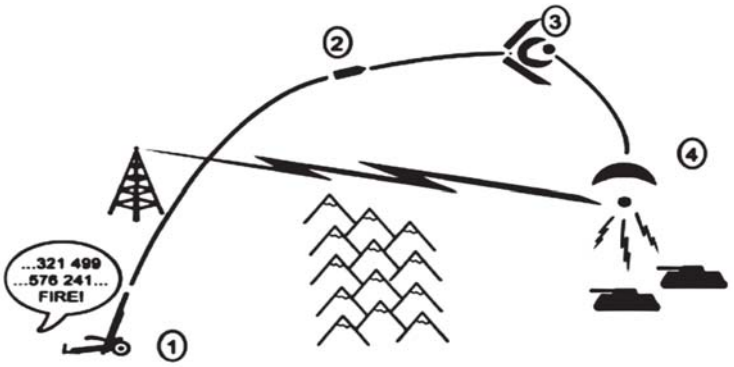


Fig. 13.16. On deploying an SE above an enemy territory: 1) Fire Support Order is issued; 2) The SEM is airborne, 3) The SEM opens and ejects the SE, 4) The SE starts to transmit gathered data from the enemy territory and targets.

Once the critical data have been collected they have to be quickly analyzed to be used for evaluating different Courses of Action (COAs). Success depends on an accurate mission analysis and a timely evaluation process of the accrued data. Improved SA results in optimal time for mission execution and simultaneous minimizing of casualties, which increases efficiency and leads to minimum recovery times improving the overall efficiency and performance capabilities of the troops utilized.

Once commanders have access to more current reconnaissance data for mission execution, they are able to analyze different COAs and calculate the pros and cons to evaluate the best possible method to operate in any scenario prevailing. As explained in Fig. 13.17, military commanders have by default value at least two different options for executing the mission in question. Having completed Military Decision Making Process (MDMP), the most effective operation can be executed to maximize the performance of the designated troops. In the described scenario below, the commander focuses the performance on incapacitating the Command Post (CP), the alternative number 2, instead of attacking against the armored enemy.

Fig. 13.18 explains the basic process of data gathering beyond the Event Horizon, especially in operations suitable for low-level troops. The deployed and hovering sensor element acts like loitering munition, sensing and measuring the prevailing electromagnetic spectrum, collecting and transmitting data to the receiver-station. An antenna can be installed both in a fighting vehicle and on the ground, depending on the prevailing combat-situation.

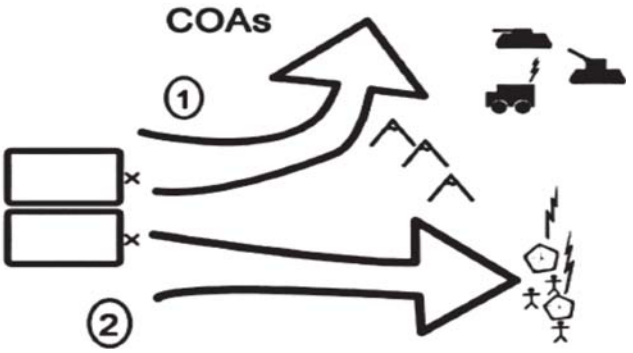


Fig. 13.17. Possibilities of COAs.

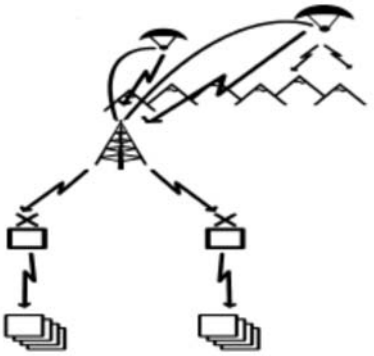


Fig. 13.18. The data gathering process for improved SA beyond the event horizon and the transmitting of these data to own troops.

Time itself is a critical resource in this type of reconnaissance process. In order to avoid wasting time, the signals and data must be transmitted reliably from the SE to the receiver-station. For this purpose, the SE utilizes smart antenna technology meaning that antenna transmission and propagation pattern can be optimized for the optimal outcome. The accrued data can be transmitted reliably to the receiver, and because of the clear Line-Of-Sight (LOS), there is only a limited number of obstacles or attenuation disturbing the coded transmission process from the sensor to the receiver-station.

13.12. On Airborne Sensors, SEMs and Communication

Data collecting systems and data transmission systems must be optimally interoperable. The functioning of these systems must be guaranteed by means of rational planning from the earliest stages to ensure forming a reliable combat combination. In any military operation, airborne sensors are important for missions, such as force protection, perimeter control and intelligence utilization [33]. Transmitting the accrued data to prevent fratricide and ensure success in operations presupposes optimal communications. WiMAX transmission offers applicable possibilities in forwarding collected data.

The distances in the transmission process are relatively short, ranging from 1 kilometer to 5 kilometers in conditions of clear Line-Of-Sight. The sensor package inside SEMs, namely the SE, can be made of existing COTS-products comprising sensors capable of sensing most of the phenomena occurring in the electromagnetic spectrum. In general, COTS-products are relatively inexpensive and reliable in terms of function, as explained in [34, 35]. An SE of a SEM comprises the following sensors: acoustic-, seismic-, magnetic-, visible image-, shortwave infrared (SWIR)-, thermal-, infrared-, low-light television (LLTV)-, and sensors for laser tracking and spotting and sensors for facial recognition.

In terms of automatized identification and verification, a facial Recognition System represents a computer application capable of automatically identifying or verifying a person from a digital image. One possibility is to compare selected facial features from the captured image with an existing facial database.

The analyzing center has the capability for the fusion of all the accrued sensor information. The sensor package comprises a short lifetime battery, which can produce energy for the sensor package for the duration of 4 – 6 minutes (hovering time). The battery can be equipped with capacitors or electric double-layer capacitors (EDLC) if the required energy level is inadequate with the selected sensor package.

Once an SE is airborne, it immediately starts to transmit the gathered data to own troops either directly or, if the transmission distance exceeds the capability of the transmission unit, the SE transmits the data to another airborne device, which acts as a relay station in relation to own troops. The SE communicates with the receiver station and other sensor element packages over a 2.4 GHz Ultra-Wideband Network system. The accrued data are encrypted for security reasons.

13.13. Comprehensive Targeting Process

In a battle space at all the command levels it is mandatory that the targeting processes be internalized. No combatant, including FFWs, can afford any mistakes, fratricide or collateral damage. The cycle of a complete targeting process can be described as Detect, Identify, Decide; Engage and Assess (DIDEA) [36]. The cycle is outlined in Fig. 13.19 below. The DIDEA provides an iterative, standardized and systematic approach supporting targeting and decision making, being generic enough to be used as a systematic process for C2 node targeting and decision making. This process is thoroughly discussed in [36].

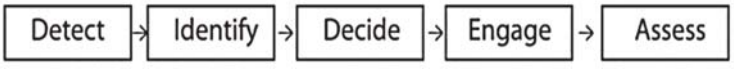


Fig. 13.19. The simplified DIDEA process.

The decision as to whether or not to open fire is based on the visual signature of a given uniform, weapon and gear as well as magnetic, seismic or acoustic signals identified by a sensor [37]. Self-evidently, the transmission of combat-critical location and identification data play a crucial role in the battle space. After the accrued data have been transmitted and received, they flow through a dissemination process, where these data are analyzed and fused to form a COP and to increase the overall SA. Fig. 13.20 explains the process of Signature Prediction Process (SPP). As described in Fig. 13.20, sensors accrue data and transmit these data for analyzing centers. The data collected with SEMs are verified with the data gathered with other sensors to in order to predict and anticipate the type of target and its actions.

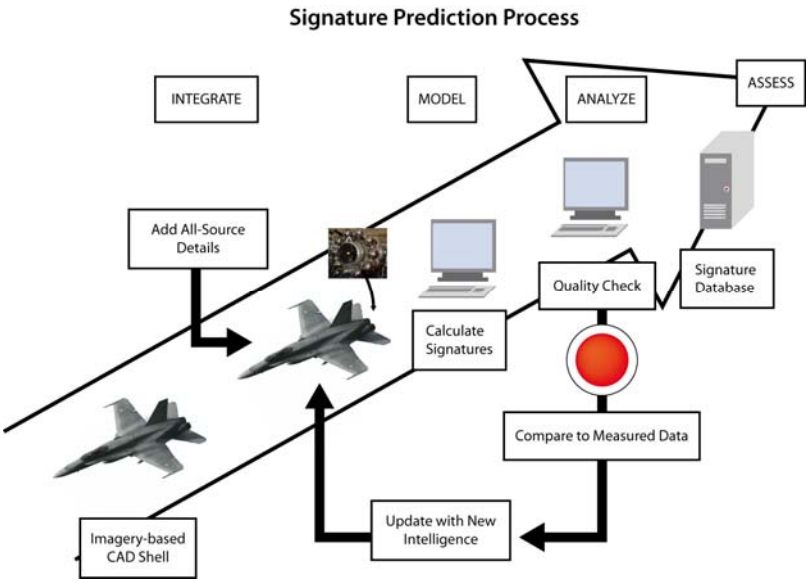


Fig. 13.20. The Signature Prediction Process, typical of several surveillance and detection systems.

The destruction power of a given weapon system has to be optimized to account for the enemy location (forest, open area, Urban Territory), the state of movement on-the Move (OTM) or at-the-halt (ATH), and the protection-level (mounted, dismounted, dug).

The cruel reality remains that an executive commander is necessarily aware of fact the there is always the possibility of fratricide and collateral damage. Fig. 13.21 emphasizes the importance of accurate and timely SA around the target area. The shooter has to be aware of the locations and status of both own troops and the enemy. It is critical to optimize the destruction power of a weapon system along the identification of a target. When the target represents a hierarchically critical enemy commander, he or she can be incapacitated by transmitting the coordinates and visual signature to the designated shooter, as indicated in Fig. 13.21.

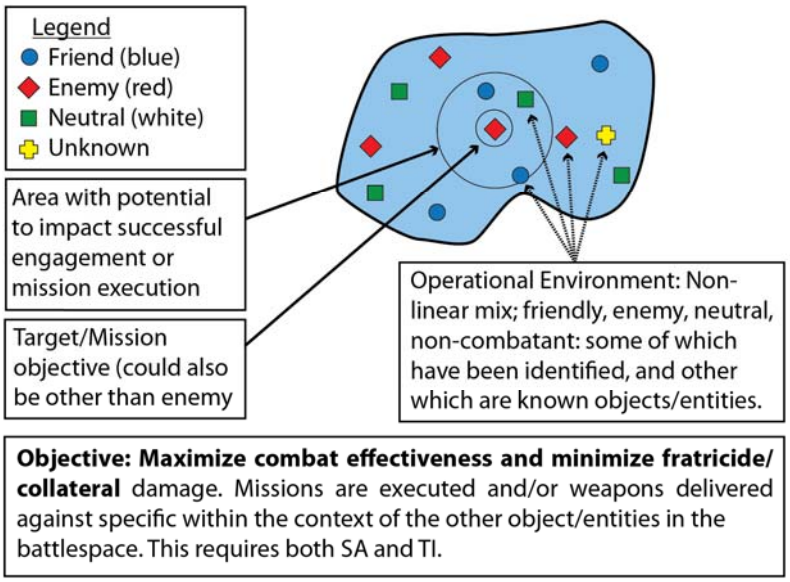


Fig. 13.21. The importance of accurate and timely Situational Awareness around the target area.

In a critical situation when the target is a very important human being, he or she can be incapacitated by transmitting the required precise data of the target and its location to the sniper or to a team of snipers. A cellular telephone can act as a receiver. The figure caption of a sniper's cellular phone is depicted in Fig. 13.22.

Fig. 13.22. Adeute target identification data transmitted to an individual member of Special Forces for eliminating purposes.

13.14. Means to Analyze Collected Data

When it comes to transmitting data, the following issues have been identified. As tested in [33], an 802.16e WiMAX testbed has provided throughputs of 5.75 Mbps upstream

with a modulation and coding of 64 QAM ¾ [33]. These amounts of data seem adequate to receive all the required sensor data.

With the assistance of automated targeting programs and classifiers, it is possible to recognize faces, find hidden and concealed targets, and look for essential information by means of computation algorithms [38]. Classifiers, such as a Support Vector Machine (SVM), K-Nearest Neighbourhood classifier (KNN), and BP neural classifier, can be utilized in battlefield target identification [39]. The recognition of an end-user can be based on visual biometrics and the most conventional identification, the computer-assisted recognition of human face [40]. The ubiquitous networks and sensor data can act as assisting tools in detection, recognition and especially in target classification [38]. Moreover, the data produced by various multi-sensors can be utilized in the data refinery process to ease the recognition and identification process with the assistance of data fusion processes by resorting to computer-programs designed for data fusion processes [41]. In fact, when using the KNearest-Neighbour (KNN) algorithm, approximately 80 % of unknown target samples can be recognized correctly, when the known target classification accuracy remains above 95 %. This enables the use of the ATR and the Automatic Target Cuer (ATC). Face recognition schemes that combine wavelet transform, SVM and clustering can be exploited identifying human beings [42].

As for object categorization, high-definition closedcircuit television (CCTV) cameras feature many compute controlled technologies that allow them to identify, track, and categorize objects in their field of view. As defined in [43], WiMAX mobile technology is a good candidate in supporting the CCTV applications in the context of mobile users.

Furthermore, Video Content Analysis (VCA) represents the capability of automatically analyzing a video to detect and determine temporal events not based on a single image. Moreover, a system utilizing a VCA can recognize changes in the environment and identify and compare objects in the database using size, speed, and color. Also, VCA analytics can be used to detect unusual patterns in a video's environment. The system can be set to detect anomalies in a crowd of people and a VCA also has the ability to track people on a map by calculating their position from the images.

13.15. Discussion

A military environment, battlespace, differs from a civilian environment. In battlespace both constant stress and uncertainty continue to dominate. The fear of losing one's life prevails. An FFW has to monitor his or her environment when fulfilling the commanded mission and stay alive. The mental capacity of an FFW must be focused on the matters at hand. The lower the number of gadgets an FFW has to monitor, the longer his or her life with an increased possibility to continue performing.

In the mission planning involving the use of unmanned aerial vehicles as parts of unmanned aircraft systems, SOA needs to be utilized. The essence of the use of UASs and UAVs translates into enabling data gathering and acting as hubs for communication systems. In the planning process of the use of frequencies, bandwidths and waveforms

available, the utilization of SOA comes in handy. It is worth paying attention to how the use of SOA results in optimizing the distribution of frequencies available in digitized battlespace and also in optimal use of transmission energy and gaining reconnaissance products, the data gathered.

It is public domain knowledge that Unmanned Aircraft Systems and Unmanned Aerial Vehicles have a tendency of losing the command and control link signal. The phenomena of electromagnetic spectrum must be recognized and the ensuing challenges must be solved. One solution relies in command and control systems. Command and control systems of all unmanned systems are versatile, as is the system configuration of UAS. Furthermore, factors which make UAV systems a low-cost alternative to a single large aircraft system, such as rapid configurability in the event of single point failures and dispensability of vehicles, also represent the main challenge from the control and design perspective. The optimization of viable command and control systems must be evaluated and created in order to maintain fleets of unmmaned aircraft systems of various types. If the data accruing systems are not in strict control, the gained reconnaissance results remain at minimal level and the data collected may not be useful for the data collector.

Militaries aim at developing SDR into a communication tool for all the troops at the tactical level. The process of embedding a functional SDR as part of military troops' communication devices is still globally ongoing in militaries, with no existing, operationally fully functional end-user devices anywhere in combat use able to transmit large amounts of data in various waveforms and frequencies.

Meeting the requirements of mobile users in a battlespace remains challenging. Issues related to SWAP-C have to be solved. One critical challenge related to military SDR use involves achieving sufficient computational capacity. This is a problem when processing wide-band high-bit rate waveforms consisting large amounts of data. In terms of SWAP-C, an FFW needs the selected communication end-user device to be reasonably tailored with optimally minimal total mass of a device, its batteries and recharging units. This means that Data Processing Units (DPUs) and Event Driven Administrative and Control Components (EDACCs) have to be carefully selected and orchestrated to meet the operational requirements of the end-user [40].

The issues related to energy cannot be over emphasized. The energy requirement of a typical handheld device can be between a few hundred milliwatts [44] to few Watts [17]. The system specifications of the SDR are significant in defining the energy needed as well as the amount of data transmitted. Challenges related to operational security are essential in reconfiguring the SDR-systems, especially as regards software. While loading a new program, new waveform or new hopping sequence, issues of transmission security during the different uploading processes have to be guaranteed. If this part becomes neglected, the SDR will not act as a useful tool in own hands in net-centric operations but rather becomes a novel tool to be exploited by the adversary.

A communication system that serves fighters' needs is creatable if we utilize the capabilities of SDR, swarms of UAVs, SDR and SON implemented in 4G networks and combine them as depicted in Fig. 13.13. This means that a Future Force Warrior's

(FFW's) end-user device, SDR, is connected with a swarm of UAVs via SON. The swarms of UAVs form an own data communication system in which the data transmission distances between UAVs are short and operationally secure.

This in turn will fulfill the requirements of LPD and LPI. The described delicate system introduced is a new one and based on ideas that can be executed by utilizing existing Commercially Off-the-Shelf (COTS) technology. The system is not yet bullet-proof and can malfunction for a number of reasons. Challenges related to creating the described system have to be solved to enable the function of different processes. The orchestration of the system can also fail because of intentional enemy action (jamming, a virus, a worm). The system needs to be equipped with an analyzing program, which indicates when the system functions properly before using the system. This asks for an easily replaceable and fault-tolerant system with inbuilt check-in routines. Otherwise, traditional methods in orchestrating services need to be adopted.

The introduced system offers an access to communication processes, which are created to support command and control systems. An FFW relies on communication services. To enhance SA and COP, it is essential to have user-friendly GUIs of some kind for presenting data. As noted, time remains a critical factor in tactical-level operations and the main function of an FFW is to fight, not to spend time browsing different databases in search for vital data.

Creating a new communication tool requires resources, such as personnel, time, money, troops and space to execute the use of the tailored device in pre-defined drills. The vendors and the end-users have to co-operate to create a functioning communication and control tool for the use of an FFW. In the development process, the use-cases and usage methods of SDRs have to be defined. This includes defining use-cases of operations, training-scenarios, types and timing of operations, training practices, data gathering during the exercises, and After Action Reviews (AARs) together with debriefing-sessions for system designers and troops after implementing the training-scenarios.

The system-creation process requires strict timing in a well-orchestrated series of field testing in which the system users and system developers have to attend the tests at the same time. The time required for field testing equals approximately a decade, the number of training drills necessary a hundred, and the number of military personnel committed to executing the drills a hundred. The complexity of the SDR system requires that a handsome number of designers and engineers from the vendor's attend the drills, ideally one on one. The estimated funding requirements equal at least a 100 M€. However, only average ballpark figures can be estimated as the actual realized costs and their approximations would by default value be labelled classified. The number of personnel and funding required described in this chapter rely on the experience accumulated over twenty years on active duty as a field-test participant. The military personnel, designers and engineers have to be fully committed to this work in order to achieve results. Table 13.2 below lists the identified resources needed for the described study. In order to create a feasible testing system, an amount of work equal to producing a dissertation is required.

Table 13.2. Estimated resources required for field-testing.

Types of resources required	Vendor side	Military side
Money	30 M€	70 M€
Personnel	100	100
Time for planning	2 years	2 years
Time for testing the system	5 years	5 years
Time for evaluating the results and the system	2 years	2 years
Reserve time	1 year	1 year
Total resources	10 years and 30 M€	10 years and 70 M€

Table 13.2 features only a rough estimate. More precise data requires pre-planning for a period of twelve months. Different sources of funding, such as industrial and / or academic contributions of personnel and / or equipment, need to be estimated before any final estimation is doable.

To sustain optimal performance, an FFW has to be able to use only one single device for command and control. This device can be an SDR with a GUI. This way there is only one single device, Software Defined Radio/Graphic User Interface (SDR/GUI), for an FFW to communicate and use controls with instead of being exposed to several communication tools. Compared to a stable civilian environment, a military environment equals a constantly altering battlespace. An FFW has to be able to monitor the events in the prevailing environment instead of needing to update the status and monitor his or her command and communication tool. An FFW has to have hands on a weapon and be ready to act when necessary. If unable to do so, the FFW will become incapacitated by a splinter or a bullet. It is essential to enable an FFW to receive and transmit data with the assistance of SDR/GUI, into which SDR can be embedded.

First, an improved SDR/GUI can act as a control-station for all of the digital systems. This allows an FFW to focus on his or her main task, to fight. He or she can monitor the prevailing environment and use his or her weaponry in a time-critical environment. Second, an FFW can concentrate on one device, SDR/GUI, instead of monitoring several screens and displays. He or she saves time and can focus on the task commanded. Third, the control units of his or her own warrior skeleton and communication controls can be found from one communication device, SDR/GUI. The fourth challenge can also be solved by adopting SDR/GUI which will take care of the various networks and waveforms and switch automatically to the free and appropriate channel to transmit or receive data.

The fifth challenge was a problem concerning the access to BMS via different communication tools. The problem is linked to the issues of bandwidth, frequency and waveform. The accessing process into BMS involves utilizing the performance provided by the SDR and SON. This means offering the frequencies, bandwidths and waveforms

required from the fighters' perspective. The swarm of UAVs serves as a secure and replaceable communication gateway for the data exchange process. This allows for keeping the transmission power low and transmission ranges relatively short. The Sixth challenge was linked to the GUIs. To comprehend the prevailing operational situation and a holistic list of events at a tactical level, an informative presentation of SA and COP is in an essential role. The presented system together with improved SDR/GUI can be seen as a feasible solution to solve all the listed challenges.

13.16. Results

The main result of this chapter is an idea-phase introduction of a tactical level (battalion and below) communication system to be used by a tactical end-user performing in a battlespace. This article focused on the use of unmanned aerial vehicles as part of unmanned aircraft systems operating in low level operations for accruing data for decision making and acting as hubs for communication purposes. As this chapter's contents represent an early idea stage of concept development, the system drafted and its features described can neither be operationalized nor field-tested.

This chapter described an idea-phase solution for utilizing Software Defined Radio (SDR) with unmanned aircraft systems and discussed the system's structure and functionalities. The description allows for creating a communication system which features Low Probability of Detection and Low Probability of Identification thereby hampering enemy action. As for advancing own capability, the proposed communication system offers wide bandwidth communication and, moreover, increased operational security as opposed to the conventional communication setting composed of fixed base-stations and different type of mobile users.

As the described system utilizes unmanned aircraft systems, it offers a battle commander command and control tool which, properly and timely used, enables its users to draw from an extensive information network that remains utilizable in the battlespace. In other words, the described communication system represents a poor man's satellite communication system in which UAS equal the base stations (sort of satellites) and FFWs are the end-users of the communication services in the battlespace. The proposed communication system propagates as its fundamental benefit the ability to rely only on one end-user device (SDR/GUI) instead of forcing the battle commander to have to use several communication devices.

An FFW performs in a battlespace filled with ubiquitous networks and communication systems. He or she has to cope with actions involving humans and machines, such as databases and UVs. Equipped with a reliable communication tool, an FFW can perform tasks with improved speed and efficiency. Bespoke SDR can enhance the performance of an FFW by answering the defined challenges listed in Section 13.5.

Fig. 13.23 features a possible view of a functional SDR/GUI for the FFW operating at low level (company and below) tactical operations.

Fig. 13.23. An example of tactical SDR/GUI [1].

Tactical SDR/GUIs can serve as command and control tools offered for the FFWs. SDR/GUIs can be utilized as tools for possibly enhancing the overall performance of the soldier. An FFW needs only a short period of time to take a look at the SDR/GUI and notice if something significant has changed in the overview. An FFW can concentrate on his or her main mission, which is to fight instead of constantly monitoring all the controlling units of his or her gear.

13.17. Conclusions

When moving towards tactical military SDR/GUI, the system presented requires funding and field testing to be able to create a functional end-product. The rough estimate of the resources required is evaluated above in Section 13.15 and in Table 13.2 based on the experience gained in twenty years spent as a participant in different military tests. Automated systems and allocation of diminishing resources force militaries to consider the facilitated performance offered by means of exploiting SDR/GUI. The result could be an agile and modular military performer with ever-improved capabilities and SA completed with the capability to utilize the diminishing resources more optimally with decreased instances of collateral damage.

Further work related to creating a functional system based on the idea-phase description outlined in this chapter needs to pay attention to operational security issues of using software and hardware in a digitized battlespace. Issues such as adequate level of constant energy flow and protection against violations caused by electronic warfare must be studied, tested and solved before the adoption of the system in any type of operational use.

The utilization of SOA can optimize the use of frequencies, bandwidths, the flight route planning of UASs and UAVs. When doing so, the end-result can involve saving operationally critical resources, such as time and energy. This may in turn result in saving lives of own troops and minimizing instances of fratricide. The process of accruing vital data from the designated areas may also enable receiving quantitatively and qualitatively improved data which are more easily and quickly utilizable. This results in an improved

decision making process and increased executing tempo of military operations which may facilitate meeting the set objective efficiently. The utilization of SOA can be benefitted in analyzing the findings of data accruing processes. SOA is usable when optimizing the use of algorithms and filters according to use cases in data analyzing processes. SOA can also be identified as a means in allocating human labour and other time critical resources.

The introduced idea-phase description of a communication system aims at being battle-proof from the perspective of using a swarms of UAVs in particular in that it guarantees the usability of the system functionality as the swarms of UAVs can automatically recreate a functional communication network and maintain an adequate distance between each UAV in all the circumstances and situations: once a UAV is destroyed or shot down, it has been programmed to destroy itself mechanically and electronically. The remaining fleets of UAVs corrects the formation of flying UAVs automated, to maintain a functioning and reliable communication system to ensure the communication system remains intact. SON supports the communication system formed by the swarms of UAVs together with SDRs.

References

[1]. T. Saarelainen, Towards Tactical Military Software Defined Radio, in *Proceedings of the 8th International Conference on Sensor Technologies and Applications (SENSORCOMM' 2014)*, November 16 – 20, 2014, Lisbon, Portugal, pp. 96 - 106.

[2]. K. A. Hempenius, A. R. Wilson, M. J. Kumar, N. Hosseini, M. E. Cordovez, and M. S. Sherriff, A More Cost-Effective Unattended Ground Sensor Using Commercial Off-the-Shelf Products, in *Proceedings of the IEEE Systems and Information Engineering Design Symposium,* 2012, pp. 62 – 67.

[3]. R. A. Syed and R. S. Wexler, Army Warfighter Information Network-Tactical (WIN-T) Theory of Operation, in *Proceedings of the Military Communications Conference, MILCOM'13*, 18–20 Nov. 2013, San Diego CA, pp. 1453 – 1461.

[4]. S. L. Cotton and W. G. Scanlon, Millimeter-Wave Soldier-to-Soldier Communications for Covert Battlefield Operations, *IEEE Communications Magazine,* 47, 10, 2009, pp. 72–81.

[5]. Potential Use of Unmanned Aircraft Systems (UAS) for NASA Science Missions, *NASA,* http://airbornescience.nasa.gov/program/documents.html, May 2006. Accessed on January 1st 2015.

[6]. P. Royo, R. Cuadrado, C. Barrado, E. Salami, M. Pérez-Batlle and E. Pastor, Towards the automation of the UAS mission management, in *Proceedings of the IEEE Conference on Digital Avionics System,* 5 – 10 October 2013, East Syracuse, NY, pp. 6D6-1 – 6D6-14.

[7]. T. Saarelainen, Modeling War as a Business Process with the assistance of Service Oriented Architecture, in *Proceedings of the Ninth International Conference on Digital Telecommunications (ICDT'14),* 2014, pp. 1 – 7.

[8]. M. Panahi, W. Nie, K-J. Lin, The Design and Implementation of Service Reservations in Real-Time SOA, in *Proceedings of the IEEE International Conference on e-Business Engineering (ICEBE'09),* 21-23 Oct. 2009, Macau, pp. 129 – 136.

[9]. M. Abdullah, and Merklinghaus D-P., Mobile Battlefield Systems, *Military Technology,* Vol. XXXVIII, Issue 6, 2014, pp. 54 – 62.

[10]. A. Gravel, X. Fu, and J. Su, An Analysis Tool for Execution of BPEL Services, in *Proceedings of the 9th IEEE International Conference on E-Commerce Technology and the*

4th IEEE International Conference on Enterprise Computing, E-Commerce, and E-Services (CEC-EEE2007), Tokyo, 2007, pp. 429 – 432.

[11]. M. Medlow, Extending Service-Oriented Architectures to the Deployed Land Environment, *Journal of Battlefield Technology,* Vol. 13, No. 1, March, 2010, pp. 27 – 33.

[12]. L. Liu, D. Russell, N. Looker, D. Webster, and J. Xu, Delivering Sustainable Evolutionary Service-Oriented Architecture for Network Enabled Capability, in *Proceedings of the International Workshop on Verification and Evaluation of Computer and Communication Systems (VECoS'08),* 2008, pp. 1 – 11.

[13]. T. Ulversøy, Software Defined Radio: Challenges and Opportunities, *IEEE Communications Surveys & Tutorials,* Vol. 12, Issue 4, Fourth Quarter 2010, pp. 531–550.

[14]. J. Mölsä, J. Karsikas, A. Kärkkäinen, R. Kettunen, and P. Huttunen, Field test results and use scenarios for a WiMAX based Finnish broadband tactical backbone network, in *Proceedings of the Military Communications Conference,* San Jose, CA, Oct. 31–Nov. 3, 2010, pp. 2014 – 2019.

[15]. Scaldio, Reconfigurable Radio Transceiver IP, www2.imec.be/content/user/File/Brochures/GR2010_Leaflet Scaldio.pdf, accessed on 10th September 2014.

[16]. C-Y. Chen, F-H. Tseng, K-D. Chang, H-C. Chao, and J-L. Chen, Reconfigurable Software Defined Radio and Its Applications, *Tamkang Journal of Science and Engineering*, Vol. 13, No. 1, 2010, pp. 29 – 38.

[17]. M. Islam, M. A. Hannan, S. A Samad, and A. Hussain, Software Defined Radio for RFID Application, in *Proceedings of the World Congress on Engineering and Computer Science (WCECS'09)*, San Francisco, USA, Vol I, October 20–22, 2009, pp. 415 – 418.

[18]. G. Thomas, Situation Awareness Issues in Tactical Cognitive Radio, in *Proceedings of the IEEE International Multi-Disciplinary Conference on Cognitive Methods in Situational Awareness and Decision Support*, 6–8 March 2012, pp. 287 –293.

[19]. X. Dong, S. Wei, Y. Li, L. Wang, and L. Bai, Service-Oriented Radio Architecture: A Novel 2M Network Architecture for Cognitive Radio Systems, *International Journal of Distributed Sensor Networks*, Vol. 2012, Article ID 762953, 8 pages.

[20]. T. Saarelainen and J. Jormakka, Interfacing Collaboration and Command Tools for Crises Management Military Command and Control Systems, *International Journal of Electronic security and Digital Forensics,* Vol. 3, No. 3, 2010, pp. 249 – 264.

[21]. R. Adams, Special Operations Forces Skills, *Military Technology*, Vol. XXXVIII, Issue 5, 2014, p. 38 – 42.

[22]. T. Saarelainen and J. Jormakka, Computer-Aided Warriors for Future Battlefields, in *Proceedings of the 9th European Conference on Information Warfare and Security (ECIW'09)*, Lisbon, 6–7 July, 2009, Portugal, pp. 224 – 233.

[23]. T. Saarelainen, Improving the Performance of a Dismounted Future Force Warrior by Means of C4I2SR, Doctoral Dissertation, *National Defence University,* Helsinki, 212 pages, 2013.

[24]. S. L. Cotton and W. G. Scanlon, Millimeter-Wave Soldier-to-Soldier Communications for Covert Battlefield Operations, *IEEE Communications Magazine,* pp. 72 – 81.

[25]. P. Pancholy, J. Kushwaha, and R. Chitranshi, Smart 4G network:-Implementation of SDR & SON with Automatic EM Radiation control, *International Journal of Engineering Science and Innovative Technology (IJESIT),* Vol. 1, Issue 2, November 2012, pp. 14 – 18.

[26]. T. Saarelainen, Free Space Optics in the use of the Future Force Warriors, *Journal of Communications and Computer,* Vol. 9, Number 2, February 2012, pp. 123–133.

[27]. T. Saarelainen, White Force Tracking, *Journal of Communications and Computer,* Vol. 9, No. 1, January 2012, pp. 113 – 121.

[28]. J. Jormakka and T. Saarelainen, UAV-based Sensor Networks for Future Force Warriors, *International Journal On Advances in Telecommunications,* Vol. 4, No. 1 and 2, 2011, pp. 58–71.

[30]. M. Medlow, Extending Service-Oriented Architectures to the Deployed Land Environment, *Journal of Battlefield Technology,* Vol. 13, No. 1, March 2010, pp. 27-33.
[31]. J. L. Liu, D. Russell, N. Looker, D. Webster, J. Xu, Delivering Sustainable Evolutionary Service-Oriented Architecture for Network Enabled Capability, in *Proceedings of the International Workshop on Verification and Evaluation of Computer and Communication Systems (VECoS'08),* 2008, pp. 1-11.
[32]. T. Ulversøy, Software Defined Radio: Challenges and Opportunities, *IEEE Communications Surveys & Tutorials,* Vol. 12, No. 4, Fourth Quarter 2010, pp. 531-550.
[33]. M. Abdullah, Merklinghaus D.-P., Mobile Battlefield Systems, *Military Technology,* Vol. XXXVIII, Issue 6, 2014, pp. 54-62.
[34]. J. Mölsä, J. Karsikas, A. Kärkkäinen, R. Kettunen, P. Huttunen, Field test results and use scenarios for a WiMAX based Finnish broadband tactical backbone network, in *Proceedings of the Military Communications Conference,* San Jose, CA, 31 Oct. – 3 Nov. 2010, pp. 2014-2019.
[35]. S. L. Cotton, W. G. Scanlon, Millimeter-Wave Soldier-to-Soldier Communications for Covert Battlefield Operations, *IEEE Communications Magazine,* Special Issue on Military Comms., October 2009, pp. 72-81.
[36]. M E. Ospital, A. Wojack, A Holistic Approach to Combat Identification, *Military Technology,* Vol. XXXIV, Issue 9, 2010, pp. 68–73.
[37]. Scaldio, Reconfigurable Radio Transceiver, IP, www2.imec.be/content/user/File/Brochures/GR2010_Leaflet Scaldio.pdf, accessed on 10th September 2014.
[38]. C.-Y. Chen, F.-H. Tseng, K.-D. Chang, H.-C. Chao, J.-L. Chen, Reconfigurable Software Defined Radio and Its Applications, *Tamkang Journal of Science and Engineering,* Vol. 13, No. 1, 2010, pp. 29-38.
[39]. J. Li, C. Zhang, and Z. Li, Battlefield Target Identification Based on Improved Grid-Search SVM Classifier, in *Proceedings of the International Conference on Computational Intelligence and Software Engineering (CiSE'09),* 11-13 Dec. 2009, pp. 1 – 4.
[40]. R. Luo, Y. Chou, T. Chung, C. Liao, C. Lail, and C. Tsai, NCCU Security Warrior: An Intelligent Security Robot System, in *Proceedings of the 33rd IEEE Conference on Industrial Electronics Society (IECON'07),* 5-8 Nov. 2007, Taipei, pp. 2960 – 2965.
[41]. Libiao, T., Lu, W., Qi, D., Kai, Z., Target Recognition Based on Seismic Sensors and Neural Network, in *Proceedings of the 8th International Conference on Electronic Measurement and Instruments (ICEMI '07),* Aug. 16 2007-July 18 2007, Xi'an, pp. 1-18 – 1-21.
[42]. Luo, B, Zhang, Y., and Pan, Y-H., Face recognition based on wavelet transform and SVM, Conference on Information Acquisition, in *Proceedings of the IEEE International Conference on Information Acquisition,* 27 June-3 July 2005.
[43]. M. Aguado, E. Jacob, J. Matias, C. Conde, and M Berbineau, Deploying CCTV As an Ethernet Service Over the Wimax Mobile Network in the Public Transport Scenario, in *Proceedings of the International Conference on Communications Workshops, (ICC Workshops),* 14-18 June 2009, pp. 1 – 5.
[44]. Y. Lin, H. Lee, M. Who, Y. Harel, S. Mahlke, and T. Mudge, SODA: A Low-Power Architecture for Software Radio, in *Proceedings of the 33rd Annual international Symposium on Computer Architecture,* pp. 89 – 101.

Chapter 14
A Survey on Wireless Sensor Networks Simulation Tools and Testbeds

Abdelrahman Abuarqoub, Mohammad Hammoudeh, Fayez Al-Fayez and Omar Aldabbas

Abstract

This chapter presents a comprehensive survey of simulation tools, test-beds and hardware platforms currently used in assessing a number of different parameters required by a target Wireless Sensor Network (WSN) application. We describe the main features and limitations of each assessment methodology. The survey shows that while WSN hardware platforms become more powerful and cheaper, the number and support to simulation tools is decreasing. The advanced capabilities of current sensor hardware platforms and their low cost make large-scale test-beds a feasible option for the design and evaluation of WSN systems. Accordingly, we advocate testbeds to become the main evaluation methodology to move the field of WSNs forward. Finally, as an interim solution, we recommend mixed-mode simulation methodology, which integrates a simulated environment with real WSN hardware to improve both the accuracy and scalability of results when evaluating different prototype designs and systems.

14.1. Introduction

A successful large-scale Wireless Sensor Network (WSN) deployment necessitates that the design concepts are checked before they are optimised for a specific hardware platform. Developing, testing, and evaluating network protocols and supporting architectures and services for WSNs can be undertaken through test-beds or simulation. In a survey on WSN evaluation tools published in 2012 [1], test-beds were found to be extremely valuable. However, three years ago implementing large-scale test-beds was not widely viable, because of the substantial cost of deploying and maintaining a large-scale WSN. Moreover, the time needed for setting up the network for experimental goals made simulation a favourable tool in developing reliable and portable WSNs applications.

Abdelrahman Abuarqoub
Faculty of Information Technology, Middle East University, Amman, Jordan

However, nowadays, tens of WSN hardware platforms exist at a very low cost. For instance, an ESP8266 node costs less than 3 USD [2]. Many of these hardware platforms are equipped with long-range, up to 1,000 Km in outdoor deployment, ultra-low power radio devices. They also come with high performance processors and large memory resources (see Table 14.2). This not only calls for relaxing the strict assumptions that were made tens of years ago about the capabilities of WSN nodes, but also for considering the methods used in evaluating exiting and new research in this field. The authors believe that this is an important step towards translating the huge amount of research in this field to real life deployments of WSN systems.

In WSNs, simulation provides a cost effective method of assessing the appropriateness of systems before deployment. It can, for example, help assess the scalability of algorithms free of the constraints of a hardware platform. Furthermore, simulators can be used to simplify the software development process for a particular WSN application. For instance, TOSSIM [3] utilises the component based architecture of TinyOS [4] and provides a hardware resource abstraction layer that enables the simulation of TinyOS applications, which can then be ported directly to a hardware platform without further modifications.

Simulation is hence the research tool of choice for the majority of the mobile ad hoc network community. An examination of research papers published in SENSORCOMM 2011 [5] reveals a significant increase in using real testbeds compared to the study published by Kurkowski et al. [6]. Yet, 53 % of the authors used simulation in their research. Apart from the self-developed simulators, there are a few widely used network simulators including NS-2 [7], OPNET [8], MATLAB [9], IFAS [10], and OMNet++ [11]. Fig. 14.1 shows the simulator usage following a survey of simulation-based papers in SENSORCOMM 2011 conference. Simulation of ad hoc wireless capabilities for WSNs have been addressed by extending existing simulators, or specifically building new ones, such as NS-3 [12]. The latter class of simulators mostly focuses on protocols and algorithms for layers of the network stack, but they do not directly support WSN.

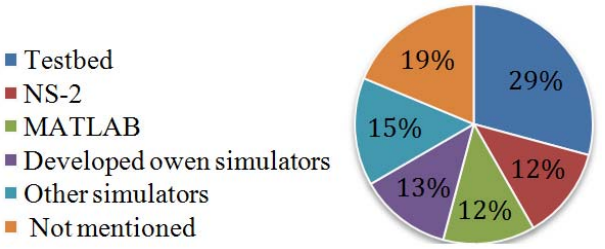


Fig. 14.1. Simulator usage results from a survey of simulation based papers in SENSORCOMM 2011.

Recently, several simulation tools have appeared to specifically address WSNs, varying from extensions of existing tools to application specific simulators. Although these tools have some collective objectives, they obviously differ in design goals, architecture, and

applications abstraction level. In the next section, we review some of the important WSNs simulation tools and explore their characteristics.

In this chapter we have addressed and discussed sixteen simulation tools, emphasizing the main features and the major drawbacks. Authors believe that these simulation platforms are the most widely used in education, research, and industrial purposes. However, some of these simulators can be useful for one purpose, while other can be adapted to all needs. For instance, CNET provide good tools for educational matters but it has limitation to support industrial use, while, for example, NS3 can work in all circumstances, as it could be simplify for education and go far in research or industrial complexity. The WSNs simulators tools reviewed in this chapter are SensorSim, TOSSIM, TOSSF, GloMoSim, Qualnet, OPNET, EmStar, SENS, J-Sim, Dingo, NS-2, NS-3, Shawn, GTSNetS, CNET and TRMSim-WSN.

The rest of the chapter is organized as follows. In Section 14.2, the most popular WSNs simulators are described and their strengths and weaknesses are discussed. In Section 14.3, the largest indoor sensor network testbeds with public remote access are reviewed. Section 14.4, presents our views about the future of WSNs testing and evaluation methods. Section 14.5 concludes the chapter.

14.2. WSNs Network Simulation Tools

14.2.1. SensorSim

SensorSim [13] builds on the NS-2 simulator providing additional capabilities for modelling WSNs. The main features of this platform are: power and communication protocol models; sensing channel and sensor models; scenario generation; and support for hybrid simulations. The public release of the SensorSim suite of tools was withdrawn due to its unfinished nature and the inability of authors to provide the needed level of support.

Georgia Tech SensorSimII [14] is written in a modular style, where sensor nodes are organized into three components: application, network, and link. The work in SensorSimII may be divided into two areas: the simulator core and the visualization tools. The simulator core essentially manages an array of independent sensor nodes throughout time. The visualization tools provide views of both individual node state and communication traffic between nodes.

Both SensorSim projects are open source and free to use. However, the simulators are limited in their realism because (apart from SensorSim's power modules) neither simulator considers the limited resources of sensor nodes such as memory, and real-time computational capability. Moreover, it is not always required by the WSN to validate the functional correctness and/or, to provide performance guarantees. SensorSim simulates the complete WSN protocol stack, although this can be regarded as overkill and adding unnecessary complexity as this is not required in order to simulate the expected behaviour. This makes the SensorSim platform complex and difficult to use.

14.2.2. TOSSIM

There are platforms specifically designed to simulate WSNs, such as TOSSIM [3] which is a part of the TinyOS development efforts [4]. TOSSIM is a discrete-event simulator for TinyOS applications [3]. It aims to assist TinyOS application development and debugging by compiling applications into the TOSSIM framework, which runs on a PC instead of compiling them for a mote. Using the TOSSIM framework, programs can be directly targeted to motes without modification. This gives users a bigger margin to debug, test, and analyses algorithms in a controlled and repeatable environment. In TOSSIM, all nodes share the exact same code image, simulated at bit granularity, and assuming static node connectivity is known in advance. Therefore, TOSSIM is more of a TinyOS emulator than a general WSN simulator. It focuses on simulating TinyOS rather than simulating the real world. This has the advantage that the developed algorithms can be tested on a target platform. However, this may place some restrictions of the target platform on the simulation. TOSSIM is not always the right simulation solution; like any simulation, it makes several assumptions about the target hardware platform, focusing on making some behaviour accurate while simplifying others [3]. TOSSIM can be used as a tool for absolute evaluation of some causes of the behaviour observed in real-world network deployments.

14.2.3. TOSSF

TOSSF [15] is a simulation framework that compiles a TinyOS application into the SWAN [16] simulation framework. It can be viewed as an improvement over TOSSIM with a primary focus on scalability. It allows simulation of a heterogeneous collection of sensor nodes and a dynamic network topology. TOSSF suffers from potentially long test-debug cycles because it does not provide a scripting framework for experimentation. Although it enables development of custom environmental models, the absence of a scripting framework requires those models to be compiled into the simulation framework. Given that both of these simulators are tightly coupled with TinyOS, they may be unsuitable for early prototyping, or developing portable WSN applications.

14.2.4. GloMoSim

GloMoSim [17] is a scalable simulation environment for wireless and wired network systems. Its parallel discrete-event design distinguishes it from most other sensor network simulators. Though it is a general network simulator, GloMoSim currently supports protocols designed purely for wireless networks. GloMoSim is built using a layered approach similar to the seven layer network architecture of the OSI model. It uses standard APIs between different simulation layers to allow rapid integration of models developed at different layers, possibly by different users.

As in NS-2, GloMoSim uses an object-oriented approach, however for scalability purposes; each object is responsible for running one layer in the protocol stack of every node. This design strategy helps to divide the overhead management of a large-scale network. GloMoSim has been found to be effective for simulating IP networks, but it is

not capable of simulating sensor networks accurately [18]. Moreover, GloMoSim does not support phenomena occurring outside of the simulation environment, all events must be gathered from neighbouring nodes in the network. Finally, GloMoSim stopped releasing updates in 2000 and released a commercial product called QualNet.

14.2.5. Qualnet

Qualnet is a commercial network simulator tool released by Scalable Network Technologies [19] that is derived from GloMoSim. Qualnet significantly extends the set of models and protocols supported by GloMoSim. It also provides a comprehensive set of advanced wireless modules and user-friendly tools for building scenarios and analyzing simulation results. Qualnet is a discrete-event simulator, as such, it is event driven and time aware. It uses a layered architecture that is run by each node. When a protocol resides in a particular layer at one node, the packets are passed down crossing the remaining layers at the sending node, across the network, and then up to the protocol stack at the receiving node. Qualnet has a modular design and an intuitive GUI that make it easy to use to learn and modify.

14.2.6. OPNET

OPNET [20] is a further discrete event, object oriented, general purpose network simulator. The engine of OPNET is a finite state machine model in combination with an analytical model. It uses a hierarchical model to define each characteristic of the system. The top hierarchy level contains the network model, where the topology is designed. The second level defines the data flow models. The third level is the process editor, which handles control flow models defined in the second level. Finally, a parameter editor is included to support the three higher levels. The hierarchical models result in event queues for a discrete event simulation engine and a set of entities that handle the events. Each entity represents a node which consists of a finite state machine which processes the events during simulation.

Unlike NS-2 and GloMoSim, OPNET supports modelling sensor-specific hardware, such as physical-link transceivers and antennas. It also enables users to define custom packet formats. An attractive feature of OPNET is its capability of recording a large set of user defined results. Furthermore, the GUI (Graphical User Interface), along with the considerable amount of documentation and study cases that come along with the license are another attractive feature of the simulator. This GUI interface can also be used to model, graph, and animate the resulting output. The network operator is provided with editors that are required to simplify the different levels of modelling. Though model parameters can be changed, the simulation accuracy is influenced because OPNET is not open source software. Similar to NS-2, the object-oriented design of OPNET causes scalability problems. It does not have a high number of protocols publicly available possibly because of source code licensing constraints. Finally, OPNET is only available in commercial form.

The second class of simulators are application-oriented simulators, including EmStar [21], SENS [22], J-Sim [23], Shawn [24], and Dingo [25].

14.2.7. EmStar

EmStar [21] is a component based, discrete-event framework that offers a range of run-time environments, from pure simulation, distributed deployment on iPAQs [26], to a hybrid simulation mode similar to SensorSim. Emstar supports the use of simulation in the early stages of design and development by providing a range of simulated sensor network components, including radios, which provide the same interfaces as actual components. It supports hybrid mode with some actual components and some simulated components, and full native mode with no simulated components. As in TOSSIM, EmStar uses the same source code that runs at each of these levels to run on actual sensors. Amongst other simulators, such as TOSSIM, EmStar provides an option to interface with actual hardware while running a simulation. EmStar is compatible with two different types of node hardware. It can be used to develop software for Mica2 motes [27] and it also offers support for developing software for iPAQ based microservers. The development cycle is the same for both hardware platforms. The next step in the development cycle following the simulation is data replay. In this model, EmStar uses data collected from actual sensors in order to run its simulation. Leading directly from this, Emstar uses the half-simulation methodology similar to SensorSim's, where the software is running on a host machine and interfacing with a real physical communication channels. The final step in the development cycle is deployment.

EmStar combines many of the features of other WSNs simulators. Its component based design allows for fair scalability. Moreover, each aspect of the network can be logically fine-tuned due to its development cycle design. Because it targets a particular platform, many protocols are already available to be used. At the deployment step in the development cycle, only the configuration files have to be designed. This potentially adds constraints on the user as they must either ensure that the hardware configuration being used matches the existing configuration file, or they must write their own files.

The main goal of Emstar is to reduce design complexity, enabling work to be shared and reused, and to simplify and accelerate the design of new sensor network applications. While not as efficient and fast as other frameworks like TOSSIM, Emstar provides a simple environmental model and network medium in which to design, develop and deploy heterogeneous sensor network applications. When used as a migration platform from code to real sensor environment, the environment model may be sufficient for most developers. Another drawback of Emstar is that the simulator supports only the code for the types of nodes that it is designed to work with.

14.2.8. SENS

SENS [22] is a customizable component-based simulator for WSN applications. It consists of interchangeable and extensible components for applications, network communication, and the physical environment. In SENS, each node is partitioned into

four main components: application, simulates the software application of the sensor node; network, handles incoming and outgoing packets; physical, reads sensed information; and environment, network propagation characteristics. Multiple different component implementations offer varying degrees of realism. For example, users can choose between various application-specific environments with different signal propagation characteristics. As in TOSSIM, SENS source code can be ported directly into actual sensor nodes, enabling application portability. Moreover, it provides a power module for development of dependable applications.

SENS defines three network models that can be used. The first successfully forwards packets to all neighbours, the second delivers with a chance of loss based on a fixed probability, and the third considers the chance of collision at each node. The physical component includes the non-network hardware for the sensor such as the power, sensors, and actuators. At a lower level, the environment component models the physical phenomena and the layout. The layout model includes different types of surfaces, each affecting radio and sound propagation in a different way.

SENS is less customizable than many other simulators, providing no chance to alter the MAC protocol, along with other low level network protocols. SENS uses one of the most sophisticated environmental models and implements the use of sensors well. However, the only measurable phenomenon is sound.

14.2.9. J-Sim

J-Sim [23] is a component-based discrete event simulator built in Java and modelled after NS-2. The design of this simulator aims at solving many of the shortcomings of comparable object-oriented simulators like NS-2. J-Sim uses the concept of components instead of the concept of having an object for each individual node. J-Sim uses three top level components: the target node which produces stimuli, the sensor node that reacts to the stimuli, and the sink node which is the ultimate destination for stimuli reporting. Each component is broken into parts and modelled differently within the simulator; this eases the use of different protocols in different simulation runs.

J-Sim claim has several advantages over NS-2 and other simulators. First its component based architecture scales better than the object oriented model used by NS-2 and other simulators. Second, J-Sim has an improved energy model and the ability to simulate the use of sensors for phenomena detection. Like SensorSim, there is support for using the simulation code for real hardware sensors. However, J-Sim is comparatively complicated to use. While no more complicated than NS-2, the latter simulator is more popular and accepted in the sensor network research community and more community support is available, therefore, more people are keen to spend the time to learn how to use it.

Though it is scalable, J-Sim has a set of inefficiencies. First, there is unnecessary overhead in the intercommunication model. The second problem is inherited by most sensor networks simulators that are built on top of general purpose simulators, 802.11 is the only MAC protocol that can be used in J-Sim. Finally, Java is possibly less efficient than many other languages.

14.2.10. Dingo

Dingo [28] provides a workbench for prototyping algorithms for WSNs taking a top-down design methodology. Having no target platform means the full functionality of a programming language can be used. This eases the design process as prototype algorithms can be tested before optimization for the target platform. Dingo consists of a fixed API, with customizable internals. It has a simple graphical user interface and a set of base classes, which are extended by the user to create simulation. Each simulated sensor node runs in its own thread and communicates using the same protocols that would be deployed on a physical node. Sensors are modelled using a pool of concurrent, communicating threads. Individual sensors are able to: (1) Gather and process data from a model environment; (2) Locate and communicate with their nearest neighbours; (3) Determine whether they are operating correctly and act accordingly to alter the network topology in case of faulty nodes being detected. Nodes may be configured differently to simulate a heterogeneous sensor network. Dingo comes with a set of application level routing packages including simple multi-hop flooding, MuMHR [29] and LEACH [30].

Dingo features a significant improvement in the simulation performance by giving the option to split the visualization from the simulation. It provides tools for the simulation and deployment of high-level, Python code on real sensor networks. For example, Dingo-boom provides a two-way interface between MoteIV's Boomerang class motes and Dingo. Dingo-top is another tool which is used to dump network topology data to a text file and generate a graphical representation of that topology. Furthermore, Dingo has several features in the form of plugins. These can be activated/deactivated on the plugin menu.

As with SensorSimII, Dingo provides an extensible visualization framework that aims at easing the life for sensor network debugging, assessment, and understanding of the software by visualizing the sensor network topology, the individual node state, and the transmission of the sensed data. Dingo comes with an interface between the simulation environment and different hardware platforms, for example the Gumstix [31] platform. Also, Dingo allows mixed-mode simulation using a combination of real and simulated nodes. In Dingo, nodes have the ability to obtain their sensed data from a database or graphical objects like maps; this improves the fidelity of simulations as it makes it possible to check the simulation results against the real data.

Dingo focuses on the protocols and algorithms for higher layers of network state but it does not directly support sensor networks at the physical layer. It has major drawbacks which limit its functionality. Most of these drawbacks are due to the incomplete nature of the tool. These drawbacks are: (1) The lack for Media Access Control or MAC layer, communications to be handled by point-to-point systems. (2) No collision management procedure, partly due to the absence of the MAC layer. (3) Dingo is still available to use, however, it is no longer supported by the owners of the project.

14.2.11. NS-2 and NS-3

NS-2 [32] is an object-oriented discrete event simulator targeted at networking research. It is an open source network simulator originally designed for wired, IP networks. The NS-2 simulation environment offered great flexibility in studying the characteristics of WSNs because it includes flexible extensions for WSNs. NS-2 has a number of limitations: (1) It puts some restrictions on the customization of packet formats, energy models, MAC protocols, and the sensing hardware models, which limits its flexibility; (2) the lack of an application model makes it ineffective in environments that require interaction between applications and the network protocols; (3) It does not run real hardware code; (4) It has been built by many developers and contains several inherent known and unknown bugs. (5) It does not scale well for WSNs due to its object-oriented design; (6) Using C++ code and oTcl scripts makes it difficult to use.

To overcome the above drawbacks the improved NS-3 simulator [12] was developed. NS-3 supports simulation and emulation. It is totally written in C++, while users can use python scripts to define simulations. Hence, transferring NS-2 implementation to NS-3 require manual intervention. Besides the scalability and performance improvements, simulation nodes have the ability to support multiple radio interfaces and multiple channels. Furthermore, NS-3 supports a real-time schedule that makes it possible to interact with a real systems [12]. For example, a real network device can emit and receive NS-3 generated packets.

14.2.12. Shawn

Shawn is an open source discrete event simulator for WSNs. It is written in C++ and can be run in Linux/Unix and Windows environments. Shawn aims to simulate large- scale WSNs, where physically accurate simulations fail. The idea behind Shawn is to use abstract models to simulate the affects of a phenomenon rather than the phenomenon itself [24]. Users of Shawn can adapt the simulation to their needs by selecting the application preferred behaviour. The authors claim that Shawn provides a high abstraction level that hides a lot of the simulation details. Users are given full access to the communication graph, which allows them to observe nodes and their data [24]. However, there are some limitation in Shawn, for instance: Visualization output is not supported, MAC module is not extent, and also users need to do much programming [33].

14.2.13. GTSNetS

An open-source, written in C++, and it is event-driven simulator for large scale WSNs. Users can evaluate different strategies and hierarchical options with its effect on lifetime and performance on particular segments. It can also adapt a new approaches such as algorithms and protocols. Khan et.al. (2011) claims the it is the only simulator can simulate network with large number of nodes(up to 200.000) [34].

GTSNetS has been designed in a way to allow users to use any architecture or design to simulate a particular required WSNs. Alternatives such as: tracing options, different

energy models, network protocols, and the ease of extended for a specific need; will provide various implementation for the developers. Other advantages of GTSNetS that the overall lifetime of the simulated network is trackable, the energy consumption of each unit can be measured, supporting mobility such as mobile sensor nodes, mobile base station, and the movement of sensed objects [35]. The serious drawback in this simulation tools is that not been updating since 2008.

14.2.14. CNET

CNET [36] is an open source network simulator developed at the University of Western Australia for research and learning purposes. This simulation platform allows various experimentation at the data-link, the transport and the network layers. It implements both the IEEE 802.112 and IEEE 802.3 protocols, which makes it capable of simulating mobile and wireless networks. At the application layer, CNET provides advanced users with the capability to develop their own applications; however, users must write protocols to fill in any necessary internal layer to cope with any lost frame that might be introduced by CNET Physical layer.

CNET comes with a set of tutorials and comprehensive documentation that describes the process from installation stage up to final stage of simulation. Users can have full simulation vision and graphical presentation of the network under Tcl/Tk CNET 's comes with a graphical interface that allow users to relocate nodes by mouse and change nodes attributes using a simple checkbox interface. The major drawback of CNET is that it can only run a maximum of few hundreds nodes, which limit its usefulness in studding large-scale networks. It also does not support Windows operating systems.

14.2.15. TRMSim

Trust and Reputation Models Simulator for WSNs (TRMSim-WSN) is a special purpose Java-based simulator to study trust and reputation models for a wide range of WSNs. TRMSim-WSN is a highly optimizable simulator. It allows users to set several simulation parameters, including, the percentage of malicious nodes, the probability of collusions, etc. It offers some of the most common trust and reputation models found in the literature. New models can be easily added using the provided API, which implements a template to carry out this task. TRMSim checks the correctness and accuracy of trust and reputation only; it does not allow users to study the performance aspects of their network. For instance, users of TRMSim-WSN are unable to measure the energy consumption imposed by trust/reputation models on the network. Another limitation is the inability to choose a specific node and change its properties, e.g., services offered and goodness. Finally, only static nodes can be simulated.

14.3. Testbd as a Service

Experimental evaluation have become an essential part of WSN innovations. The WSNs research community needs networking research “Laboratories” in which to develop and

test novel concepts. Currently, the number of algorithms evaluated through testbeds is relatively low. This is due to the high deployment cost and application dependence of WSNs.

Recently, a new class of WSNs testbed facilities, called Testbed as a Service (TaaS) has emerged to overcome the difficulties in testbed-based evaluation. TaaSprovides online laboratories to facilitate research in sensor network programming environments, communication protocols, system design, and applications. It makes the testbed accessible through a web based user interface (WUI), opening it to users across the globe. Therefore, TaaS provides reliable and cost effective work environment for the researcher enabling broad range of innovations. TaaS offers the ability to rapidly prototype new ideas as well as the ability to duplicate real world scenarios such that research results are realistic and valid.

Compared to the conventional testbeds, TaaS has many advantages including: First, it reduces the costs of both deployment and maintenance of a large-scale testbed. Second, a variety of platforms are available with different capabilities, and hence, the deployment of heterogeneous sensor nodes becomes feasible. Third, it prevents unexpected or errant behaviour from interfering with production services or other testing. Fourth, since TaaS provides easy-to-use testbeds infrastructure, the level of expertise needed to set up experiments is less compared to that in conventional testbeds. Fifth, developers can focus development and evaluation efforts on the area of contribution without worrying about other network tasks. For instance, a variety of MAC and routing protocols may be available to use saving the developer time and effort to build the entire network stack. Finally, running large-scale WSN experiments is hard. Programming a single sensor node can take up to five minutes depending on the size of the binary it is being programmed with. Without a testbed to streamline the process of programming a large number of sensor nodes, programming each sensor node individually can quickly become a significant bottleneck. Furthermore, as programming typically takes place by physically attaching a sensor node to the local machine, deploying them afterwards also takes a significant amount of time.

The rest of this section presents some of the largest indoor sensor network testbeds with public remote access.

INDRIYA[37] is a large-scale, low-cost WSN deployed across three floors of the School of Computing, at the Singapore. The Testbed comprises of 139 TelosB sensor "motes"[38] and a cluster of Arduinodevices [39] to support the requirements of the future Internet of Things (IoT). TelosB nodes in INDRIYA run the TinyOS [40] operating system and are programmed in the NesC programming language, a component-oriented variant of C. Typically, users will be able to prototype their application either using the TOSSIM simulation environment [3] or with a handful of motes on their desktop. Users then use the INDRIYA's web interface to upload their program to the building-wide network. In INDRIYA, the wireless connectivity among nodes is three-dimensional, which allows experimentation of protocols that are sensitive to placement and connectivity, such as geographical routing protocols. INDRIYA’s heterogeneous nodes allow the evaluation of various WSN applications. However, INDRIYA has a number of

limitations: First, the number of nodes is still not sufficient for a very large-scale WSN (i.e., thousands of nodes). Second, it does not consider nodes mobility, which limits the deployment of mobile WSN systems. Finally, in terms of heterogeneity of hardware setup, INDRIYA has only two platforms.

Washington University in St. Louis is currently in the process of deploying a WSN testbed called The WUSTL WSNs Testbed [41]. It currently consists of 79 wireless sensor nodes placed throughout several office areas. The testbed deployment is based on the TWIST architecture [42]. It is hierarchical in nature, consisting of three different levels of deployment: sensor nodes, microservers, and a desktop class host/server machine. The entire TWIST architecture is based on cheap, off-the-shelf hardware and uses open-source software at every tier of its deployment. The testbed software has a client/server architecture, plus a set of user applications that interact with the wireless testbed. The client runs on the microservers and manages finding, programming, and communicating with the attached motes. The server coordinates the clients and the user applications. The user applications help the user to program, list, and communicate with the motes in the system. The client, server, and user application software is written mostly in Python, with some performance-critical sections written in C. In addition, the server application interacts with a PostgreSQL database, which lists the active microservers and motes. The PostgreSQL backend is also available for user applications to store experimental data. The testbed features a localization service where motes in the testbed act as anchor nodes to help localize a beacon node roaming throughout the network. However, the number of sensor nodes in the WUSTL testbed is still relatively low and the homogenous hardware platform does not allow developers to capture some aspect of such complex systems.

The WISEBED project is a consortium of nine academic and research institutes across Europe. The project aims to build a multi-level infrastructure of interconnected testbeds for research purposes. Such a testbed allows applying an interdisciplinary approach to WSNs development by integrating the aspects of hardware, software, algorithms, and data. It consists of over 1500 wireless sensor nodes spread across the nine sites. A variety of sensor nodes is available allowing heterogeneous network setup. The testbed has more than eight platforms including iSense, Pacemate, and TelosB. The nodes can be allocated various types of mobility and environments. The WISEBED supports nodes mobility by deploying senor nodes on a robot such as, Roomba [43], and NXT [44]. The Wiselib C++ template-based algorithm library allows WISEBED users to build cross-platform, highly efficient WSN applications. They can also use the GNU C based on Lorien OS to build highly dynamic component oriented applications. WISEBED is compatible with Shawn simulator; therefore, WISEBED can derive models of real-life scenarios for simulations.

IoT-LAB [45] is an open testbed that evolved from SENSLAB testbed. It offers a very large-scale infrastructure suitable for testing small wireless sensor devices and heterogeneous communicating objects. IoT-LAB is made of over 2700 wireless sensor nodes that are distributed across six sites in France. A range of wireless sensor hardware platforms are available, with different processor architectures (MSP430, STM32 and Cortex-A8) and different wireless chips (802.15.4 PHY at 800 MHz or 2.4 GHz). IoT-LAB features an “open node” that can receive custom wireless sensors for inclusion

in IoT-LAB testbed. Nodes are either stationary or mobile and can be organized in different topologies. IoT-LAB include mobile nodes with either controlled or uncontrolled mobility. Similar to WISEBED , IoT-LAB controlled mobile nodes are carried on robot, such asTurtlebot2 [46] or Wifibot [47], which have advanced navigation features. IoT-LAB allows full control of network nodes and direct access to the gateways to which nodes are connected. This allows researchers to monitor nodes energy consumption and network-related metrics, e.g., end-to-end delay, throughput and overhead. This testbed offers quick experiments deployment, along with easy evaluation, results collection and analysis. It also allows, implementing testbeds with different node types, topologies and environments to coverage of a broad range of real-life WSNs applications. IoT-LAB offers web-based reservation and tooling for applications development, along with direct command-line access to the platform.

Table 14.1 summarizes the main features of the above reviewed testbed services. Table 14.2 gives a summary of the hardware and software resources that are commonly used in WSNs testbed facilities.

Table 14.1. Summary of the reviewed testbed services.

TaaS	Open Access	Number of Nodes	Mobility Support	Heterogeneity (Number of Platforms)
INDRIYA	✓	139		2 Platforms: TelosB and Arduino
WUSTL	✓	79		1 Platform: TelosB
WISEBED	✓	>1500	✓	> 8 Platforms: including iSense, Pacemate and TelosB
IoT-LAB	✓	>2700	✓	> 3 Platforms: including Contiki WSN430 node, M3 node and A8 node

14.4. Discussion

The high cost of sensor node hardware has significantly delayed the deployment of WSNs. However, during the past few years, the cost of sensor nodes has decreased rapidly due to increased production volume and technological advancements. In some cases, e.g. RFID hardware, sensor costs have declined by as much as 100 times over the past decade. As the cost of a WSN system plunges, one can expect to see larger systems built and deployed. The focus then will shift to the availability of scalable networking and software platforms to support WSN systems deployment.

Generally, real WSNs testbeds provide a more accurate, realistic, and replicable validation mechanism for algorithms and protocols. However, the wide variety of available sensor hardware can make it rather difficult to replicate any results produced by real testbeds. Besides, in some applications, where dangerous conditions are being studied, e.g. chemical pollution, a real testbed is an unwanted choice. Finally, there is a handful TaaS environments that can satisfy some complex application needs. Out of these restrictions came the need for simulation as a tool for validating and testing algorithms/protocols. As

shown in Section 14.2, simulation tools are widely available and used by WSNs researchers. However, most of the existing simulators are incomplete and follow different approaches to investigate different problems. The variety of existing simulation tools has led to accuracy and authenticity issues that concern even the best simulators available today. Such issues also make it even more difficult to replicate and compare evaluation results from competing simulation systems. Simulation drawbacks also include the lack of visualization tools, GUI's, poor documentation, absence of examples, amongst others.

There are some factors that affect the accuracy of results among different simulators, even when applying the same scenario in different platforms, according to [34, 48-50]. These factors including: protocol stack, application's nature, parameter sitting, simulation assumption, and traffic generator parameters.

To solve the dilemma of having an accurate but scalable and low-cost prototyping solution, we suggest the use of mixed-mode simulation as an effective interim solution. Mixed-mode simulation is the integration of a simulated environment and a real testbed to improve both the accuracy and scalability of testing results. In other words, the mixed-mode simulation enables the simulation of algorithms partially in software and partially in a real hardware WSN testbed. A small number of simulation tools like NS-3 and Dingo already support this mode of simulation. This simulation mode allows researchers to compare the results of running the same algorithm in both simulation and on physical sensor hardware, the comparison allows the inclusion or the modelling of more realistic conditions in the simulation environment. A flexible mixed-mode simulator should support integration of heterogeneous sensor devices. Also, the simulation-testbed interaction remains a challenging task that needs to be addressed. For instance, the authors of Dingo describe in [51] a new Python library that implements synchronous message-passing concurrency to improve coordination between many hosts.

Yet, the choice of a suitable simulator is a difficult decision. There is no 'best' simulator, each simulator has specific features that work well in certain circumstances. The selection of a simulator depends mostly on the algorithmic feature to be evaluated. High level simulators like NS-2 gives an estimation about the applications and some middleware behaviour. Mid-level simulators, e. g. OMNET, provides more information about the physical layer components that are simulated without giving too much details. Low-level simulators provide accurate bit level estimations of the hardware as well as software performance. Regardless of the simulator, any simulations will always have weaknesses either due to non-realistic assumptions or modelling errors that may be present in the algorithm itself. Therefore, developing formal methods, e. g. using graph theory [52], to verify the correctness of new algorithms and protocols is also part of the testing or evaluation research. Table 14.3 summarizes the main features of the reviewed simulation tools.

Table 14.2. Summary of WSNs hardware platforms.

Node	Operating System	CPU	Memory Flash, (RAM)	Wireless Interface	Interfaces	Sensors
iSense	iSense	32-bit (16 MHz) ARM	128 KB, (96 KB)	IEEE 802.15.4 (2.4 GHz)	USB, RS- 232	Temperature, Light, PIR, AMR, Accelerometer
iMote2	TinyOS	32-bit (13-416 MHz) Intel PXA271	32 MB, (32 MB)	IEEE 802.15.4 CC2420 (2.4 GHz)	USB, JTAG I2C, 2 x (SPI) 3xUART, GPIOs, AC97 etc.	Accelerometer, Temperature, Humidity, Light
Scatter Node MSB430	ScatterWeb	16-bit(4 MHz) MSP 430F1612	55 KB, (5 KB)	CC1020 (868 MHz)	JTAG, GPIO	Temperature, Humidity, Accelerometer
MicaZ	Contiki	8-bit(16 MHz) Atmega128L	128 KB, (4 KB)	CC2420 (2.4 GHz)	JTAG, USB	Temp., Light, Microphone, Accelerometer, Magnetometer
Tmote Sky	TinyOS	16-bit (8 MHz) MSP430F1611	48 KB, (10 KB)	CC2420 IEEE 802.15.4 (2.4 GHz)	USB	Temperature, Humidity, Light
ESB	ScatterWeb	16-bit (8 MHz) MSP430F149	60 KB, (2 KB)	TR1001 868.35 MHz	RS232, JTAG	Temp., Light, Noise, Vibration, Infrared
Sun SPOT	None (JVM)	32-bit(180 MHz) ARM920T	4 MB, (512 KB)	IEEE 802.15.4 (2.4 GHz)	USB, GPIOs	Accelerometer, Temperature, Light
Scatter Net MSBA2	Contiki	32-bit (72 MHz) ARM7	512 KB, (98 KB)	Chipcon CC1100 (864-970MHz)	miniUSB 2.0, GPIOs	Temperature, Humidity
GumStIx	Linux	32-bit (400 MHz) Intel PXA255	16 MB, (64 MB)	IEEE 02.11b/g (2.4 GHz)	USB, Bluetooth	Water Depth
Pacemates	iSense	32-bit (60 MHz) LPC2136	256 KB, (64 KB)	Xemnics RF (868 MHz)	Serial, I2C, Radio Interface	Heart rate monitor
TelosBs	TinyOS/ Octopus	16-bit (8 MHz) MSP430F1611	48 KB, (10 KB)	CC2420 IEEE 802.15.4 (2.4 GHz)	USB	Temperature, Humidity, Light

Table 14.3. Summary about reviewed simulation tools.

Simulator	Programing Language	GUI	General or Specific Simulator	Open Source	Main Features	Limitations
SensorSim	C++	No	Specifically designed for WSNs	Yes	Power and communication protocol models Sensing channel and sensor models. Scenario generation. Support for hybrid simulations	Limited in SensorSim project realism. Consider limited resources of sensor nodes. Simulates the complete WSN protocol stack
TOSSIM	C++	Yes	Specifically designed for WSNs	Yes	Can be targeted to motes without modification. Nodes share the exact same code image. The developed algorithms can be tested on a target platform.	Makes several assumptions about the target hardware platform. Focusing on making some behaviour accurate while simplifying others
TOSSF	C++	Yes	Specifically designed for WSNs	Yes	Primary focus on scalability. Support heterogeneous nodes and dynamic topology	Long test-debug cycles
GloMoSim	C/Parsec	Yes	General	Yes	Supports protocols designed purely for wireless networks .Built using a layered approach. Uses standard APIs between different simulation layers.	Not scalable of simulating sensor networks accurately. Does not support phenomena occurring outside of the simulation environment
Qualnet	C/C++	Yes	General	Comm-ercial	Comprehensive set of advanced wireless modules and user-friendly tools	The annual license is expensive
Simulator	Programing Language	GUI	General or Specific Simulator	Open Source	Main Features	Limitations
OPNET	C/C++	Yes	General	Comm-ercial	Uses a hierarchical model to define each characteristic of the system. Capability of recording a large set of user defined results.	Scalability problems
EmStar	C	Yes	Specifically designed for WSNs	Yes	Supports hybrid mode. Provides an option to interface with actual hardware while running a simulation. Compatible with two different types of node hardware	Supports only the code for the types of nodes that it is designed to work with
SENS	C++	No	Specifically designed for WSNs	Yes	Multiple different component implementations	Less customizable. Only measurable phenomenon is sound

Table 14.3. (Continued).

Simulator	Programing Language	GUI	General or Specific Simulator	Open Source	Main Features	Limitations
J-Sim	Java	Yes	Specifically designed for WSNs	Yes	Ability to simulate the use of sensors for phenomena detection. Support for using the simulation code for real hardware sensors	Comparatively complicated to use. Unnecessary overhead in the intercommunication model
Dingo	Python	Yes	Specifically designed for WSNs	Yes	Full functionality of a programming language can be used. Option to split the visualization from the simulation	Does not directly support sensor networks at the physical layer. Incomplete nature of the tool. Not supported anymore.
NS-3	C++	No	General	Yes	Supports simulation and emulation. Supports a real-time schedule. Ability to support multiple radio interfaces and multiple channels.	Some restrictions on the customization. Lack of an application model. Does not run real hardware code. Does not scale well for WSNs.
Simulator	Programing Language	GUI	General or Specific Simulator	Open Source	Main Features	Limitations
Shawn	C++	No	Specifically designed for WSNs	Yes	Able to simulate large- scale WSNs. Ability of selecting the application preferred behaviour. Full access to the communication graph.	Does not support visualization output. MAC module is not extent. Lots of programing is required
GTSNetS	C++	Yes	General	Yes	can also adapt a new approaches such as algorithms and protocols. The only simulator can simulate network with large number of nodes.	No update since 2008
CNET	C	Yes	General	Yes	For research and learning purposes. Allows various experimentation at the data-link, the transport and the network layers. Full simulation vision and graphical presentation.	It can only run a maximum of few hundreds nodes. Does not run under Windows
TRMSim-WSN	Java	Yes	Specific for WSNs trust/reputation testing	Yes	Check the correctness and accuracy of trust and reputation models. Provide an API to add new models.	It cannot simulate mobile networks. No support for performance evaluation.

14.5. Conclusion

This chapter provides a comprehensive review of simulation tools that are widely used in the field of WSNs. The aim is to help researchers choosing the most appropriate simulation tools to evaluate their work. There are a variety of simulation tools with different capabilities. However, the authors believe that they are insufficient for testing and evaluating WSNs algorithms. This is because the simulation results can be unrealistic due to the incomplete or inaccurate simulation models. An immediate measure is to develop unified models, e. g. energy, for different simulators. This allows realistic comparisons between results produced by different simulators to be made. To improve authenticity and accuracy of simulation results, it is important that researchers make their simulation code available for download by other researchers. Moreover, researchers should dedicate more space in their papers to clearly describe their simulation setup. On the other hand, large-scale real testbeds are still infeasible due to their cost and complexity. It can be easily observed that the trend in the WSNs field is to use mixed-mode simulation as an interim solution. Finally, we believe that theoretical validation of algorithms can serve as a good means for evaluating many WSNs algorithms.

References

[1]. F. Al-Fayez, *et al.,* Wireless Sensor Network Simulation: The Current State and Simulation Tools, *Sensors & Transducers,* Vol. 17, Special Issue, December 2012, pp. 145-155.

[2]. E. Systems, (2015, 2 Nov, 2015), *ESP8266EX Datasheet*, Available: https://www.adafruit.com/images/product-files/2471/0A-ESP8266__Datasheet__EN_v4.3.pdf

[3]. P. Levis, *et al.,* TOSSIM: accurate and scalable simulation of entire TinyOS applications, in *Proceedings of the 1st International Conference on Embedded Networked Sensor Systems (SenSys'03)*, 2003, pp. 126-137.

[4]. P. Levis, *et al.,* TinyOS: An Operating System for Sensor Networks, in Ambient Intelligence, *Springer,* 2005.

[5]. IARIA. (2006-2011, Retrieved 19 April 2015). *The 5th International Conference on Sensor Technologies and Applications, SENSORCOMM 2011*. Available: http://www.iaria.org/conferences2011/SENSORCOMM11.html.

[6]. S. Kurkowski, *et al.,* MANET Simulation Studies: The Current State and New Simulation Tools, *The Colorado School of Mines*, February 2005.

[7]. E. Larsen, *et al.,* iOLSR: OLSR for WSNs Using Dynamically Adaptive Intervals, in *Proceedings of the 5th International Conference on Sensor Technologies and Applications (SENSORCOMM'11)*, Nice/St. Laurent du Var, France, 2011, p. 18 - 23.

[8]. K. Shi, *et al.,* TinyMQ: A Content-based Publish/Subscribe Middleware for Wireless Sensor Networks, in *Proceedings of the 5th International Conference on Sensor Technologies and Applications (SENSORCOMM'11)*, Nice/Saint Laurent du Var, France, 2011, p. 12 - 17.

[9]. R. Behnke, *et al.,* HDLS: Improved Localization via Algorithm Fusion, in *Proceedings of the 5th International Conference on Sensor Technologies and Applications (SENSORCOMM'11)*, Nice/Saint Laurent du Var, France, 2011, p. 36 - 44.

[10]. S. Feldman and M. Feldman, Tree-Based Organization for Very Large Scale Sensor Networks, in *Proceedings of the 5th International Conference on Sensor Technologies and Applications (SENSORCOMM'11),* Nice/Saint Laurent du Var, France, 2011, p. 45-50.

[11]. F. Derogarian, *et al.,* A Routing Protocol for WSN Based on the Implementation of Source Routing for Minimum Cost Forwarding Method, in *Proceedings of the 5th International Conference on Sensor Technologies and Applications (SENSORCOMM'11)*, Nice/Saint Laurent du Var, France, 2011, p. 85 to 90.
[12]. nsnam, *NS-3*, Available: from http://www.nsnam.org/(2011, Retrieved 18 April 2015).
[13]. S. Park, *et al.,* SensorSim: a simulation framework for sensor networks, in *Proceedings of the 3rd ACM International Workshop on Modeling, Analysis and Simulation of Wireless and Mobile Systems (MSWIM'00)*, 2000, pp. 104-111.
[14]. C. Ulmer, 2004, *Wireless Sensor Probe Networks-SensorSimII*.
[15]. L. F. Perrone and D. M. Nicol, A Scalable Simulator For TinyOS Applications, in *Proceedings of the Winter Simulation Conference,* Vol. 1, 2002, pp. 679 - 687.
[16]. A. Gahng-Seop, *et al.,* Supporting Service Differentiation for Real-Time and Best-Effort Traffic in Stateless Wireless Ad Hoc Networks (SWAN), *IEEE Transactions on Mobile Computing,* Vol. 1, 2002, pp. 192-207.
[17]. X. Zeng, *et al.,* GloMoSim: a library for parallel simulation of large-scale wireless networks, in *Proceedings of the 12th Workshop on Parallel and Distributed Simulation (PADS '98)*, 1998, pp. 154-161.
[18]. D. Curren, 2005, A survey of simulation in sensor networks, https://www.researchgate.net/publication/228642170_A_survey_of_simulation_in_sensor_networks.
[19]. QualNet Simulator, *S. N. Technologies,* Available: from http://www.scalable-networks.com/products/qualnet/, Retrieved 02/01/2015.
[20]. X. Chang, Network simulations with OPNET, in *Proceedings of the 31st Conference on Winter Simulation (WSC '99)*, 1999, pp. 307-314.
[21]. L. Girod, *et al.,* Emstar: A software environment for developing and deploying heterogeneous sensor-actuator networks, *ACM Trans. Sen. Netw.,* Vol. 3, p. 13, 2007.
[22]. S. Sundresh, *et al.,* SENS: A Sensor, Environment and Network Simulator, in *Proceedings of the 37th Annual Simulation Symposium (ANSS'37),* April 2004.
[23]. A. Sobeih, *et al.,* J-Sim: A Simulation Environment for Wireless Sensor Networks, in *Proceedings of the 38th Annual Symposium on Simulation (ANSS '05),* pp. 175-187, 2005.
[24]. S. P. Fekete, *et al.,* Shawn: The fast, highly customizable sensor network simulator, in *Proceedings of the 4th International Conference on Networked Sensing Systems (INSS '07)*, 2007, pp. 299-299.
[25]. S. Mount, *et al.,* SenSor: an Algorithmic Simulator for Wireless Sensor Networks, *in Proceedings of the Eurosensors XX,* Vol. II, 2006, pp. 400-411.
[26]. hp, iPAQs, 2000.
[27]. xbow, Mica Mote, 2012, Retrieved 19-April-2012s.
[28]. S. Mount, Dingo Wireless Sensor Networks Simulator, 2008.
[29]. M. Hammoudeh, *et al.,* MuMHR: Multi-path, Multi-hop Hierarchical Routing, in *Proceedings of the International Conference on Sensor Technologies and Applications (SENSORCOMM'07)*, 2007, pp. 140-145.
[30]. W. Heinzelman, *et al.,* Energy-Efficient Communication Protocol for Wireless Microsensor Networks, in *Proceedings of the 33rd International Conference on System Sciences,* Vol. 2, January 2000, p. 10.
[31]. Gumstix.com, Gumstix way small computing, 2007.
[32]. NS-2, The Network Simulator, 2007.
[33]. E. Kolega, *et al.,* Assessment of network simulators for real world WSNs in forest environments, in *Proceedings of the IEEE International Conference on Networking, Sensing and Control (ICNSC),* 2011, pp. 427-432.
[34]. M. Z. Khan, *et al.,* Limitations of Simulation Tools for Large-Scale Wireless Sensor Networks, in *Proceedings of the IEEE Workshops of International Conference on Advanced Information Networking and Applications (WAINA'11)*, 2011, pp. 820-825.

[35]. C. J. Fung and Y. E. Liu, Lifetime Estimation of Large IEEE 802.15.4 Compliant Wireless Sensor Networks, in *Proceedings of the IEEE International Symposium on Modeling, Analysis and Simulation of Computers and Telecommunication Systems, (MASCOTS'08)*, 2008, pp. 1-4.
[36]. CNET, An introduction to the cnet network simulator, ed. *The University of Western Australia*, 2015.
[37]. M. Doddavenkatappa*, et al.,* Indriya: A Low-Cost 3D Wireless Sensor Network Testbed, *TRIDENTCOM,* 2011.
[38]. I.n.c. Crossbow Technology, 2005, 12 Nov 2015, TelosB Mote Platform, Available: http://www.willow.co.uk/TelosB_Datasheet.pdf.
[39]. Arduino, 2015, 13 Nov 2015, Arduino, Available: https://www.arduino.cc/
[40]. TinyOS, 2015, 12 Nov 2015, TinyOS, Available: http://www.tinyos.net/
[41]. W. U. i. S. Louis, 2015, 13 Nov 2015, The WUSTL Wireless Sensor Network Testbed. Available: http://wsn.cse.wustl.edu/index.php/The_WUSTL_Wireless_Sensor_Network_Testbed.
[42]. T. Group, 2015, 13 Nov 2015, TKN WIreless NetworkS Testbed. Available: https://www.twist.tu-berlin.de/index.html
[43]. iRobot, 2015, 12 Nov 2015, Roomba Robot, Available: http://www.irobot.co.uk/Store/Robots/roomba
[44]. LEGOengineering, 2014, 12 Nov 2015, NXT sensors, Available: http://www.legoengineering.com/nxt-sensors/
[45]. T. F. consortium, 2014, 12 Nov 2015, FIT IoT-LAB, Available: https://www.iot-lab.info/
[46]. I. Open Source Robotics Foundation, 2015, 12 Nov 2015, TurtleBot 2. Available: http://www.turtlebot.com/
[47]. N. Robotics, 2015, 15 Nov 2015, http://www.wifibot.com/. Available: http://www.wifibot.com/.
[48]. T. R. Andel and A. Yasinsac, On the credibility of manet simulations, *Computer,* Vol. 39, 2006, pp. 48-54.
[49]. S. Mehta*, et al.,* A Case Study of Networks Simulation Tools for Wireless Networks, in *Proceedings of the 3rd Asia International Conference on Modelling & Simulation (AMS '09)*, 2009, pp. 661-666.
[50]. M. Korkalainen*, et al.,* Survey of Wireless Sensor Networks Simulation Tools for Demanding Applications, in *Proceedings of the 5th International Conference on Networking and Services (ICNS '09)*, 2009, pp. 102-106.
[51]. M. Hammoudeh, Modelling Clustering of Sensor Networks with Synchronised Hyperedge Replacement, in *Proceedings of the 4th international Conference on Graph Transformations (ICGT '08)*, 2008*,* pp. 490-492.
[52]. S. Mount*, et al.,* CSP as a Domain-Specific Language Embedded in Python and Jython, in Communicating Process Architectures*, Eindoven*, Netherlands, 2009.

Chapter 15
CWSN: A Graph-based Model for Collaborative Wireless Sensor Networks

Laura Margarita Rodríguez Peralta,
Lina M. Pestana Leão Brito and Eduardo Ismael Hernández

15.1. Introduction

WSNs are a special case of wireless ad hoc networks, but characterized by specific constraints. Besides energy restrictions, sensor nodes suffer from other resource limitations: they have reduced memory and processing capabilities; and, due to short transmission range, nodes can only communicate locally, with a certain number of local neighbors [1-4]. In many cases, these networks are subject to highly dynamic conditions, caused by nodes' mobility, hardware failures, lack of battery, or other factors. To overcome these limitations, nodes have to collaborate in order to accomplish their tasks: sensing, signal processing, computing, routing, localization, security, etc. Therefore, WSNs are, by nature, collaborative networks [5].

There are quite a few works in the literature concerning collaboration in WSNs; however, they only focus a specific type of collaboration, which is associated with the accomplishment of a specific task. In [6], we proposed a formal and hierarchical model of cooperative work, the Collaborative Wireless Sensor Networks (CWSN) model, which is designed specifically for WSNs. It allows not only the modeling of collaborative work (based in CSCW - Computer Supported Cooperative Work [7] - concepts), but also the modeling and visual representation of all the entities that can compose a WSN, as well as its properties. Moreover, CWSN is a generic model since it can be applied to heterogeneous networks. We have used first-order logic to formalize and describe the proposed model. We further employ graph-theory to describe how communication occurs within a cluster. This formal description of clusters' communication is based on directed graphs; nodes are labeled with the signal +/-, respectively indicating if a node is active or not [6].

Laura Margarita Rodríguez Peralta
Engineering Department, UPAEP University, Puebla, Mexico

In this chapter, we focus on using the graphical representation of the CSWN model to represent the state of the network and its properties. This graphical representation allows the user to comprehend what is occurring in a certain moment of the WSN lifetime, by easily visualizing the communication interactions, the state of the nodes (active or not), the state of the links, etc. This will allow enriching the proposed model and giving the user a better understanding of the components and the state of the WSN as well, through a more complete visual representation of the WSN. This is the main contribution of this chapter.

This chapter is organized as follows. In Section 15.2, we briefly describe the related work. In Section 15.3, the CWSN model and its entities are briefly presented. Then, a comparison between CWSN and other models is presented. The advantages of the model are, also, outlined. Section 15.4 presents the SAVER project, an application example of our model. Section 15.5 provides some conclusions and future work perspectives.

15.2. Related Works

There are several works that try to model some aspects of WSNs. We have observed that the great majority of works focus on modeling of connectivity or mobility problems, or even on both problems. Moreover, we have identified other modeling concerns, such as: communication models, interference models, data aggregation models, coverage models, and signal processing models. On the contrary, the CWSN model intends to model a whole WSN, i.e., it tries to consider the most complete set possible of entities that can exist in a WSN, and their respective attributes.

Regarding the works focusing collaboration in WSNs, the great majority of them covers a specific type of collaboration, which is associated with the accomplishment of a certain task, such as: signal processing [8], sensing [9], computing [10], routing [11], localization [12], security [13], task scheduling [14], heuristics [15], calibration [16], resource allocation [17], time synchronization [18], transmission [19], etc., and also works concerning collaboration between wireless sensor nodes and other kind of devices (heterogeneous groupware collaboration) [20, 21] to support some specific applications (for example, collaboration between sensor nodes and PDAs, in a fire fighting scenario).

According to the literature, the only work that presents a model for collaborative work, in sensor networks, was proposed by Liu et al. [22]. It is the Sensor Networks Supported Cooperative Work (SNSCW) model, a hierarchical model that essentially divides cooperation in sensor networks in two layers; the first one relates to cooperation between humans and sensor nodes; the second one relates to cooperation between the sensor nodes. This model was designed for sensor networks.

However, the SNSCW model only allows the modeling of collaboration itself. On the contrary, the CWSN model, which has been presented in [6], is a formal model that was created specifically to describe WSNs. However, the CWSN model allows not only the modeling of collaborative work (based in CSCW concepts), but also the modeling, formalization and visual representation of the entities that can constitute a WSN (different

types of nodes, clusters, relationships, sessions, obstacles, etc.), as well as its attributes. Moreover, it allows the representation of the WSN's hierarchy and of the network evolution.

The CWSN model formalizes the most significant properties of each entity through first-order logic. Even though the CWSN model is a graph-based model, it includes other objects [6] in order to make the modeling of the various entities of a WSN possible. This is of paramount importance to completely represent a WSN.

WSNs are extremely dynamic systems, both in the sense that their characteristics change over their lifetime and for the fact that sensor networks' technology (hardware and software) is subject to fast changes. To overcome this issue, the CWSN model can be updated or extended, through the introduction of new entities and/or new attributes. Therefore, another key point of this model is its scalability, since it can easily evolve.

15.3. The CWSN Model

The CWSN model is a formal model of collaborative work that was specifically created to describe WSNs. This model allows the representation of the entities (different types of nodes, clusters, relationships, sessions, obstacles, etc.) and properties of a WSN, of its hierarchy, and of the network evolution; therefore, it allows not only the modeling of collaborative work (based in CSCW concepts), but also the modeling, formalization and visual representation of a whole WSN.

The CWSN model formalizes all the properties of each entity through first-order logic. Also, CWSN is a graph-based model; however, it includes other objects in order to make possible the modeling and visual representation of all the entities that can compose a WSN. This is of paramount importance to completely represent a WSN.

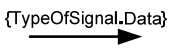
15.3.1. CWSN Model Definitions

We define entities as all the components that might exist in a WSN. The symbol, the concept and the description of all the entities included in the proposed model are illustrated in Table 15.1.

A WSN can have different types of nodes: ordinary wireless sensor nodes, anchor nodes, one or more sink nodes (also known as base stations) and a gateway. The sink node and the anchor node are wireless sensor nodes with special functions.

A cluster is a group of nodes, created according to: geographical area, type of sensor nodes, type of phenomenon, task to be performed, etc., providing the WSN with a hierarchical structure. If nodes are grouped in clusters, one of the members of each cluster becomes the cluster head (there is only one cluster head per cluster). In this case, all nodes in the cluster have to send collected data to the cluster head (for instance, the more powerful node or the router, in case of a ZigBee-based WSN), which, in turn, is responsible for sending data to a sink node.

Table 15.1. Definition of the entities can constitute a Wireless Sensor Network.

Symbol	Concept	Description
	Sensor node	Nodes can be either stationary or mobile. Also, they can be in one of three possible states: active, sleep mode (in order to save energy), or inactive.
	Sink node/ Base Station	Node to which data collected by ordinary nodes is sent; being responsible to send data to the gateway. If there is only one sink node, all data collected by sensor nodes has to be sent to it. Otherwise, data may be sent to any sink node and, in this case, sink nodes must be able to communicate to each other.
	Anchor node	Node with known localization, which support the other sensor nodes in the localization process
	Cluster	Group of nodes, created according to: geographical area, type of sensor, type of phenomenon, task, etc.
	Cluster Head	Sensor node to whom all sensor nodes in the cluster send the collected data; it is responsible for sending the received data to the Sink node.
	Relationship	The arrow represents a relationship between nodes A and B. It also represents and adjacency relation between nodes A and B; nodes A and B are neighbors. A relationship can be established based on: localization, phenomenon, type of sensor node, etc.
	Data flow	This label identifies both the type of signal being used (radio frequency, ultrasound, acoustical or light) and the type of data being transmitted between nodes (temperature, humidity, light, sound, video, internal voltage, etc.).
	Gateway	Device responsible to send the data to the user, through the Internet or satellite.
	Obstacle	An object (building, tree, rock, etc.) which may obstruct the line-of-sight between two or more nodes; depending on the type of signal that is being used by nodes (radio frequency, optical, acoustical, etc.), the obstacles may even not allow for communication between nodes.
	Session	In a certain moment, there may be several collaborative sessions in a WSN. A session can be established based on the objective (type of phenomenon to monitor, geographical area to monitor, etc.) of the WSN.
	Battery	It represents the percentage of the sensor node's remaining battery.
	User	Person that interacts with the WSN, querying the network, visualizing data, etc. The user customizes the work of the sensor nodes; the data collected by sensor nodes is used by the users' application.

If two nodes collaborate, there is a relationship between them. Associated with a relationship there is always an exchange of data, which corresponds to the data flow entity.

Collected data (temperature, humidity, light, acceleration, etc.) can be sent to other nodes using one or more types of signals (radio, ultrasound, acoustical, etc.).

Obstacles are objects (for, e.g., building, tree, rock, etc.) that may obstruct the line-of-sight between two or more nodes, not allowing for direct communication between them. So, they can influence the relationships created.

The user is the entity who interacts with the WSN, defining the application, querying the network, visualizing data, customizing the work of the sensor nodes, etc.

Several collaborative sessions can be established when monitoring a WSN, and they can exist simultaneously or not. Basically, new sessions may be established based on new queries a user makes on the WSN.

And, last but not least, as the battery is the most critical resource of a sensor node, it is really important that the user knows the state of the battery of each sensor. That is why the battery is also an entity of our model.

In this section, we formalize the model's entities and their main properties, using both first-order logic and graph theory.

1) Definitions: We can formulate the sensor network as a graph G (V, E). V (vertices) represents the set of sensor nodes, and E (edges) describes the adjacency relation between nodes. That is, for two nodes u, $v \in V$:

$(u, v) \in E$, if, and only if, v is adjacent to u. If a node u is within a node v's transmission range, we say that u is adjacent to v, or equivalently, that u is a neighbor of v. In the absence of interference, this relation is typically symmetric (or bidirectional), i.e., if a node u can hear a node v, also v can hear u.

Two nodes are able to communicate if they are adjacent and if the distance (d) between them is less than their maximum transmission range (R). This is known as the reachability property and is presented bellow.

$$d(a,b) < R(a) \wedge d(a,b) < R(b) \Rightarrow a \text{ and } b \text{ are reachable}$$

This property verifies that:

$$Adj(a,b): bool$$
$$Adj(a,b) \Rightarrow reachable(a,b)$$
$$\exists\, c: reachable(a,c),\ Adj(c,b) \Rightarrow reachable(a,b)$$

An arrow between two nodes represents a relationship between them. A relationship can be established based on: localization, phenomenon, type of sensor node, etc. The arrow represents a producer-consumer relationship. Let us consider two nodes A and B; the arrow Ⓐ⟶Ⓑ means that node A transmits data to node B. So, node B consumes

information from node A. The transmission of data between both nodes follows the format TypeOfSignal.Data ($\xrightarrow{\{TypeOfSignal.Data\}}$), verifying the consumer-producer property.

So, according to the specifications of the CWSN model, a WSN can be represented using labeled and directed graphs; the labels are associated to the edges and are designated by data flow. TypeOfSignal.Data identifies these labels. The labels inform the user about the type of signal that is being used by the sensor nodes for transmitting data (for, e.g., radio frequency, ultrasounds, acoustical, etc.), and about the type of data that is being collected and sent to the sink node (for, e.g., temperature, humidity, light, acceleration, etc.). These labels are important because they allow the user to become more aware of the state and the behavior of the WSN, since the labels add information that goes beyond the mere representation of the communication interactions between nodes.

Let's represent the total number of sensor nodes that constitute the WSN by Nr, with $N = \{1, 2, \ldots, Nr\}$ and a wireless sensor node by Ni, with $i \in N$. The WSN has a limited lifetime, which can vary from some hours to several months or years. Let us denote the lifetime of the network (in seconds) by LT, with $T = \{1, 2, \ldots, LT\}$, and the j^{th} second of life of the network by tj, with $j \in T$.

2) Sensor Node (Ni): A sensor node (Ni) is defined by:

Ni = {ID, TS, CM, CT, R, PS, L, TM, S, PD, CN}

Table 15.2 defines and formalizes the properties that are important to identify a sensor node (Ni). This Table serves as an example for the type of formalization that has been proposed for the remaining entities, presenting a formal description of their most important properties [6].

3) Sink Node (S_K): The sink node is the node to which data collected by ordinary sensor nodes is sent. It is responsible for sending data to the gateway being the only node that can do it, what verifies the flow control property.

Basically, a sink node is defined and characterized by the same set of attributes as a sensor node is. So,

```
SK = {ID, TS, CM, CT, R, PS, L, S, PD, CN, TM, Da}
```

Regarding mobility, two cases must be distinguished: the Stationary Sink Node (StS_K), with the localization of the sink being well-known and independent of time; and the Mobile Sink Node (MbS_K), where the localization of the sink node varies as it moves along the WSN.

In the case of stationary sink nodes, what is the case in the great majority of application scenarios, sink nodes are usually supplied by an external power supply source, which means that they do not rely on batteries. This ensures that the sink nodes have a bigger lifetime when comparing to the other nodes in the WSN what is a crucial aspect since the failure of sink nodes will conduct to the failure of an entire WSN.

Table 15.2. Properties definition of Sensor Node (Ni) entity.

◯	Properties	Description / Formalization
Sensor Node (N_i)	Identifier (I_D)	Each sensor node has a unique identifier (I_D) I_D (N_i) = i , i ∈ ℕ
	Types of sensors (TS)	A sensor node (N_i) can have several types of sensors, each one measuring a different phenomenon: light (Li), temperature (Te), humidity (Hu), sound (Sd), internal voltage (Iv), acceleration (Acc), pressure (Pr), vibration (Vb), received signal strength indicator (RSSI), etc. So, TS (N_i) ⊆ {Li, Te, Hu, Sd, Iv, Acc, Pr, Vb, RSSI, ...}
	Communication modality (CM)	A number of communication modalities can be used, such as: radio (RF), light (Li), ultrasound (US), acoustical (Aco), optical (Opt), hybrid (Hy). So, CM (Ni) ⊆ {RF, Li, US, Aco, Opt, Hy}
	Communication Technology (CT)	A sensor node (N_i) can use different communication technologies. So far, three technologies have been proposed: ZigBee (ZB), which corresponds to IEEE 802.15.4; Bluetooth (BT); IEEE 802.11 (Wi-Fi). So, CT (Ni) ⊆ {ZB, BT, Wi-Fi}
	Transmission Range (R)	Let P_t be the nominal transmission power of a node. P_R;j ← i is the received power of a signal propagated from node i to node j. A received power P_R;j ← i above a given threshold P_{th} will provide sufficient SNR (*Signal to Noise Ratio*) in the receiver to decode the transmission. The nominal transmission range for successful communication can be defined as [19]: $R = P_t/P_{th}$ Note that due to the instability in the transmission range, the area a wireless sensor node can reach is not necessarily a circle and the range can vary between $r=(1-\varepsilon).R$ and R, $\varepsilon>0$ [19].
	Power Supply (PS)	Energy can be supplied by batteries (that are, usually, of very limited capacity), solar cells or an external and unlimited power supply (only possible if nodes are stationary and in indoor applications). Battery (B); Solar cells (SC); External and unlimited power supply (VDC); Hybrid (Hy) – for, e.g., battery and solar cells; Etc. PS (StS_K) ⊆ {B, SC, VDC, Hy, etc.} However, the great majority of sensor nodes are equipped with batteries. The lifetime of a sensor node (N_i) is limited by its battery, depending on its capacity and type. The battery can be defined by: Type of battery: T_B, with T_B (N_i) ∈ {lithium, alkaline, li-ion, AA, external power supply, solar cells, electromagnetic and piezoelectric transducers, etc.} Capacity (voltage): C_B (N_i) [V] Remaining capacity at time t_j: P_{BNi} (t_j) [%] BNi (tj) = {TB (Ni), CB (Ni), PBNi (tj)}

Table 15.2. (Continued).

	Properties	Description / Formalization
	Localization (L)	Let $L_{Ni}(t_j)$, with $i \in \mathcal{N}$ and $j \in \mathcal{T}$, denote the location of node N_i at time t_j. The type of deployment affects important properties of the network (node density, node locations, etc.). The deployment of sensor nodes may be: Random (ad hoc deployment, for, e.g. dropped by an aircraft). In this case, the localization of a node is unknown: $L_{Ni}(t_j) = (x, y, z)$, where $x, y, z \in \mathbb{R}$ are unknown. Manual: sensor nodes are deployed in pre-determined positions. In this case, the localization of a node is well-known: $L_{Ni}(t_j) = (a, b, c)$, where $a, b, c \in \mathbb{R}$ are known.
	Product Description (PD)	Alphanumeric that identifies the brand and the model of the sensor node. PD (Ni) = {Brand (Ni), Model (Ni)}
	Consumer Node (CN)	Often referred to as parent node, is the node to which the sensor node N_i is sending data at time t_j. Considering for example a relationship between nodes A and B; if node A is transmitting to node B, then B is the consumer node since it is receiving the data.
	Type of Mobility (TM)	A sensor node (Ni) can be: Stationary (St): $L_{Ni}(t_1) = L_{Ni}(t_2) = \ldots = L_{Ni}(t_{LT})$ Mobile (Mb): The period of mobility can be occasional or continuous: Occasional (Oc), when long periods of immobility occur: $\exists j, l \in \mathcal{T}: L_{Ni}(t_j) \neq L_{Ni}(t_l)$, and $j \neq l \ \wedge \ \exists r, s \in \mathcal{T}: L_{Ni}(r) = L_{Ni}(r+1) = \ldots = L_{Ni}(s)$, and $s >> r$ Continuous (Cont): $\forall j \in \mathcal{T} \setminus \{LT\}\ L_{Ni}(t_{j+1}) \neq L_{Ni}(t_j)$ Mobility can still be classified in: Incidental (Inc), for, e.g., due to environmental influences ≈ Occasional Desired (Des), whether active or passive, which can be applied to any period of mobility (occasional or continuous). So, TM (Ni) ∈ {St, {OcMb, Inc}, {OcMb, Des}, {ContMb, Inc}, {ContMb, Des}}
	State (S)	Depending on its power mode, the node N_i can be in one of two states (S): ● Active (Ac): Node that is in the active state. Its color will depend on the cluster it belongs to, since each cluster will be represented by a different color. By default, the green color will be used. ○ Sleep Mode (Sm): Node that is in the sleep mode, in order to save energy. Colored in grey color. ● Inactive (In): Node that is damaged, or has some failure or has run out of energy. Colored in black color. So, S (Ni) ∈ {Ac, Sm, In}

4) Anchor Node (A): If the localization (L) of wireless sensor nodes is unknown (usually, due to an ad hoc deployment), it may be necessary to have some anchor nodes to help these sensor nodes to determine their own localization. So, an anchor node differs from a sensor node because its localization is always well known. This can be achieved either by equipping the anchor node with a GPS receiver or by manually configuring its position prior to deployment. Regarding mobility, an anchor node (A) can be:

Stationary (StA): In this case: TM (StA) = St

Mobile (MbA). In this case: TM (MbA) = {ContMb, Des} or TM (MbA) = {OcMb, Des}

5) Network (WSN). In what concerns to the network, the attributes that we consider more important to define and characterize this entity are: topology (To), mobility (M), homogeneity (H), number (Nr), area (A), coverage (C), density (D), number of sink nodes (NSK), number of anchor nodes (NA), number of clusters (NC), number of obstacles (NO), hierarchy (Hi), and lifetime (LT). So, a WSN can be defined by the following properties:

$$WSN = \{To, M, H, Nr, A, C, D, Hi, NS_K, NA, NC, NO, LT\}.$$

This set of attributes is defined in Table 15.3.

6) Session (Sei): A session is the essential unit of a collaborative activity, which can be created based on different queries posed by the user. Depending on the WSN specific application, sessions can take place in parallel or in sequence; or they can be synchronous or asynchronous. Thus, in a certain moment, there may be several collaborative sessions in a WSN. A session (Sei) can also be formulated as a sub graph, g, of the WSN, with $g(V, E) \subseteq G(V, E)$. Accordingly, some properties of the entities network and the sensor node are inherited. Similarly to a sensor node (Ni), a session (Sei) can be in one of two states: Active (Ac), or Inactive (In) when its objective is fulfilled. So:

S(Sei) = Ac or S(Sei) = In.

Besides, similarly to the entity network, each session can have a group of active sensor nodes, a group of inactive sensor nodes and a group of relationships and data flows. So, a session (Sei) is defined by the following properties:

Sei = {SeiID, SeObj, S, DTx, SeTLife, To, M, H, A, C, D, Hi, Nr, NSK, NA, NC, NO}.

Note that the topology (To) has the same definition as in Table 15.2; however, considering a specific instant of time, the topology of the session (Sei) may be different from the topology of the WSN.

Also, considering the number of nodes, number of anchor nodes, number of sink nodes, number of clusters and number of obstacles, note that:

Nr (Sei) ≤ Nr (WSN), NA (Sei) ≤ NA (WSN), NSK (Sei) ≤ NSK (WSN), NC (Sei) ≤ NC (WSN), NO (Sei) ≤ NO (WSN), and SeTLife (WSN) ≤ SeTLife (Sei)

Table 15.3. Attributes Definition of entity Network (WSN).

	Attributes	**Description / Formalization**
Network (WSN)	Topology (To)	The WSN can have different topologies (To): Single-hop, Star, Networked stars, Tree, Grid, etc. (Aspnes et *al.*, 2004). So, To ∈ {Single-hop, Star, Net-Stars, Tree, Grid, etc.}
	Mobility (M)	There are some different possible scenarios, regarding mobility of sensor nodes: All nodes are stationary: $\forall i \in \mathcal{N}, \forall j, l \in \mathcal{T}\ L_{Ni}(t_j) = L_{Ni}(t_l)$ All nodes are mobile: $\forall i \in \mathcal{N}, \exists j, l \in \mathcal{T}: L_{Ni}(t_j) \neq L_{Ni}(t_l)$, and $j \neq l$. Note that this does not imply that all nodes move the same amount within a time interval. Also, no specific mobility model is considered. Only some nodes move: $\exists i \in \mathcal{N}, \exists j, l \in \mathcal{T}: L_{Ni}(t_j) \neq L_{Ni}(t_l)$, and $j \neq l \ \wedge \exists p \in \mathcal{N}: L_{Np}(t_1) = L_{Np}(t_2) = \ldots = L_{Np}(t_{LT})$
	Homogeneity (H)	A WSN can be: Homogeneous (Ho), when it is composed of homogeneous devices, which means that sensor nodes are mostly identical from a hardware and a software point of view. Ho: $\forall$ i, p $\in \mathcal{N}$, Md (N_i) = Md (N_p) and $i \neq p$ Heterogeneous (He), when it is composed of heterogeneous devices, which means that sensor nodes are mostly different from both hardware and a software point of view, e.g., in type and number of attached sensors (TS). He: $\exists$ i, p ∈ N : Md (N_i) ≠ Md (N_p) So, H = Ho or H= He
	Number (Nr)	Total number of sensor nodes (N_i) that constitute the WSN, which may vary from a few nodes to thousands of sensor nodes. Nr ∈ $\mathbb{N}$
	Area (A)	This corresponds to the area of deployment (m^2). A ∈ $\mathbb{R}^+$
	Coverage (C)	A WSN can have different types of coverage: Sparse (Sp), when the network coverage is much smaller than its deployment area. Dense (De), when the network coverage coincides with its deployment area, or comes close to it. Redundant (Re), when multiple sensor nodes cover the same area. So, C ∈ {Sp, De, Re}
	Density (D)	The network density can be defined in terms of number of nodes per nominal coverage area (Bulusu et *al.*, 2001): $D = (Nr \times \pi \times R^2) / C_A$, C_A (m^2) ∈ $\mathbb{R}^+$, where C_A is the area that is covered by the whole network and R is the transmission range of the sensor node. Note that the coverage area (C_A) may be different from the deployment area (A).

Table 15.3. (Continued).

	Attributes	Description / Formalization
Network (WSN)	Number of sink nodes (NS_K)	A WSN has one or more sink nodes. A sink node can be stationary (StS_K) or mobile (MbS_K). $NS_K < Nr$
	Number of anchor nodes (NA)	Anchor nodes are nodes with known location. They can be stationary or mobile: Stationary anchor node (StA) Mobile anchor node (MbA) NA < Nr
	Number of Clusters (NC)	The number of clusters can be: Preset (Pr): In some approaches, the set of CHs are predetermined and thus the number of clusters (NC) is preset. Variable (Var): Randomly picking CHs from the deployed sensor nodes usually leads to a variable number of clusters. So, NC = Pr or NC = Var
	Number of Obstacles (NO)	$NO \in \mathbb{N}$
	Hierarchy (Hi)	Regarding data transmission, there is an inherent hierarchy in WSNs, since sensor nodes cannot transmit directly to the user; first, they have to transmit to the sink node, which in turn transmits to the user. If clusters exist, it provides the WSN with a hierarchical structure comprising more levels; all nodes in the cluster must transmit to the cluster head, which has to transmit to the sink node that in turn transmits to the user.

Also, considering the number of nodes, number of anchor nodes, number of sink nodes, number of clusters and number of obstacles, note that:

Nr (Sei) ≤ Nr (WSN), NA (Sei) ≤ NA (WSN), NSK (Sei) ≤ NSK (WSN), NC (Sei) ≤ NC (WSN), NO (Sei) ≤ NO (WSN), and SeTLife (WSN) ≤ SeTLife (Sei)

7) Cluster (C): If a clustering algorithm is applied [23], clusters will be formed. Sensor nodes are grouped into clusters, mainly to support scalability (for managing a high number of nodes). But, besides supporting scalability, clustering can have several different objectives, such as: load balancing, fault tolerance, network connectivity, maximal network longevity, etc. Each cluster has a leader, the cluster head (CH). So, a cluster (C) is defined by:

C = {CID, CH, Stb, NrC, IaC-To, IeCH-Con, CMet}.

Even though clustering is influenced by the network and link layer protocols, some attributes can be identified.

8) Cluster Head (CH): The cluster head (CH) can be elected by the sensors in a cluster or pre-assigned by the network designer. Also, CHs may be the richest nodes in resources of the whole network. So, a cluster head (CH) is defined by [23]:

```
CH = {CHID, TM, TN, Ro}
```

9) Obstacle (Oi)

To characterize an obstacle (Oi) it is important to consider the following attributes: identifier (OID), localization (L), Dimension (Dm), Scale (Sc), Rotation (Rt), and Description (Desc). These attributes are formalized in Table 15.2.

Hence, an obstacle (Oi) is defined by:

```
Oi = {OID, L, Dm, Sc, Rt, Desc}.
```

15.3.2. Main Properties Represented by the CWSN Model

The CWSN enables the graphical representation of several important properties, like nodes' mobility, connectivity and communication, network heterogeneity and stability, network coverage, consumer-producer and flow control, as well as the graphical representation of other important aspects, like, the occurrence of failures, the network topology, the established routing paths, or the communication modality used by nodes and the type of sensed data. However, some aspects like signal interference was not considered in the CWSN model.

The representation of some of these network properties, like mobility of nodes or topology changes, is possible through the representation of the network evolution. In other words, the model represents a screenshot of a WSN in a specific moment of time. As time goes by, several aspects of a particular WSN can change: (i) The state of the nodes can change; (ii) New nodes can be deployed; (iii) The topology can suffer modifications; (iv) New clusters can be created; (v) New obstacles can appear; etc. In the CWSN model, these network changes are naturally represented through a succession of figures as exemplified in Fig. 15.1.

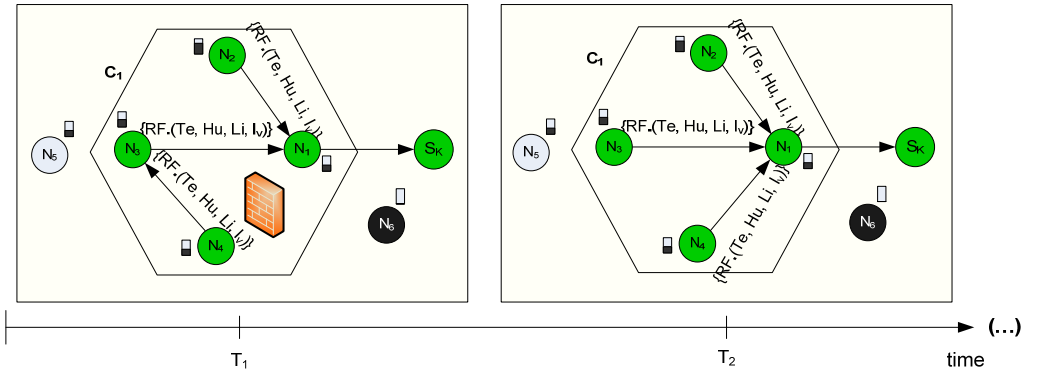


Fig. 15.1. Modeling a change in the network topology, applying the CWSN Model.

Fig. 15.1 represents a modification in the topology of the network caused by the elimination of an obstacle. The obstacle was located between nodes N4 and N1, impeding direct communication between these two nodes. For some reason (the obstacle may be static, like a tree or building, or it may move, like an animal or a car), the obstacle has disappeared. Consequently, connectivity and, therefore, communication has been established between the referred nodes.

The consumer-producer is another property that can also be deduced from the CWSN model, and it is denoted through the use of directed graphs, i.e., through the direction of the relationship between nodes. Formally, the consumer-producer property can be described as follows:

Considering that the WSN is represented as a graph constituted by vertices (nodes) and edges (relationships), and that the vertices use these edges to transmit data, with a set of data being represented by Da , then a WSN can be described as:

$$G_D(V_D, E_D),$$

where V_D represents the group of Nr participant nodes (N_i), and

$$E_D = (N_{i,prod}, Da, N_{j,cons}), \text{ with } i \neq j,$$

and Da is the set of data that is shared by the producer (N_{prod}) and the consumer nodes (N_{cons}).

For example, in Fig. 15.1, N1 is the cluster head and, thus, the consumer node, whereas the remaining nodes of the cluster, namely N2, N3, and N4, are the producer nodes. Similarly, the sink node is the consumer of all data produced in the WSN. Also, a cluster head consumes data from all the nodes of the cluster.

This property allows assuring that $G_D(V_D, E_D)$ is a connected graph, given that at least one relationship between two nodes exists.

Another property that can be inferred from the CWSN model is the flow control property. The sink node is the only node that can send data to the user. Therefore, all nodes have to send collected data to the sink node. Therefore, the sink node verifies flow control property; it controls the flow of data. This means that, regardless the considered moment in the whole lifetime of the WSN, only the participant sink node can send data to the user.

Similarly, if clustering is implemented, all the nodes that belong to the cluster have to send data to the cluster head, which in turn forwards data to the sink node (using other cluster heads, if necessary). Consequently, the cluster head verifies flow control property; it controls the flow of data.

Since both the sink node and the cluster head verify the flow control property, this property can formally be described as:

$$E_D = (S_{k,prod}, Da, User_{cons}) \text{ or } E_D = (CH_{prod}, Da, S_{k,cons}),$$

where Da is the set of data that is shared by the producer (Sk or CH) and the consumer nodes (Usercons or Sk).

Thus, the CWSN model also addresses some of the most important CSCW concepts but, more importantly, it addresses the analysis of temporal information. This can be used as an input for creating a real-time tool that allows visualization and representation of a WSN, as we have demonstrated in [25], where the CWSN model was translated into XML language, which was used as an input for the tool.

15.3.3. Comparing the CWSN Model with Other Models for WSNs

Table 15.4 allows a quick overview of the main differences between the CWSN model and other models that have been applied to model these networks.

The remaining state-of-the-art, models mostly address the modeling of a single issue of WSNs, addressing up to two issues (like mobility and connectivity) in the most complete modeling cases. Contrarily to these models, instead of focusing on modeling some specific problem of WSNs, the CWSN model is used to systematically describe and represent the features and properties of all the components that might constitute a WSN. So far, no other model has covered so many aspects of a WSN as the CWSN model does.

For example, unlike the SNSCW model [22], our model does not only model the cooperation within the network, but also the entire WSN. The CWSN model also allows the representation of the network hierarchy, from the collected data to the user (passing through the clusters, the session and the WSN). Moreover, the CWSN model is generic, in the sense that it can model heterogeneous networks and that it can be applied to any type of wireless sensors, (regardless its size, its hardware characteristics, the types of signals it can measure, etc.). It can also be applied to any WSN despite its specific application. So, it is possible to use all the entities defined in the model to represent a specific scenario of any application (monitoring a forest, a vineyard, a volcano, a museum, a natural catastrophe, a building, etc.).

Besides, any changes that might occur on a certain application scenario (new collaborative sessions, new clusters, nodes moving, etc.) can be represented by a sequence of figures; hence, the CWSN model allows the representation of each state of the network and its evolution.

Regarding collaboration, the model includes some fundamental CSCW [7] concepts (for, e.g., session, relationship, data flow and groups) and properties (for, e.g., consumer-producer and flow control). Thus, analyzing Table 15.4, it is possible to conclude that the CWSN model presents important advantages over the state-of-the-art models presented in Section 15.2.

Table 15.4. Comparative analysis of the models created for WSNs.

Model	Modeling issues	Graph-based	CSCW concepts	Collaboration levels	Graphical representation of the WSN	Formal model
Kenniche and Ravelomananana (2010)	Topology	YES	NO	NO	YES (topology)	NO
SNSCW (Liu et al., 2006b)	Cooperation	NO	YES	YES (two)	NO	YES
Wang (2011)	Coverage	no	no	no	no	Deployment and topology control
Bonaci et al. (2010)	Behavior of the WSN under attack	yes	no	no	no	Network security
Wu and Chung (2009)	Sensing and Coverage	no	no	no	no	Deployment and topology control
Ni et al. (2009)	Sensor faults	no	no	no	no	Fault detection
Wüchner et al. (2010)	Performance	no	no	no	no	Evaluation of performance and energy efficiency
CWSN	The whole WSN and collaboration hierarchy	YES	YES	YES (five)	YES (whole WSN)	YES

15.3.4. Contributions of the CWSN Model

WSNs are extremely dynamic systems, both in the sense that their characteristics change over their lifetime and for the fact that sensor networks' technology (hardware and software) is subject to fast changes. To overcome this issue, the CWSN model can be updated or extended, through the introduction of new entities and/or new properties. Therefore, another key point of this model is scalability, since it can easily evolve.

The CWSN model itself models the most important components of the WSN. As such, their advantages are:

The CWSN model provides a grammar for formally modeling an entire WSN, i.e., all the entities that can exist in a WSN, and their respective attributes.

It also allows to visually and graphically representing a whole WSN, including not only its entities, but a very important aspect to users and network managers, the network topology. Basically, the CWSN model provides a common framework for describing and representing any WSN.

Moreover, the CWSN model allows representing the network hierarchy, from the collected data to the user, passing through the sensor nodes, the clusters, the sessions and, finally, the whole WSN.

Besides, any changes that might occur on a certain application scenario (mobility of nodes, failure of nodes, topology changes, new collaborative sessions, new clusters, etc.) can be represented by a sequence of sub-graphs; hence, the CWSN model allows the representation of each state of the network and its evolution.

Furthermore, the CWSN model is generic, since it can model heterogeneous networks; it can be applied to any type of wireless sensor nodes (regardless their size, their hardware characteristics, the types of signals they can measure, etc.), and to describe and represent any WSN, despite of its specific application. So, it is possible to use all the entities defined in the model to represent a specific network scenario of any application (monitoring a forest, a vineyard, a volcano, a museum, a natural catastrophe, etc.).

15.4. The CWSN Model Applied to Structural Health Monitoring

The Structural Health Monitoring (SHM) is an emerging methodology that can be used to establish the structural performance in civil engineering buildings. For monitoring the SHM methodologies use different kind of sensors, some of them can to record movement, corrosion, deformation and other parameters that are correlated with the structural health condition into the building. This is particularity important when a building can be exposed to loads and ambient conditions that could change the original state of health. On the other hand, the main objective of the structural engineer is to avoid the occurrence of structural damage and the collapse of a building, and now a day the use of sensors into the buildings can help to identify changes in structural properties in order to identify some problems related with structural damage. In this way SHM systems are useful tools for establishment of damage condition or the structural vulnerability, which can be defined as the no health condition.

We consider that WSNs will be the new generation of systems that can be very useful for monitoring civil engineering buildings in the context of the Internet of things technologies (IOT).

The use of WSNs has brought several advantages in structural monitoring and the establishment of structural health compared to conventional methods where computers connected to accelerometers are used. In conventional methods, it is necessary to install cables through the structure, disturbing its normal operation and generating maintenance cost. Compared with conventional methods, WSNs provide the same functionality at a much lower price and a more flexible monitoring. Another problem is the high equipment and wiring installation and maintenance cost.

The cost of a conventional system with a computer and a piezoelectric accelerometer is about USD 40000 per sampling point. The estimated cost of the proposed system, in this work is less than USD 200 per point. In WSNs no wiring is required, making installation

and maintenance much easier and inexpensive. Moreover, the use of WSNs allows Structural Analysis of VulnerabilitiEs of buildings through wiReless sensor networks (SAVER) platform to be deployed and operate even if the building is in operation. It does not cause further visual impact due to its small size, low power consumption and installation flexibility. The advantage of structural health monitoring based on WSNs can be extended if the MEMS acceleration sensor type is used. The MEMS accelerometer is a silicon chip, which is very compact in size, low power consumption and cheap. Without MEMS, a small WSN, even low-power and low-cost accelerometer, would be degraded.

Thus, the SAVER platform will aim at gathering information to establish the vulnerability level of structural health of buildings. Such information will be used in decision making for two schemes: i) Prevention programs; and ii) Post-seismic evaluation.

The SAVER platform will be able to monitor and display information in real-time. It will determine from the implementation of several methods for estimating seismic response, damage detection and establishment of the structural vulnerability level on buildings. In addition, our platform will offer several services that will notify users about potential risks of the structure through alarms, email and SMS. Besides, it will have a Web based monitoring platform and a mobile app for Android and IOS. Also, this platform will generate graphs, reports and statistics. Some preliminary results of the SAVER project was publish in [26].

The expected results, in SAVER project, intend to give the basis for the analysis of buildings and gather instrumental data that can be useful for decision-making of institutions and users that are responsible for infrastructure and buildings. Furthermore, in this project, we pretend to provide the necessary information to implement methods of vulnerability analysis and therefore, to estimate the seismic risk of buildings, such as hospitals or schools.

SAVER project will be validated in the building B (Fig. 15.2) of UPAEP University, located in Puebla city in México. This building is structured based on reinforced concrete frames. Furthermore, the building is regular in plan and consists of four levels with a height of 3.15 m each one, so it has a total height of 12.60 m. In the transversal direction, it has a bay of 10.50 m. In the longitudinal direction, the building has eight bays of 6 m each one, so that it has a total length of 48 m. The structural elements are composed of beams and columns. The beams, in transversal direction, have square cross section at all levels. In the longitudinal direction at level 1 and 4, they have variable prismatic section, while in level 2 and 3 are rectangular. All columns in each level have variable hexagonal section. Floor system has 0.25 m thickness and is prefabricated. The building has masonry walls with 0.15 m thickness.

This building was built in 1984. In recent studies about its structural performance, we have determined a high level of structural vulnerability. But, these studies were made using limited information because we use only three wired sensors. In order to obtain a better vulnerability estimation, we intend to instrument this building using the SAVER platform. The proposed topology for this building is shown in Fig. 15.3.

Fig. 15.2. Building B of the UPAEP University.

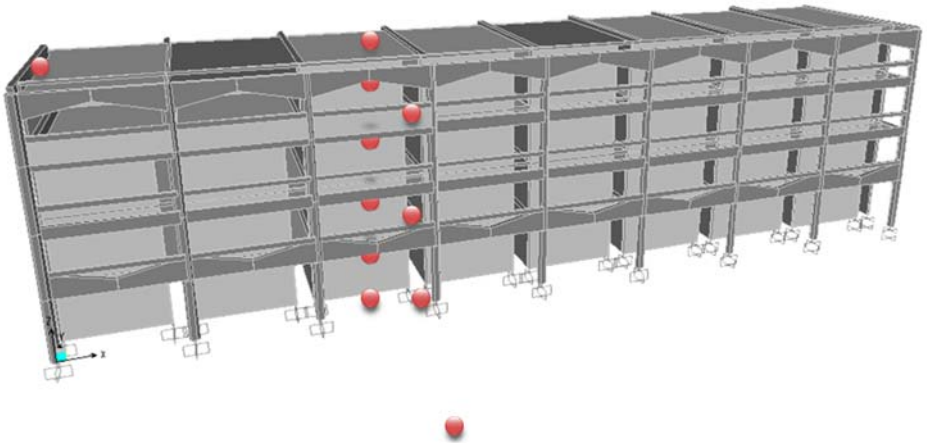


Fig. 15.3. Topology of the WSN installed in the Building B.

The spatial distribution of the sensors is established from the geometry of each building. But it is necessary to deploy each sensor at least one in the geometric center of each level, and one sensor on the corner of the roof. If the longitudinal dimension of the building is large, it is suggested deploy some sensors in one border of the building. It is important to monitor also the ground response using a free field sensor.

After a propagation study of RF signal, we are planning to install 11 sensor nodes. Each node has two sensors, a temperature sensor and an acceleration sensor. The sensor node that is located out of the building, free field sensor node includes a temperature sensor, an acceleration sensor and a GPS sensor. This node sensor synchronizes the whole WSN. Using CWSN model, we can visualize the interaction among the sensors and their relationship.

Fig. 15.4 shows the WSN deployed in Building B, but represented using the entities proposed by the CWSN model. Basically, this representation clearly shows that only one

session is established, but more importantly, it shows the structure of the WSN and the interactions that will take place between the different types of nodes that compose it. For instance, it shows that nodes are grouped in three clusters.

Using the CWSN model, it becomes evident the type of data collected by each node.

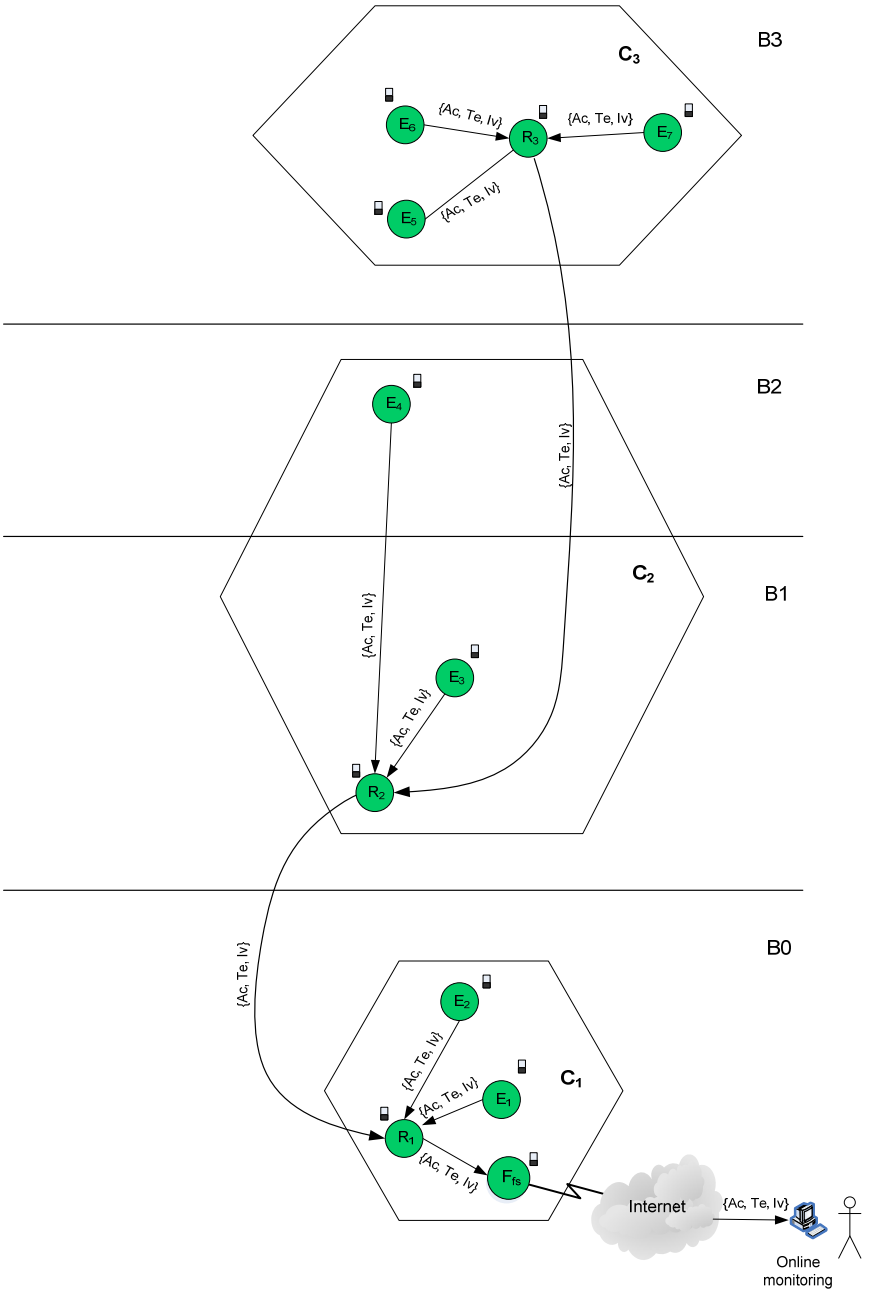


Fig. 15.4. Modeling Building B's WSN.

The problem to establish the structural health on buildings will be reduced if we use the CWSN model. The model can provide a tool for getting instrumental information about the structural properties, like acceleration on each storey as well as some structural dynamic parameters. If we use a sensor that can measure the acceleration (Acc) at each storey and considering that each sensor can provide the record of (Acc), in three orthogonal directions, respect to the building (longitudinal, transversal and vertical), the CWSN model can be implemented in order to estimate the structural vulnerability level. For this reason, we need to compute a parameter correlated with the lateral displacement and the structural damage.

The structural vulnerability level can be estimated using the algorithm presented in [26], where this level is associated with a damage parameter. The vulnerability can be estimated using a function, which considers the physical damage at any story and is defined as:

$$d(u) = 1 - \exp(-au^m), \tag{15.1}$$

where a and m are the parameters to be determined according with the structural system features (frame, walls, etc.) and u is the local deformation of interest, normalized with respect to its peak value at failure (total loss). The damage function is obtained as function of the corresponding inter-story drift. In this way the parameter u is related to the lateral displacement. The lateral displacement u can be determined considering two criteria: 1) using actual seismic records, and 2) using ambient vibration records.

The first approach is direct, because we use the record acceleration time history on different floors along the building. A double integration of the acceleration time-history can be used in order to obtain the response displacement. For this is necessary to apply a numerical procedure for integrating the corrected and filtered acceleration time-history, assuming that it has a linear variation between each time increment. For the velocity time-history this procedure is repeated in order to estimate the displacement time-history. The maximum inter-story drift can be determined by the following expression:

$$\Psi_{maxj} = \left| \frac{u_{j+1}(t) - u_j(t)}{h_{j+1} - h_j} \right|, \tag{15.2}$$

where Ψ_{maxj} is the maximum inter-story drift, $u_j(t)$ is the lateral displacement at level "j" for a time t, and h_j is the vertical distance between each level.

The second approach considers several steps that are described in the follow. First, it is necessary to synchronize the signals with a common time reference and carry out the polarization procedure according to the sensor´s orientation and the reference system. The baseline correction of the original records also is needed. In order to eliminate the undesirable components of frequency a signal filtering procedure is recommended for this we can use a Butterworth. For the ambient vibration records in three directions we can apply the Fast Fourier Transform (FFT), in order to obtain the Amplitude Fourier Spectra. With this information we can estimate the transfer functions, the vibration periods and mode shapes. The vibration period and the mode shapes can be used for generating a Simplified Reference System (SRS) using the criteria proposed in [27]. The SRS has dynamic properties that represent the behaviour of the actual building, however is

necessary introduce the corresponding transform response factors. These factors are also defined in [28]. In order to obtain the non-linear response of the SRS, in terms of lateral displacement, an adequate hysteretic model will be adopted. The non-linear responses can be related with a specified seismic scenario. In Fig. 15.5 the SRS model is shown, where m, k and c are the mass, stiffness and damping coefficients of the building, respectively. These parameters are correlated with the dynamic properties. The mechanics properties are represented by relationship between shear lateral force v and the corresponding lateral displacement u. The parameter v_y defines the yielding force that is a measure of the lateral resistance in the building, and u_y represents the corresponding yielding displacement. k_1 and k_2 are the initial (linear and elastic behavior) and the post-yielding (non-linear and inelastic behavior) stiffness, respectively.

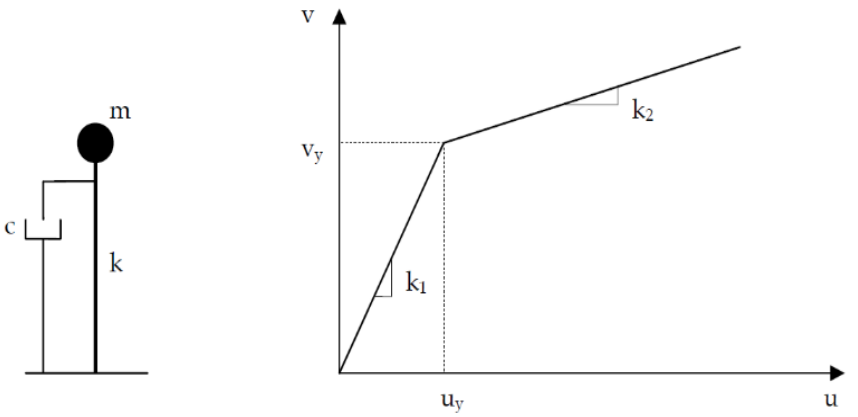


Fig. 15.5. SRS model.

In this implementation we can define the specific properties for each Sensor Node (Ni), for example, we can use three Types of sensors (TS): Ac, Te and lv; Communication modality (CM) can be defined as radio; for the Communication Technology (CT), we can use ZigBee (ZB); the Power Supply (PS) can be a Hybrid type (Hy); the Localization (L) can be defined as Manual.

In order to show the advantages of the SAVER platform, we present a structural vulnerability function for building B (Fig. 15.6). The vulnerability function describes the damage level d(Ψ)in terms of inter-storey drift d(Ψ). The values of damage are from 0 to 1. A damage equals to 0 indicates fully health condition and damage equals to 1.0 indicates collapse of the building. The vulnerability function was estimated using only one accelerometer located in three different points on the building. We recorded three acceleration records (ambient vibration), but these records were in three different intervals of time. This is a big limitation that can be covered by the SAVER platform, because the acceleration records must be at the same interval of time. The CWSN model can represent this easily, since it defines the evolution of the network, which can be represented through a succession of graphs.

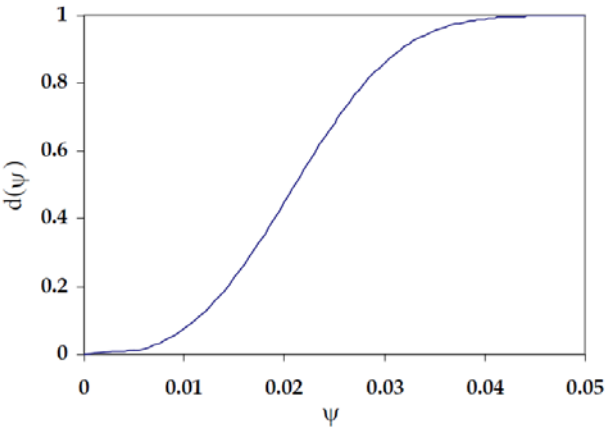


Fig. 15.6. Typical vulnerability function to establish the structural health condition on the Building B.

The modeling and representation of the deployed WSN using the CWSN model can bring several advantages from the user and network manager's points of view. The main contribution of the CWSN model is to standardize ways to model a WSN and provide a unified view of such a network regardless of what aspects are considered. Moreover, it allows the user and the network manager to become more aware of the composition and state of the whole network. That is, the CWSN model allows for visually representing several details about the WSN that has been deployed, what provides them with a more intuitive and prompt understanding of the WSN.

15.5. Conclusions

The CWSN model also defines the evolution of the network, which can be represented through a succession of graphs.

In this extended version of the paper [29], we presented the CWSN model, which is based on the CSCW methodology and specifically designed for WSNs. The great advantage of using this model to represent WSNs is that, besides modeling collaboration, it can also model the entire WSN. Also, this model allows for the representation of each state of the network and its evolution. Moreover, the CWSN model is generic, in the sense that it can be applied to heterogeneous WSNs, and scalable, as it can be updated if any modifications need to be introduced. The CWSN model was formalized in first-order logic. This attribution allows knowing which edges are active and which are not. In other words, we are able to identify which nodes are transmitting information. If some failure occurs on the process of transmission, our approach allows identifying this situation.

We consider that some advantages of the CWSN model arise from the fact of being formal and based on graphs.

The attributes defined for each entity of the CWSN model cover several dimensions. Thus, this model can be used as a framework for developing more generic software solutions for WSNs. Given the fact of being a broad and generic model, also confers the CWSN model with the ability of being applied to describe any WSN regardless of its application scenario.

We also believe that this model can assist network designers in making better decisions regarding the organization and management of the network. This contribution becomes more significant given that the CWSN model was used as a basis to implement an awareness tool and a sessions' managing tool for WSNs, which will be described in the next chapter.

Moreover, the CWSN model can be used to automatically generate some graphs of the WSN that will allow for identifying routing paths, detecting damaged/failed nodes or links, etc.

Acknowledgements

This project has been supported by CONACyT- Consejo Nacional de Ciencia y Tecnología within the Project Scope: 249071, UPAEP Research Department and FCT – Fundação para a Ciência e Tecnologia, within the Project Scope: UID/EEA/50009/2013.

References

[1]. I., Su, W. Akyildiz, Y. Sankarasubramaniam and E. Cayirci, Wireless sensors networks: a survey, *Computer Networks,* 38, 2002, pp. 393-422.

[2]. L. Brito and L. M. Rodríguez Peralta, Wireless Sensor Networks: Basic Concepts, in Encyclopedia of Networked and Virtual Organizations, *Idea Groups,* Vol. I, 2008, pp. 57-64.

[3]. S. Tilak, N. Abu-Ghazaleh, and W. Heinzelman, A taxonomy of wireless micro-sensors network models, *Mobile Computing and Communication Review,* Vol. 6, No. 2, April 2002, pp. 28-36.

[4]. M. Tubaishat and S. Madria, Sensor Networks: an Overview, *IEEE Potentials,* May 2003, Vol. 22, No. 2, pp. 20-23.

[5]. K. Römer and F. Mattern, The Design Space of Wireless Sensor Networks, *IEEE Wireless Communications,* 11, 6, December 2004, pp. 54-61.

[6]. L. Brito and L. M. Rodríguez Peralta, A Collaborative Model for Wireless Sensor Networks Applied to Museums' Environmental Monitoring, in *Proceedings of the 5th International Conference on Cooperative Design, Visualization and Engineering (CDVE'08),* 5220/2008, Mallorca, Spain, Lecture Notes in Computer Science, Springer Verlag, September 2008, pp. 107-116.

[7]. K. Mills, Computer-Supported Cooperative Work Challenges, Encyclopedia of Library and Information Science, 2003.

[8]. P. Ramanathan, K. Saluja and Y. Hu, Collaborative Sensor Signal Processing for Target Detection, Localization and Tracking, in *Proceedings of the 23rd Army Science Conference*, Orlando, USA 2002.

[9]. K. C. Wang and P. Ramanathan, Collaborative Sensing Using Sensors of Uncoordinated Mobility, Int. Conference on Distributed Computing in Sensor Systems, LNCS, Vol. 3560, *Springer*, Marina del Rey, USA (2005), pp. 293—306.

[10]. L. Iftode, C. Borcea and P. Kang, Cooperative Computing in Sensor Networks, Handbook of Sensor Networks: Compact Wireless and Wired Sensing Systems, Mohammad Ilyas (Ed.), *CRC Press,* 2004.

[11]. G. Chen, T. D. Guo, W. G. Yang and T. Zhao, An improved ant-based routing protocol in Wireless Sensor Networks, in *Proceedings of the International Conference on Collaborative Computing: Networking, Applications and Worksharing (CollaborateCom' 06),* Atlanta, USA, 2006, pp. 1—7.

[12]. D. Dardari, and A. Conti, A Sub-Optimal Hierarchical Maximum Likelihood Algorithm for Collaborative Localization in Ad-Hoc Networks, in *Proceedings of the IEEE Communications Society Conference on Sensor and Ad Hoc Communications and Networks (SECON'04),* Santa Clara, USA 2004, pp. 425—429.

[13]. A. Chadha, Y. Liu, and S. Das, Group Key Distribution via Local Collaboration in Wireless Sensor Networks, in *Proceedings of the IEEE Communications Society Conference on Sensor and Ad Hoc Communications and Networks (SECON'05),* Santa Clara, USA, 2005, pp. 46 - 54.

[14]. H. Sanli, R. Poornachandran and H. Cam, Collaborative Two-Level Task Scheduling for Wireless Sensor Nodes with Multiple Sensing Units, in *Proceedings of the IEEE Communications Society Conference on Sensor and Ad Hoc Communications and Networks (SECON'05),* Santa Clara, USA, 2005, pp. 350- 361.

[15]. R. Reghelin, and A. Fröhlich, A Decentralized Location System for Sensor Networks Using Cooperative Calibration and Heuristics, in *Proceedings of the 9th ACM Int. Symposium on Modeling Analysis and Simulation of Wireless and Mobile Systems (MSWiM'06),* Torremolinos, Spain, 2006, pp. 139-146.

[16]. V. Bychkovskiy, S. Megerian, D. Estrin and M. Potkonjak, A Collaborative Approach to In-Place Sensor Calibration, in *Proceedings of the 2nd International Workshop on Information Processing in Sensor Networks (IPSN'03),* Palo Alto, USA, 2003, pp. 301—316.

[17]. S. Giannecchini, M. Caccamo, and C. S. Shih, Collaborative resource allocation in wireless sensor networks, in *Proceedings of the 16th Euromicro Conference on Real-Time Systems (ECRTS'04),* Catania, Italy, 2004, pp. 35—44.

[18]. A. Hu and S. Servetto, Algorithmic Aspects of the Time Synchronization Problem in Large-Scale Sensor Networks, *Mobile Networks and Applications,* Vol. 10, 2005, pp. 491--503.

[19]. A. Krohn, M. Beigl, C. Decker, T. Riedel, T. Zimmer and D. Varona, Increasing Connectivity in Wireless Sensor Network using Cooperative Transmission, in *Proceedings of the 3rd International Conference on Networked Sensing Systems (INSS'06)*. Chicago, USA, 2006.

[20]. L. Cheng, T. Lian, Y. Zhang and Q. Ye, Monitoring Wireless Sensor Networks by Heterogeneous Collaborative Groupware, in *Proceedings of the Sensors for Industry Conference (Sicon/04),* New Orleans, USA, 2004.

[21]. Z. Chaczko, F. Ahmad, and V. Mahadevan, Wireless Sensors in Network Based Collaborative Environments, in *Proceedings of the 6th International Conference on Information Technology Based Higher Education and Training (ITHET 2005),* 2005, pp. F3A-7- F3A-13.

[22]. L. Liu, H. Ma, D. Tao and D. Zhang, A Hierarchical Cooperation Model for Sensor Networks Supported Cooperative Work, in *Proceedings of the 10th International Conference on Computer Supported Cooperative Work in Design (CSCWD'06),* 2006, pp. 1-6.

[23]. B. Krishnamachari, D. Estrin, and S. Wicker, The impact of Data Aggregation in Wireless Sensor Networks, in *Proceedings of the IEEE 22nd International Conference on Distributed Computing Systems Workshops,* 2002, pp. 575-578.

[24]. S. Hussain, U. Farooq, K. Zia, and M. Akhlaq, An Extended Topology for Zone-Based Location Aware Dynamic Sensor Networks, in *Proceedings of the National Conference on Emerging Technologies (NCET)*, Szabist Karachi, Pakistan, Dec. 2004.

[25]. L. M. Rodríguez Peralta, L. M. P. Leão Brito, Teixeira Gouveia B. A., Sousa D. J. G., and Alves C. S., Automatic monitoring and control of museums' environment based on Wireless Sensor Networks, *Electronic Journal of Structural Engineering (EJSE),* Special Issue: Wireless Sensor Networks and Practical Applications, 2010, pp. 12 – 34.

[26]. L. M. Rodríguez Peralta, E. Ismael Hernández, S. A. Cardeña Moreno, D. Martínez Jiménez and A. E. Muñoz Guarneros, Towards to a platform of monitoring based in WSN to estimate the structural health of buildings, in *Proceedings of the 2nd European Conference on Earthquake Engineering and Seismology (2ECEES)*, Istanbul, Turkey, August 2014, pp. 24-29.

[27]. E. Ismael-Hernández, O. Díaz-López and L. Esteva, Seismic vulnerability analysis for optimum design of multistory reinforced concrete buildings, in *Proceedings of the Thirteenth World Conference on Earthquake Engineering (WCEE'04),* Vancouver, Canada, August, 2004, Paper 514.

[28]. E. Ismael-Hernández, Seismic design based on performance and reliability of wall-frame systems, PhD Thesis, (In Spanish), *National University of Mexico (UNAM),* 2010.

[29]. L. M. Rodríguez Peralta, Lina M. P. Brito, E. Ismael Hernández, A Formal Graph-Based Model Applied to Cluster Communication in Wireless Sensor Networks, in *Proceedings of the 8th International Conference on Sensor Technologies and Applications (SENSORCOMM'14),* Lisbon, Portugal, November 16 – 20, 2014, pp. 137-146.

Chapter 16
Target Localization in Cooperative Wireless Sensor Networks Using Measurement Fusion

Slavisa Tomic, Marko Beko, Rui Dinis and Milan Tuba

16.1. Introduction

Wireless localization technologies play essential rolls in many practical applications that facilitate everyday life and make saving lives possible worldwide. Emergency services, weather forecasting, navigation, social networking, and targeted advertising, all rely on wireless localization technologies, which are dominant in virtually all wireless network deployments for a myriad of commercial, industrial, governmental and military uses [1, 2].

In a wireless sensor network (WSN), the locations of the unknown sensors (targets) are determined by using a kind of localization schemes that typically rely on the locations of the reference sensors (anchors) and range measurements between them. Range measurements can be extracted from different characteristics of the radio signal, such as time-of-arrival (ToA) [3], time-difference-of-arrival (TDoA) [4], round-trip time (RTT) [5], angle-of-arrival (AoA) [6] or received signal strength (RSS) [7], depending on the available hardware. Recently, hybrid systems that fuse two measurements of the radio signal have been investigated [8-15]. Hybrid systems benefit from the use of combined measurements (more available information), taking advantage of the strongest points of each technique and minimizing their drawbacks. On the other hand, the price to pay for using such systems is the increased complexity of network devices, which increases the network implementation costs [1, 16].

The approaches in [8-10] are based on the fusion of RSS and ToA measurements. A hybrid system that merges range and angle measurements was investigated in [11]. The authors in [11] proposed two estimators to solve the non-cooperative target localization problem in a 3-D scenario: linear least squares (LS) and optimization based. The LS estimator is a relatively simple and well known estimator, while the optimization based estimator was solved by Davidson-Fletcher-Powell algorithm [17]. Although both estimators are simple

Slavisa Tomic
ISR/IST, LARSyS, Lisbon, Portugal

to implement, they do not tightly approximate the maximum likelihood (ML) estimator, and their estimation accuracy can be further improved. In [12], the authors derived an LS and an ML estimator for a hybrid scheme that combines RSS difference (RSSD) and AoA measurements. The work in [12] employed non-linear constrained optimization to estimate the target's location from multiple RSS and AoA measurements. Their results indicate that measurement fusion can significantly improve the estimation accuracy. A selective weighted LS (WLS) estimator for RSS/AoA localization problem was proposed in [14]. The authors determined the target location by exploiting weighted ranges from two nearest anchor measurements, which were combined with the serving base station AoA measurement. The hybrid WLS approach in [14] was derived without requiring information about the statistical properties of RSS and AoA measurements, which might lead to significant performance degradation in practice. In [12-14], authors investigated the centralized non-cooperative hybrid RSS/AoA localization in a 2-D WSN only. A WLS estimator for a 3-D RSSD/AoA non-cooperative localization problem when the transmitted power is unknown was presented in [15]. However, similar as the method proposed in [14], the WLS method does not take the statistical properties of RSS and AoA measurements into account. Also, the authors in [15] only investigated a small-scale WSN, with extremely low noise power.

All mentioned approaches investigate non-cooperative localization problem only, where the location of a single target, which communicates with anchors exclusively, is determined at a time. Contrary to existing approaches, in this paper we investigate the target localization problem in a large-scale WSN, where the number of anchors is scarce and the communication range of all sensors is restricted (*e.g.*, to prolong sensor's battery life). In such settings, only some targets can directly communicate with anchors; therefore, cooperation between any two sensors within the communication range is required in order to acquire sufficient amount of information to perform localization. We design novel distributed hybrid localization algorithm based on second-order cone programming (SOCP) relaxation to estimate the locations of all targets in WSN. The new method is based on a hybrid system that fuses RSS and AoA measurements. By using the RSS propagation model and simple geometry, we derive a novel local non-convex estimator based on the LS criterion, which tightly approximates the local ML one for small noise levels. Then, we show that the derived non-convex estimator can be transformed into a convex SOCP estimator which can be solved efficiently by interior-point algorithms [18]. To the best of the authors' knowledge, distributed localization algorithms for hybrid RSS/AoA systems in cooperative WSNs are yet to be published.

Throughout the work, upper-case bold type, lower-case bold type and regular type is used for matrices, vectors and scalars, respectively. $\mathbb{R}^n$ and $\mathbb{C}^n$ respectively denote the n dimensional real and complex Euclidean space. The operators $(\cdot)^T$ and $(\cdot)^H$ denote transpose and Hermitian, respectively. The normal (Gaussian) distribution with mean μ and variance σ^2 is denoted by $\mathcal{N}(\mu, \sigma^2)$. The N-dimensional identity matrix is denoted by $\boldsymbol{I}_N$ and the $M \times N$ matrix of all zeros by $\mathbf{0}_{M\times N}$ (if no ambiguity can occur, subscripts are omitted). $\|\boldsymbol{x}\|$ denotes the vector norm defined by $\|\boldsymbol{x}\| = \sqrt{\boldsymbol{x}^H\boldsymbol{x}}$, where $\boldsymbol{x} \in \mathbb{C}^n$. For Hermitian matrices $\boldsymbol{A}$ and $\boldsymbol{B}$, $\boldsymbol{A} \succcurlyeq \boldsymbol{B}$ means that $\boldsymbol{A} - \boldsymbol{B}$ is positive semidefinite.

The remainder of this work is organized as follows. In Section 16.2, the RSS and AoA measurement models are introduced and the target localization problem is formulated. Section 16.3 presents the development of the proposed distributed estimator. In Section 16.4 we provide an analysis about the computational complexity, while in Section 16.5 we discuss the performance of the proposed algorithm. Finally, Section 16.6 summarizes the main conclusions.

16.2. Problem Formulation

We consider a large-scale WSN with N anchors and M target nodes, randomly deployed over a region of interest. The considered network can be seen as a connected graph, $\mathcal{G}(\mathcal{V}, \mathcal{E})$, with $|\mathcal{V}| = M + N$ vertices and $|\mathcal{E}|$ edges (connections), where $|\cdot|$ represents the cardinality of a set, *i.e.*, the number of elements in a set. The set of targets and the set of anchors are respectively labeled as $\mathcal{T}$ ($|\mathcal{T}| = M$) and $\mathcal{A}$ ($|\mathcal{A}| = N$), and their locations are denoted by $\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_M$ and $\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_N$, ($\boldsymbol{x}_i, \boldsymbol{a}_j \in \mathbb{R}^3, \forall i \in \mathcal{T}$ and $\forall j \in \mathcal{A}$), respectively. To save power (battery duration conditions the lifetime of a network), it is assumed that all sensors have limited communication range, R. Hence, two sensors, i and j, can exchange information if and only if they are within the communication range of each other. The sets of all target/anchor and target/target edges are defined as $\mathcal{E}_{\mathcal{A}} = \{(i,j): \|\boldsymbol{x}_i - \boldsymbol{a}_j\| \leq R, \forall i \in \mathcal{T}, \forall j \in \mathcal{A}\}$ and $\mathcal{E}_{\mathcal{T}} = \{(i,k): \|\boldsymbol{x}_i - \boldsymbol{x}_k\| \leq R, \forall i,k \in \mathcal{T}, i \neq k\}$, respectively.

For ease of expression, let us define a matrix $\boldsymbol{X} = [\boldsymbol{x}_1, \cdots, \boldsymbol{x}_M]$ ($\boldsymbol{X} \in \mathbb{R}^{3\times \mathrm{M}}$) as the matrix of all unknown target locations. We determine these unknown locations by using a hybrid system that fuses range and angle measurements.

Throughout this work, it is assumed that the range measurements are extracted from the RSS information exclusively, since ranging based on RSS requires the lowest implementation costs [1]. However, the RSS measurement model can be replaced with the path loss model by using the relationship $L_{ij} = 10\log_{10}\frac{P_T}{P_{ij}}$ (dB), where L_{ij} and P_{ij} are respectively the path loss and received power between two sensors i and j which are within the communication range of each other (from the transmitting sensor) and P_T is the transmission power of a sensor, as:

$$L_{ij}^{\mathcal{A}} = L_0 + 10\gamma \log_{10}\frac{\|\boldsymbol{x}_i - \boldsymbol{a}_j\|}{d_0} + n_{ij}, \forall (i,j) \in \mathcal{E}_{\mathcal{A}}, \tag{16.1a}$$

$$L_{ik}^{\mathcal{T}} = L_0 + 10\gamma \log_{10}\frac{\|\boldsymbol{x}_i - \boldsymbol{x}_k\|}{d_0} + n_{ik}, \forall (i,k) \in \mathcal{E}_{\mathcal{T}}, \tag{16.1b}$$

(see [19], [20]) where L_0 denotes the path loss value at a short reference distance d_0 $\left(\|\boldsymbol{x}_i - \boldsymbol{a}_j\| \geq d_0, \|\boldsymbol{x}_i - \boldsymbol{x}_k\| \geq d_0\right)$, γ is the path loss exponent (PLE), and n_{ij} and n_{ik} are the log-normal shadowing terms modeled as zero-mean Gaussian random variables

with variances $\sigma_{n_{ij}}^2$ and $\sigma_{n_{ik}}^2$, *i.e.*, $n_{ij} \sim \mathcal{N}\left(0, \sigma_{n_{ij}}^2\right)$ and $n_{ik} \sim \mathcal{N}(0, \sigma_{n_{ik}}^2)$. We assume that the target/target path loss measurements are symmetric[1].

To obtain the AoA measurements (both azimuth and elevation angles), we assume that either antenna arrays or a directional antenna is implemented at anchors [11, 21], or that the anchors are equipped with video cameras [22]. In order to make use of the AoA measurements from different sensors, the orientation information is required, which can be obtained by implementing a digital compass at each sensor [11, 21]. However, a digital compass introduces an error in the AoA measurements due to its static accuracy. For the sake of simplicity and without loss of generality, the angle measurement error and the orientation error are modeled as one random variable in the rest of this paper.

Fig. 16.1 gives an illustration of a target and anchor locations in a 3-D space. As shown in Fig. 16.1, $x_i = [x_{i1}, x_{i2}, x_{i3}]^T$ and $a_j = \left[a_{j1}, a_{j2}, a_{j3}\right]^T$ are respectively the unknown coordinates of the i-th target and the known coordinates of the j-th anchor, while $d_{ij}^{\mathcal{A}}$, $\phi_{ij}^{\mathcal{A}}$ and $\alpha_{ij}^{\mathcal{A}}$ represent the distance, azimuth angle and elevation angle between the i-th target and the j-th anchor, respectively. The ML estimate of the distance between two sensors can be obtained from the RSS measurement model (1) as follows [1].

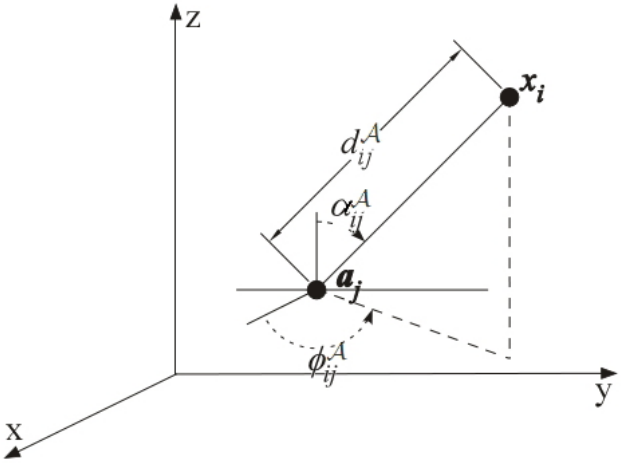


Fig. 16.1. Illustration of a target and anchor locations in a 3-D space.

[1] This assumption is made without loss of generality; it is readily seen that, if $L_{ik}^{\mathcal{T}} \neq L_{ki}^{\mathcal{T}}$, then it is enough to replace $L_{ik}^{\mathcal{T}} \leftarrow (L_{ik}^{\mathcal{T}} + L_{ki}^{\mathcal{T}})/2$ and $L_{ki}^{\mathcal{T}} \leftarrow (L_{ik}^{\mathcal{T}} + L_{ki}^{\mathcal{T}})/2$ when solving the localization problem.

$$\hat{d}_{ij} = \begin{cases} d_0 10^{\frac{L_{ij}^{\mathcal{A}} - L_0}{10\gamma}}, \text{if } j \in \mathcal{A}, \\ d_0 10^{\frac{L_{ij}^{\mathcal{T}} - L_0}{10\gamma}}, \text{if } j \in \mathcal{T}. \end{cases} \tag{16.2}$$

Applying simple geometry, azimuth and elevation angle measurements[1] can be modeled respectively as [11]:

$$\phi_{ij}^{\mathcal{A}} = \tan^{-1}\left(\frac{x_{i2}-a_{j2}}{x_{i1}-a_{j1}}\right) + m_{ij}, \forall (i,j) \in \mathcal{E}_{\mathcal{A}}, \tag{16.3}$$

and

$$\alpha_{ij}^{\mathcal{A}} = \cos^{-1}\left(\frac{x_{i3}-a_{j3}}{\|\boldsymbol{x}_i-\boldsymbol{a}_j\|}\right) + v_{ij}, \forall (i,j) \in \mathcal{E}_{\mathcal{A}}, \tag{16.4}$$

where m_{ij} and v_{ij} are the measurement errors of azimuth and elevation angles, respectively, modeled as zero-mean Gaussian random variables with variances $\sigma_{m_{ij}}^2$ and $\sigma_{v_{ij}}^2$, *i.e.*, $m_{ij} \sim \mathcal{N}\left(0, \sigma_{m_{ij}}^2\right)$ and $v_{ij} \sim \mathcal{N}(0, \sigma_{v_{ij}}^2)$.

Given the observation vector $\boldsymbol{\theta} = [\boldsymbol{L}^T, \boldsymbol{\phi}^T, \boldsymbol{\alpha}^T]^T$ ($\boldsymbol{\theta} \in \mathbb{R}^{3|\mathcal{E}_{\mathcal{A}}|+|\mathcal{E}_{\mathcal{T}}|}$), where $\boldsymbol{L} = \left[L_{ij}^A, L_{ik}^T\right]^T, \boldsymbol{\phi} = \left[\phi_{ij}^A\right]^T, \boldsymbol{\alpha} = \left[\alpha_{ij}^A\right]^T$, the conditional probability density function (PDF) is given as:

$$p(\boldsymbol{\theta}|\boldsymbol{X}) = \prod_{i=1}^{3|\mathcal{E}_{\mathcal{A}}|+|\mathcal{E}_{\mathcal{T}}|} \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(-\frac{\left(\theta_i - f_i(\boldsymbol{X})\right)^2}{2\sigma_i^2}\right), \tag{16.5}$$

where

$$\boldsymbol{f}(\boldsymbol{X}) = \begin{bmatrix} \vdots \\ L_0 + 10\gamma \log_{10}\dfrac{\|\boldsymbol{x}_i - \boldsymbol{a}_j\|}{d_0} \\ \vdots \\ L_0 + 10\gamma \log_{10}\dfrac{\|\boldsymbol{x}_i - \boldsymbol{x}_k\|}{d_0} \\ \vdots \\ \tan^{-1}\left(\dfrac{x_{i2}-a_{j2}}{x_{i1}-a_{j1}}\right) \\ \vdots \\ \cos^{-1}\left(\dfrac{x_{i3}-a_{j3}}{\|\boldsymbol{x}_i-\boldsymbol{a}_j\|}\right) \\ \vdots \end{bmatrix}, \boldsymbol{\sigma} = \begin{bmatrix} \vdots \\ \sigma_{n_{ij}} \\ \vdots \\ \sigma_{n_{ik}} \\ \vdots \\ \sigma_{m_{ij}} \\ \vdots \\ \sigma_{v_{ij}} \\ \vdots \end{bmatrix}.$$

Maximizing the log of the likelihood function (16.5) with respect to $\boldsymbol{X}$ produces the ML estimate, $\widehat{\boldsymbol{X}}$, of the unknown locations, as:

[1] Note that we consider here the case where only anchors have the necessary equipment to perform the respective angle measurements. An alternative approach would be to provide the necessary equipment to all sensors. However, our simulations showed that there is no gain for such a setting, and it would severely raise the overall network implementation costs.

$$\begin{aligned}\widehat{\boldsymbol{X}} &= \arg\max_{\boldsymbol{X}} \ln\{p(\boldsymbol{\theta}|\boldsymbol{X})\} \\ &= \arg\max_{\boldsymbol{X}} \ln\left\{\prod_{i=1}^{3|\mathcal{E}_{\mathcal{A}}|+|\mathcal{E}_{\mathcal{T}}|} \exp\left(-\frac{\left(\theta_i - f_i(\boldsymbol{X})\right)^2}{2\sigma_i^2}\right)\right\} \\ &= \arg\min_{\boldsymbol{X}} \sum_{i=1}^{3|\mathcal{E}_{\mathcal{A}}|+|\mathcal{E}_{\mathcal{T}}|} \frac{1}{\sigma_i^2}[\theta_i - f_i(\boldsymbol{X})]^2.\end{aligned} \tag{16.6}$$

Although the ML estimator is approximately the minimum variance unbiased estimator [23], the LS problem in (16.6) is non-convex and has no closed-form solution. In the remainder of this work, we will show that the LS problem in (16.6) can be solved in a distributed manner by applying certain approximations. More precisely, we propose a convex relaxation technique leading to a distributed SOCP estimator that can be solved efficiently by interior-point algorithms [18].

16.2.1. Assumptions

We outline here some assumptions for the WSN (made for the sake of simplicity and without loss of generality):

(1) The network is connected and it does not vary in time;

(2) Measurement errors for RSS and AoA models are independent, and $\sigma_{n_{ij}} = \sigma_n, \sigma_{m_{ij}} = \sigma_m$ and $\sigma_{v_{ij}} = \sigma_v, \forall (i,j) \in \mathcal{E}_{\mathcal{A}} \cup \mathcal{E}_{\mathcal{T}}$;

(3) All sensors have identical P_T;

(4) A coloring scheme of the network is available.

In assumption (16.1), we assume that the sensors are static and that there is no sensor/link failure, and that there exists a path between each sensors $i, j \in \mathcal{V}$. Assumptions (16.2) and (16.3) are made for the sake of simplicity. Finally, assumption (16.4) implies that a coloring scheme is available in order to color (number) the sensors and establish a working hierarchy in the network. More precisely, we assume that a second-order coloring scheme is employed, meaning that no sensor has the same color (number) as any of its neighbors nor its neighbors' neighbors [24 - 26]. In this way, we avoid message collision and reduce the execution time of the algorithm, since nodes with the same color can work in parallel.

16.3. Distributed Localization

The objective function in (16.6) depends only on the locations and pair-wise measurements between the adjacent sensors. Hence, assuming that the initial location estimations of the targets, $\widehat{\boldsymbol{X}}^{(\mathbf{0})}$, are known, problem in (16.6) can be solved locally by each target using only the information gathered from its neighbors and iterative approach.

Therefore, target i updates its position estimate in each iteration, t, by minimizing the following local ML problem:

$$\hat{\boldsymbol{x}}_i^{(t+1)} = \arg\min_{\boldsymbol{x}_i} \sum_{j=1}^{3|\mathcal{E}_{\mathcal{A}_i}|+|\mathcal{E}_{\mathcal{T}_i}|} \frac{1}{\sigma_j^2}\left[\theta_j - f_j(\boldsymbol{x}_i)\right]^2, \forall i \in \mathcal{T}. \tag{16.7}$$

where $\mathcal{E}_{\mathcal{A}_i} = \{j: (i,j) \in \mathcal{E}_{\mathcal{A}}\}$ and $\mathcal{E}_{\mathcal{T}_i} = \{k: (i,k) \in \mathcal{E}_{\mathcal{T}}\}$ represent the set of all anchor and all target neighbors of the target i respectively, and the first $|\mathcal{E}_{\mathcal{A}_i}| + |\mathcal{E}_{\mathcal{T}_i}|$ elements of $\boldsymbol{f}(\boldsymbol{x}_i)$ are given as:

$$f_j(\boldsymbol{x}_i) = L_0 + 10\gamma \log_{10} \frac{\|\boldsymbol{x}_i - \hat{\boldsymbol{a}}_j\|}{d_0}, \text{for } j = 1, \ldots, |\mathcal{E}_{\mathcal{A}_i}| + |\mathcal{E}_{\mathcal{T}_i}|,$$

with

$$\hat{\boldsymbol{a}}_j = \begin{cases} \boldsymbol{a}_j, \text{if } j \in \mathcal{A}, \\ \boldsymbol{x}_j^{(t)}, \text{if } j \in \mathcal{T}. \end{cases}$$

16.3.1. The Proposed Distributed SOCP Algorithm

Assuming that $\hat{\boldsymbol{X}}^{(0)}$ is given, when the noise power is sufficiently small, from (16.1) we can write:

$$\lambda_{ij}\|\boldsymbol{x}_i - \hat{\boldsymbol{a}}_j\| \approx d_0, \forall i \in \mathcal{T}, \forall j \in \mathcal{E}_{\mathcal{A}_i} \cup \mathcal{E}_{\mathcal{T}_i}, \tag{16.8}$$

where

$$\lambda_{ij} = \begin{cases} 10^{\frac{L_0 - L_{ij}^{\mathcal{A}}}{10\gamma}}, \text{if } j \in \mathcal{A}, \\ 10^{\frac{L_0 - L_{ij}^{\mathcal{T}}}{10\gamma}}, \text{if } j \in \mathcal{T}. \end{cases}$$

Similarly, from (16.3) and (16.4) we respectively get:

$$\boldsymbol{c}_{ij}^T(\boldsymbol{x}_i - \boldsymbol{a}_j) \approx 0, \forall i \in \mathcal{T}, \forall j \in \mathcal{E}_{\mathcal{A}_i}, \tag{16.9}$$

and

$$\boldsymbol{k}_{ij}^T(\boldsymbol{x}_i - \boldsymbol{a}_j) \approx \|\boldsymbol{x}_i - \boldsymbol{a}_j\| \cos(\alpha_{ij}^{\mathcal{A}}), \forall i \in \mathcal{T}, \forall j \in \mathcal{E}_{\mathcal{A}_i} \tag{16.10}$$

where $\boldsymbol{c}_{ij} = \left[-\sin\phi_{ij}^A, \cos\phi_{ij}^A, 0\right]^T$ and $\boldsymbol{k}_{ij} = [0,0,1]^T$. According to the LS criterion and (16.8), (16.9) and (16.10) each target updates its location by solving the following problem:

$$\hat{\boldsymbol{x}}_i^{(t+1)} = \arg\min_{\boldsymbol{x}_i} \sum_{j\in\mathcal{E}_{\mathcal{A}_i}\cup\mathcal{E}_{\mathcal{T}_i}} \left(\lambda_{ij}\left\|\boldsymbol{x}_i - \hat{\boldsymbol{a}}_j\right\| - d_0\right)^2 + \sum_{j\in\mathcal{E}_{\mathcal{A}_i}} \left(\boldsymbol{c}_{ij}^T(\boldsymbol{x}_i - \boldsymbol{a}_j)\right)^2 + \sum_{j\in\mathcal{E}_{\mathcal{A}_i}} \left(\boldsymbol{k}_{ij}^T(\boldsymbol{x}_i - \boldsymbol{a}_j) - \left\|\boldsymbol{x}_i - \boldsymbol{a}_j\right\| \cos(\alpha_{ij}^{\mathcal{A}})\right)^2. \tag{16.11}$$

The LS problem in (16.11) is non-convex and has no closed-form solution. To facilitate the design of a convex estimator, we first replace $f(\cdot) = \|\cdot\|^2$ in (11) by another penalty function $f(\cdot) = \|\cdot\|_1 = \sum_i |[\cdot]_i|$, *i.e.*, the Taxicab norm, which is also known as the l_1 norm. In other words, instead of minimizing the l_2 norm of the residual error, for sufficiently small residual error, we can approximate (16.11) by using l_1 norm [18], as:

$$\hat{\boldsymbol{x}}_i^{(t+1)} = \arg\min_{\boldsymbol{x}_i} \sum_{j\in\mathcal{E}_{\mathcal{A}_i}\cup\mathcal{E}_{\mathcal{T}_i}} \left|\lambda_{ij}\left\|\boldsymbol{x}_i - \hat{\boldsymbol{a}}_j\right\| - d_0\right| + \sum_{j\in\mathcal{E}_{\mathcal{A}_i}} \left|\boldsymbol{c}_{ij}^T(\boldsymbol{x}_i - \boldsymbol{a}_j)\right| + \sum_{j\in\mathcal{E}_{\mathcal{A}_i}} \left|\boldsymbol{k}_{ij}^T(\boldsymbol{x}_i - \boldsymbol{a}_j) - \left\|\boldsymbol{x}_i - \boldsymbol{a}_j\right\| \cos(\alpha_{ij}^{\mathcal{A}})\right|. \tag{16.12}$$

Introduce auxiliary variables $r_{ij} = \left\|\boldsymbol{x}_i - \hat{\boldsymbol{a}}_j\right\|$, for $j = 1, \ldots, 3\left|\mathcal{E}_{\mathcal{A}_i}\right| + \left|\mathcal{E}_{\mathcal{T}_i}\right|$ and $\boldsymbol{z} = [z_{ij}]$, where $z_{ij} = \lambda_{ij} r_{ij} - d_0$, for $j = 1, \ldots, \left|\mathcal{E}_{\mathcal{A}_i}\right| + \left|\mathcal{E}_{\mathcal{T}_i}\right|$, $z_{ij} = \boldsymbol{c}_{ij}^T(\boldsymbol{x}_i - \boldsymbol{a}_j)$, for $j = \left|\mathcal{E}_{\mathcal{A}_i}\right| + \left|\mathcal{E}_{\mathcal{T}_i}\right| + 1, \ldots, 2\left|\mathcal{E}_{\mathcal{A}_i}\right| + \left|\mathcal{E}_{\mathcal{T}_i}\right|$, and $z_{ij} = \boldsymbol{k}_{ij}^T(\boldsymbol{x}_i - \boldsymbol{a}_j) - \left\|\boldsymbol{x}_i - \boldsymbol{a}_j\right\| \cos(\alpha_{ij}^{\mathcal{A}})$, for $j = 2\left|\mathcal{E}_{\mathcal{A}_i}\right| + \left|\mathcal{E}_{\mathcal{T}_i}\right| + 1, \ldots, 3\left|\mathcal{E}_{\mathcal{A}_i}\right| + \left|\mathcal{E}_{\mathcal{T}_i}\right|$, $\forall i \in \mathcal{T}$. From (16.12), we get:

$$\underset{\boldsymbol{x}_i, \boldsymbol{z}, \boldsymbol{r}}{\text{minimize}} \quad \sum_{j=1}^{3\left|\mathcal{E}_{\mathcal{A}_i}\right| + \left|\mathcal{E}_{\mathcal{T}_i}\right|} \left|z_{ij}\right|$$

subject to

$$\begin{aligned} r_{ij} &= \left\|\boldsymbol{x}_i - \hat{\boldsymbol{a}}_j\right\|, \text{for } j = 1, \ldots, 3\left|\mathcal{E}_{\mathcal{A}_i}\right| + \left|\mathcal{E}_{\mathcal{T}_i}\right|, \\ z_{ij} &= \lambda_{ij} r_{ij} - d_0, \text{for } j = 1, \ldots, \left|\mathcal{E}_{\mathcal{A}_i}\right| + \left|\mathcal{E}_{\mathcal{T}_i}\right|, \\ z_{ij} &= \boldsymbol{c}_{ij}^T(\boldsymbol{x}_i - \boldsymbol{a}_j), \text{for } j = \left|\mathcal{E}_{\mathcal{A}_i}\right| + \left|\mathcal{E}_{\mathcal{T}_i}\right| + 1, \ldots, 2\left|\mathcal{E}_{\mathcal{A}_i}\right| + \left|\mathcal{E}_{\mathcal{T}_i}\right|, \\ z_{ij} &= \boldsymbol{k}_{ij}^T(\boldsymbol{x}_i - \boldsymbol{a}_j) - \left\|\boldsymbol{x}_i - \boldsymbol{a}_j\right\| \cos(\alpha_{ij}^{\mathcal{A}}), \text{for } j \\ &\quad = 2\left|\mathcal{E}_{\mathcal{A}_i}\right| + \left|\mathcal{E}_{\mathcal{T}_i}\right| + 1, \ldots, 3\left|\mathcal{E}_{\mathcal{A}_i}\right| + \left|\mathcal{E}_{\mathcal{T}_i}\right|. \end{aligned} \tag{16.13}$$

Introduce epigraph variable $\boldsymbol{e} = [e_{ij}], \forall \mathrm{i} \in \mathcal{T}, \mathrm{j} = 1, \ldots, 3\left|\mathcal{E}_{\mathcal{A}_i}\right| + \left|\mathcal{E}_{\mathcal{T}_i}\right|$ and apply SOCP relaxation of the form $\left\|\boldsymbol{x}_i - \hat{\boldsymbol{a}}_j\right\| \leq r_{ij}, \forall \mathrm{i} \in \mathcal{T}, \mathrm{j} = 1, \ldots, 3\left|\mathcal{E}_{\mathcal{A}_i}\right| + \left|\mathcal{E}_{\mathcal{T}_i}\right|$, to obtain:

$$\underset{\boldsymbol{x}_i, \boldsymbol{z}, \boldsymbol{r}, \boldsymbol{e}}{\text{minimize}} \quad \sum_{j=1}^{3\left|\mathcal{E}_{\mathcal{A}_i}\right| + \left|\mathcal{E}_{\mathcal{T}_i}\right|} e_{ij} \tag{16.14}$$

subject to

$$\|\boldsymbol{x}_i - \widehat{\boldsymbol{a}}_j\| \leq r_{ij}, \text{for } j = 1, \dots, 3|\mathcal{E}_{\mathcal{A}_i}| + |\mathcal{E}_{\mathcal{T}_i}|,$$

$$z_{ij} = \lambda_{ij} r_{ij} - d_0, \text{for } j = 1, \dots, |\mathcal{E}_{\mathcal{A}_i}| + |\mathcal{E}_{\mathcal{T}_i}|,$$

$$z_{ij} = \boldsymbol{c}_{ij}^T(\boldsymbol{x}_i - \boldsymbol{a}_j), \text{for } j = |\mathcal{E}_{\mathcal{A}_i}| + |\mathcal{E}_{\mathcal{T}_i}| + 1, \dots, 2|\mathcal{E}_{\mathcal{A}_i}| + |\mathcal{E}_{\mathcal{T}_i}|,$$

$$z_{ij} = \boldsymbol{k}_{ij}^T(\boldsymbol{x}_i - \boldsymbol{a}_j) - \|\boldsymbol{x}_i - \boldsymbol{a}_j\| \cos(\alpha_{ij}^{\mathcal{A}}), \text{for } j = 2|\mathcal{E}_{\mathcal{A}_i}| + |\mathcal{E}_{\mathcal{T}_i}| + 1, \dots, 3|\mathcal{E}_{\mathcal{A}_i}| + |\mathcal{E}_{\mathcal{T}_i}|,$$

$$-e_{ij} \leq z_{ij} \leq e_{ij}, \text{for } j = 1, \dots, 3|\mathcal{E}_{\mathcal{A}_i}| + |\mathcal{E}_{\mathcal{T}_i}|,$$

$\forall i \in \mathcal{T}$. The problem in (16.14) is an SOCP problem, which can be efficiently solved by the CVX package [27] for specifying and solving convex programs.

Assuming that $\mathcal{C}$ represents the set of colors of the sensors, Algorithm 16.1 summarizes the proposed distributed SOCP algorithm. Algorithm 16.1 is distributed in the sense that no central sensor coordinates the network, all communication occur exclusively between two incident sensors, and the data associated with each sensor is processed locally. Lines 5-7 are executed simultaneously by all targets $i \in \mathcal{C}_c$, which may decrease the execution time of the algorithm. The only information exchange occurs at Line 7, when targets broadcast their location updates $\hat{x}_i^{(t+1)}$ to their neighbors. Since $\hat{x}_i^{(t+1)} \in \mathbb{R}^3$, we can conclude that the proposed algorithm requires at most a broadcast of $3 \times T_{\max} \times M$ real values. In the further text, we will refer to Algorithm 16.1 as "SOCP".

Algorithm 16.1. The proposed distributed SOCP algorithm.

Require: $\widehat{\boldsymbol{X}}^{(0)}, T_{\max}, \mathcal{C}, \boldsymbol{a}_j, \forall j \in \mathcal{A}$

1. **Initialize:** $t \leftarrow 0$
2. **repeat**
3. **for** $c = 1, \dots, \mathcal{C}$ **do**
4. **for all** $i \in \mathcal{C}_c$ (in parallel) **do**
5. Collect $\widehat{\boldsymbol{a}}_j, \forall j \in \mathcal{E}_{\mathcal{A}_i} \cup \mathcal{E}_{\mathcal{T}_i}$
6. $\widehat{\boldsymbol{x}}_i^{(t+1)} \leftarrow$ solve (14)
7. Broadcast $\widehat{\boldsymbol{x}}_i^{(t+1)}$ to $\widehat{\boldsymbol{a}}_j, \forall j \in \mathcal{E}_{\mathcal{A}_i} \cup \mathcal{E}_{\mathcal{T}_i}$
8. **end for**
9. **end for**
10. $t \leftarrow t + 1$
11. **until** $t < T_{\max}$

16.4. Complexity Analysis

In order to evaluate the overall performance of a localization algorithm, it is necessary to analyze the trade off between the estimation accuracy and computational complexity. In this section, we investigate the computational complexity of the proposed algorithm. The formula for computing the worst case computational complexity of a SOCP problem [28], given below, is used to calculate the complexity of the proposed SOCP approach:

$$\mathcal{O}\left(\sqrt{S}\left(m^2\sum_{i=1}^{S}n_i+\sum_{i=1}^{S}n_i^2+m^3\right)\right), \tag{16.15}$$

where S is the number of the second-order cone constraints, m is the number of the equality constraints, and n_i is the dimension of the i-th second-order cone.

According to (16.15), we can see that the worst case computational complexity of the proposed SOCP algorithm is $\approx T_{\max}\times M\times\mathcal{O}\left(\left(\max_i\{3|\mathcal{E}_{\mathcal{A}_i}|+|\mathcal{E}_{\mathcal{T}_i}|\}\right)^{3.5}\right)$. From this result, one can see that the computational complexity of a distributed algorithm mainly depends on the size of neighborhood fragments, rather than the total number of sensors in a WSN. Although it is theoretically possible to have that $\max_i\{|\mathcal{E}_{\mathcal{A}_i}|+|\mathcal{E}_{\mathcal{T}_i}|\}=$ $= M+N-1$, for $\forall i\in\mathcal{T}$, in practice, the size of neighborhood fragments are much smaller, due to limited R. Therefore, distributed algorithms are a preferable solution in large-scale and highly-dense networks, since adding more sensors into the network will not have a significant impact on the size of neighborhood fragments.

16.5. Simulation Results

In this section, we present a set of results in order to asses the performance of the proposed approaches in terms of the estimation accuracy and convergence. All of the presented algorithms were solved by using the MATLAB package CVX [27], where the solver is SeDuMi [29]. In order to demonstrate the benefit of fusing two radio measurements versus traditional localization systems, we include also the performance results of the proposed method when only RSS measurements are employed, called here "SOCP$_{\text{RSS}}$".

A random deployment of M targets and N anchors inside a cube region of length B in each Monte Carlo (M_c) run is considered. Random deployment of the sensors is of particular interest, since a common practical requirement for a WSN is that it is flexible in topology; hence, the localization algorithms need to be robust to various scenarios. In order to make the comparison of the considered approaches as fair as possible, we first obtained $M_c=500$ targets' and anchors' positions, as well as noise realizations between two sensors $\forall(i,j)\in\mathcal{E}_{\mathcal{A}}\cup\mathcal{E}_{\mathcal{T}}$ in each M_c run. Furthermore, we made sure that the network graph is connected in each M_c run. We then solved the localization problem with the considered approaches for those scenarios. In all simulations presented here, the reference distance was set to $d_0=1$ m, the reference path loss to $L_0=40$ dB, the communication range of a sensor to $R=6.5$ m, and the PLE was fixed to $\gamma=3$. However, in practice it is almost impossible to perfectly estimate the value of the PLE. Thus, in order to account for a realistic measurement model mismatch and test the robustness of the considered approaches to imperfect knowledge of the PLE, the true PLE for each link was drawn from a uniform distribution on an interval $\gamma_{ij}\in[2.7,3.3],\forall(i,j)\in\mathcal{E}_{\mathcal{A}}\cup\mathcal{E}_{\mathcal{T}}$. Finally, we assumed that the initial guess of the targets' locations, $\widehat{\boldsymbol{X}}^{(0)}$, is in the intersection of the big diagonals of the cube area, since the biggest possible error for this case is half of the big diagonal of the area.

The performance metric is the normalized root mean square error (NRMSE), defined as:

$$\text{NRMSE} = \sqrt{\frac{1}{MM_c}\sum_{i=1}^{M_c}\sum_{j=1}^{M}\left\|\boldsymbol{x}_{ij} - \hat{\boldsymbol{x}}_{ij}\right\|^2},$$

where $\hat{\boldsymbol{x}}_{ij}$ denotes the estimate of the true location of the j-th target, $\boldsymbol{x}_{ij}$, in the i-th M_c run.

Figs. 16.2, 16.3 and 16.4 illustrate the NRMSE versus t performance when $R = 6.5$ m, $\sigma_{n_{ij}} = 3$ dB, $\sigma_{m_{ij}} = 6$ deg, and $\sigma_{v_{ij}} = 6$ deg, for $N = 20, M = 50, N = 30,$ $M = 50,$ and $N = 20, M = 60$, respectively. From these figures, we can see that the performance of all considered algorithms betters as t grows, as anticipated. Also, it can be seen from Figs. 16.2 and 16.3 that the performance of all algorithms improves significantly as more anchors are added into the network. This behavior is expected, since when N grows more reliable information is available in the network. Furthermore, Figs. 16.2 and 16.4 reveal that all approaches require a slightly higher number of iterations to converge when M is increased. However, the estimation accuracy of the considered algorithms does not deteriorate when more targets are added in the network; it actually betters when M is increased. Moreover, from Figs. 16.2, 16.3 and 16.4, one can perceive that all major changes in the performance for the considered algorithms take place in the first few iterations ($t \leq 10$), and that the performance gain is negligible afterwards. This result is very important because it shows that our approache requires a low number of signal transmissions, which might enhance the utilization efficiency of the radio spectrum, a precious resource for wireless communications. It also shows that our algorithms are energy efficient; the communication phase is much more expensive (in terms of energy) than the data processing one [26]. Finally, Figs. 16.2, 16.3 and 16.4 exhibit that the proposed hybrid method outperforms considerably its traditional counterpart that employs RSS measurements only.

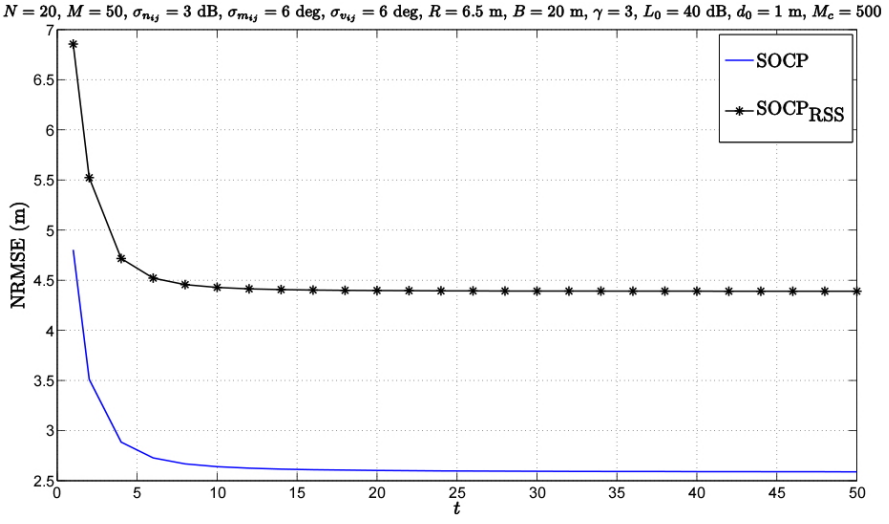


Fig. 16.2. NRMSE versus t comparison, when $N = 20, M = 50, R = 6.5$ m, $\sigma_{n_{ij}} = 3$ dB, $\sigma_{m_{ij}} = 6$ deg, $\sigma_{v_{ij}} = 6$ deg, $\gamma_{ij} \in [2.7, 3.3], \gamma = 3, B = 20$ m, $L_0 = 40$ dB, $d_0 = 1$ m, $M_c = 500$.

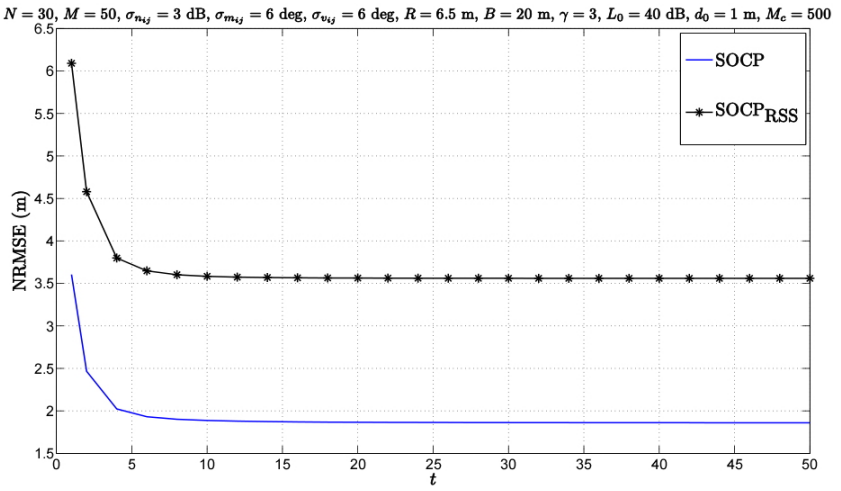


Fig. 16.3. NRMSE versus t comparison, when $N = 30, M = 50, R = 6.5$ m, $\sigma_{n_{ij}} = 3$ dB, $\sigma_{m_{ij}} = 6$ deg, $\sigma_{v_{ij}} = 6$ deg, $\gamma_{ij} \in [2.7, 3.3], \gamma = 3, B = 20$ m, $L_0 = 40$ dB, $d_0 = 1$ m, $M_c = 500$.

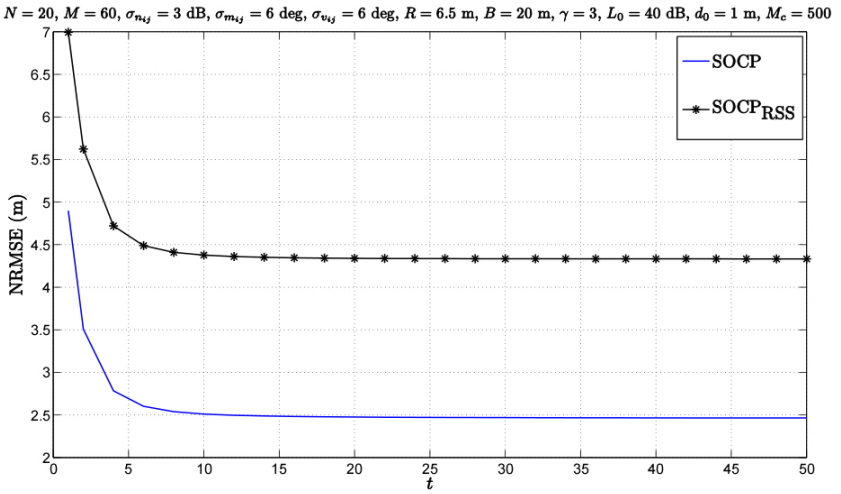


Fig. 16.4. NRMSE versus t comparison, when $N = 20, M = 60, R = 6.5$ m, $\sigma_{n_{ij}} = 3$ dB, $\sigma_{m_{ij}} = 6$ deg, $\sigma_{v_{ij}} = 6$ deg, $\gamma_{ij} \in [2.7, 3.3], \gamma = 3, B = 20$ m, $L_0 = 40$ dB, $d_0 = 1$ m, $M_c = 500$.

In Figs. 16.5, 16.6 and 16.7 we investigate the impact of the quality of RSS and AoA measurements on the performance of the considered approaches. More precisely, Figs. 16.5, 16.6 and 16.7 respectively illustrate the NRMSE versus $\sigma_{n_{ij}}$ (dB), $\sigma_{m_{ij}}$ (deg) and $\sigma_{v_{ij}}$ (deg) comparison, when $N = 20$, $M = 50$, $R = 6.5$ m, and $T_{\max} = 25$. In these figures, we can observe that the performance of the proposed algorithm degrades as the quality of a certain measurement decreases, as expected. It can also be seen that the quality of the RSS measurements has the most significant impact on the performance of the proposed algorithm, while the error in the azimuth and elevation

angle measurements have marginal influence on the performance. This is not surprising, since the error of a few degrees in AoA measurements does not impair considerably their quality on a fairly short distance (communication range of all sensors is restricted to $R = 6.5$ m), as shown in Figs. 16.6 and 16.7. On the other hand, RSS measurements are notoriously unpredictable [1]. Nonetheless, we can see from Fig. 16.5 that the performance loss is lower than 15 % for the proposed algorithm, which is relatively low for the considered noise range.

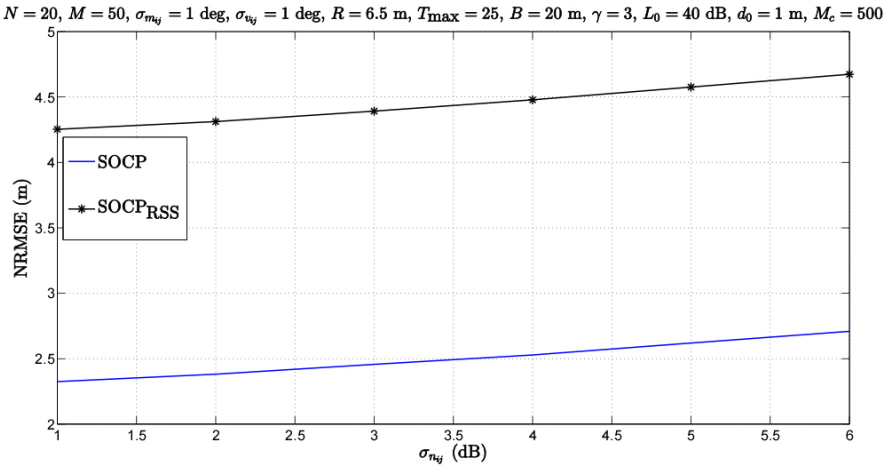


Fig. 16.5. NRMSE versus $\sigma_{n_{ij}}$ (dB) comparison, when $N = 20, M = 50, R = 6.5$ m, $\sigma_{m_{ij}} = 1$ deg, $\sigma_{v_{ij}} = 1$ deg, $\gamma_{ij} \in [2.7, 3.3], \gamma = 3, T_{\max} = 25, B = 20$ m, $L_0 = 40$ dB, $d_0 = 1$ m, $M_c = 500$.

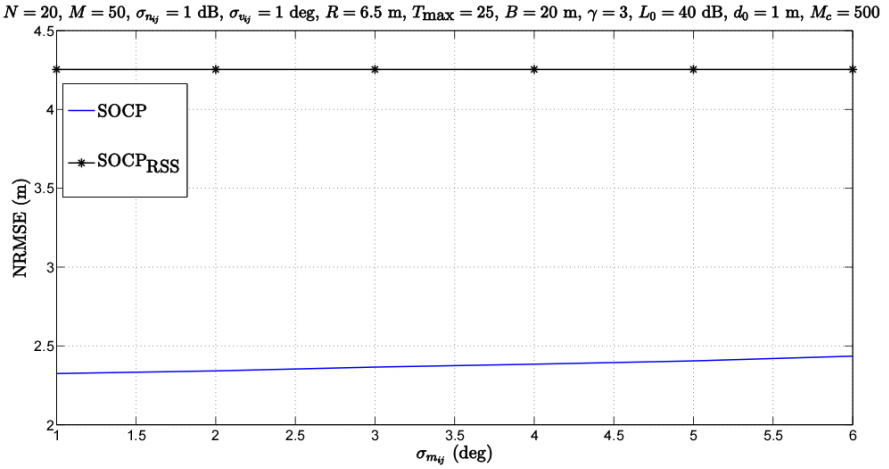


Fig. 16.6. NRMSE versus $\sigma_{m_{ij}}$ (deg) comparison, when $N = 20, M = 50, R = 6.5$ m, $\sigma_{n_{ij}} = 1$ dB, $\sigma_{v_{ij}} = 1$ deg, $\gamma_{ij} \in [2.7, 3.3], \gamma = 3, T_{\max} = 25, B = 20$ m, $L_0 = 40$ dB, $d_0 = 1$ m, $M_c = 500$.

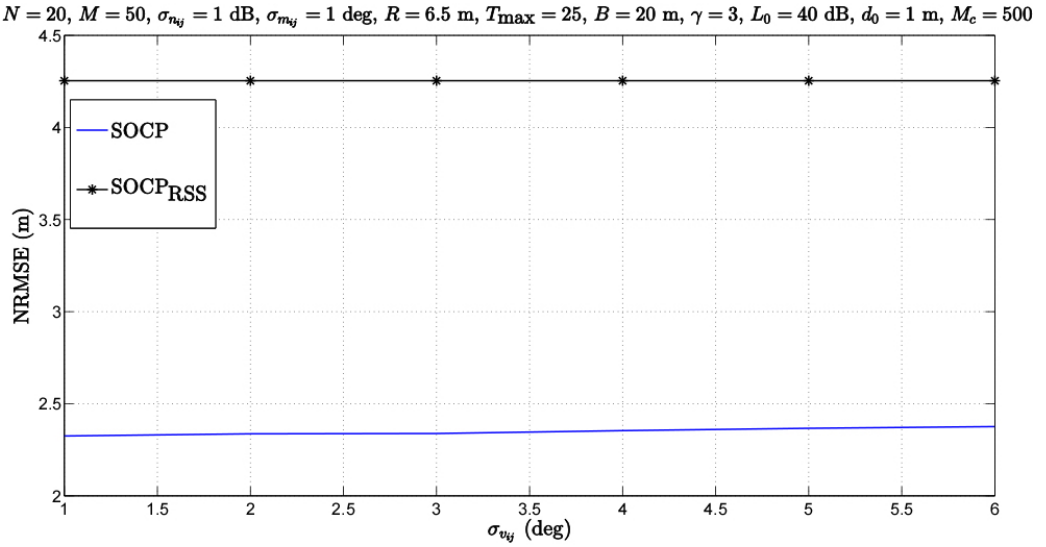


Fig. 16.7. NRMSE versus $\sigma_{v_{ij}}$ (deg) comparison, when $N = 20, M = 50, R = 6.5$ m, $\sigma_{n_{ij}} = 1$ dB, $\sigma_{m_{ij}} = 1$ deg, $\gamma_{ij} \in [2.7, 3.3], \gamma = 3, T_{\max} = 25, B = 20$ m, $L_0 = 40$ dB, $d_0 = 1$ m, $M_c = 500$.

16.6. Conclusions

In this work, we addressed the hybrid RSS/AoA target localization problem in a cooperative 3-D WSN. A new formulation based on a nonconvex LS joint formulation fusing information from RSS and AoA and a tight convex relaxation for this problem were devised. We showed that the derived novel estimator is suitable for distributed implementation, and we proposed a novel distributed algorithm based on SOCP relaxation technique. The simulation results showed that the new algorithm can provide accurate localization in a variety of scenarios, exhibiting excellent performance both in terms of the estimation accuracy and convergence. Moreover, the simulation results confirmed the effectiveness of fusing RSS and AoA radio measurements in comparison with traditional RSS localization system, showing a remarkable improvement in the estimation accuracy.

Acknowledgements

This work was partially supported by Fundação para a Ciência e a Tecnologia under Projects PEst-OE/EEI/UI0066/2014, and PEst-OE/EEI/LA0008/2013 (IT pluriannual founding and HETNET), PEst-OE/EEI/UI0066/2011 (UNINOVA pluriannual founding), ADIN PTDC/EEI-TEL/ 2990/2012, COPWIN PTDC/EEI-TEL/1417/2012 and PTDC/EEITEL/6308/2014-HAMLeT, as well as the grants SFRH/BPD/108232/ 2015, SFRH/BD/91126/2012 and Ciência 2008 Post-Doctoral Research grant.

References

[1]. N. Patwari, Location Estimation in Sensor Networks, Ph. D. Thesis, *University of Michigan,* Ann Arbor, MI, USA, 2005.

[2]. G. Destino, Positioning in Wireless Networks: Noncooperative and Cooperative Algorithms. Ph. D. Thesis, *University of Oulu,* Oulu, Finland, 2012.

[3]. Y. Liua, F. Guoa, L. Yangb, and W. Jianga, Source Localization Using a Moving Receiver and Noisy TOA Measurements, *Signal Process.,* Vol. 119, February 2016, pp. 185--189.

[4]. X. Qua and L. Xie, An Efficient Convex Constrained Weighted Least Squares Source Localization Algorithm Based on TDOA Measurements, *Signal Process,* Vol. 119, February 2016, pp. 142--152.

[5]. J. Prieto, S. Mazuelas, A. Bahillo, P. Fernandez, R. M. Lorenzo, and E. J. Abril, On the Minimization of Different Sources of Error for an RTT-based Indoor Localization System Without Any Calibration Stage, *IPIN*, September 2010, pp. 1-6.

[6]. Y. Wang and K. C. Ho, An Asymptotically Efficient Estimator in Closed- Form for 3D AOA Localization Using a Sensor Network, *IEEE Trans. Wirel. Commun*., Vol. PP, No. 99, July 2015, pp. 1.

[7]. S. Tomic, M. Beko, and R. Dinis, RSS-based Localization in Wireless Sensor Networks Using Convex Relaxation: Noncooperative and Cooperative Schemes, *IEEE Trans. Veh. Technol.,* Vol. 64, No. 5, May 2015, pp. 2037--2050.

[8]. A. Bahillo, S. Mazuelas, R. M. Lorenzo, P. Fernández, J. Prieto, R. J. Durán, and E. J. Abril, Hybrid RSS-RTT Localization Scheme for Indoor Wireless Networks, *EURASIP J. Advan. Sig. Process.,* Vol. 2010, No. 1, March 2010, pp. 1--12.

[9]. U. Hatthasin, S. Thainimit, K. Vibhatavanij, N. Premasathian, and D. Worasawate, The Use of RTOF and RSS for a One Base Station RFID system, *IJCSNS*, Vol. 10, No. 7, July 2010, pp. 862-865.

[10]. T. Gadeke, J. Schmid, M. Kruger, J. Jany, W. Stork, and K. D. Muller- Glaser, A Bi-Modal Ad-Hoc Localization Scheme for Wireless Networks Based on RSS and ToF Fusion, *WPNC*, March 2013, pp. 1-6.

[11]. K. Yu, 3-D Localization Error Analysis in Wireless Networks, *IEEE Trans. Wirel. Commun.,* Vol. 6, No. 10, October 2007, pp. 3473-3481.

[12]. S. Wang, B. R. Jackson, and R. Inkol, Hybrid RSS/AOA Emitter Location Estimation Based on Least Squares and Maximum Likelihood Criteria, *IEEE QBSC*, June 2012, pp. 24-29.

[13]. L. Gazzah, L. Najjar, and H. Besbes, Selective Hybrid RSS/AOA Approximate Maximum Likelihood Mobile Intra Cell Localization, *EW*, April 2013.

[14]. L. Gazzah, L. Najjar, and H. Besbes, Selective Hybrid RSS/AOA Weighting Algorithm for NLOS Intra Cell Localization, *IEEE WCNC*, April 2014, pp. 2546-2551.

[15]. Y. T. Chan, F. Chan, W. Read, B. R. Jackson, and B. H. Lee, Hybrid Localization of an Emitter by Combining Angle-of-Arrival and Received Signal Strength Measurements, *IEEE CCECE*, May 2014, pp. 1-5.

[16]. N. Patwari, J. N. Ash, S. Kyperountas, A. O. Hero III, R. L. Moses, and N. S. Correal, Locating the Nodes: Cooperative Localization in Wireless Sensor Networks, *IEEE Sig. Process. Mag.,* Vol. 22, No. 4, July 2005, pp. 54-69.

[17]. R. Fletcher, Practical Methods of Optimization, *John Wiley & Sons*, Chichester, UK, 1987.

[18]. S. Boyd and L. Vanderberghe, Convex Optimization, *Cambridge University Press,* Cambridge, UK, 2004.

[19]. T. S. Rappaport, Wireless Communications: Principles and Practice, *Prentice-Hall,* Upper Saddle River, NJ, USA, 1996.

[20]. M. L. Sichitiu and V. Ramadurai, Localization of Wireless Sensor Networks with a Mobile Beacon, *IEEE MASS*, October 2004, pp. 174-183.

[21]. Z. Xiang and U. Ozguner, A 3-D Positioning System for Off-road Autonomous Vehicles, *IEEE IV*, June 2005, pp. 130-134.
[22]. M. B. Ferreira, Hybrid Indoor Localization Based on Ranges and Video, M. Sc. Thesis, *Universidade de Lisboa, Instituto Superior Técnico,* Lisbon, Portugal, 2014.
[23]. S. M. Kay, Fundamentals of Statistical Signal Processing: Estimation Theory, *Prentice-Hall,* Upper Saddle River, NJ, USA, 1993.
[24]. S. C. Ergen and P. Varaiya, TDMA Scheduling Algorithms for Wireless Sensor Networks, *The Journal of Mobile Communication, Computation and Information*, Vol. 16, No. 4, May 2010, pp. 985-997.
[25]. J. F. C. Mota, J. M. F. Xavier, P. M. Q. Aguiar, and M. Puschel, DADMM: A Communication-Efficient Distributed Algorithm for Separable Optimization, *IEEE Trans. Signal Process*., Vol. 61, No. 10, May 2013, pp. 2718-2723.
[26]. S. Tomic, M. Beko, and R. Dinis, Distributed RSS-Based Localization in Wireless Sensor Networks Based on Second-Order Cone Programming, *Sensors*, Vol. 14, No. 10, October 2014, pp. 18410-18432.
[27]. M. Grant and S. Boyd, CVX: Matlab Software for Disciplined Convex Programming, Version 1.21, Available online: http://cvxr.com/cvx (accessed on 15 April 2010).
[28]. I. Pólik and T. Terlaky, Interior Point Methods for Nonlinear Optimization, Nonlin. Optim., G. Di Pillo, F. Schoen (Eds.), 1st Edition, *Springer*, 2010.
[29]. J. F. Sturm, Using SeDuMi 1.02, a MATLAB Toolbox for Optimization Over Symmetric Cones, *Optim. Meth. Softw.,* 1998.

Chapter 17
Clustering Approach Based on the Redundancy in a Linear Sensor Network Using a Token-based MAC Protocol

El Hadji Malick NDOYE, Frédérique JACQUET, Michel MISSON and Ibrahima NIANG

17.1. Introduction

A wireless sensor network is a large number of sensor nodes deployed in a fixed or random manner over a wide area for environmental monitoring applications. Wireless sensors communicate via wireless links and are powered by batteries. They collect and provide information to the base station usually called sink. The information collected is generally of physical, chemical or biological nature. For some of these applications, as pipeline or road monitoring, wireless sensor nodes have to be deployed in a linear manner. We refer to these WSNs as Linear Sensor Networks (LSNs) [1]. The example is shown in Fig. 17.1.

In LSN, the neighborhood is really dependent on the physical characteristics of the nodes. The number of neighbors for a given node is function of the transmission range and distance between two adjacent nodes. In this chapter, we consider a Linear Token Based MAC protocol for linear sensor network. The access to the transmission channel is guaranteed by the token which is created and generated toward the sink node. The token contains temporal information on the periods of activity and inactivity of nodes.

In the case of many tokens circulating, it is important to define a minimal distance between two adjacent tokens in order to avoid interferences and collisions. This distance enables the definition of an isolation pattern around the token holder which we will call linear cluster, and which allows us a spatial reutilization of the activities in the linear network. In this chapter we focus on a clustering approach in a linear sensor network in order to protect exchanges in the network.

El Hadji Malick Ndoye
Clermont Université / LIMOS CNRS - Complexe scientifique des Cézeaux, 63172 Aubière cedex, France
Laboratoire d'Informatique, Université Cheikh Anta Diop de Dakar (UCAD), B.P. 5005 Dakar-Fann, Sénégal

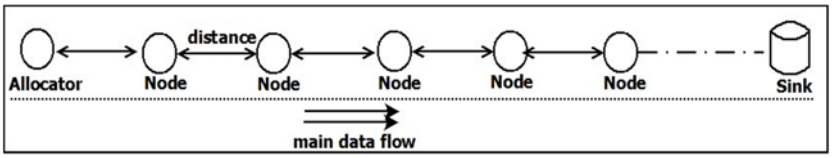


Fig. 17.1. Example of LSN.

The rest of the chapter is outlined as follows: in Section 17.2 we present a state of the art on the clustering used wireless networks. Section 17.3 gives the hypothesis and the network topology; Section 17.4 presents the mechanism of redundancy. In Section 17.5, we are going to calculate the minimal distance between two node token holders. Section 17.6 defines the notion of linear cluster and we evaluate its impact on the throughput in Section 17.7. Finally, we end this chapter with a conclusion and perspectives in Section 17.8.

17.2. State of Art

We find linear sensor networks in the pipeline monitoring applications [2]-4]. In this case, sensor networks are used to detect the toxic products contained in those pipelines (gas, chemicals, radioactive elements, oil, etc.) or leakage. Another application of the linear topology of wireless sensor networks is the monitoring of mines [5, 6] or volcanoes [7]. In this case, the sensor nodes are particularly responsible for monitoring environmental temperature changes to detect abnormalities in a mine. LSNs are also found in the monitoring of trains, rails for detecting a malfunction of hardware components such as wagons, the quality of railways, etc. These wireless networks with linear topology are also used in some monitoring applications of bridges or roads.

The clustering technique [8] is often used to optimize performance parameters such as the throughput, the delivery ratio or the end-to-end delay. It consists in subdividing the network in patterns to ensure spatial reuse i.e. each of these patterns can use a common resource for those reasons without generating harmful interferences. Some clustering algorithms are based on the physical characteristics of the nodes as the transmission range [9, 10], the residual energy [11, 12]. Other algorithms are based on network characteristics such as mobility [13, 14], the nature of data collected using data selection function [15] or the network topology based on the deployment model of the nodes [16]. Generally inter-cluster communication is done in minimum hops. Intra-cluster interferences are managed by the access mechanism to the transmission channel. The TDMA technique is most often used to assign the nodes access channel [11, 17]. For inter-cluster interference, several management methods are used as in the technique used in FDMA [18]. LEACH uses DS-SS technique [19] in all clusters. This technique involves a spread spectrum and is used in satellite communications, wireless networks to make them more resistant to radio interference signals. In [20], the authors present a method for managing inter-cluster interference based on TDMA technology.

However, these techniques no matter how effective they are in WSN networks are not adapted to linear networks for several reasons. First, they do not fully guarantee that absence of interference. Indeed, these techniques offer the access management mechanisms to the transmission channel but do not unequivocally focus on the management of interferences which are diverse in nature and therefore are not just related to access to transmission channel. In the case of linear networks the interferences can be very detrimental and significantly impact on the packet delivery rate to the sink. Then, the use of both methods of access to the transmission channel one between nodes in a cluster and the other between the different cluster nodes may not be an optimal solution for end-to-end delay in linear wireless sensor networks. Indeed, because of the usual long length in this type of network, it is important that access to the transmission channel reduces the end-to-end delay.

Therefore, the concept of clustering as defined in WSN is not adapted for linear sensor array. A cluster in a linear network should take into account its particular characteristics such as linearity, limited neighborhood, the propagation conditions in the given setting, interferences and reusability factor to optimize network performances defined above. Some work has recently focused on the clustering in the linear sensor networks, but without meeting the listed conditions. In [21], the authors have presented a method to calculate the optimal size of a cluster in a linear sensor network based on connectivity of the sensor nodes. However, this calculation is done regardless of reusability factor of clusters or interferences generated by the activity of the nodes.

In this chapter, we propose a clustering method in a linear sensor network to enhance performance parameters. We base our proposition on propagation conditions and characteristics of sensor nodes.

17.3. Hypothesis

We consider a linear network of N nodes placed uniformly on the same line. The distance between two adjacent nodes equals to ***d***. The MAC protocol considered is defined in [22] and we assume that the existence of several simultaneous access to the transmission channel whenever network length allows it. In this case, several tokens can circulate through the linear network at the same time. The propagation model considered in this study is the Tworayground. In this case, for a given radio link, the path loss is considered to be constant. Therefore, the reception power for a given link does not vary a lot between the transmitter and the receiver. So, in the linear topology, we neglect the environmental fluctuations that could make complex the environment and difficult to model. We focus our study on this strength to define the concept of logic and static cluster in the following sections.

Under these conditions, each node has an identifier ranging from 1 to N from the Token Allocator to the sink. The possibility of downlink traffic is assumed in the network. The physical properties of the sensors are such that a node is in range of one or more other nodes in its neighborhood according to the distance d. In this case, a node can relay its data to its neighbors in one hop or at its more distant neighbors in its transmission range.

For several reasons as the maximum number of retransmissions (occurs when a node returns 5 times data frame without receiving acknowledgment) or overflow of the queues, some packets do not arrive at the sink. If we denote by t a period of time including several token periods, if ***n*** is the number of packets reaching the sink during ***t***, we can define the aggregate throughput ***D*** of the LSN as:

$$D = \frac{n * PacketSize}{t}$$

We also define the end-to-end delay as the time between the packet creation at the source, and its successful delivery at the sink.

We defined ***SDur*** (Shuttle Duration) as the shuttle duration and ***SC*** (Shuttle Capacity) as the shuttle capacity that defines the number of packets that a given node can send during the passage of the shuttle [22].

17.4. The Mechanism of Redundancy

In this section we study the concept of redundancy according to the neighborhood in the linear network. Then we analyze the impact of redundancy on the distance to the sink and network performance indicators.

17.4.1. Definition of Redundancy

Let be ***p*** the radio transmission range expressed in units of length and ***d*** the distance between nodes. The relationship between ***p*** and ***d*** defines the neighborhood by determining the number of neighbors for a given node. This neighborhood is important for the data routing. Redundancy is defined to indicate the number of neighbors of a current node in the network. This redundancy is different from network density because it allows the exploitation to take advantage advantage of linearity of the topology. This redundancy occurs in a network to:

- Route packets: redundancy allows multiple choices on relay nodes. If a node has a range of ***n*** neighboring nodes in the direction of the sink, then it has ***n*** choices to relay the packets;
- Ensure connectivity in the network: redundancy increases network connectivity and helps improve fault tolerance in case of broken links or nodes. Indeed, the conditions of use of the sensors due to the needs of surveillance applications or environmental conditions monitored are often hostile (temperature, weather conditions, etc.) cause that some nodes or links are often broken;
- Reduce congestion areas in the linear topology: the drainage of data towards the sink in a one hop manner is a source of congestion in the network when approaching the sink. This congestion is manifested by queues overflow, loss and packet drops [23]. Redundancy helps relay data in a multi-hop which has the advantage of reducing overflow queues.

Depending on the relationship between ***p*** and ***d***, we may define several levels of redundancy in a linear topology. A linear sensor network is R-redundant when for a given current node, there is at least ***R*** left neighbors and ***R*** right neighbors. The integer ***R*** is called the redundancy factor and determines the number of neighbors in each direction (right and left) as shown in Fig. 17.2. This contribution is based on the strong following assumption: for R-redundant topology, the relation $R * d < p < (R+1) * d$ is stable, and therefore the length of the radio links of such a topology is relatively constant and e over the line [24].

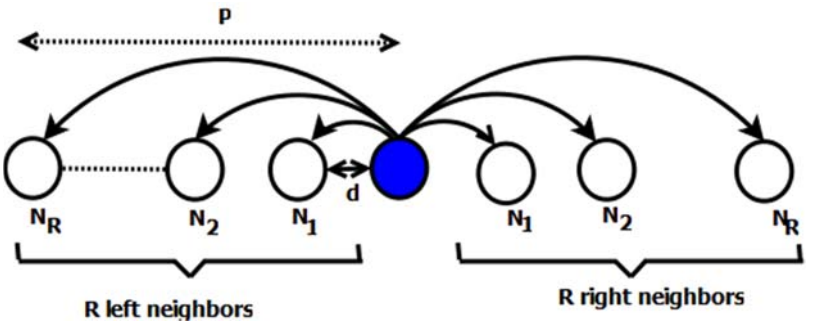


Fig. 17.2. Connectivity of an R-Redundant LSN.

17.4.2. Some Examples of R-redundant Networks

Depending on the value of ***R***, we can define several types of R-redundant networks. In this study, we limit our study to the description of the 1, 2, 3-redundant networks.

17.4.2.1. Case of 1-redundant LSN

It is characterized by the relation: $d < p < 2 * d$. In this case, each node has two neighbors: one on the left and another on the right. The network is also called strictly linear since the vicinity of the network is very low. The drainage of the flow is done from node to node in one hop manner to the sink. For such topology, if a link is broken or a node is down, the whole network is affected because there is no alternative route. Therefore, a break causes an overall malfunction of the linear topology.

17.4.2.2. Case of 2-redundant LSN

It is characterized by the relation: $2 * d < p < 3 * d$. In this case, each node has four neighbors: two on its left and two on the right. In this network, when there is a faulty node, it is still possible to transmit data to the next node in range. In other words, for a given node, there are two possibilities of relaying data in each direction. This gives it greater reliability in the delivery data as in 1-redundant network.

17.4.2.3. Case of 3-redundant LSN

In this network the neighbourhood relationship is given by the formula: $3 * d < p < 4 * d$. Each node has three neighbors to the left and three to the right and therefore has three possible choices for relaying node of a given direction. This topology offers more solutions for fault tolerance than the topology 2-redundant by compensating for the failure of two nodes or two consecutive nodes.

17.4.3. Impact of Redundancy on the Distance to the Sink

In this part we will analyze the impact of redundancy on the flow of data at the sink. To take advantage of redundancy, it is assumed that a node of an R-redundant network relays data to its neighbors at R hops following node. Let's consider the flow to the sink.

Fig. 17.3 and Fig. 17.4 provide an understanding of the interest to exploit redundancy to route traffic. In both figures, the dotted arrows indicate the path of the generated tokens that systemically follow the path *A -> B -> C ... -> sink* as relayed from node to node in one hop manner. The solid arrows indicate the path taken by the data packets in the direction of the sink.

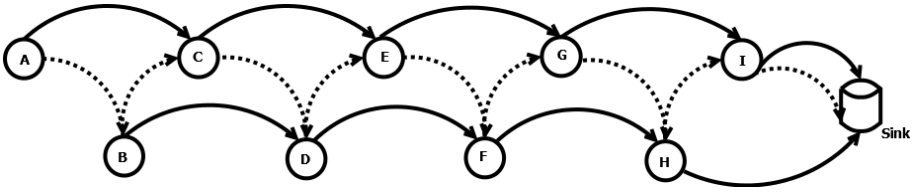


Fig. 17.3. Branches formation in a 2-Redundant LSN.

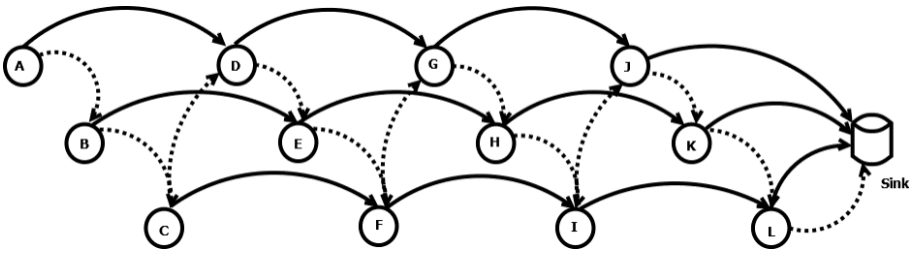


Fig. 17.4. Branch formation in a 3-Redundant LSN.

For Fig. 17.3 for the case of a 2-redundant LSN, traffic generated by A is transmitted directly to its neighbors at two hops namely the C node in order to be concatenated with the traffic generated locally. This concatenated traffic is then forwarded to E and spread out to the others. In such a network, it can be seen that there are two branches (A, C, E, G, I) and (B, D, F, H) which converge towards the sink. In this case, the sink directly receives traffic concatenated by nodes I and H.

For Fig. 17.4 for the case of 3-redundant, traffic from A is directly transmitted to D to be concatenated with the locally generated traffic. This concatenated traffic is then transmitted to the R right neighbor and so on. In such a network, it can be seen that there are three branches (A, D, G, J), (B, E, H, K) and (C, F, I, L) which converge towards the sink. This receives the traffic generated by node J, K and H. This way of relaying the traffic on each of these branches has an impact on the performance in terms of throughput, delivery ratio or end-to-end delay.

Referring to Fig. 17.3 showing a situation in which the stations A and K are in possession of a token. A will transmit its data to D then its token to B. K will transmit its data to the sink and then the token to L.

We can notice two important things:

- The existence of a token allows the sink to receive when its neighbors are token holders. For example for a 3-redundant: 3 times (J, K and L);

- The existence of several branches allows the stations next to the sink to spread the cumulative traffic at each hop. Thus, K accumulates only the traffic from A to D and G.

We can generalize this analysis to an R-redundant network. Indeed, the traffic transmitted by a node X towards the sink is relayed between R hopneighbors towards the sink. The X node relays the data to its R + X hop neighbors, which relays it to the neighbor X + 2 * R and carries to the wells. Therefore, X is associated with the branch (X, X + R, R * X + 2, ..., X + n * R) where n is the number of relay nodes associated with X. In addition, with the token, the sink receives traffic from R nodes neigboring nodes. The impact of redundancy on the formation of branches can be seen on the fair distribution of the load on the nodes. The overall network load is distributed over the various branches which reduces the cumulative effect, which is well known in linear networks. This allows better management of congestion.

17.5. Estimation of the Distance between Token Holders

As indicated in our assumptions, we consider a linear network consisting of ***N*** nodes sensors. If ***N*** is large enough, i.e. the linear network is long enough. Therefore, multiple tokens can circulate simultaneously. This can have several advantages in the linear network performances.

- Increase throughput at the sink: the passage of several tokens at the same time in the network is used to route data over the sink. This has a significant impact on the throughput at the sink in the sense that it receives more data per time units.

- Reduce end-to-end delay: when the tokens are frequent frames stay less time in queues.

- Increase delivery ratio to the sink: the passage of several tokens reduces frame accumulation effect in the queues. This is to better manage the impact of traffic clipping. Accordingly, the delivery rate increases for a given node.

This induces several distinct activity areas on the network as introduced in Fig. 17.5. These areas of activities, if they are not managed properly, can cause interferences between the nodes causing thereafter collision risks. In other words, it is necessary that the activity of a node with the token (or N_i N_j in Fig. 17.5) does not affect other token holder nodes at the same time. This question is equivalent to setting a minimum distance between two consecutive areas of activities to avoid interference effects, or more precisely to define the distance between two consecutive token holder nodes to protect the exchanges initiated by them.

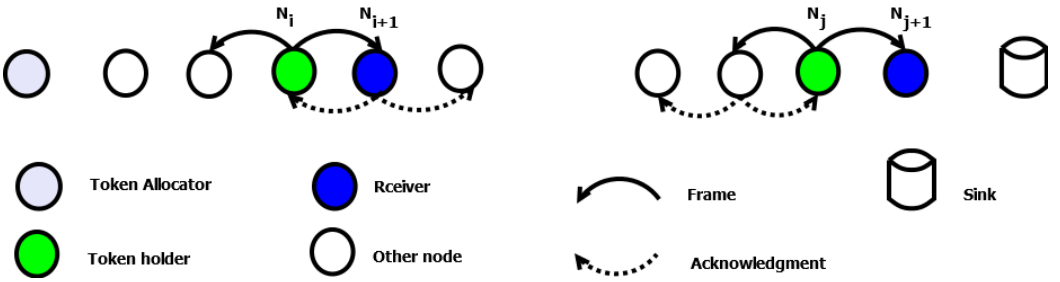


Fig. 17.5. Example of activity areas in a 1-Redundant LSN.

This situation described in Fig. 17.5 allows us to specify that traffic is mainly aimed at the sink (like the frame transmitted by N_i for N_{i+1}) but a reverse traffic would also be possible (such as the frame transmitted by N_j for N_{j-1}).

The distance between two token holdernodes is logically calculated and is static in terms of number of hops without taking into account an evolutionary behavior of the transmission range of a node. In other words, it is calculated based on the considerations without network connectivity fluctuations that could generate variations in path loss on the transmitted signal.

17.5.1. Case of a 1-redundant LSN

This logical approach assumes that the condition $d < p < 2 * d$ means that a frame transmitted is heard by the neighboring node but does not cause significant interference to neighbor two-hop as indicated in the assumptions. Fig. 17.6 shows a portion of the network with downlink traffic. Consider the case of transmission of an uplink frame from A to B and the existence of downlink traffic from E to D when A and E are two consecutive token holders. It is considered the worst-case that A sends uplink traffic to B and E sends downlink traffic to D. In this case, the node C is exposed to interferences produced by nodes B and D acknowledgments but this has no impact since C is not the receiver of the exchanged frames. This assumption applies when B acknowledges a frame transmitted by

A without disturbing the reception by D of the frame sent by E. Thus, in this process the exchanges between A and B are protected from interference caused by the exchanges between E and D. It is then easy to see that the minimum distance in terms of hops between two node token holders (here A and E in Fig. 17.6) is 4 hops.

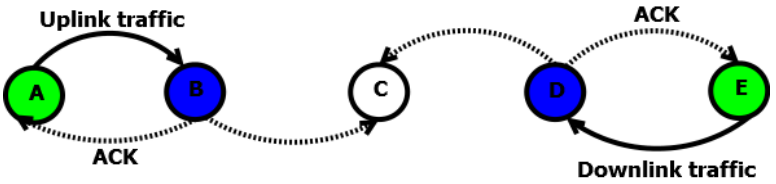


Fig. 17.6. Distance between two token holders in a 1-Redundant LSN.

17.5.2. Case of a 2-redundant LSN

Let's consider the uplink traffic from A to C and downlink traffic from H to F when A and H are simultaneously tokens holders as shown in Fig. 17.7. When C and F and H respectively acknowledge nodes A and H, nodes B, D, E and G are therefore involved by the exchanges. Nodes D and E are the most exposed to radio interference from the linkage produced by C and H. But interferences will not impact on the nodes D and E which are not receivers of the exchanged traffic. As in the case of 2-redundant network, this assumption is valid only when the acknowledgments by C to A do not interfere with the reception of the frame transmitted by H to F. Therefore, the activities of the nodes A and H token holders are simultaneously protected to the risks of interference due to acknowledgments respectively of the nodes C and F. Considering the more general case where two token holder nodes can generate uplink and downlink traffic, the Fig. 17.7 is be used to show that the minimum distance is 7 hops.

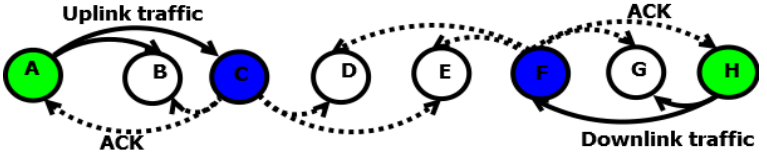


Fig. 17.7. Distance between two token holders in a 3-Redundant LSN.

17.5.3. Case of a 3-redundant LSN

Consider the case where nodes A and K token holders respectively transmit their uplink and downlink traffic to nodes D and H as shown in Fig. 17.8. Here, the intermediate nodes involved in the interference are E, F and G. However, they are not receivers of A and K traffic. Therefore, the risk of interference is not detrimental to this portion of the network. Indeed, with the access method, the concerned nodes are either in sleep mode or in idle and therefore are not engaged in exchanges of the token holders. Fig. 17.8 shows that for a 3-redundant network, the distance between two token holder nodes is 10 hops.

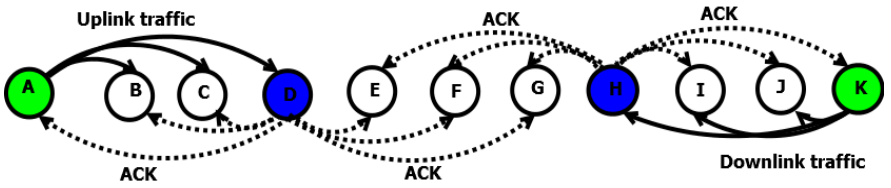


Fig. 17.8. Distance between two token holders in a 3-Redundant LSN.

17.5.4. Generalization

We can extend this study to an R-redundant LSN to estimate the distance between two token holder nodes. The number of nodes between the token holders can be divided into three groups according to the impact that the nodes' activities may have on those nodes.

- Receiving nodes: each node token holder sends its traffic to its neighbor at R hops. Thus, between these two nodes are two receiving ones when downlink traffic exists in the LSN.

- The nodes between the transmitter and the receiver: these nodes have the particularity of being concerned only with trade between the transmitter and the receiver. They are not concerned with the activity of the other transmitter. In addition, they are usually in token standby mode. For each transmitter, there are R-1 nodes separating it from the receiver node. Thus, in total there are 2 * R-2 nodes between the two token holdernodes in the case of R-redundant network. We call this set of nodes the relay area as shown in Fig. 17.9.

- The nodes between the two receivers: these are the nodes that are within range of the two receivers. These nodes are affected by the acknowledgments of receivers and therefore can be exposed to radio interference. However, this possibility does not exist due to the fact that these nodes are waiting for the next token and are therefore in sleep mode. Since all nodes have the same characteristics, we can count the number R nodes, between two receivers. We call this set of nodes the isolation area as shown in Fig. 17.9.

Fig. 17.9 shows the distribution of nodes between two node token holders.

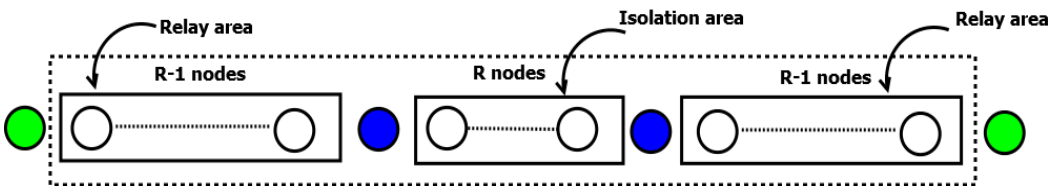


Fig. 17.9. Distance between two token holders in an R-Redundant LSN.

If we call Dist the distance between two token holder nodes in terms of number of hops, then the distance according to R is given by the formula:

$$Dist(R) = 3 * R + 1$$

17.6. Definition of a Logical Cluster

A node N token can communicate with all the nodes in its neighborhood whose size depends on the factor R. This set of nodes is defined by $\{N_R, N_{R-1}, ..., N, ..., N_{R-1}, N_R\}$. Consider $\{N_{A-R,}. ..,N_A , ...,N_{A+R}\}$ and $\{N_{B-R,}. ..,N_B , ...,N_{B+R}\}$ whose two central nodes N_A and N_B are simultaneous token holders We can identify the risk of interference on the frame transmitted by the node N_A. These interferences are caused by transmission activities of nodes of the bloc $\{N_{B-R,}. ..,N_B , ...,N_{B+R}\}$.

In the previous section, we evaluated the distance in number of hops between N_{A+R} and interfering node N_{int} which logically guarantees a good reception. This distance has identified a pattern of isolation around the token holder that we call linear cluster and allows us a spatial reuse activities on the linear network. The cluster is then composed of a central that is the token holder and set of nodes called resources that protect the frames transmitted by the central node. Fig. 17.10 shows an example of linear cluster.

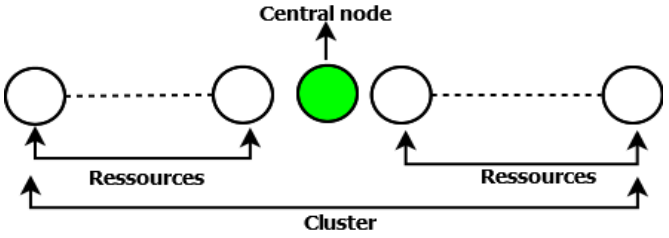


Fig. 17.10. Representation of a linear cluster.

It is known that propagation conditions such as radio links and transmission range of the nodes remain constant over the entire line. So, for a given central node we can associate two relay areas and one isolation areas in order to protect its transmitted frames. This is shown in Fig. 17.11. If we call LCSize () the size of the linear cluster in terms of number of nodes, we can say that according to the propagation conditions that LCSize is given by the formula:

$$LCSize\ (R) = 3 * R + 1$$

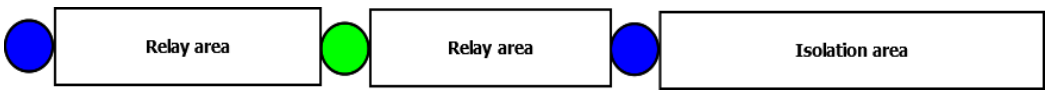


Fig. 17.11. Representation of a linear cluster size.

The size of the linear cluster depends only on R and increases linearly as a function of this parameter. Consider the cluster $\{N_{A-R,}..,N_A, ...,N_{A+R}\}$. At time T_0 the node A is the token. After the passage of the token at the instant $T_0 + t$ the node N_{A+1} becomes the token holder and therefore the cluster moves and becomes $\{N_{A-R+1},....,N_A....N_{A+R+1}\}$. The process is repeated until the sink. We can then consider the cluster $\{N_{A-R,}..,N_A, ...,N_{A+R}\}$ as a cluster moving toward the well by regular time intervals equivalent to SDur.

17.7. Impact of the Clustering on the Throughput

In this part we will evaluate the maximum throughput at the sink for an R-redundant network. We rely on the following assumptions.

The size of the shuttle is just large enough to allow a neighboring node of the sink to empty its queue during the passage of the shuttle. This is the optimal case, because all the shuttle duration is used for transmission. The size of the FIFO is large enough to store aggregated data of the various branches of the network.

We have shown that in the 1-redundant network the distance between two token owner nodes is 4. A node receives the token each *4 * SDur* and sends an equivalent *SC* packet traffic during a time *SDur*. Since in a 1-redundant network one node is within range of the sink, then it receives a traffic packet *SC* every 4 * SDur. Therefore, the throughput at the sink is given by the formula: $D = \frac{1 * SC}{4 * SDur}$.

For a 2-redundant network, it is known that the distance between token holders is 7 hops. Thus, the sink receives *2*SC* traffic each *7 * SDur*. So, $D = \frac{2 * SC}{4 * SDur}$.

For the same reasons as those used previously in the 3-redundant network the sink receives *3*SC* traffic of nodes within range during the passage of the shuttle every *10 * SDur*. Therefore, the throughput at the sink given by: $D = \frac{3 * SC}{10 * SDur}$.

Passing to the more general case of an R-redundant network, we can say that all sink node receives (3 * R + 1) * SC traffic each *(3 * R + 1) * SDur*. So, the throughput is given by: $D = \frac{R * SC}{(3 * R + 1) * SDur}$.

According to the cluster size defined previously in section 17.7, we can say that throughput is given by: $D = \frac{R * SC}{LCSize\ (R) * SDur}$.

Let's consider the ratio $\frac{R}{LCSize(R)}$. It shows the impact of the clustering on the throughput.

The curve in Fig. 17.12 gives a representation of the impact of redundancy factor R and the cluster size on the throughput. It shows that the sequence is increasing. It starts at 0.26 and converges to 0.33. The *SC / SDur* ratio is constant and independent of the redundancy factor. Indeed, it only depends on the characteristics of sensor nodes, the needs of the application and the monitored environment. Therefore, we can say that the optimum throughput is obtained for a 3-redundant network which corresponds to a cluster size of 10. Indeed, this optimum obtained for a 3-redundant LSN is 12/10 higher than the one of a 1-redondant LSN, so this function can be considered as a clustering gain.

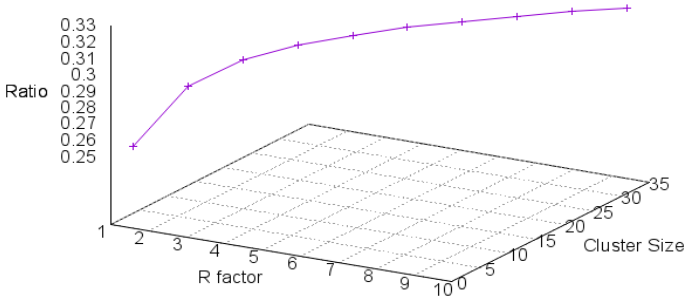


Fig. 17.12. Impact of clustering on the throughput.

17.8. Conclusion

In this chapter we presented a clustering technique of a linear network by considering the transmission range as invariant for all nodes. We have shown that the size of the cluster in the chosen propagation conditions depends on the redundancy factor. This defines the neighborhood in a linear sensor network i.e. the number of neighbors to the left or right for a given node. We have defined an R-redundant network in which a given node is in range of 2 * R nodes distributed equitably to its left and right. We have shown the impact of clustering on network performance. Network parameter such as the throughput increases with the size of the linear cluster.

However, the deduction of the size of the logical cluster based just on knowledge of the neighborhood i.e. the list of one, two and /or three hops of the token holder node does not guarantee good modelling of the medium. In the future we envisage to see the impact of a more realistic propagation model on the consistency of the assumptions of this chapter. It is to master interactions between the characteristics of a linear sensor network of sensors deployed in a medium whose propagation conditions are adequately represented by a model of Log-normal type Shadowing [25].

References

[1]. Z. Wang, X. Zhao, and X. Qian, The application and issues of linear wireless sensor networks, in *Proceedings of the International Conference on System Science, Engineering Design and Manufacturing Informatization (ICSEM'11),* 2011, Vol. 2, pp. 9–12.

[2]. I. Jawhar, N. Mohamed, and K. Shuaib, A framework for pipeline infrastructure monitoring using wireless sensor networks, in *Proceedings of the Wireless Telecommunications Symposium (WTS'07)*, 2007, pp. 1–7.

[3]. S. Yoon, W. Ye, J. Heidemann, B. Littlefield, and C. Shahabi, SWATS: Wireless sensor networks for steamflood and waterflood pipeline monitoring, *Netw. IEEE*, Vol. 25, No. 1, 2011, pp. 50–56.

[4]. T. T.-T. Lai, W.-J. Chen, K.-H. Li, P. Huang, and H.-H. Chu, Triopusnet: Automating wireless sensor network deployment and replacement in pipeline monitoring, in *Proceedings of the ACM/IEEE 11th International Conference on Information Processing in Sensor Networks (IPSN'12),* 2012, pp. 61–71.

[5]. M. Li and Y. Liu, Underground coal mine monitoring with wireless sensor networks, *ACM Trans. Sens.netw., Vol.* 5, No. 2, March 2009, pp. 1–29.

[6]. M. Li and Y. Liu, Underground Structure Monitoring with Wireless Sensor Networks, in *Proceedings of the 6th International Conference on Information Processing in Sensor Networks*, New York, NY, USA, 2007, pp. 69–78.

[7]. G. Werner-Allen, J. Johnson, M. Ruiz, J. Lees, and M. Welsh, Monitoring volcanic eruptions with a wireless sensor network, in *Proceedings of the 2nd European Workshop on Wireless Sensor Networks,* 2005, pp. 108–120.

[8]. A. A. Abbasi and M. Younis, A survey on clustering algorithms for wireless sensor networks, *Comput. Commun., Vol.* 30, No. 14, 2007, pp. 2826–2841.

[9]. D. J. Baker and A. Ephremides, The Architectural Organization of a Mobile Radio Network via a Distributed Algorithm, *IEEE Trans.Commun., Vol.* 29, No. 11, November 1981, pp. 1694–1701.

[10]. K. Xu and M. Gerla, A heterogeneous routing protocol based on a new stable clustering scheme, in *Proceedings of the IEEE Military Communications Conference (MILCOM'02), Vol.* 2, 2002, pp. 838–843.

[11]. W. R. Heinzelman, A. Chandrakasan, and H. Balakrishnan, Energy-efficient communication protocol for wireless microsensor networks, in *Proceedings of the 33rd Annual Hawaii International Conference on System Sciences, Vol.* 2, 2000, p. 10.

[12]. A. Manjeshwar and D. P. Agrawal, APTEEN: A hybrid protocol for efficient routing and comprehensive information retrieval in wireless sensor networks, in *Proceedings of the International Parallel and Distributed Processing Symposium (IPDPS'02),* 2002, p. 0195b.

[13]. L. Subramanian and R. H. Katz, An architecture for building self-configurable systems, in *Proceedings of the 1st Annual Workshop on Mobile and Ad Hoc Networking and Computing (MobiHOC)*, 2000, pp. 63–73.

[14]. M. Demirbas, A. Arora, and V. Mittal, FLOC: A fast local clustering service for wireless sensor networks, in *Proceedings of the Workshop on Dependability Issues in Wireless Ad Hoc Networks and Sensor Networks (DIWANS/DSN'04)*, 2004, pp. 1–6.

[15]. B. Krishnamachari, D. Estrin, and S. Wicker, Modelling data-centric routing in wireless sensor networks, in *Proceedings of the IEEE INFOCOM*, 2002, Vol. 2, pp. 39–44.

[16]. S. Lindsey and C. S. Raghavendra, PEGASIS: Power-efficient gathering in sensor information systems, in *Proceedings of the IEEE Aerospace Conference*, Vol. 3, 2002, pp. 3-1125 - 3-1130.

[17]. S. G. Santhi and K. Venkatachalpathy, Multiple Cluster Tree Routing and Scheduling for Collision Avoidance in 802. 15. 4. Sensor Networks, *Research Journal of Applied Sciences, Engineering and Technology,* 7, 15, January 2014, pp. 3075-3082.

[18]. S. Waharte and R. Boutaba, Performance Comparison of Distributed Frequency Assignment Algorithms for Wireless Sensor Networks, in *Proceedings of the Network Control and Engineering for QoS, Security and Mobility, III*, D. Gaïti, S. Galmés, and R. Puigjaner (Eds.), Springer US, 2005, pp. 151–162.

[19]. Y.-S. Shiu, S. Y. Chang, H.-C. Wu, S. C.-H. Huang, and H.-H. Chen, Physical layer security in wireless networks: a tutorial, *IEEE Wirel. Commun., Vol.* 18, No. 2, Apr. 2011, pp. 66–74.

[20]. M. Al-Shawaqfeh, A. Abu-El-Haija, and M. J. Abdel Rahman, Collision Avoidance Slot Allocation Scheme for Multi-cluster Wireless Sensor Networks, *Wirel Netw*, Vol. 19, No. 6, Aug. 2013, pp. 1187–1201.

[21]. G. Chakraborty, Optimum Cluster Size for Cluster-Based Communication in Wireless Sensor Network, in *Proceedings of the 4th International Conference on Mobile Ubiuitous Computing, System, Services and Technologies (UBICOMM'10)*, 2010, pp. 328-333.

[22]. E. H. M. Ndoye, F. Jacquet, M. Misson, and I. Niang, A Token-based MAC Protocol for Linear Sensor Networks, *Sens. Transducers, Vol.* 189, No. 6, 2015, pp. 31-40.

[23]. E. H. M. Ndoye, F. Jacquet, M. Misson, and I. Niang, Evaluation of RTS/CTS with unslotted CSMA/CA algorithm in linear sensor networks, *NICST*, France, 2013.

[24]. T. S. Rappaport and others, Wireless Communications: Principles and Practice, Vol. 2. *Prentice Hall PTR*, New Jersey, 1996.

[25]. A. Goldsmith, Wireless Communications, *Cambridge University Press*, 2005.

Chapter 18
Performance Study of Wireless Sensor Databases

Abderrahmen Belfkih, Bruno Sadeg, Claude Duvallet and Laurent Amanton

18.1. Introduction

In this last decade, Wireless Sensor Networks (WSNs) have been used in many application domains (military, agriculture, industry, transport), in order to monitor the environment and to prevent accidents. WSNs can be considered as a type of ad hoc wireless networks, composed of small wireless nodes, which are manually or randomly deployed in a region of interest to sense different physical characteristics of the environment, like temperature, humidity, pressure, light, etc. Sensor nodes periodically transmit sensed data to a sink node referred to as Base Station (BS). Data collection must be able to meet certain deadlines, i.e. to update data before the expiration of their validity intervals, in order to reflect the current state of the environment. WSN applications, such as industrial process monitoring and control, can be considered as time critical. They require strict deadlines to send data to the sink nodes. WSNs are often deployed without giving some importance to the temporal constraints including the data freshness constraint, and the deadline constraints [1]. Data freshness constraint imposes a limit on the data validity period from the time it is read from a sensor. The constraint of data deadline also requires that data have to be collected within a deadline [2]. In real-time applications, data freshness is a key performance factor. It is important to study the efficiency of data management mechanisms in WSNs. Nowadays, researchers are interested in data processing techniques to offer an efficient solution to increase the WSN lifetime. They have proposed new techniques for data aggregation, data collection and query processing, to simplify the data extraction and management in WSN. The query processing systems such as Cougar [3] and TinyDB [4] use the abstract database approach which considers the WSN as a relational database. They aim to improve the data extraction mechanisms from WSNs, to save energy and to reduce the communication losses [5]. In this chapter, we are investigate two approaches for data processing in WSN: (i) A query processing approach using an abstract database and (ii) A periodic data collection using a traditional database. We simulate temporal constraints in order to identify which method gives better data arrival

Abderrahmen Belfkih
University of Le Havre, BP 1123, 76063 Le Havre cedex, France

time to the user and that ensures data freshness in WSN. We consider the following performance parameters: (i) The network convergence time, which is the time that a sensor node takes to connect to the base station, to build and to update its routing tables. It can affect the arrival time of data, (ii) The data collection time, and (iii) The database response time. The data freshness can be improved, once these parameters are faster. The remainder of this chapter is structured as follows. Section 18.2 reviews sensor database approaches in WSN. In Section 18.3, we present the study of temporal aspects in wireless sensor databases. In Section 18.4, we carry out simulations, and we present and comment the results obtained. Finally, in Section 18.5, we conclude the chapter by showing how to effectively exploit the two techniques and we give some perspectives to this work.

18.2. Sensor Database Approaches in WSN

The data storage in WSN is classified into three different approaches: warehousing approach, distributed approach and abstract databases.

18.2.1. Warehousing Approach

This approach is based on a centralized system. Data are collected from sensor nodes and sent to a database via a base station. The base station receives the user queries and forwards them to the network using dissemination methods. Once the base station have received data from sensors, it stores them in a persistent repository which can be accessed at any time. This approach has some limitations [6]: (a) Access and processing of queries are separated, (b) The absence of load balancing between nodes, (c) The data transmission time increases with the density of the network and it requires a lot of energy, leading the data to miss their temporal constraints. Among the works based on warehousing approach, we find the data stream processing system proposed by Kitagawa et al. [7] for event-based continuous queries system. It's used for data obtained from multiple wireless sensor networks. Authors have solved the problem of continuous data storage in RDBMS[1] by reading data objects into memory for query processing. However, this method needs one disk access for each arrival data, which increases response time of storage. We notice that the authors have used PostgreSQL [25] as database management system in their experiments. Kanzaki et al. discussed a wireless sensor network testbed based on warehousing approach, named X-Sensor, which manages multiple wireless sensor networks on different sites [8]. It's composed of a single testbed server and multiple sensor sites where everyone contains a gateway and a sensor network. The gateway collects data from sensors and stores them in a PostgreSQL database. This proposal is useful for research that uses a large-scale sensor network.

Elias et al. have proposed a model for WSNs monitoring based on REST1 Web service and XML2 messages [9], to provide a mobile ubiquitous approach for WSNs monitoring. Sensed data are collected and stored in a database to be requested by mobile clients.

[1] Relational DataBase Management System

MySQL [26] RDBMS was used by authors in order to support a large amount of collected data and to reduce redundant and null values.

18.2.2. Distributed Approach

This approach allows the user to query the network through a base station using declarative queries and to get data stored in a database server. Data can be stored in a remote database and in the sensors themselves. Users send queries to the base station which broadcasts them to all the nodes in the network. When a sensor node receives a query, it generates responses to send them to parent nodes. The parent nodes will forward them to a base station, combining them with their own data [10]. This approach allows to exploit the computational capabilities of sensors that will be considered as a database [11], and the query workload will be determined based on the data that should be extracted from sensors. This reduces the number of data transmissions and increases the network lifetime.

In distributed approach, we distinguish two classes of data storage: data centric storage and fully distributed data storage. Data centric storage presents a new approach, based on data storage inside the WSN. It uses a geographic hash table to map an event type into a geographic location [30]. Data are collected from regions and stored in sensor nodes according to their location, type, or a specific event. Hash values are used to distinguish between data types and the corresponding storage locations [31]. In the fully distributed data storage, sensor nodes are involved in sensing and storing data in the same manner. They store locally their readings and when their local memories are filled up, they delegate storage to other nodes [31].

Among the works based on a distributed approach, we cite the paper of Matsuura et al. [29], in which the authors have discussed a new distributed mechanism for managing ubiquitous sensors. They proposed a geographical-based overlay network, where they have used SQLite RDBMS [27] to store sensing data. Becker et al. proposed a new logistic application with Wireless Sensor Networks [12]. They have used a decentralized storage data system. Each sensor node uses flash memory to write regularly sensing data. Furthermore, the gateway uses embedded database file (SQLite file) to write the acquired data. Ratnasamy et al. [32] have proposed a data-centric storage model based on a Geographic Hash Table for WSN. They have used a geographic hash table to map an event type into a geographic location [30]. The data concerning the same event are stored together, in the same node in the corresponding location of the event. This method facilitates access to the data and avoids a query flooding throughout the network. Queries having the event name can be forwarded directly to the respective node.

18.2.3. Abstract Database

Abstract databases for WSN have been cited for the first time by Bonnet et al. [13]. The authors consider the WSN as a relational database that can be queried using SQL-like queries. Abstract databases present a declarative approach to facilitate the description and query processing in WSNs. Data required by the base station represent a virtual table, in which columns represent the data requested by the user.

18.2.3.1. Cougar

The Cougar system [33] presents an infrastructure dedicated to the sensor data management. It is based on preprogrammed sensors and BS, to which the data is aggregated and stored using query processing techniques. Cougar considers the sensor network as a large distributed database where data are assimilated to a relational table. The attributes of this table are the attributes of the sensors and the values are the measures. Sensor nodes are organized into groups (clusters), and each group has a cluster head, which aggregates sensor data. Cougar defines a sensor as an object of abstract data type (ADT). The sensor database contains a single type of relation R (loc point, floor int, s sensorNode), where "loc" defines a point ADT, that stores the coordinates of the sensor, "floor" represents the floor where the sensor is located and "s" is the sensorNode which defines the Sensor ADT and which provides sensor data [13]. The sensor ADT supports different methods like "getTemp ()" to get temperature. For example, to return the temperature on the second floor every minute, we write the following query Q1:

Q1: SELECT R.s.getTemp ()
FROM R
WHERE R.floor = 2 **AND**
EVERY (60)

18.2.3.2. TinyDB

Madden et al. have presented a new query processing system, named TinyDB, to extract information from a network using TinyOS [14] as an operating system. TinyDB is composed of two subsystems: (i) TinyDB application, running on each node of the network, and (ii) Java-based client interface, allowing the user to describe the data using a declarative SQL- LIKE queries, i.e., without knowing how the data are processed in the network. The user writes the query that will be compiled and split into tasks, then injected into the network. Once a node receives the query, it processes, and returns the results to the user via the base station. For example, to calculate the average temperature for all sensors when the light is greater than 400 for a period of 10 seconds, we write the following query Q2:

Q2: SELECT AVG (temperature)
FROM sensors
WHERE light > 400
EPOCH DURATION 10 s

18.2.3.3. TikiriDB

TikiriDB is another declarative database abstraction [20] using ContikiOS [19]. It provides functionality to collect data from WSN without having any information about programming knowledge of sensor nodes. TikiriDB provides a TikiriSQL library, to receive queries from client, to parse them and to generate query packets. The results are sent to the gateway computer via the Serial Forwarder (SF) application. The node

connected to the gateway, receives the queries, then it injects them in WSN. It receives thereafter the results, which will be forwarded to the user through the serial forwarder application. Q3 is an example of TikiriDB query, which permits to get the light and the temperature from all sensors whose temperature is above 50. Each new reading will be delivered to the user every 2 seconds for a duration of 10 seconds.

Q3: SELECT light, temperature
FROM sensors
WHERE temperature > 50
SAMPLE PERIOD 2s FOR 10s

18.2.3.4. MaD-WiSe

Amato et al. have proposed MaD-WiSe (Management of Data in Wireless Sensor networks) [16], which considers a WSN as a highly distributed and dynamic database. MaD-WiSe takes account of various aspects related to database system design. Users can express queries using an SQL-like language, named MW-SQL, to collect, to filter, and to manage sensor data. MW-SQL uses the concept of the source of data (sensor id, room, etc.) to specify the data that will be retrieved from the network. It also includes temporal aggregates “EVERY” and “EPOCH” clauses. “EVERY” clause is used to specify the periodic sensor data acquiring rate and the “EPOCH” clause is used to define the time interval of data collection from WSN. Q4 is an example of MaD-WiSe query:

Q4: SELECT min (1.Light), avg (2.Temperature)
FROM 1.Light, 2.Temperature
WHERE 2.Temperature > 20
EPOCH 10 SAMPLES
EVERY 30000

In Q4, the user requests the minimum light from sensor with id=1 and the temperature average from sensor with id=2 whenever the corresponding temperature readings exceed 20°. Data will be taken every 30 seconds and aggregation will be performed every 10 samples (i.e., every 20 seconds).

18.2.3.5. Corona

Corona [17] is a distributed in-network query processing system which provides a declarative query SQL-Like language to formulate queries. Corona introduces the notion of freshness into WSN, allowing the user to obtain data from a sensor network with data freshness guarantees. It uses multiple aggregate queries to reduce processing delays and costly communication in WSN. It is able of running multiple applications simultaneously. Q5 is an example of Corona query.

Q5: SELECT temperature
FROM sensors
WHERE temperature > 50

EPOCH 60s
FRESHNESS 10s

In Q5, the user queries all sensors to get a temperature greater than 50 for an epoch period of 60 seconds. It defines also the freshness constraint clause of 10 seconds to fix the time allowed between two acquisitions.

18.2.3.6. BBQ

Amol et al. have proposed a data acquisition model named "BBQ" [18], based on probabilistic modeling techniques to optimize data acquisition from sensor network. This model is based on time varying multivariate Gaussians. User sends a SQL query translated into probabilistic computations over the model including specific parameters like error bound indicating allowing the user to specify the approximation value tolerance and confidence threshold. Q6 is an example of probabilistic query used in BBQ model:

Q6: **SELECT** nodeid, temp ± 0.1 °C, Confidence (0.95)
FROM Temperature AS T
WHERE T.temp > 30 °C

In Q6, the user requests the node identifier and the temperature values with an error bound more or less 0.1 °C, i.e. the difference between the real value of temperature and the value detected, must not exceed 0.1 °C, also with a confidence equal to 95 %, and greater than 30 °C. Based on the model and sensor readings values, the system decides the efficient way to answer the query with the requested confidence. BBQ model is updated over the time, based on new reading values which are reported as a probability density function. Thereafter, BBQ generates an observation plan and sends it over the network in order to select observations that minimize cost and maximize confidence. We note that temporal constraints have not been considered in the most of abstract databases. They have been studied only in the recent Corona work [17]. In the next section, we study of temporal aspects in wireless sensor databases through two scenarios: a periodic data collection and a query processing using TinyDB. Then, we present the different parameters selected to discuss them in Section 18.4.

18.3. Study of Temporal Aspects in Wireless Sensor Databases

In this section, we present a study of timing properties through two data processing techniques for WSN. The first is based on a warehousing approach, in which data are collected and stored on a remote database. The second is based on query processing using an abstract database like TinyDB. This study has allowed us to identify some factors which enhance the respect of temporal constraints.

18.3.1. Network Model

We have used a network model composed of four components: (1) Sensor nodes to report information about the environment; (2) A base station to collect periodically data from sensor nodes; (3) A remote database used to store the received data; (4) A user which can connect to the database to get information about WSN. In the second scenario, we have added an abstract database for WSN, named TinyDB, which allows the user to send queries to the base station and to get information.

18.3.2. First Scenario: Data Collection with Remote Database

In this scenario, sensor nodes are randomly deployed over an area and the base station is placed outside the network. Sensor nodes communicate amongst themselves through wireless communications. They retrieve data values like temperature and humidity. Then, they send them periodically to a base station.

The data received are transmitted to the remote database to be stored. Finally, they can be displayed through an interface to the users (Fig. 18.1). The remote database has been deployed in order to manage and storage the data collected from WSN. The user can query the database using SQL-like queries to get information about the WSN (cf. Fig. 18.2).

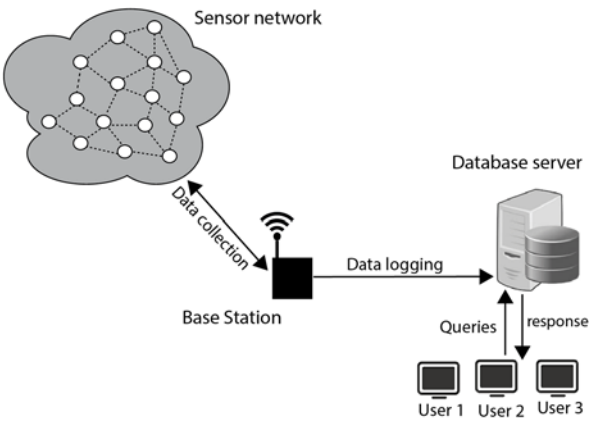


Fig. 18.1. Data collection with remote database.

18.3.3. Second Scenario: Query Processing with WSN Abstract Database

This scenario is based on a query processing technique, in which we use an abstract database system, named TinyDB. User specifies the data that will be retrieved from WSN, through SQL-like queries. It sends them to the base station via the abstract database interface. The base station injects user queries over the network. Each sensor receives a query, responds with the requested values for a duration fixed by the user. The base station sends the responses to the user via the abstract database system. Sensor data can be stored

in a remote database. We have considered some performance parameters for the two approaches and we discuss them in next section.

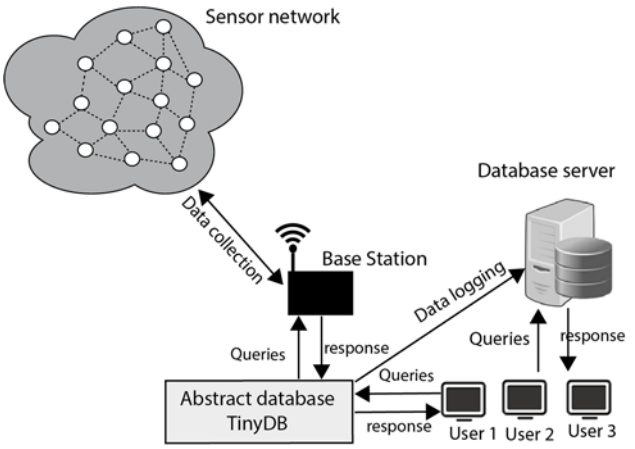


Fig. 18.2. Query Processing with WSN Abstract Database.

18.4. Simulation Environment

18.4.1. Data Collection Scenario

We have used COOJA [15], the network simulator for ContikiOS [19], to create the first scenario. We have created one sink node as base station and a set of sender nodes of type Tmote Sky with 48 KB of flash memory. The nodes are randomly distributed in an area within 100 m × 100 m and the sink is placed outside of the deployment area. The number of sensors varies from 10 to 100 with step of 10. Each node sends temperature and humidity values to the base station every 60 seconds. Table 18.1 summarizes the simulation parameters.

Table 18.1. Simulation Parameters.

Parameters	Values
Operating system	ContikiOS 2.6
Simulator	Cooja
Mote type	Tmote Sky
Simulation area (m)	100×100
Number of nodes	10-100
Number of sink Nodes	1
Source data rate	Random
Random Radio range (m)	80
Data packet size (bytes)	15

We use RPL routing protocol [21] to provide communication between sensor nodes and the base station. RPL is a routing protocol designed for low power and lossy networks and for large networks. It is more efficient for data delivery time than other existing known protocols such as LOAD (LOAD is derived from AODV), DSR (Dynamic Source Routing) and DSDV (Destination-Sequenced Distance Vector routing) [22]. It provides efficient routing paths guaranteeing data delivery before deadline in WSN. We have used ContikiMac radio duty cycling protocol [23] inspired from existing duty cycling protocols. It has shown efficient time management because it uses a wake-up mechanism to define a precise timing between packet transmissions. We have used the default ContikiMac layer CSMA (carrier sense multiple access). Table 18.2 below summarizes the different protocol layers used in simulation.

Table 18.2. Simulation layer protocol parameters.

Layer	Protocol
Transport layer	UDP
Network layer	IPV6/RPL
MAC layer	CSMA
Radio Duty Cycling	ContikiMac
Physical layer	IEEE 802.15.4

18.4.2. Query Processing Scenario

We have used TinyDB as abstract database in TinyOS. TinyDB runs on Mica2 motes with 128 KB flash memory because its code execution requires a flash memory upper than 48 KB. We have chosen to use TinyDB because it is the most widely used system by researchers and it has shown its effectiveness in maintaining energy and for data management. Sensor nodes are deployed randomly, and their number varies from 10 to 100 by step of 10. They are queried on demand by the user via the base station.

Table 18.3. Simulation Parameters.

Parameters	Values
Operating system	TinyOS 1.x
Simulator	Tossim
Mote type	Mica2
Simulation area (m)	100×100
Number of nodes	10-100
Number of sink Nodes	1
Source data rate	Random
Random Radio range (m)	80
Data packet size (bytes)	20

TinyDB uses a tree-Based Routing Protocol [24] to provide the query dissemination and the data collection from the network. Each node forwards the query to other nodes in the

network, to form the routing tree. This process ends when all nodes have received the query. TinyDB uses CSMA protocol, which is the already existing default MAC protocol used by TOSSIM simulator. It uses also the Low Power Listening (LPL) for low power communication and the Drip/Drain routing. Drip is a transport-layer component for disseminating messages throughout a network and Drain is a collection routing layer for TinyOS 1.x. Table 18.4 summarizes the different protocol layers used in simulation.

Table 18.4. Simulation layer protocol parameters.

Layer	Protocol
Transport layer	UDP
Network layer	IPV6/RPL
MAC layer	CSMA
Radio Duty Cycling	ContikiMac
Physical layer	IEEE 802.15.4

18.4.3. Simulation Description

18.4.3.1. Data Collection Scenario

Each sender node uses the UDP sender algorithm to send its data to the base station according to a transmission period. We have defined a transmission period equal to 60 seconds. We have chosen a random waiting time (SEND TIME), before sending packets to avoid that the senders sending packets at the same time. Thereafter, we have activated the digital SHT11 of sensor, which is relative to sensor Humidity and temperature. Sender node listens to its neighbors using TCP/IP handler function and continues to send data currently detected to the base station using the "send packet" function. The base station receives data from a sensor network, then it stores them in remote database using INSERT queries. Algorithm 18.1 presents the data collection process included in each sensor.

Algorithm 18.1. UDP sender process.

```
1: define period 60
2: define send time random Value
3: sensor activate (sht11 sensor)
4: while 1 do
5:       if ev = TCP/IP event then
6:             TCP/IP handler ()
7:       end if
8:       if etimer expired(period) then
9:             Etimer reset (period)
10:          Ctimer set (backoff time, send time, send packet, null)
11:      end if
12: end while
```

18.4.3.2. Query Processing Scenario

When TinyDB receives the query from the user, it parses and translates it to SQL query into an execution plan. Before injecting the query in the network, TinyDB uses a method query named Semantic Routing Trees (SRT) [24] to determine which node in the routing tree will participate and has capabilities to produce query results. SRT is composed of two phases: the first phase consists in building initial tree formation allowing parent nodes to know about the capabilities of its children. SRT disseminates a query including the type of information requested. Each node selects its parent and responds with requested information. Then, each parent receives the id and the range of values from its children. In the same way, each parent chooses its own parent until the root node, which has the information of its children.

In the second phase, the real query is disseminated to nodes having relevant responses. Once queries have been disseminated. Each node receives a query begins the query execution, which includes the filtering and the aggregation of results according to the query plan, then the data are delivered to parent nodes, thereafter, to the base station. TinyDB adjusts the result transmission rates to reduce power consumption. When the result arrives, TinyDB calls the result listener method and displays the query result for the user using a Java-based client interface.

18.4.4. Simulation Results

18.4.4.1. Impact of Number of Nodes on Network Convergence Time

The network convergence time is defined as the time needed for the sensors to connect to the base station and to build routing tables. This step comes just before the process of data collection. It can affect the data arrival time. The shorter the convergence time, the faster availability of data.

In Fig. 18.3, the network convergence time increases when the number of nodes increases. The network convergence time depends on the DODAG building process, which begins at the root (base station). The root starts the dissemination of information about the structure using DIO (DODAG Information Object) messages. The nodes, which are not connected to the tree (without communication with the root) receive the message DIO and process it. Then they decide to join or not the structure. Once the node has joined the structure, it has a route to the root of the DODAG structure. Building the final routing table depends on the number of nodes and the path cost between nodes. We note also in Fig. 18.4, that the convergence time for TinyDB, which is based on the acquisition on demand, is less than that of periodical data collection.

TinyDB constitutes a routing tree or Directed Acyclic Graph (DAG) with all sensors in which the root node (base station). TinyDB uses a selecting multi-hop tree based on efficiency of routing which depends on energy consumption. Sensors make intelligent routing decisions, where every node selects a parent based on link quality to the base station. The nodes keep both a short list of neighbors which have a recently heard

transmission, and some routing information about the connectivity of those neighbors of the rest of the network. They associate also a link quality with each of their neighbors.

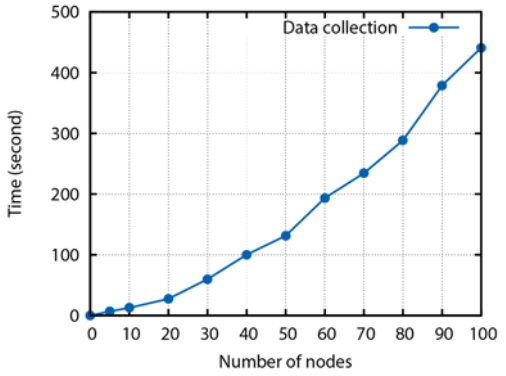


Fig. 18.3. Network convergence time for data collection.

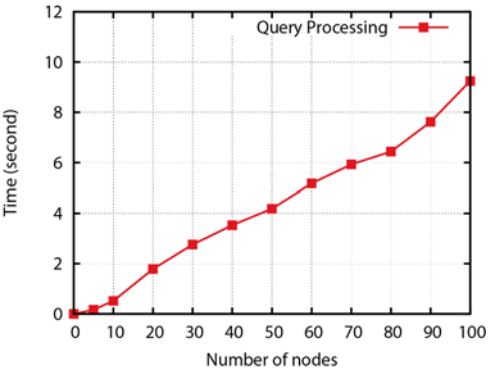


Fig. 18.4. Network convergence time for query processing system (TinyDB).

18.4.4.2. Impact of Number of Nodes on Data Collection Time

Data collection time represents the time taken to get the responses from all sensors connected to the base station. It can affect the data validity time. In result, data do not reflect the current state of the environment. The time required to send the data from sensor to the base station depends of the sensors capacity, its position, and the path quality which depends on the number of hops until the base station. Generally, the sender uses multi-hop paths to send its data when the base station is far located, which leads to transmission delay and which can causes a failure to respect data validity time.

Fig. 18.3 shows the time spent by the nodes to complete the sending of their data, using both data collection technique and query processing technique. We observe that the curve increases when the number of sensors increases for the two techniques. The time required to send the data depends of the number of hops and the availability of parent nodes to send data received from children. Their availability is not guaranteed all time because the choice of parent node depends on the node decision, which consist on joining or not the structure. We also notice in Fig. 18.3 that the time required to obtain results from all sensors using the technique of processing request is less than that of data collection. With TinyDB, the data is regularly reported and aggregated by a tree or a directed acyclic graph from the nodes to an access point network. It includes aggregation and filtering operations to maintain all routing information. The parent nodes, located near the root, put agreements with their children, on a time interval to receive data. The whole process is repeated for each period and query.

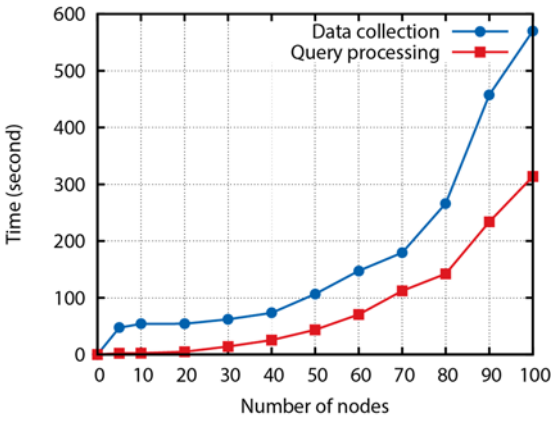


Fig. 18.5. Completed Cycle Time.

18.4.4.3. Impact of the MAC Layer Protocols

The goal of this test is to observe the Complete Cycle Time (CCT) compared to the. Duplicate Packets Sending (DPS) using the two scenarios. We have varied the number of sensors in each time.

We note in Table 18.5 that the number of packet transmissions with TinyDB is greater than periodic data collection.

Table 18.5. Duplicate sends with periodic data collection.

	Periodic data collection		Query processing (TinyDB)	
Number of nodes	**CCT(sec)**	**DPS**	**CCT(sec)**	**DPS**
5	47.430	0	2.203	0
10	53.955	0	2.434	0
20	54.157	0	4.766	21
30	62.082	0	14.060	102
40	73.302	0	25.441	121
50	106.609	0	435.43	174
60	147.357	2	705.02	294
70	179.478	1	112.261	308
80	266.357	15	142.114	405
90	266.357	78	233.737	420
100	570.100	72	313.832	633

The MAC-layer protocol can be considered as one of the causes that explains these results. TOSSIM implements CSMA based MAC protocol by default. The basic principles of this protocol are “listen before talk and contention”. Each node listens on the channel and checks if it is clear before each sending. If the channel is not clear, it waits for a randomly chosen period of time called the back-off, until the channel becomes clear. When the number of sensor nodes increases, the nodes compete for the channel, which make the

channel occupied. This competition causes a transmission delays and a back off time accumulation, which increases the time needed to collect data from all sensors. On the other hand, each sender repeatedly sends its packet, during the full wake-up, until it receives a link layer acknowledgment from the receiver (sink node), meaning that the receiver has correctly received the packet.

In a periodic data collection, the sky motes implement a radio duty cycling protocol, named ContikiMac using a wake-up mechanism to listen for packet transmissions from neighbors and to keep the transceiver turned off, in order to reduce power consumption in WSN [23]. ContikiMac timing includes the interval between each packet transmission, the interval between each CCA (Clear Channel Assessment), the time between receiving a packet and sending the acknowledgment packet, the time required for successfully detecting an acknowledgment from the receiver and the time required for a stable RSSI (Received Signal Strength Indicator), needed for a stable CCA indication. ContikiMac applies for a set of timing constraints to precise timing between transmissions which explains the decrease in duplicate packets sending.

18.4.4.4. Impact of Choosing the Database on Average Response Time

This experience consists to collect data from WSN and insert them in a remote database. We use for this purpose three DBMS[1] (PostgreSQL, MySQL, and SQLite) to evaluate queries response time, which represents the total time required to complete query execution.

PostgreSQL [25] is an object-relational database management system (ORDBMS), based on an open source research project at the University of California called POSTGRES. It supports a large part of the SQL standard and offers many advanced features as transactional integrity, complex queries support, table inheritance. New data types, functions, operators, aggregate functions, and index methods may be added by the user. MySQL [26] is an Open Source SQL database management system developed, distributed, and supported by Oracle Corporation. It can support foreign key, multi-threaded, partial stored procedures and transactional and non-transactional storage engine. It supports several storage engines that act as handlers for different table types: MyISAM is the default storage engine of MySQL, InnoDB storage has been designed for maximum performance and it supports foreign key referential-integrity constraints. SQLite [27] is a relational database management system and represents a simply scaled-down version of MySQL, PostgreSQL, and other popular database systems. It provides an embedded SQL database engine and it reads and writes directly to ordinary disk files. SQLite does not require server means and setup. It can support tables, indexes, triggers, views and the most of the query language features found in the standard SQL. SQLite transactions are fully ACID-compliant, allowing safe access from multiple processes or threads.

[1] DataBase Management System

Table 18.6 shows that SQLite database gives the best average response time for the SELECT queries, then MySQL and finally PostgreSQL. For the INSERT queries, PostgreSQL database has the less average response time compared to MySQL and SQLite. SQLite takes a lot of time to insert data, because it does not have a central server to coordinate accesses. It must close and re-open the database file, and invalidate its cache, for each transaction, which increases the response time. SQLite use a locking mechanism, when a write operation happens, the entire database must literally be locked, written to and then unlocked. It does not allow multiple concurrent writes. Whereas, PostgreSQL allows multiple transactions to proceed with inserts concurrently. MySQL is better than PostgreSQL for readings tables because it has a "Query Cache" to make queries faster. Also, it uses InnoDB table that makes users searches more efficient. SQLite is faster than MySQL because it uses a disk access to read and to write directly from the database files on disk. It not need to make a connection to server.

Table 18.6. Query response time (ms).

Query type	PostgreSQL	MySQL	SQLite
Insert queries	9.397	48.626	72.788
Select queries	0.992	0.690	0.225

18.4.4.5. Impact of the Nodes Positions and the Number of Hops on Data Collection Time

In this section, we show the performances of network by changing the position of the nodes. In all experiences, the sink node is 40 meters from each node. In the first experience, we place the sink in the center surrounded by 9 nodes. Each node sends data through single-hop transmission. The second experiment consists of groups of nodes, in which the closest nodes to the base station are selected as parents. Sensors nodes send their data through parent nodes, by making two hops. In the last experience, we have placed the nodes in the form of a tree and observed if this form will improve the data arrival time. All nodes use a multi-hop transmissions to send their data.

We can see from Table 18.7 that the completed cycle time result changes from simulation to another. Usually, the position of the sink and the position of nodes affect the data collection time. We can explain it by the number of hops used to forward data to the sink: higher is the number of hops, longer is the time for data collection.

Table 18.7. Impact of number of hops on data collection time.

Hop transmission	Completed cycle time (ms)	Average response time (ms)
Single-hop transmission	32.931	3.659
Two-Hop transmission	34.740	3.860
Multi-Hop transmission	53.528	5.947

18.4.4.6. Impact of Network Topologies on Data Collection Time

Wireless Sensor Network includes different topologies which have a greatly impact on the data collection time. In this experience, we have tested 15 sender nodes and one sink node with four network topologies (Tree, Star, Mesh, and Grid). In tree-based topology, nodes are organized by hierarchy levels, composed by children nodes, parent nodes and relay nodes, which communicate together and transmit their data to a root node (base station) via multi-hop routes. In star-based topology, the base station is in the center of the network and the other nodes are in the one-hop neighborhood of the base station. In this topology, all nodes communicate via the base station [28]. In mesh-based topology, one node communicates with multiple neighbors and the message can take several paths to achieve the destination. Also, when sensor node failed, the network reorganize itself to maintain the communication with the base station. In grid-based topology, the network is divided into square grid of size divided by user. One node is elected as the coordinator of grid and the responsible of forwarding information and transmitting data packets. Nodes send data periodically to the coordinator, which will be transmitted thereafter to the base station. The result in Table 18.8 shows the time spent to collect data from all sensors for the two technologies. We note that the tree network topology for data collection technology is faster than the other topologies. In fact, RPL forms a tree-like topology rooted at the sink, reflecting the above results. Star, mesh and grid topologies have not shown good results. RPL takes a time to define DODAG networks and to build path to the root. It cannot fully exploit the network, which affects the quality of paths.

Table 18.8. Complete cycle time (ms).

Topologies	Data collection	Query processing
Star	62.312	1.832
Mesh	56.002	1 .362
Grid	54.121	2.163
Tree	53.515	2.143

For the query processing technology, the mesh topology gives the best complete cycle time compared the other topologies. In mesh topology, all nodes cooperate in the distribution of data in the network. The same technology is used by TinyDB. The nodes make an intelligent routing decisions, where every node selects a parent based on link quality to the base station.

18.5. Conclusions

In this work, we have studied and compared the timing properties of the data collection from a sensor network using a static database versus an abstract in-network database, named TinyDB. We have evaluated temporal constraints such as data collection time, database response time, and network convergence time in both approaches. We have found, according to the tests performed, that many factors can affect the temporal constraints in a WSN. We have determined that the network topology and the routing

protocol together may play an important role on data collection time. The convergence time also has an impact on the process of data collection. We have shown clearly the timing-response advantage, using abstract database compared to remote databases to access the data. So, we can conclude that the great choice of the network topology and the routing protocol with the right approach can improve the temporal constraints in WSN. We plan to work in this direction in a future works by taking into account data temporal consistency.

References

[1]. Abderrahmen Belfkih, Bruno Sadeg, Claude Duvallet, Laurent Amanton, Study of Timing Properties on Data Collection and Query Processing Techniques in Wireless Sensor Networks, in *Proceedings of the International Conference on Sensor Networks (SENSORNETS'15),* Angers, France, 2015, pp. 77–84.

[2]. Bhattacharya, S., Achieving Application Quality of Service in Resource-constrained Wireless Sensor Networks, PhD thesis, *Washington University*, St. Louis, MO, USA, 2008.

[3]. Fung, W. F., Sun, D., and Gehrke, J., Cougar the Network is the Database, in *Proceedings of the International Conference on Management of Data*, New York, NY, USA, 2002, pp. 621–621.

[4]. Madden, S., J. Franklin, M., M. Hellerstein, J., and Hong, W., The Design of an Acquisitional Query Processor for Sensor Networks, in *Proceedings of the International Conference on Management of Data*, 2003, pp. 491-502.

[5]. García-Hernando, A.-B., Martínez-Ortega, J.-F., López-Navarro, J.-M., Prayati, A., Redondo-López, L., Problem Solving for Wireless Sensor Networks, Computer Communications and Networks, *Springer*, 2008.

[6]. Arati Manjeshwar, Agrawal, D. P., TEEN: a routing protocol for enhanced efficiency in wireless sensor networks, in *Proceedings of the International Parallel and Distributed Processing Symposium.,* 2000, pp. 2009-2015.

[7]. H. Kitagawa, Y. Watanabe, H. Kawashima, and T. Amagasa, Stream-based real world information integration framework, Chapter in Wireless Sensor Network Technologies for the Information Explosion Era, Vol. 278, *Springer Berlin Heidelberg*, 2010, pp. 173-204.

[8]. A. Kanzaki, T. Hara, Y. Ishi, N. Wakamiya, S. Shimojo, X-Sensor: A Sensor Network Testbed Integrating Multiple Networks, in *Proceeding of the International Conference on Complex, Intelligent and Software Intensive Systems, (CISIS '09),* 2009, pp. 1082-1087.

[9]. A. Elias, J. Rodrigues, L. Oliveira, and B. Zarpelao, A Ubiquitous Model for Wireless Sensor Networks Monitoring, in *Proceedings of the International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing*, 2012, pp. 835-839.

[10]. A. Agrawal, L. Gupta, B. Kr. Behera, R. Balaji, and V. Bhootna. Query processing in wireless sensor network, *Advances in Electronic and Electric Engineering (AEEE),* Vol. 3, 1, 2013, pp. 1-6.

[11]. B. Philippe and S. Praveen, Device database systems, in *Proceedings of the IEEE Computer Society International Conference on Data Engineering,* 2000, p. 194.

[12]. M. Becker, B.-L. Wenning, C. Görg, R. Jedermann, and A. TimmGiel, Logistic applications with Wireless Sensor Networks, in *Proceedings of the Workshop on Hot Topics in Embedded Networked Sensors,* 2010, pp. 1-5.

[13]. Bonnet, P., Gehrke, J., and Seshadri, P, Towards Sensor Database Systems, in *Proceedings of the International Conference on Mobile Data Management,* London, UK, 2001, pp. 3–14.

[14]. Levis, P., Madden, S., Polastre, J., Szewczyk, R., Whitehouse, K., Woo, A., Gay, D., Hill, J., Welsh, M., Brewer, E., and Culler, D., TinyOS: An Open Operating System for Wireless Sensor Networks. Ambient Intelligence, Vol. 35, *Springer Berlin Heidelberg*, pp. 115–148.

[15]. F. Osterlind, A. Dunkels, J. Eriksson, N. Finne, and T. Voigt, Cross-Level Sensor Network Simulation with COOJA, in *Proceedings of the Annual IEEE Conference on Local Computer Networks,* Tampa, Florida, USA, 2006, pp. 641- 648.

[16]. Amato, G., Chessa, S., and Vairo, C, Mad-wise: A Distributed Stream Management System for Wireless Sensor Networks, *Software: Practice and Experience*, Vol. 40, Issue 5, 2010, pp. 431–451.

[17]. Khoury, R., Dawborn, T., Gafurov, B., Pink, G., Tse, E., Tse, Q., Almi'Ani, K., Gaber, M. M., Rohm, U., and Scholz, B. Corona: Energy-Efficient Multiquery Processing in Wireless Sensor Networks, in *Proceedings of the International Conference on Database Systems for Advanced Applications,* 2010, pp. 416–419.

[18]. Amol Deshpande, Carlos Guestrin, Samuel R. Madden, Joseph M. Hellerstein, and Wei Hong, Model-Driven Data Acquisition in Sensor Networks, in *Proceedings of the International Conference on Very Large DataBases*, 2004, pp. 588-599.

[19]. Dunkels, A., Gronvall, B., Voigt, T., Contiki - a Lightweight and Flexible Operating System for Tiny Networked Sensors, in *Proceedings of the Annual IEEE Conference on Local Computer Networks,* 2004, pp. 455-462.

[20]. N. M. Laxaman, J. S. Goonethilake, Kasun de Zoysa, TikiriDB: Shared Wireless Sensor Network Database for Multi-User Data Access, in CSSL, Colombo, Sri Lanka, 2010.

[21]. Winter, T., Thubert, P., Brandt, A., Hui, J., Kelsey, R., Levis, P., Pister, K., Struik, R., Vasseur, J., and Alexander, R., RPL: IPV6 Routing Protocol for Low-Power and Lossy Networks, in *Proceedings of the Seminar of Network Architectures and Services*, 2012, pp. 59-66.

[22]. Vucinic, M., Tourancheau, B., and Duda, A., Performance Comparison of the RPL and Loadng Routing Protocols in a Home Automation Scenario, in *Proceedings of the Wireless Communications and Networking Conference,* 2013, pp. 1974–1979.

[23]. Dunkels, A., The ContikiMAC Radio Duty Cycling Protocol. Technical Report T2011:13, *Swedish Institute of Computer Science*, 2011.

[24]. Gehrke, J. and Madden, S., Query Processing in Sensor Networks, *IEEE Pervasive Computing,* 2004, pp. 46–55.

[25]. C. S. R. Prabhu, Object – Oriented Database Systems: Approaches and Architectures, *Prentice-Hall of India,* 2010.

[26]. S. K. Singh, Database Systems: Concepts, Design and Applications, Second ed., *Pearson Education,* 2011.

[27]. Sibsankar Haldar, SQLite Database System: Design and Implementation, *Motorola Mobility, Inc,* 2011.

[28]. Bulent Tavli, Wendi B. Heinzelman, Mobile Ad Hoc Networks - Energy-Efficient Real-Time Data Communications, *Springer*, 2006, pp. 1-265.

[29]. S. Matsuura, K. Fujikawa, H. Sunahara, Applying Overlay Networks to Ubiquitous Sensor Management, Chapter in Wireless Sensor Network Technologies for the Information Explosion Era, Vol. 278, *Springer Berlin Heidelberg*, 2010, pp. 231-247.

[30]. Thang Nam Le, Wei Yu, Xiaole Bai, Dong Xuan, A dynamic geographic hash table for data-centric storage in sensor networks, in *Proceedding of the Wireless Communications and Networking Conference,* 2006, Vol. 4, No., pp. 2168-2174.

[31]. Pietro Gonizzi, Gianluigi Ferrari, Jérémie Leguay and Paolo Medagliani, Distributed Data Storage and Retrieval Schemes in RPL/IPv6-Based Networks, in Wireless Sensor Networks From Theory to Applications, Chapter 19, *CRC Press*, 2013, pp. 567–594.

[32]. S. Ratnasamy, B. Karp, S. Shenker, D. Estrin, and L. Yin, Data-centric storage in sensornets with GHT, a geographic hash table, *ACM Mobile Networks and Applications,* 2003, Vol. 8, Issue 4, pp. 427–442.
[33]. Y. Yao, J. Gehrke, The Cougar approach to in-network query processing in sensor networks, *SIGMOD Record,* Vol. 31 Issue 3, 2002, pp. 9–18.

Chapter 19
Review of Resolver to Digital Converter (RDC) Soft Computing Techniques

Chandra Mohan Reddy Sivappagari and Nagabhushan Raju Konduru

19.1. Introduction

Every machine, process and monitoring system has a rotating shaft in its mechanism. One of the critical point for the motion control system is to determine the rotor angular position accurately in order to enable its control and allow phase inversions at the right time. The accuracy of the angular position influences the system efficiency, but also the torque control for optimum driving sensation. Such angle sensors need to be able to work in harsh environments, be accurate, safe and reliable. Machine tool and robotics manufacturers use resolvers and synchros to provide accurate angular and rotational information. These devices excel in demanding factory and aviation applications requiring small size, long term reliability, absolute position measurement, high accuracy, and low noise operation. Resolvers are extensively used in applications that demand instantaneous, accurate and high resolution information angular position or speed. A resolver's analog outputs have been modulated by rotor excitation signal and Resolver to Digital Converter (RDC) is always adopted to recover the angular position in digital form. RDCs are widely used in automotive and industrial applications to provide motor shaft position and/or velocity feedback. RDC performs two basic functions: demodulation of the resolver signals to remove the carrier and angle determination to provide a digital representation of the rotor angle. All RDC techniques use the two analog signals to produce a digital output. The differences between the various converter methods is in the resolution available, the speed at which the shaft can be rotated, the designed resolution and the sensitivity of the system to the unwanted distortion of the resolver signals [1]. The main drawback of RDC is its cost, which is about the same price as that of the resolver [2, 3]. Recently, researchers have paid attention on RDCs with soft computing techniques to improve the linearity, resolution and accuracy of the rotor shaft angle of the resolver.

Chandra Mohan Reddy Sivaapagari
JNTUA College of Engineering Pulivendula, J.N.T.University Anantapur, Andhra Pradesh, India

19.2. RDC Techniques

John Pezzlo and Chong LohTsiang [4] discussed the disadvantages and deficiencies to measure the angular position using contacting transducers and presented a design method to overcome the disadvantages by providing a novel non-contacting method. According to this design, the resolver is operated in a phase shift mode whereby the alternating current output voltage is constant and the phase of the output signal with respect to a reference varies in direct proportional to the shaft position of interest. Robert M. Kay [5] developed a simple structure for RDC based on phase locked loop using analog electronic hardware. This method converts a predefined frequency signal to a given waveform sequence by an electronic logic circuit, band pass filtered and phase splitted into orthogonal components. This method reduces the system components by 90 %.

George W. Miller and Larry A. Meyer [6] proposed an RDC circuit to provide the accurate control of the resolver stator winding excitation signal using digital signals. The proposed design method improves the accuracy of the digital output. Richard W. Cording and Daniel D. Morton [7] designed an electronic circuit to measure the angular position of a rotatable device that has the capability of recognizing the fault conditions of the resolver. The designed circuit provides a digital output that can be directly interfaced with a microprocessor control unit. Control logic and fault detection logic are provided in the proposed design to measure the periodic angle of a plurality of rotatable device and to provide the erroneous angle indications in a resolver respectively. Edward L. Denham and Michael J. Tuso [8] proposed the design of a resolver control system to provide the feedback information to monitor the position of a controlled machine and to measure the true monitored position continuously. The developed system compensates the output value of resolver position detection from systematic phase errors.

Peter. G. Serev and Roger M. Bogin [9] developed a microcomputer control system for sensing the shaft angle of a resolver and controlling the programmable limit switches using resolver to digital angle converter. Tadahiro Ono [10] proposed a resolver type rotational positioning arrangement system to provide an improved resolution without resorting to increasing clock pulse frequency. The design of highly accurate, high resolution multiple self-corrected synchro/resolver presented by Ross D. Wellburn and Santa Rosa [11]. This design requires minimal maintenance, very much immune to distortions and it can also be used as an independent positioning device. James A. Blackburn, et. al. [12] experimented on a driven damped pendulum to observe the range of dynamical modes and presented a state diagram for the system. The pendulum coordinate was measured with an angular resolver, in combination with an integrated RDC with fourteen bit precision. The presented experimental results emphasized that this constitutes a direct and accurate measurement of chaos in a real physical system.

Peter Serev [13] developed a microcontroller based RDC with synchronous sample and hold demodulator. This method optimizes the time taken for the sample and hold circuit and minimizes the quadrature and even harmonic effects in the resolver output. The shaft angle of the resolver was extracted using digital signal processing techniques. Duane Hanselman [14-16] has analyzed the effects of the most common non-ideal resolver signal characteristics on the position accuracy using RDC.

The proposed RDC is based on tracking algorithm. The expressions for angular position accuracy due to inductive harmonics, inductive DC component feed through, quadrature error, amplitude imbalance, and imperfection quadrature component rejection were developed. Various methods for reducing the position error caused by the existence of non-ideal resolver signal characteristics are presented by Duane Hanselman [16]. Yasuhiro Ezuka [17] described that an RDC with high accuracy can be achieved if tracking type signal processing technique is employed. Different techniques to reduce the position error caused by existence of resolver signal characteristics are discussed by Duane Hanselman [18]. This paper introduced a signal processing technique to eliminate the position error particularly due to resolver quadrature imperfection. The proposed method eliminates the quadrature error by simple algebraic manipulation of resolver signals. The paper also showed that all the even harmonics in the resolver signals can be cancelled if the resolver was constructed with complementary phases. Choong Hyuk Yim, et. al. [19] proposed a fast tracking method for RDC with a bang-bang type phase comparator. To reject carrier signal and noise, the proposed method replaces the low pass filter with two pre filters outside the RDC. Bruce N. Eyerly and Donald R. Cargille [20] invented a phase compensation circuit to compensate resolver angle measurement error without the use of special compensation windings on the resolver. The proposed invention provides increased accuracy and reduced circuit complexity.

A. Murray and W. D. Li [21] designed RDC using TMS320C14 digital signal processor to provide absolute angular position and velocity information for digital servo control systems. The converter algorithm and the error calculation hardware were also developed. The demodulation and error calculation functions were performed in hardware to optimize the low speed performance of the system. Donald K. Taylor, et. al. [22] invented a device and method for recording the position reached by a moving part moved by a rotary shaft in a resolver and tracks the position of a group of servomotors during power down conditions. The proposed method provides for avoiding false motion conditions in detecting motion. According to the method, when a failure in the external power supply occurs, the proposed system was activated automatically without any loss of resolver position information. Dong Il Kim and Jin Won Lee [23] developed fuzzy based control algorithm to estimate the absolute rotor position of the permanent magnet AC servo motor with an incremental encoder coupled to the shaft. The proposed algorithm also enables the servo motor with incremental encoder always controlled with maximum generated torque per ampere of stator current without pulsation. Dean C. Alhorn [24] designed digital IC based multi-phase resolver converter used with angular resolver system. The alternative versions that employ incremental or absolute encoders were also discussed. David T. Robinson [25] proposed a low cost design method for inductive transducer to measure absolute position of ac and brushless DC servomotors as well as flux vector control of ac induction motors.

L. Harnefors [26] presented a Kalman filter method for estimation of the rotor speed of an AC motor. This design was based on trigonometric operation and was implemented on a digital signal processor. The performance of the system was validated both in simulation and hardware. Saso P. Vlahu [27] proposed a direct RDC based on tangent algorithm to obtain the angular position of the shaft. The tangent values were converted to linear values

using an eight bit look up table. Martin Piedl, et. al. [28] designed digital signal processor based a low cost resolver system to provide close loop motor velocity and position. The paper also discussed the ability to use the system to generate closed loop motor position and/or velocity control of both brushless motors and motors which use commutating brushes. Sheng Ming Yang and Shuenn Jenn Ke [29] presented accurate velocity estimation method for servo motor drives. The performance of the proposed method was investigated by both analysis and experiments and at normal and very low speeds. The quantization error in the velocity feedback signal can be reduced when a closed loop observer is used instead of a backward difference to estimate the motor velocity.

Lennart Harnefors and Hans Peter Nee [30] designed and analyzed an efficient speed and position algorithm applicable to ac motor drives. The algorithm has dynamics corresponding to a phase locked loop and ideally stable. TMS320F240 DSP solution for obtaining the angular position and speed of a resolver was presented by Martin Staebler [31]. This method utilized the under sampling and an inverse tangent algorithm to decode the resolver output signals and oversampling technique was added to achieve high angular resolution. George Ellis and Jens OhnoKrah [32] addressed the tracking loop technique used to measure the position of the rotor shaft of a resolver which causes phase lag between the actual and measured positions. This paper also discussed the problem of phase lag that causes instability in the control loop and reduces the performance of the servo system. The paper concludes the methods to reduce the problems with mechanical resonance and improvement of the dynamic stiffness of the control system. The design of surface micro-machined rotations sensor for angular position detection was presented by Winston Sun and Wen J. Li [33]. The sensor was designed to detect the angular position of a rotating element by measuring the resistance change due to stress induced by centrifugal force on the seismic mass using piezo-resistive effects. A wireless transmission scheme for the rotation sensing system was also evaluated.

Sung Jun Park, et. al. [34, 38] proposed a low cost linear encoder suitable for Switched Reluctance Motor (SRM) and also presented a control algorithm to generate switching signals using a simple digital logic. Full implementation of a low cost RDC on a combined analog and digital board was presented by Attaianese, et. al. [35]. The experimental results were presented and a comparison with the performances achieved by means of an incremental encoder. George Ellis and Jens OhnoKrah [36] outlined the procedure to reduce the phase lag caused by sensors using observers and RDCs. This paper explained the advantages of RDCs over observers that include providing position and velocity feedback with little or no phase lag, and estimations of motor acceleration and torque disturbance. Aengus Murray, et. al. [37] introduced the design of a high resolution position sensing system, variable reluctance resolver and RDC integrated circuit. The new resolver position sensing system addressed both the cost issues and reliability issues associated with safety automotive applications. The fault detection systems meet the requirements of safety auto motive systems.

The design of a low cost software based tracking RDC was proposed by A.O.Di Tommaso and Miceli [39]. The comparison between the proposed RDC and a commercial encoder was also presented. The software approach makes no parameter variations due to

component drifts, temperature variations, etc. and the output signals result quite immune to noise and to external electromagnetic disturbances. Mohieddine Benammar, et. al. [40] described a new scheme for the measurement of mechanical angle of a RDC. The proposed converter produces an output voltage proportional to the shaft angle by using a linear technique. The converter was implemented using analog circuitry. A design method was proposed by Andreas Bunte and Stephan Beineke [41] to suppress systematic errors of resolvers and optical encoders with sinusoidal line signals. Though the proposed method does not require any additional hardware, the dynamics of the speed control was not affected and it will not cause any time delay. A fundamental impact of the speed measurement system and the dynamic behaviour of modern servo controlled drives are also presented. N. Nowlin, et. al. [42] designed a radiation hardened high precision RDC. The designed RDC has a maximum of 16 bits precision and was manufactured in a total dose hardened 0.6μm CMOS process. Single event latch up and dose rate latch up hardening were designed using guard rings and DICE latches.

L. Z. Sun, et. al. [43] presented a new structure of variable reluctance resolver for integration with motor systems. The proposed resolver directly utilized the salient pole effects with only simple winding patterns on the stator. Based on the theoretical analysis, it is concluded that the errors were mainly caused by the third harmonics and non-effective EMFs existing in the signals. A RDC system was designed that is capable of outputting a good resolver signals without being affected by the motor speed and switching noise was presented by A. Balkovoy and E. Kallenbach [44]. The proposed design was validated through experimental setup. The processed resolver data was compared with the incremental encoder data to estimate the accuracy of the position measurement. Don Payne [45] presented and described the analysis method of accurately measuring a reference position at high rpm engine. The paper also explained the construction and operation of a measurement system to enable convenient determination of angle position. An RDC was described for the linearization of the sine and cosine signals to enable the angle to be determined using simple linear equations was presented by Mohd. A. Avlamadi, et. al. [46]. The converter was implemented using analog electronic circuitry and the practical performance of the converter was also evaluated using PC based testing.

Victor D. Aksenenko and Sergey I. Matveyev [47] suggested and discussed two approaches for self-calibration of digital angle sensors based on integration of two conversion channels with the errors. The methods for determining the rotary angle orientation of a motor using resolver signal were described by Robert Herb [48]. This proposed method utilized a single control system that is arranged both for triggering and for evaluating the resolver signal. A solution for obtaining the estimations of actual angle and speed of the resolver was described by Freescale semiconductors [49]. A theoretical analysis and proposal of the RDC hardware interface and a design of the device software driver were also explained. Mohieddine Benammar, et. al. [50] presented the design of RDC that provides pseudo linear voltage proportional to the shaft angle. The proposed converter was based on the concept of absolute values of the resolver demodulated signal together with a dedicated linearization technique. The converter was implemented using an analogue circuitry.

Gabriel Gross, et. al. [51] implemented an accurate and fast tracking all digital RDC using oversampling and frequency shifting technique along with synchronous rotating reference frame based phase locked loop. The frequency shifting technique was used to demodulate the incoming signals. The input signals were oversampled 32 times to increase their resolution. The proposed system was implemented in a 16 bit digital signal processor. Masayuki Katakura, et. al. [52] has developed a 12-bit RDC with LSI complex twin PLL architecture that tracks a mechanical angle offset by the carrier frequency. The proposed architecture works based on analog oriented mixed signal processing technique. Lizhi Sun, et. al. [53] presented the realization of a new variable reluctance resolver by varying air gap reluctance in certain waveform. The proposed resolver was one type of rotor position sensor for the inverter driven motors. This resolver was designed into compact size and was suitably integrated into Permanent Magnet Synchronous Motor (PMSM) or brushless motors.

A novel hybrid design method for angle tracking observer with a combination of closed loop LTI observer and quadrature encoder was introduced by Reza Hoseinnezhad and Peter Harding [54] and Reza Hoseinnezhad [55]. Finite gain stability of this hybrid design was proved based on circle theorem and the simulation studies comprised two cases where an LTI-ATO and an extended Kalman filter were unstable due to high acceleration and speed. The proposed observer was stable with finite tracking errors. Yoshi Ishizuka et al. [56] presented a method related to compensating resolver detected position and also developed a system that can compensate the dynamic errors that vary with a change in rotational speed of the resolver rotor. According to the design, even when the resolver rotor speed becomes high, the resolver position detection accuracy can be enhanced. Based on the use of pseudo linearity of sinusoidal signals around zero crossing, a novel low cost design technique for RDC with basic analog electronics was presented by Mohieddine Benammar, et. al. [57].

Armando Bellini and Stefano Bifaretti [58] explained the necessity of using the filters at lower speeds in order to remove the noise. The paper also proposed a phase locked loop based steady state linear Kalman filter to obtain the filtered speed signal starting from the signals supplied by an electromagnetic resolver. The proposed Kalman filter was based on third order linear time invariant model. Jens Onno Krah, et. al. [59] described a new FPGA based method to convert an analog resolver signals to a digital position signal using delta sigma ADC technology. It is possible to increase the resolution by two bits with second order delta sigma modulator compared to sampling converters. Kamel Bouallaga, et. al. [60] listed the advantages and drawbacks of demodulation methods for resolver signals. They proposed an ATO based algorithm and realized the same by Fusion Field Programmable Gate Array. Jin Woo Ahn, et. al. [61] developed a low cost analog encoder with a proper control method suitable for SRM. The proposed encoder uses a simple structure with an optical analog gradation for high performance of rotor position.

Weera Kaewjinda and Mongkol Konghirun [62] focused on the detection of rotor position of PMSM by the resolver sensor. A resolver algorithm was proposed and implemented in the vector control drive system of PMSM. The algorithm was verified by both simulation and experiment using MATLAB®/Simulink and the TMS320F2812 based digital signal

processor respectively. Konstantin Veselinov Dimitrov [63] presented a 3-D silicon Hall effect sensor for precision angular position measurement over 360^0 rotation. The z-axis sensor was introduced to compensate the misalignment of a magnet above the sensor. The angle measurements at this moment were almost performed with the 2-D Hall sensors. Reza Hoseinnezhad, et. al. [64] proposed a new technique to develop the resolver parameters for real time tracking with varying speed and long resting periods. A new recursive and adaptive estimator was designed to track the parameters of characteristic ellipse. The proposed technique is a modified version of recursive weighted least square estimator.

Andrzej Michlski, et. al. [65] created a model of the magnetic circuit for high resolution, multi-pole and two-speed resolvers. They performed an analysis for the influence of manufacturing errors on the resolver accuracy. The created model concludes that high-precision resolver with electrical error well below one minute of arc was possible if an appropriate magnetic material was used and high precision manufacturing was assured. M. Benammar [66] described a novel converter for linearizing the sine and cosine signals of an angle of a resolver sensor. The converter was implemented using ordinary analog electronic components. The theoretical and simulation performances of the proposed converter were also explained. S. K. Kaul, et. al. [67] presented a software based error compensation method for improving the accuracy of low cost resolver based 16 bit encoders. The error profiles of ten encoders were calibrated repeatedly on a high precision rotary table and suggested that error profiles are unique and predominantly systematic in nature for a particular resolver decoder combination. These errors can be corrected by using an appropriate compensation procedure. Non-ideal characteristics of a resolver such as amplitude imbalance, quadrature error, inductive harmonics and excitation signal distortion are also discussed.

Ciro Attainanese and Giuseppe Tomasso [68] proposed the design and implementation of a low cost fully integrated board for RDC based on dedicated analog and digital assembly. This designed board was tested for different resolver speeds. Douglas W. Brown, et al. [69] proposed a new design for a real time fault detection and accommodation routine for a resolver position sensor. The identified fault detection and accommodation routines were evaluated using Simulink model of an electro mechanical actuator. The proposed design can be applied to already existing Commercial Off–The- Shelf (COTS) resolver sensors without any internal hardware modifications or additional sensors. Lizhi Sun [70] presented a review of various variable reluctance resolver structures and proposed a new variable reluctance resolver structure that is capable of outputting an absolute position signal of the same electrical period as the inverter driven motors. The paper concluded that the position errors mainly come from the odd time harmonics and the null voltages in the signals and proposed several improvement methods including the number of stator poles, changing the stator tooth shape to the salient one and adopting a sinusoidally distributed winding pattern.

Lazhar Ben-Brahim, et. al. [71, 72] developed a new low cost feed forward technique based RDC without look up tables. The proposed method was implemented using low cost analog electronic components and it has the advantage of robustness to amplitude

fluctuations of the resolver excitation. Santanu Sarma, et. al. [73] proposed a cost effective software based method for RDC using digital signal processor. The proposed method incorporated software generation of resolver carrier using a digital filter in such a way that there was a substantial savings on costly carrier oscillators and amplitude demodulators. A mathematical model based d-q axis theory and dynamic performance characteristics of brushless resolvers were discussed by D. Arab- Khaburi, et. al. [74]. The impact of rotor eccentricity on the accuracy of position in precise applications was investigated. The proposed model takes the stator currents of brushless resolver into account and the model was used to compute the dynamic and steady state equivalent circuit of resolvers. A complete software implementable scheme for position and speed sensing using a DSP based RDC was presented by S. Sarma, et. al. [75]. The amplitude demodulators and the measurement of position and speed do not cause any time delay and also the dynamics of the proposed system was not affected. The correct functioning and outstanding performance of the proposed RDC was shown both by simulation and experimental results.

The optical encoders do not provide robustness comparable to electrical motors and resolver provides better mechanical robustness but their resolution was not sufficient for good speed control behavior. So, the capacitive encoders are an attempt to develop to combine good robustness with higher resolution [76]. Analog Devices [77] has proposed a monolithic RDC IC AD2S1210 which completes a 10 bit to 16 bit resolution tracking. A type II servo loop was employed in the proposed AD2S1210 to track the inputs and convert the input sine and cosine information into a digital representation of the input angle and velocity. Lazhar Ben-Brahimet al. [78] developed feed forward technique based a new low cost RDC to convert the amplitude of sin and cosine resolver output signals into a measure of the input angle without using look up tables. The proposed design method offered the advantage of robustness to amplitude fluctuations of the resolver excitation signal. Seon Hwan Hwang, et. al. [79–81] proposed a new compensation algorithm to reduce rotor position errors between the resolver output signals caused by amplitude imbalance in vector control drive system of PMSM. The presented method does not require any additional hardware and reduces computation time with simple integral operation according to rotor position.

Zhang Haixial and Yanlan [82] described the working principle, hardware configuration and software design of synchro RDC based on PXI bus. The digital converter of synchro RDC converts analog angle signals into digital signals then connected to computer through PXI. Thus the intelligence of the instrument was realized. Zhuangzhi Han, et. al. [83] proposed tangent algorithm based RDC to digitize the angle information of the two speed resolvers. The proposed algorithm is simpler than traditional tracking algorithm. An analog shaping network for the linearization of resolver output signals and linear determination of rotor shaft angle was proposed by Mohieddine Benammar, et. al. [84] and is based on tangent/cotangent technique. The optimal breakpoint positions of the shaping network were determined experimentally and LabVIEW based setup in order to minimize the non-linearity of the converter output.

V. K. Dhar, et. al. [85] developed an Artificial Neural Network (ANN) based error compensation method to improve the accuracy of low cost resolver based 16-bit encoders for their respective systematic error profiles. The method allows to use the existing resolvers at an accuracy which is within the limits of the encoder resolution. The proposed method was implemented for four encoders by training the ANN with their respective error profiles. Ilpakurty Ravi and K. Nagabhushan Raju [86] discussed about interfacing of a resolver based motor to a servo drive with an incremental encoder interface and developed the hardware to convert resolver interface to incremental encoder interface based on AD2S80 and microcontroller 8051F310. The hardware interface was tested with permanent magnet synchronous motor at speeds up to 1000 rpm. A new technique for angular position and speed sensing of an imperfect resolver angular sensor was presented by Santanu Sarma and A. Venkateswaralu [87]. The proposed method was compared with PLL based RDC. This design provides accurate estimation of the imperfect phase quadrature and magnitudes. The correct functioning and performance of the proposed RDC has explained both by simulation and experimental results.

Mohieddine Benammar, et. al. [88] described an analog converter for the linearization of sine and cosine signals and linear computation of mechanical shaft angle of a resolver. The designed converter was based on a simple two breakpoint shaping network used as linearization scheme. Nicolas Javahiraly, et. al. [89] proposed a new optical fiber angular position sensor connected to an automotive power steering column. The developed sensor was based on the coupling between the guided modes and the radiated modes of the fiber during the light transmission. Multimode step index fiber was used for the design and the sensor allows the measurement of angular position of a car steering wheel over a large range. Sinusoidal amplitude detector based demodulator for resolver converters was presented by Anucha Kaewpoonsuk, et. al. [90]. The designed circuit produces two output signals proportional to sine and cosine envelopes of resolver shaft angle without low pass filter.

A single CMOS type II tracking resolver to digital converter RDC5028 was designed by Aeroflex [91]. This monolithic chip was implemented using precision analog circuitry and digital logic.

H. Hwang, et. al. [92] proposed a new compensation algorithm for the gain and offset errors of the sinusoidal encoder signals. The effectiveness of the proposed method was verified experimentally. Lazhar Ben-Brahim and Mohieddine Benammar [93] presented a low cost closed loop design method for resolver to analog conversion based on classical phase locked loop. The proposed design was implemented without the use of voltage controlled oscillator, digital to analog converter, counter and look up tables. Zhu Yi et al. [94] developed a method to improve the software approach of using the resolver to digital conversion. In the original approach, the samples were taken at positive peak values of excitation signal which increases system complexity. In the proposed method, the sample information was taken at other positions in an excitation period.

Ralph Kennel [95] proposed a scheme for encoder-less control of synchronous machines with permanent magnets. The proposed scheme has no limitations with respect to a minimal speed and the drive was able to provide full torque in encoder-less operation even

at stand still. Cheon Soo Park [96] presented a method and apparatus to minimize the magnetic interference in a variable reluctance resolver. The proposed apparatus comprised a source generation unit for generating a uni-phase source signal to excite a resolver. Joan Bergaset al. [97] implemented high accuracy all digital RDC. The basic components of the conventional tracking RDC was implemented in software by using frequency shifting technique and a decoupled synchronous reference frame based phase locked loop. Zhu Ming, et. al. [98] presented a design method for RDC in frequency domain. The proposed design method was based on transforming the complex signal into frequency domain and the components at the carrier frequency are used to calculate the angular position of the resolver.

Jiebin Zhang, et. al. [99] introduced the operating principle, circuit design, algorithm structure, and feasibility analysis of a high precision shaft angle acquisition system used in the solar panel. The system was implemented with absolute round in due to syn as angle sensor, the direct digital synthesizer, RDC and the AVR microcontroller. H. Loge and L. Angerpointner [100] explained the overview of the sources of angular errors and how they will be affect the resolver signals. There are a couple of mechanical, magnetic and electrical reasons that causes perceptible distortions of the resolver signals with results in noise and harmonic ripple on the angular information as well as derived velocity information. A trained artificial neural network algorithm was proposed to replace the demodulation of RDC by Prerna Gaur, et. al. [101]. The proposed resolver algorithm was implemented in the current controlled drive system of PMSM.

Qi Xun Zhou [102] analyzed the working principle and a couple of typical fault of sine-cosine resolver. High reliable sine signal generator for the sine-cosine resolver was designed and the RDC circuit with its mathematical model was also presented. Software based Mallat algorithm was proposed for the fault diagnosis. Ruijie Zhao, et. al. [103] described the use and working principle of resolver and explained the decoding arithmetic of resolve to digital converter. The decoding software arithmetic method was based on angle tracking observer. The proposed method has the smooth ability but also can track the motor rotor position and rotor speed in the same time compared to the inverse trigonometric function method. Anna K. S. Baasch, et. al. [104] presented digital Finite Impulse Response (FIR) filter methods to measure the speed and position of a RDC. The proposed algorithm was implemented in a fixed point digital signal processor based control.

Kazuya Sakai [105] invented a rotational angle sensor that detects the rotational angle with a high degree of accuracy and in which the number of magnetic poles may be flexibly changed. The invented system has rotational angle sensor, a motor, a rotational angle detector and an electric power steering system. Davood Arab Khaburi [106] proposed a modified software based angle tracking observer method for RDC to extract the rotor angle in high speeds as well as in low speeds. The proposed estimated algorithm was based on the sign and absolute values of sine and cosine of the rotor angle. Demodulation of resolver signals using synchronous demodulator is proposed by Idkhajine, et. al. [107, 108]. A fully integrated Field Programmable Gate Array (FPGA) board is used for a synchronous motor drive. One sole ADC is used for performing sampling and ADC.

A DSP-based controller for PMSM servo drives using a resolver is presented by Kaewjinda and M. Konghirun [109]. The resolver is excited by a separate signal source at 1 kHz. The resolver outputs and excitation signal are sampled with a sampling frequency of a 10-kHz. The rotor shaft position and speed are computed by using closed loop angle tracking algorithm. As the frequency of the excitation is low, this method is limited to low speeds.

The design of RDC converter without a processor or LUT was reported by A. Kaewpoonsuk, et. al. [110]. This RDC employs the OTA-based inverse- sine function circuit to generate angular output. A steady-state linear Kalman filter-based PLL is proposed by A. Bellini, et. al. [111, 112], to obtain velocity information by reducing noise from the derivative operation. The Kalman filter has the expected angular acceleration of the shaft, which is not always available, as input. Moreover, the Kalman filter has the disadvantage that the gain vector for correcting the predicted state, which plays an important role in the dynamic characteristics of the speed control loop, includes a trial-and error selection procedure, making this technique difficult to implement.

19.3. Conclusions

This Chapter provides various design techniques available in the literature using analog, digital and hybrid circuitry for measurement of resolver rotor shaft angle. From the literature, two basic methods are used for the measurement of angle, namely inverse tangent method and angle tracking observer or Type-II tracking loop method. Inverse tangent method is an open loop method so the design of RDC using this technique is simpler but it gives poor angle tracking performance. Whereas angle tracking observer is a closed loop method, tracks the resolver angle every time and gives good angle performance. The researchers have paid more attention to reduce the complexity in the hardware by implementing software based excitation source to the resolver. It is also proved that the software based RDCs improve the linearity, accuracy, resolution of the angular position.

References

[1]. Jens Onno Krah, Heiko Schmirgel and Marcel Albers, FPGA based resolver to digital converter using delta-sigma technology, *in Proceedings of the International PCIM Europe Conference,* Europe, 2006, pp. 931-936.

[2]. Davood Arab Khaburi, Software based resolver-to-digital converter for DSP-based drives using an improved angle-tracking observer, *IEEE Transactions on Instrumentation and Measurement,* Vol. 61, No. 4, April 2012, pp. 922-929.

[3]. Ciro Attaianese and Giuseppe Tomasso, Position measurement in industrial drives by means of low-cost resolver-to-digital converter, *IEEE Transactions on Instrumentation and Measurement*, Vol. 56, No. 6, December 2007, pp. 2155-2159.

[4]. John Pezzlo and Chong-Loh Tsiang, Resolver to pulse width converter, *U. S. Patent* 3,803,567, April 9, 1974, USA.

[5]. Robert M. Kay, Phase locked loop resolver to digital converter, *U. S. Patent* 4,010,463, March 1, 1977, USA.

[6]. George W. Miller and Larry A. Meyer, Resolver to digital converter, *U. S. Patent* 3,990,062, November 2, 1976, USA.
[7]. Richard W. Cording and Daniel D. Morton, Resolver processor with error detection, *U. S. Patent* 4,355,305, October, 19, 1982, USA.
[8]. Edward l. Denham and Michael J. Tuso, Compensated resolver feedback, *U. S. Patent 4,472,669*, September 18, 1984, USA.
[9]. Peter G. Serev and Roger M. Bogin, Programmable limit switch system using a resolver-to-digital angle converter, *U. S. Patent 4,511,884*, April 16, 1985, USA.
[10]. Tadahiro Ono, Resolver-type rotational positioning arrangement, *U. S. Patent 4,529,922*, July 16, 1985, USA.
[11]. Ross D. Wellburn and Santa Rosa, Self-corrected synchro/resolver, *U. S. Patent 4,568,865*, February 4, 1986, USA.
[12]. James A. Blackburn, Yang Zhou-Jing, S. Vik, H. J. T. Smith And M. A. H. Nerenberg, Experimental study of chaos in a driven pendulum, *Physica*, Vol. 26d, 1987, pp. 385-395.
[13]. Peter G. Serev, Microcontroller based resolver-to-digital converter, *U. S. Patent* 4,989,001, January 29, 1991, USA.
[14]. Duane Hanselman, Resolver signal requirements for high accuracy resolver-to-digital conversion, in *Proceedings of the 15th IEEE Annual Conference on Industrial Electronics Society (IECON'89),* 1989, pp. 486-493.
[15]. Duane Hanselman, Resolver signal requirements for high accuracy resolver-to-digital conversion, *IEEE Transactions on Industrial Electronics*, Vol. 31, No. 6, December 1990, pp. 556-561.
[16]. Duane Hanselman, Signal processing techniques for improved resolver-to-digital conversion accuracy, in *Proceedings of the 16th IEEE Annual Conference on Industrial Electronics Society (IECON'90),* 1990, pp. 6-10.
[17]. Yasuhiro Ezuka, Resolver system, *U. S. Patent 5,189,353*, February 23, 1993, USA.
[18]. Duane C. Hanselman, Techniques for Improving Resolver-to-Digital Conversion Accuracy, *IEEE Transactions on Industrial Electronics*, Vol. 38, No. 6, December 1991, pp. 501-504.
[19]. Choong-Hyuk Yim, In-Joong Ha and Myoung-Sam Ko, A resolver-to-digital conversion method for fast tracking, *IEEE Transactions on Industrial Electronics*, Vol. 39, No. 5, 1992, pp. 369–378.
[20]. Bruce N. Eyerly and Donald R. Cargille, Phase compensation for electromagnetic resolvers, *U. S. Patent* 5,134,397, July 28, 1992.
[21]. B. A. Murray and W. D. Li, A digital tracking r/d converter with hardware error calculation using a TMS32OC14, *The European Power Electronics Association*, 1993, pp. 472-477.
[22]. Donald K. Taylor, Richard J. Maczka, Russell III and Carl H, Software controllable circuit for resolver excitation switching in a motion control system, *U. S. Patent 5,198,739*, March 30, 1993.
[23]. Dong-Il Kim and Jin-Won Lee, Commutation of permanent magnet a. c. servo motors with incremental encoders via fuzzy reasoning, *Mechatronics*, Vol. 4, No. 5, 1994, pp. 455-469.
[24]. Dean C. Alhorn and David E. Howard, Multi-speed multi-phase resolver converter, *U. S. Patent 5,451,945*, September 19, 1995.
[25]. David T. Robinson, A new absolute inductive transducer for brushless servomotors, in *Proceedings of the International Intelligent Motion Conference,* in Long Beach CA, September, 1995, pp. 188-194.
[26]. L. Harnefors, Speed estimation from noisy resolver signals, in *Proceedings of the Sixth International Conference on Power Electronics and Variable Speed Drives*, publication No. 42, September 23-25, 1996, pp. 279-282.
[27]. Saso P. Vlahu, Direct resolver to digital converter, *U. S. Patent 5,912,638*, June 15, 1999, USA.

[28]. Martin Piedl, Moe Barani and Ron Flanary, Low cost resolver system, *U. S. Patent 6,084,376*, July 4, 2000, USA.

[29]. S. Buchner, L. Tran, J. Mann, T. Turflinger, D. McMorrow, A. Campbell and C. Dozier, Single-event effects in resolver-to-digital converters, *IEEE Transactions on Nuclear Science*, Vol. 46, No. 6, pp. 1445-1452, December, 1999.

[30]. Sheng-Ming Yang and Shuenn-Jenn Ke, Performance evaluation of a velocity observer for accurate velocity estimation of servo motor drives, *IEEE Transactions on Industry Applications,* Vol. 36, No. 1, January/February, 2000, pp. 98-104.

[31]. Lennart Harnefors and Hans-Peter Nee, A general algorithm for speed and position estimation of AC motors, *IEEE Transactions on Industrial Electronics*, Vol. 47, No. 1, February 2000, pp. 77-83.

[32]. Martin Staebler, TMS320F240 DSP solution for obtaining resolver angular position and speed, *Texas Instruments DSP Application Report*, SPRA605, February 2000, pp. 1-22.

[33]. Winston Sun and Wen J. Li, A MEMS high-speed angular-position sensing system with RF wireless transmission, in *Proceedings of the SPIE*, Vol. 4334, 2001, pp. 244-251.

[34]. Sung-Jun Park, Jin- Woo Ahn, Man-Hyung Lee, and T. A. Lipo, Novel encoder for SRM drive with high resolution angle control, in *Proceedings of the IEEE International Conference ISIE'01*, Pusan, KOREA, 2001, pp. 1781-1785.

[35]. C. Attaianese, G. Tomasso and D. DeBonis, A low cost resolver-to-digital converter**,** in *Proceedings of the IEEE International Electrical Machine Drives Conference,* Cambridge, MA, June 2001, pp. 917–921.

[36]. George Ellis and Jens Ohno Krah, Observer-based resolver conversion in industrial servo systems, in *Proceedings of the Power Converter and Intelligent Motion Conference PCIM'01*, Nuremberg, Germany, June 19-21, 2001.

[37]. Aengus Murray, Bruce Hare and Akihiro Hirao, 42.3: Resolver position sensing system with integrated fault detection for automotive applications, in *Proceedings of the IEEE Sensors Conference*, 2002, pp. 864-869.

[38]. Sung-Jun Park and Jin-Woo Ahn, A novel encoder for switching angle control of SRM, in *Proceedings of the IEEE International Electric Machines and Drives Conference, (IEMDC'03)*, Vol. 3, 2003, pp. 1726-1731.

[39]. A. O. Di Tommaso and R. Miceli, A new high accuracy software based resolver-to-digital converter, in *Proceedings of the 29th IEEE Annual Conference (IECON '03),* 2003, pp. 2435-2440.

[40]. Mohieddine Benammar, Lazhar Ben-Brahim, and Mohd A. Alhamadi, A novel resolverto-360^0 linearized converter, *IEEE Sensors Journal*, Vol. 4, No. 1, 2004, pp. 96–101.

[41]. Andreas Bünte and Stephan Beineke, High-performance speed measurement by suppression of systematic resolver and encoder errors, *IEEE Transactions on Industrial Electronics*, Vol. 51, No. 1, February 2004, pp. 49-53.

[42]. N. Nowlin, S. McEndree and D. Butcher, A Radiation-Hardened High-Precision Resolver-to-Digital Converter (RDC), in *Proceedings of the IEEE Radiation Effects Data Workshop,* 2004, pp. 96-103.

[43]. L. Z. Sun, J. B. Zou and Y. P. Lu, New Variable-reluctance resolver for Rotor-position sensing, in *Proceedings of the IEEE Region 10 Conference (TENCON'04),* 2004, pp. 5-8.

[44]. A. Balkovoy and E. Kallenbach, A low cost resolver-to-digital converter, in *Proceedings of the 49th Internationales Wissenschaftliches Kolloquium*, Technische Universität Ilmenau, September 27-30, 2004.

[45]. Don Payne, Accurate measurement of angle position at high angular velocities, in *Proceedings of the 5th European Magnetic Sensors & Actuators Conference (EMSA'04)*, July 7, 2004.

[46]. Mohd. A. Avlamadi, M. Benammar and L. Ben-brahim, Precise method for linearizing sine and cosine signals in resolvers and quadrature encoders applications, in *Proceedings of the*

30th Annual Conference of the IEEE Industrial Electronics Society, Busan, Korea, November 2-6, 2004, pp. 1935-1940.

[47]. Victor D. Aksenenko and Sergey I. Matveyev, Digital angle sensor self-calibration: Two approaches to accuracy increasing, in *Proceedings of the IEEE Instrumentation and Measurement Technology Conference*, Canada, May 2005, pp. 17-19.

[48]. Robert Herb, Device and method for determining the rotary orientation of a motor through use of Resolver signal derived from the rotary orientation, *U. S. Patent 6,931,918 B2,* August 23, 2005.

[49]. 56F80x resolver driver and hardware interface, Application note, AN1942, *Freescale Semiconductors,* Rev. 1, August, 2005, pp. 1-28. [Online] Available: www.freescale.com.

[50]. Mohieddine Benammar, Lazhar Ben-Brahim, and Mohd A. Alhamadi, A high precision resolver-to-DC converter, *IEEE Transactions on Instrumentation and Measurement*, Vol. 54, No. 6, pp. 2289-2296, December, 2005.

[51]. Gabriel Gross, Miquel Teixid, Antoni Sudria and Joan Bergas, All-digital resolver-to-digital conversion, in *Proceedings of the European Conference on Power Electronics and Applications (EPE'05)*, Dresden, pp. P.1-P.8.

[52]. Masayuki Katakura, Asako Toda, Yuichi Takagi, Norihito Suzuki, Takahide Kadoyama and Hiroshi Kushihara, A 12-bits resolver-to-digital converter using complex twin pll for accurate mechanical angle measurement, in *Proceedings of the VLSI Circuits Symposium*, 2005, pp. 236-239.

[53]. Lizhi Sun, Jing Shang, and Jibin Zou, New absolute rotor-position sensors for inverter-driven motors, in *Proceedings of the IEEE Macnetic Conference INTERMAG Asia 2005*, 4-8 April 2005, pp. 975 – 976.

[54]. Reza Hoseinnezhad and Peter Harding, A novel hybrid angle tracking observer for resolver to digital conversion, in *Proceedings of the 44th IEEE Conference on Decision and Control, and the European Control Conference*, Seville, Spain, December 12-15, 2005, pp. 7020-7025.

[55]. Reza Hoseinnezhad, Position sensing in brake-by-wire capillers using resolvers, *IEEE Transactions on Vehicle Technology*, Vol. 55, No. 3, pp. 924–932, May 2006.

[56]. Yoshi Ishizuka, Kazuhiro Makiuchi and Toru Miyajima, Compensation method of resolver detected position, *U. S. Patent US 7,047,145 B2*, May 16, 2006.

[57]. Mohieddine Benammar, Lazhar Ben-brahim, Mohd. A. Alhamadi and Mohamed El-Naimi, A novel converter for sinusoidal encoders, in *Proceedings of the 5th IEEE Conference on Sensors,* Daegu, Korea, October 22-25, 2006, pp. 1415-1418.

[58]. Armando Bellini and Stefano Bifaretti, A digital filter for speed noise reduction in drives using an electromagnetic resolver, *International Journal of Mathematics and Computers in Simulation*, Vol. 71, 2006, pp. 476-486.

[59]. Jens Onno Krah, Heiko Schmirgel and Marcel Albers, FPGA based resolver to digital converter using delta-sigma technology, in *Proceedings of the PCIM*, Europe, 2006, pp. 931-936.

[60]. Kamel Bouallaga, Lahoucine Idkhajine, Antonio Prata and Eric Monmasson, Demodulation methods on fully FPGA-based system for resolver signals treatment, in *Proceedings of the European Conference on Power Electronics and Applications,* Denmark, September 2007, pp. 1–6.

[61]. Jin-Woo Ahn, Sung-Jun Park and Dong-Hee Lee, Novel encoder for switching angle control of SRM, *IEEE Transactions on Industrial Electronics*, Vol. 53, No. 3, June 2006, pp. 848-854.

[62]. Weera Kaewjinda and Mongkol Konghirun, Vector control drive of Permanent Magnet Synchronous Motor using resolver sensor, *ECTI Transactions on Electrical Engineering, Electronics and Communications*, Vol. 5, No. 1, February 2007, pp. 134-138.

[63]. Konstantin Veselinov Dimitrov, A 3-D Hall sensor for precise angular position measurements, *Turk Journal of Physics*, Vol. 31, 2007, pp. 97 – 101.

[64]. Reza Hoseinnezhad, Alireza Bab-Hadiashar and Peter Harding, Calibration of resolver sensors in electromechanical braking systems: a modified recursive weighted least-squares approach, *IEEE Transactions on Industrial Electronics*, Vol. 54, No. 2, April 2007, pp. 1052-1060.

[65]. Andrzej Michalski, Jan Sienkiewicz and Zbigniew Watral, Universal magnetic circuit for resolvers with different speed ratios, *IEEE Instrumentation & Measurement Magazine*, 10, 5, October 2007, pp. 58-68.

[66]. M. Benammar, A novel amplitude-to-phase converter for sine/cosine position transducers, *International Journal of Electronics*, Vol. 94, No. 4, April 2007, pp. 353–365.

[67]. S. K. Kaul, A. K. Tickoo, R. Koul and N. Kumar, Improving the accuracy of low-cost resolver-based encoders using harmonic analysis, *International Journal of Nuclear Instruments and Methods in Physics Research*, A 586, 2008, pp. 345–355.

[68]. Ciro Attaianese and Giuseppe Tomasso, Position measurement in industrial drives by means of low-cost resolver-to-digital converter, *IEEE Transactions on Instrumentation and Measurement*, Vol. 56, No. 6, December 2007, pp. 2155-2159.

[69]. Douglas W. Brown, Derek L. Edwards, George Georgoulas, Bin B. Zhang, and George J. Vachtsevanos, Real-time fault detection and accommodation for cots resolver position sensors, in *Proceedings of the IEEE International Conference on Prognostics and Health Management*, 2008, pp. 1-6.

[70]. Lizhi Sun, Analysis and improvement on the structure of variable reluctance resolvers, *IEEE Transactions on Magnetics*, Vol. 44, No. 8, August 2008, pp. 2002-2008.

[71]. Lazhar Ben-Brahim, Mohieddine Benammar, Mohd. Alhamadi, Nasser Al-Emadi and Mohammed Al-Hitmi, A new angle determination method for resolvers, in *Proceedings of the 10th IEEE International Workshop on Advanced Motion Control (AMC'08),* 2008, pp. 126-131.

[72]. Lazhar Ben-Brahim, Mohieddine Benammar, Mohd A. Alhamadi, Nasser A. Al-Emadi and Mohammed A. Al-Hitmi, A new low cost linear resolver converter, *IEEE Sensors Journal*, Vol. 8, No. 10, October 2008, pp. 1620-1627.

[73]. S. Sarma, V. K. Agrawal and S. Udupa, Software-based resolver-to-digital conversion using a DSP, *IEEE Transactions on Industrial Electronics*, Vol. 55, No. 1, January 2008, pp. 371-379.

[74]. D. Arab-Khaburi, F. Tootoonchian and Z. Nasiri-Gheidari, Dynamic Performance Prediction of Brushless Resolver, *Iranian Journal of Electrical & Electronic Engineering*, Vol. 4, No. 3, July 2008, pp. 94- 103.

[75]. S. Sarma, V. K. Agrawal, S. U. Dupa and K. Parameswaran, Instantaneous angular position and speed measurement using a DSP based resolver-to-digital converter, *Measurement*, Vol. 41, No. 1, 2008, pp. 788-796.

[76]. R. M. Kennel and St. Basler, New developments in capacitive encoder for servo drives, in *Proceedings of the IEEE International Symposium on Power Electronics Electrical Drives, Automation and Motion (SPEEDAM'08)*, 2008, pp. 190-195.

[77]. Variable resolution 10-bit to 16-bit R/D converter with reference oscillator, catalogue of AD2S1210, *Analog Devices,* pp. 1-19, 2008. [Online] Available: www.analog.com

[78]. Lazhar Ben-Brahim, Mohieddine Benammar and Mohd. A. Alhamadi, A resolver angle estimator based on its excitation signal, *IEEE Transactions on Industrial Electronics,* Vol. 56, No. 2, February 2009, pp. 574-580.

[79]. Seon Hwan Hwang, Young Hwa Kwon, Jang Mok Kim and Jin Seok Oh, Compensation of position error due to amplitude imbalance in resolver signals, *Journal of Power Electronics*, Vol. 9, No. 5, September 2009, pp. 748-756.

[80]. S. H. Hwang, H. J. Kim, J. M. Kim, Hui Li and Liming Liu, Compensation of amplitude imbalance and imperfect quadrature in resolver signals for PMSM drives, in *Proceedings of the IEEE,* 2009, pp. 1720-1725.

[81]. Young-Hwa Kwon, Seon-Hwan Hwang, Jang-Mok Kim and Jin-Woo Ahn, Compensation of amplitude imbalance of resolver signal for PMSM drives, in *IEEE Proceedings IPEMC2009*, 2009, pp. 1827-1831.

[82]. Zhang Haixia and Yanlan, Design of synchro resolver-to-digital converter based on PXI bus, in *Proceedings of the IEEE Pacific-Asia Conference on Knowledge Engineering and Software Engineering*, 2009, pp. 194-196.

[83]. Zhuangzhi Han, Heng Zhang, Qiang He and Chaoxuan Shang, Resolver-to-digital converter based on tangent algorithm, in *Proceedings of the IEEE International Symposium on Industrial Electronics (ISIE'09)*, Korea, July 5-8, 2009, pp. 329-332.

[84]. Mohieddine Benammar, Mohamed Bagher and Mohammed Al Kaisi, Novel linearizer for tangent/cotangent converter, in *Proceedings of the IEEE International Conference*, 2009, pp. 575-578.

[85]. V. K. Dhar, A. K. Tickoo, S. K. Kaul, R. Koul and B. P. Dubey, Artificial neural network-based error compensation procedure for low-cost encoders, *Measurement Science and Technology,* 21, 1, 2009. [Online] Available: http://arxiv.org/abs/0911.3717v1

[86]. Ilpakurty Ravi and K. Nagabhushan Raju, Converting resolver interface to incremental encoder interface, *International Journal of Electronic Engineering Research*, Vol. 1, No. 4, 2009, pp. 345-348.

[87]. Santanu Sarma and A. Venkateswaralu, Systematic error cancellations and fault detection of resolver angular sensors using a DSP based system, *Mechatronics*, Vol. 19, 2009, pp. 1303-1312.

[88]. Mohieddine Benammar, Mohamed Bagher and Mohamed Al Kaisi, Digitally-tuned resolver converter, in *Proceedings of the Eurosensors XXIII Conference*, 2009, pp. 449-452. [Online] Available: www.sciencedirect.com

[89]. Nicolas Javahiraly, Cédric Perrotton, Ayoub Chakari and Patrick Meyrueis, Design, study and achievement of a fiber optic amplitude modulation sensor for angular position detection: application to an automotive steering system, *Proceeding of SPIE,* Vol. 7314, 2009, pp. 731405-1-8.

[90]. Anucha Kaewpoonsuk, Ratchanoo Katman, Thawatchai Kamsri, Apinai Rerkratn and Vanchai Riewruja, A simple amplitude detector-based demodulator for resolver converters, in *Proceedings of the International Conference on Control, Automation and Systems*, Korea, 2010, pp. 370-373.

[91]. Aeroflex datasheet catalogue, RDC5028C 16-bit monolithic tracking rad tolerant resolver-to-digital converter, SCD5028-2 Rev D, October 22, 2010. [Online] Available: www.aeroflex.com/RDC

[92]. S. H. Hwang, J. H. Lee, J. M. Kim and C. Choi, Compensation of analog rotor position errors due to non-ideal sinusoidal encoder output signals, in *Proceedings of the IEEE International Conference*, 2010, pp. 4469-4473.

[93]. Lazhar Ben-Brahim and Mohieddine Benammar, A new PLL method for resolvers, in *Proceedings of the IEEE International Power Electronics Conference*, 2010, pp. 299-305.

[94]. Zhu Yi and Wang Jian Ming, An approach based on AD converted resolver demodulation, in *Proceedings of the IEEE 3rd International Conference on Advanced Computer Theory and Engineering (ICACTE)*, 2010, pp. V5-192-195.

[95]. Ralph Kennel, Encoderless control of synchronous machines with permanent magnets-impact of magnetic design, in *Proceedings of the IEEE 12th International Conference on Optimization of Electrical and Electronic Equipment*, 2010, pp 19-24.

[96]. Cheon Soo Park, Minimizing magnet interference in a variable reluctance resolver, *U. S. Patent US 2011/0068960 A1*, March 24, 2011, USA.

[97]. Joan Bergas Jané, Coia Ferrater Simón, Gabriel Gross, Rodrigo Ramírez Pisco, Samuel Galceran Arellano and Joan Rull Duran, High-accuracy all-digital resolver-to-digital conversion, *IEEE Transactions on Industrial Electronics*, Vol. 59, No. 1, January 2012, pp. 326-333.

[98]. Zhu Ming, Wang Jianming, Ding Ling, ZhuYi and Dou Ruzhen, A software based robust resolver-to-digital conversion method in designed in frequency domain, in *Proceedings of the IEEE International Symposium on Computer Science and Society*, 2011, pp. 244-247.

[99]. Jiebin Zhang, Sun Hua, Zhao Qi and Wenquan Feng, The design and implementation of the shaft angle acquisition system used in the solar panel, in *Proceedings of the IEEE 4th International Symposium on Computational Intelligence and Design*, 2011, pp. 287-290.

[100]. H. Loge and L. Angerpointner, The best way how to use resolvers, in *Proceedings of the 1st International Electric Drives Production Conference (EDPC'11),* 2011, pp. 208-213.

[101]. Prerna Gaur, Sumit Bhardwaj, Naveen Jain, Nipun Garg, Prashant A, A. P. Mittal, and Bhim Singh, A novel method for extraction of speed from resolver output using neural network in vector control of PMSM*,* in *Proceedings of the India International Conference on Power Electronics (IICPE'10),* 2011, pp. 1-7.

[102]. Qi-Xun ZHOU, Research on the signal process circuit and fault diagnosis of sine-cosine resolver, in *Proceedings of the International Conference Electric Information and Control Engineering (ICEICE'11),* 2011, pp. 2581-2584.

[103]. Ruijie Zhao, Xuejun Tao, Dawei Wang and Suli Tian, Research on the decoding method of resolver, in *Proceedings of the International Conference on Electrical and Control Engineering (ICECE'11),* 2011, pp. 329-333.

[104]. Anna K. S. Baasch, Elisabeth C Lemos, Felipe Stein, Aleksander S. Paterno, José de Oliveira and Ademir Nied, Resolver-to-digital conversion implementation–a filter approach to PMSM position measurement, in *Proceedings of the IEEE International Conference on Power Engineering, Energy and Electrical Drives*, May 2011, Spain.

[105]. Kazuya Sakai, Rotational angle sensor, motor, rotational angle detector, and electric power steering system, *U. S. Patent US 2011/0068780 A1*, March 24, 2011, USA.

[106]. Davood Arab Khaburi, Software based resolver-to-digital converter for DSP-based drives using an improved angle-tracking observer, *IEEE Transactions on Instrumentation and Measurement,* Vol. 61, No. 4, April 2012, pp. 922-929.

[107]. L. Idkhajine, E. Monmasson, M. W. Naouar, A. Prata, and K. Bouallaga, Fully integrated FPGA-based controller for synchronous motor drive, *IEEE Transactions on Industrial Electronics*, Vol. 56, No. 10, October 2009, pp. 4006–4017.

[108]. L. Idkhajine, A. Prata, E. Monmasson, K. Bouallaga, and M.-W. Naouar, System on chip controller for electrical actuator, in *Proceedings of the IEEE International Symposium on Industrial Electronics (ISIE'08)*, Cambridge, U. K., July 2008, pp. 2481–2486.

[109]. W. Kaewjinda and M. Konghirun, A DSP-based vector control of PMSM servo drive using resolver sensor, in *Proceedings of the IEEE Region 10 Conference (TENCON)*, November 14–17, 2006, pp. 1–4.

[110]. A. Kaewpoonsuk, W. Petchmaneeluka and A. Perkratn, A novel rosolver-to-DC converter based on OTA-based inverse sine function circuit, in *Proceedings of the SICE Annual Conference*, 2008, pp. 609-614.

[111]. A. Bellini and S. Bifaretti, Implementation of a digital filter for speed noise reduction in drives with electromagnetic resolver, in *Proceedings of the European Conference on Power Electronics Applications*, September 11–14, 2005, pp. P.1–P.10.

[112]. A. Bellini, S. Bifaretti, and S. Costantini, A PLL-based filter for speed noise reduction in drives using a resolver transducer, in *Proceedings of the IEEE International Symposium on Industrial Electronics (ISIE'02),* July 8–11, 2002, Vol. 2, pp. 529–534.

Index

Sensors, Transducers, Signal Conditioning and Wireless Sensors Networks

Sergey Y. Yurish, Editor

The third volume titled *'Sensors, Transducers, Signal Conditioning and Wireless Sensors Networks'* contains nineteen chapters with sensor related state-of-the-art reviews and descriptions of latest achievements written by 55 authors from academia and industry from 19 countries: Botswana, Canada, China, Finland, France, Germany, India, Jordan, Mexico, Portugal, Romania, Russia, Senegal, Serbia, South Africa, South Korea, UK, Ukraine and USA.

This book ensures that our readers will stay at the cutting edge of the field and get the right and effective start point and road map for the further researches and developments. By this way, they will be able to save more time for productive research activity and eliminate routine work. Built upon the series *Advances in Sensors: Reviews* - a premier sensor review source, it presents an overview of highlights in the field and becomes. Coverage includes current developments in physical sensors and transducers, chemical sensors, biosensors, sensing materials, signal conditioning energy harvesters and wireless sensor networks. Sure, we would have liked to include even more topics, but it is difficult to cover everything due to reasonable practical restrictions. With this unique combination of information in each volume, the Advances in Sensors book Series will be of value for scientists and engineers in industry and at universities, to sensors developers, distributors, and users.

Like the first two volumes of this book Series, the third volume also has been organized by topics of high interest. In order to offer a fast and easy reading of the state of the art of each topic, every chapter in this book is independent and self-contained. All chapters have the same structure: first an introduction to specific topic under study; second particular field description including sensing applications. Each of chapter is ending by well selected list of references with books, journals, conference proceedings and web sites.

BN-20160415-XX
www.sensorsportal.com
ISBN 978-84-608-7705-9