Smart Sensors, Measurement and Instrumentation 12

Alex Mason
Subhas Chandra Mukhopadhyay
Krishanthi Padmarani Jayasundera
Editors

Sensing Technology: Current Status and Future Trends IV

Springer

Smart Sensors, Measurement and Instrumentation

Volume 12

Series editor

Subhas Chandra Mukhopadhyay
School of Engineering and Advanced Technology (SEAT)
Massey University (Manawatu)
Palmerston North
New Zealand
E-mail: S.C.Mukhopadhyay@massey.ac.nz

More information about this series at http://www.springer.com/series/10617

Alex Mason
Subhas Chandra Mukhopadhyay
Krishanthi Padmarani Jayasundera
Editors

Sensing Technology: Current Status and Future Trends IV

Springer

Editors
Alex Mason
Liverpool John Moores University
Liverpool
United Kingdom

Krishanthi Padmarani Jayasundera
Massey University (Manawatu Campus)
Palmerston North
New Zealand

Subhas Chandra Mukhopadhyay
Massey University (Manawatu)
Palmerston North
New Zealand

ISSN 2194-8402 ISSN 2194-8410 (electronic)
Smart Sensors, Measurement and Instrumentation
ISBN 978-3-319-12897-9 ISBN 978-3-319-12898-6 (eBook)
DOI 10.1007/978-3-319-12898-6

Library of Congress Control Number: 2013953628

Springer Cham Heidelberg New York Dordrecht London

Printed on acid-free paper

[Springer International Publishing AG Switzerland] is part of Springer Science+Business Media
(www.springer.com)

Guest Editorial

The applications of Sensing Technology include gas sensing, medical diagnostics, industrial manufacturing, defense, national security, prevention of natural disaster and terrorism. The proper detection and reporting of events by high performance sensor systems delivers huge cross-sector benefits through increasing use of continuous real-time systems for improved automation and control. The continuing development of sensors, and the systems which support them, will undoubtedly continue to bring great value, and will assist in tackling global issues such as water quality monitoring, ageing populations and natural disasters. The cross-cutting nature of sensor systems demands interaction between researchers and industry from technologically advanced and developing countries in order to drive toward the sensors of the future. ICST 2013, as in previous years, brings together such people, and so we hope that it can provide a useful platform for interaction.

This book contains a collection of selected works stemming from the 2013 International Conference on Sensing Technology (ICST), which was held in Wellington, New Zealand. This was the seventh time the conference had been held, and over the years it has become an incredibly successful event – in 2013 it attracted over 220 participants and provided a forum for interaction between researchers across technologically advanced and developing countries working on design, fabrication and development of different sensors.

The conference was organized by the School of Engineering and Advanced Technology, Massey University, and we thank the department's management for extending their continued support for the conference. We would also like to thank all of the authors as they are main ingredient for any conference to succeed. In addition, the Technical Programme Committee has done a tremendous and wonderful job. We are very much indebted to everybody in the Technical Programme Committee for accepting the invitation to lend their help, support, time and effort to make ICST a great success. We would also like to extend our special thanks to the keynote and invited speakers for their time and support during the event.

The purpose of these books is to distill the highlights of the conference, and therefore track the latest developments in sensing technologies. The book contents are quite broad, since sensors can be applied in many areas as was made apparent previously.

Chapters 1-3 of the book discuss the development of gas based sensing. In particular, Chapter 1 introduces solid state vapor sensors, Chapter 2 looks at the development of low cost sensors for gas flow measurement in domestic boiler systems and Chapter 3 looks at the use of light scattering for ultrafine particulate measurement in combustion engine exhaust gases.

The theme of optical based sensors then continues through Chapters 4-6. Chapter 4 introduces a system for measurement of blood glucose using near infrared technology, while Chapter 5 demonstrates a fusion of optical and radar imaging. Chapter 6 presents a novel application of thermal imaging which hopes to improve toilet systems used particularly in care homes.

In Chapter 7 an innovative electromagnetic sensor design is demonstrated for the detection of liquids, with the work demonstrating a clear method for differentiation between various liquids such as water and oils.

Chapter 8 details a new wireless mote device developed by researchers in New Zealand, with the aim of achieving both small size and ultra-low power. Chapter 9 presents work also of relevance in the wireless sensing domain, namely discussing the implementation of a fault detection mechanism, named Intellectus, which seeks to improve autonomous node resilience.

Chapter 10 presents work which has potential across a wide spectrum of industrial applications in the area of 3D sensors to assist in the automation of logistics; this particular work concentrates on sensing the shape and arrangement of palletized goods.

Chapter 11 discusses work in the area of acoustics utilizing modulation effects to produce signal direction of arrival (DoA) estimation. The work aims to deliver performance similar to distributed array systems at a fraction of the cost.

The topic of amorphous alloy ribbons is presented in Chapter 12, which are well known to be very effective for the measurement of magnetic fields. In particular, the work demonstrates that stress annealed ribbons exhibit a significantly higher quality factor than their typical as-cast counterparts, which gives a strong potential for future sensing applications where high sensitivity is required.

Chapter 13, the final in this book, presents a hybrid intelligent control for outer rotor spherical actuators. Results of modelling in this work prove the effectiveness of an adaptive neuro-fuzzy inference system as a feedback control method.

This book is written for academic and industry professionals working in the field of sensing, instrumentation and related fields, and is positioned to give a snapshot of the current state of the art in sensing technology, particularly from the applied perspective. The book is intended to give broad overview of the latest developments, in addition to discussing the process through which researchers go through in order to develop sensors, or related systems, which will become more widespread in the future.

We would like to express our appreciation to our distinguished authors of the chapters whose expertise and professionalism has certainly contributed significantly to this book.

Alex Mason, Guest Editor
Built Environment and Sustainable Technologies Research Institute,
School of Built Environment,
Liverpool John Moores University
Liverpool, United Kingdom
e-mail: A.Mason1@ljmu.ac.uk

Subhas Chandra Mukhopadhyay, Guest Editor
School of Engineering and Advanced Technology (SEAT),
Massey University (Manawatu Campus)
Palmerston North, New Zealand
e-mail: S.C.Mukhopadhyay@massey.ac.nz

Krishanthi Padmarani Jayasundera, Guest Editor
Institute of Fundamental Sciences,
Massey University (Manawatu Campus)
Palmerston North, New Zealand
e-mail: K.P.Jayasundera@massey.ac.nz

Dr. Alex Mason graduated from the University of Liverpool, UK, with a first class honors degree in Computer and Multimedia Systems, after which he went on to complete a PhD in Wireless Sensor Networks and their Industrial Applications at Liverpool John Moores University (LJMU), UK. Upon completing his PhD in 2008, he concentrated for 2 years solely on research, working on aspects of non-invasive and non-destructive sensing for the healthcare, automotive and defense sectors.

Dr. Mason currently holds the position of Reader in Smart Technologies within the School of Built Environment at LJMU, after holding a Senior Lecture post prior to this since 2010. He has continued research in healthcare and defense, in addition to new areas such as water quality monitoring. Since becoming more involved in Built Environment issues, he has also developed an interest in Structural Health Monitoring and is currently working closely with the UK Defense Science and Technology Laboratories in this field.

Dr. Mason is responsible for supervising a number of PhD students in the areas of sensing and renewable energy technologies, has coauthored over 130 publications (including 4 patents), has helped to organize national and international conferences, and gave a number of invited talks on his work. He is also an active member of the IET in the UK, and has achieved Chartered Engineer status with them.

Dr. Subhas Chandra Mukhopadhyay graduated from the Department of Electrical Engineering, Jadavpur University, Calcutta, India in 1987 with a Gold medal and received the Master of Electrical Engineering degree from Indian Institute of Science, Bangalore, India in 1989. He obtained the PhD (Eng.) degree from Jadavpur University, India in 1994 and Doctor of Engineering degree from Kanazawa University, Japan in 2000.

Currently, he is working as a Professor of Sensing Technology with the School of Engineering and Advanced Technology, Massey University, Palmerston North, New Zealand. His fields of interest include Smart Sensors and Sensing Technology, Wireless Sensors Network, Electromagnetics, control, electrical machines and numerical field calculation etc.

He has authored/co-authored over 300 papers in different international journals and conferences, edited eleven conference proceedings. He has also edited ten special issues of international journals and twenty books with Springer-Verlag as guest editor. He is currently the Series editor for the Smart Sensing, Measurements and Instrumentation of Springer-Verlag.

He is a Fellow of IEEE, a Fellow of IET (UK), a Topical editor of IEEE Sensors journal. He is also an Associate Editor for IEEE Transactions on Instrumentation and Measurements and a Technical Editor of IEEE Transactions on Mechatronics. He was a Distinguished Lecturer of IEEE Sensors council. He is Chair of the Technical Committee 18, Environmental Measurements of the IEEE Instrumentation and Measurements Society. He is in the editorial board of many international journals. He has organized many international conferences either as a General Chair or Technical Programme Chair.

Dr. Krishanthi P. Jayasundera graduated from University of Peradeniya, Sri Lanka with honors degree in Chemistry. She obtained her both Master and PhD in Organic Chemistry from Kanazawa University, Japan. After completing PhD she worked as researcher nearly 14 years in New Zealand involving various projects related to organic synthesis. Currently, she is working as a Post-doctoral researcher with the Institute of Fundamental Sciences, Massey University, New Zealand. She has published over 30 papers in different international journals and conference proceedings. She has also edited a few conference proceedings and books.

Contents

Potentiality of Semiconducting Metal Oxide Nanoforms as Solid State Vapor Sensors

P. Bhattacharyya[1], B. Bhowmik[1],
A. Hazra[1], and P.P. Chattopadhyay[2]

[1] Department of Electronics and Telecommunication Engineering,
Indian Institute of Engineering Science and Technology (IIEST),
Shibpur- 711103, Howrah, India
[2] Department of Metallurgy and Materials Engineering,
Indian Institute of Engineering Science and Technology (IIEST),
Shibpur- 711103, Howrah, India
`pb_etc_besu@yahoo.com`

Abstract. Nanostructured semiconducting metal oxides have been slotted as the potential material for its applications in gas/vapor sensing systems. During the last couple of decades, extensive efforts have been devoted in maximizing the parameters like sensitivity and selectivity and minimizing the response and recovery times. The efforts are essentially directed towards modification of chemical, dimensional and morphological attributes of the metal oxides. The evolution started with the optimization of particles size and thickness of the thin films, but was accelerated, particularly since last decade, with the incorporation of several unprecedented interesting nanoforms of metal oxides like nanorods, nanowires, nanosheets, nanohollowspheres and nanotubes. It has been established that the vapour/gas sensing characteristics (e.g., operating temperature, sensitivity, response time and recovery time) of the semiconducting oxides improved dramatically if the nanostructure is changed from three dimensional to two dimensional and then to one dimensional. Thus, the domain calls for basketting of the state of knowledge with the contemporary relevance. In view of the above, the present article aims to discuss the underlying mechanisms governing the formation of metal oxide nanoforms under different process routes and correlation of the structural attributes with the semiconducting behavior with special reference to the sensing of acetone vapor. The beginning of the chapter elaborates the necessity of nanoforms in improving the sensing performance with the help of grain surface/boundary model. Subsequently, different semiconducting oxide nanostructure/nanoform (like nanorods, nanowires, nanosheets, nanohollowspheres and nanotubes) synthesis and their sensing properties with emphasis on their respective advantages and bottlenecks have been discussed. The chapter concludes with the salient features of the advancement of oxide nanoform based acetone sensors and predicting the future direction of research.

Keywords: Semiconductor metal oxide, Nanorod, Nanowires, Nanosheet, Hollowsphere, Acetone sensing.

A. Mason et al. (eds.), *Sensing Technology: Current Status and Future Trends IV*,
Smart Sensors, Measurement and Instrumentation 12, DOI: 10.1007/978-3-319-12898-6_1

1 Introduction

Efficient detection and subsequent alarming of Volatile Organic Compounds VOCs) are becoming more and more important for domestic and industrial environmental monitoring. Acetone is one of the most highly Volatile Organic Compound (VOC), often used as an organic solvent or chemical intermediate in laboratory or industrial application [1]. Easily evaporable acetone vapor is harmful to human health. This toxic VOC is generated in human bodies under starvation conditions, consumption of high-fat and low carbohydrate diets, or under uncontrolled diabetes [2]. Investigations have indicated that continuous exposure of acetone vapor (500 ppm) may cause irritation of eye and respiratory system, mood swings, nausea [3]. Likewise, breathing high levels (>500 ppm) of acetone (in industrial area) can cause sluggishness and drowsiness of nervous system as listed in Fig 1.

Human breath composed of inorganic compound like N_2, O_2, CO_2, inert gases, and water vapor. Apart from that there are more than 1000 VOCs exist with concentrations ranging from parts per trillion (ppt) to parts per million (ppm) [4]. These compounds can be used as biomarkers in medical science, such as inorganic gases (e.g. NO, CO) and volatile organics (e.g. ethane, pentane, ammonia, acetone, and ethanol) to determine illness directly. The study raveled that acetone in healthy person in between 300 - 900 ppb unexpectedly increase to 1800 ppb for diabetic patients [4]. In particular, acetone is a selective breath marker for diabetes. Recent research raveled that detection of concentration of acetone in breath is an alternative to blood analysis to diagnose diabetes in a non-invasive, painless and cost effective manner.

There are many techniques/method to measure the concentration of acetone (in atmosphere or human body), such as gas chromatography, gas chromatography-mass spectroscopy (GC–MS) analysis, chemical ionization mass spectrometry, cavity ring down spectrometry, laser spectroscopy, ion mobility spectrometry and semiconducting metal oxide based conductometric gas/vapor sensors [5-64]. Amongst these techniques/methods, except the last one, all are either time consuming or expensive. Recently, metal oxide semiconductor nanostructures proved their potentiality (as chemo resistive gas/vapor sensors) in acetone detection, owing to their high sensitivity and fast response, even at ppb level. Numerous semiconducting oxides e.g. SnO_2, Fe_2O_3, ZnO, WO_3, In_2O_3 and TiO_2 have so far been investigated for detection acetone employing various nanoforms like nanorod, nanosheets, nanowire, hollowsphere etc [4-6, 19, 28, 37].

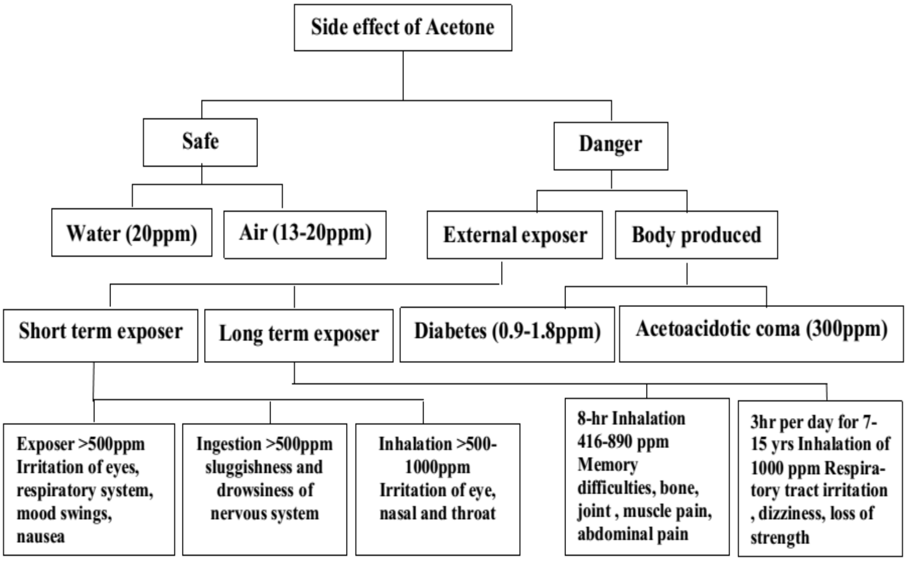


Fig. 1. Effect of acetone exposer on human body

This chapter aims to describe the technological developments of VOC sensing by employing semiconductor nanostructures (e.g. thin film, nanorods, nanowire, nanosheets, hollowsphere and nanotube) in a comprehensive and judicious manner with special emphasis on Acetone as a test case. After a short description about the various techniques for nanostructure formation, the chapter emphasizes on the role of nanostructure on acetone sensing mechanism with particular focus on development milestones and bottlenecks. Nevertheless, in spite of extensive research efforts, synthesis of repeatable and reproducible hierarchical nanostructures still remains a technological challenge. Moreover, typical vices of semiconducting oxide gas/vapor sensors like selectivity and long term stability is still a burning issue. The chapter concludes with the salient features of the advancements and predicting the future trends of research.

2 Acetone Sensing Mechanism: Role of Nanostructures

2.1 Acetone Sensing Mechanism

The acetone sensing mechanism is essentially an oxidation process of acetone on the surface of the sensing oxide which finally produces either acetic acid or CO_2 [5-6]. Due to non-stoichiometric nature and high free surface energy of nano-oxide materials, the oxygen from air gets adsorbed onto the oxide surface and subsequently produces oxygen species like O^{2-}, O^-, and O_2^- by capturing electron from the conduction band of nano metal-oxides. As a result the resistance of the oxide increases. When acetone vapor interacts with these (O^{2-}, O^-, and O_2^-) oxygen species adsorbed on semiconductor surface producing different byproducts following either of the ways based on operating temperature and reaction rate constant (k) as described below.

In the first case, at lower temperature (≤300°C), acetone gets oxidized (when it interacts with nano-oxide surface) to form acetic acid and generates vacant sites, which again helps to attract more acetone molecule on the surface [5]. The details reaction kinetics is as follows;

Acetone after interacting with O^-, forms an acetone–oxygen complex compound and release an electron. The unstable intermediate acetone–oxygen complex is then dissociated to an acetate group and methyl group. The water vapor (humidity) from air approaches the semiconducting surface and produces hydrogen and hydroxyl ions. The produced acetate reacts with hydrogen atom and generates acetic acid. On the other hand hydroxyl (OH^-) ion is converted to water and produce an oxygen ion which again get adsorbed on semiconductor oxide surface if the vacant sites are available. At elevated temperature the acetic acid from the surface decomposes to acetic acid gas and creates a vacant site [5]. These reactions chain can be expressed as:

$$O^-_{(ads)} + \rightarrow \{CH_3\text{-}CO\text{-}CH_3\text{–oxygen}\}_{(ads)} + e- \quad (1)$$

$$\{CH_3\text{-}CO\text{-}CH_3\text{–oxygen}\}_{(ads)} \rightarrow CH_3COO^-_{(ads)} + CH_3\,(gas) \quad (2)$$

$$H_2O \leftrightarrow OH_{(ads)} + H^+ + e^- \quad (3)$$

$$CH_3COO^-_{(ads)} + H_{(ads)} \rightarrow CH_3COOH_{(ads)} \quad (4)$$

$$2OH_{(ads)} \leftrightarrow H_2O + O \quad (5)$$

$$CH_3COOH_{(ads)} \rightarrow CH_3COOH_{(gas)} \quad (6)$$

In the second possibility, at relatively higher temperature (>300°C), adsorbed oxygen (O^-) helps to decompose acetone vapor into acetyl group, formaldehyde radical and subsequently release one e^- [6]. The unstable acetyl group then converts to CO and forms a methyl radical on the surface. CO subsequently converted to CO_2 and releases one electron. This chain of reactions is represented by equation 7-9 below:

$$CH_3COCH_3\,(vapor) + O^- \rightarrow CH_3C^+O + CH_3O^- + e^- \quad (7)$$

$$CH_3C^+O \rightarrow CH_3^+ + CO \quad (8)$$

$$CO + O \rightarrow CO_2 + e^- \quad (9)$$

2.2 Advantages of Nanostructured Oxides for Vapor Sensing

“The term ‘Nano’ refers to the dimension which is ~10^{-9} m precisely Particles/structures having at least one dimension ranging from 1nm to 100nm are considered as nanostructure" [7]. "With decreasing crystal/grain and more exposed surface region, the fraction of atoms at the grain boundary increases and as at the grain boundaries the concentrations of defects states (like vacancies and dangling bonds) are maximum, the electron transport properties (both transient and steady state) are predominantly governed by these nano-grain boundaries" [7]. As already evidenced by numerous researchers, the characteristics of gas sensors improve dramatically by employing nanostructured semiconducting oxide as the sensing material [13-18]. It has also been established that the sensor sensitivity/response magnitude, transient responses etc. effectively determined by how efficiently and rapidly the target species interacts with sensing surface [5-36]. More precisely, interactions of chemical species are strongly influenced by boundary between grains where the schottky barrier/ depletion layer has formed (shown in Fig 2).

The role of nanostructuring the metal oxides on gas sensor properties has already been discussed by Bhattacharyya and Basu [7]. The nano-sized grains in semiconducting oxides are generally connected by neck with neighboring grains and the grain size effectively decide the reaction mechanism with gaseous species, as the conductivity depends on how grains are connected. Three different cases can be distinguished according to the relationship between the particle size (λ_p) and the width of the space-charge layer (λ_d) is shown in Fig 2.

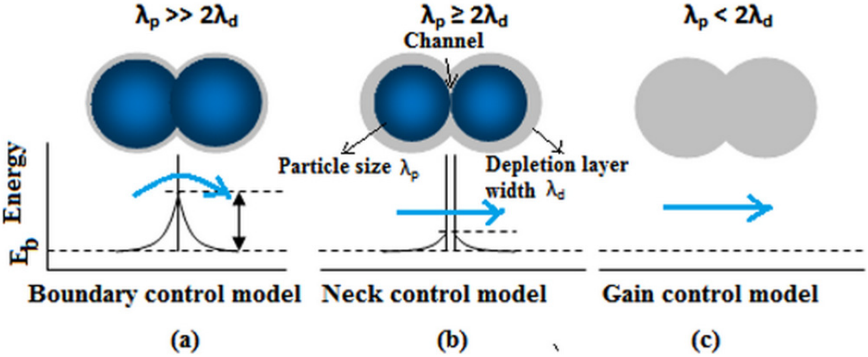


Fig. 2. Semi-quantitative model (a) Boundary controlled model (b) Neck controlled model (c) Grain controlled model (following ref. [8])

Boundary controlled model (shown in Fig.2.), where $\lambda_p >> 2\lambda_d$, is not sensitive to reaction charges. The conductivity of the whole structure depends on the inner mobile charge carriers and the electrical conductivity depends exponentially on the barrier height. In case of Neck controlled model where $\lambda_p \geq 2\lambda_d$, the barrier height reduces as well as a channel is formed between each neck shown in Fig.2.(b). As a result, the conductivity depends on barrier height and the cross section area of those channels. It is too sensitive to reaction charges. Grain control model (shown in Fig.2.(c)) where

$\lambda_p<2\lambda_d$, energy bands are completely flat throughout the whole structure of the interconnected grains. In that case depletion layer region dominate the particle size and the crystallites are almost fully depleted of mobile charge carriers.

The conductivity of thin film is effectively controlled by the inter crystallite conductivity [8]. Factors that affect the sensitivity are porosity, catalytic effect of additives, grains size and shape of the nanostructure. So controlling grain size is one of key solution to get higher sensitivities at the lower detection limit. Apart from controlling the grain size, the acetone sensitivity can also be improved by changing shape of nanostructure (nanorod, nanowire, nanosheets, hollowsphere, nanotube etc.) and controlling the porosity. The application of nanowires, nanorods for gas sensors can significantly lower the minimum detection limit, and hence the effective dynamic range of acetone sensing, since the conductance of such nanostructure is influenced even by very small quantities (ppb) of chemical species.

Further, nanostructured metal oxides are capable of reducing the operating temperature of the chemical sensors as already reviewed by Bhattacharyya [7] and Basu [9]. “The surface or the interfacial tension decreases with the decreasing particle size because of the increase in the potential energy of the bulk atoms of the particles”[10]. “Therefore the smaller particles with increased molar free energy are more prone to adsorption per unit area of molecules or ions onto their surfaces in order to decrease the total free energy and to become more stable, and therefore, smaller particles have higher adsorption coefficient for gases. Thus, the adsorption of oxygen or the reducing gases takes place relatively easily onto the nano-metal oxide surface”[7].

Response and recovery dynamics are strongly dependent on gas adsorption and desorption kinetics on metal oxide surface. Different materials have very specific adsorption/desorption kinetics for different target species. Apart from that adsorption/desorption kinetics are also strongly influenced by temperature. It is well known that adsorption of gaseous molecule at higher activation energy (higher temperature) lead to higher desorption time that is slow recovery which might be not accepted for sensor application [11]. Usually, with increase in surface to volume ratio, response and recovery time improves due to more available adsorbent site, but if the particles are too small (that is surfaces are compact in nature i.e. nonporous), then Van der walls force aggregates them and target molecule cannot penetrate deep into the sensing layer, thereby restricting the gas interaction process only at surface level providing lower sensitivity. To the contrary, for porous layer gas interaction can take place at much deeper level, so sensitivity increases, but consequently sluggish diffusion of gases immensely decreases recovery kinetics [10].

Taking into account such unprecedented structural peculiarities of nano-materials, 1D metal oxide nanoforms, like nanorods, nanowires, 2D nanosheets and 3D hollowsphere received much attention from sensor researchers since last decade. These nanostructures exhibit excellent crystallinity and clear facet with less concentration of point defects. Further, these nanostructures (thin film, nanorod, nanowire, nanosheet, hollowsphere and nanotube) offer well defined nano-porosity which appreciably speeds up the diffusion of gas molecule on the surfaces.

3 Nanostructured Metal Oxide Based Solid State Acetone Sensor

3.1 Thin Film Based Sensor

Since the early stage of gas sensor research, the gas detecting devices are fabricated by employing a thick film (sensing layer) grown on a Al_2O_3/ceramic tube with two electrode contacts and a heater coil. Later, with rapid advancement in technology, thin films consisting of different nanoforms were developed with an intention to optimize the sensor parameters (sensitivity, response/recovery time and operating temperature). Amongst them, a few have been discussed in this chapter in the context of acetone sensor. Several metal oxide nanoparticles, deposited in the form of thin film, such as ZnO, SnO_2, TiO_2, and WO_3 have so far been investigated to detect acetone. Synthesis of nanoparticles generally takes place by vaporizing the metal precursor, followed by a subsequent annealing process. For example Khan et al. fabricated a chemo resistive sensor using butyl carbitol acetate (BCA) and ethyl acetate (EA) as a coating agent to detect organic pollutant (acetone) [12]. The measured low current (0.1-1.1μA) for 1.33M acetone at room temperature possibly attributed to the small conducting channel between coating electrode surface and ZnO nanoparticles as shown in Fig 3 (a)[12]. Fig 3 (b) shows the gradual increase of current in presence of (0.3mM - 1.33M) acetone concentration and applied potential (0-1V).

It is well established that the performance of the gas sensor can be improved if a desired amount of appropriate dopants are introduced to the metal oxide surface [13, 14]. Dopants provide catalytically active sites for a particular species, and hence improve sensitivity towards a particular gas even at lower concentration. Recently influence of indium doping on ZnO nanoparticulate thin films developed by Prajapati et.al. [13], detect acetone even at lower concentration of 25 ppm with fast response and recovery. Detection of ultralow acetone concentration down to 20 ppb employing 10 mol% Si doped on WO_3 thin film has been reported by Righettoni et. al [14]. Un-equivalence doping can change material's conductivity by forming impurity defect. In that case metal oxide molecule partially replaced by dopants which minimizes the electron concentration for n-type and vice versa for p type. The phenomena was observed when doping concentration of CeO_2 increases from 1-10 at.% on ZnO surface, the sensitivity decreases gradually, this is possibly the dual effect of CeO_2 on ZnO surface [15].

Apart from doping the sensing oxides by inorganic catalytic metals, another effective way to improve the gas sensing phenomena through enriching larger amount of organic species at the metal oxide surface. Huang et. al [16] proposed a method for direct modification (Infrared reflection elements modification or IREs modification) of ZnO nanoparticles on the surface of optical sensing elements which improves high surface-to-volume ratio of zinc oxide nanoparticles and their selective interactions with target gases. Direct immobilization of ZnO nanoparticles onto the surfaces of optical sensing elements (prepared sol-gel solution of ZnO directly coated on IREs prior growing nanoparticles), lower the acetone detection limit down to 2.2 ppm.

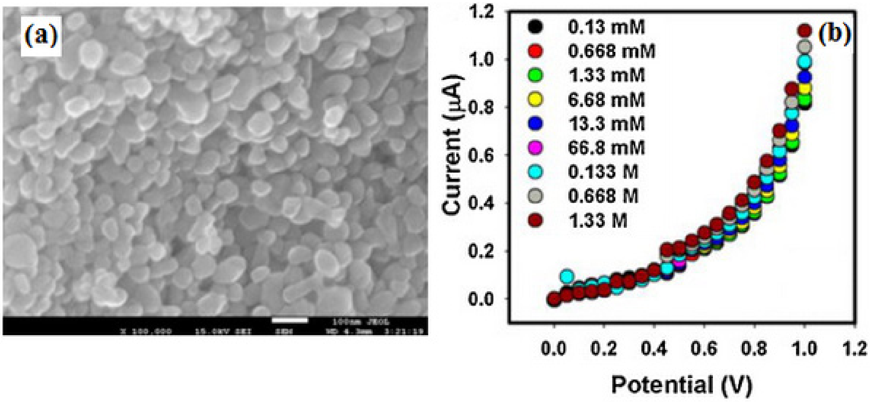


Fig. 3. **(a)** FESEM images of ZnO nanoparticles (reproduced with permission [12]) (b) I–V curves of ZnO acetone sensors (reproduced with permission [12]).

Nanocrystalline thin film, originated as a result of annealing and long term sintering, has been investigated by many research groups for acetone sensor application. An ultra low acetone detection based on nanocrystalline WO_3 thin film was investigated through simple sol-gel method by Shi et.al.[17]. WO_3 thin film after sintering at 500ºC (average crystalline size of 25-100 nm), was able to detect acetone down to 0.05 ppm at 300ºC, whereas employing same method Hazra et al. reported TiO_2 nanoparticle with crystalline size (annealed at 450ºC) of 7.5nm which was capable of detecting acetone at 100-200°C with fast response and recovery (7 s and 13 s) [18]. Reduced particle size of SnO_2 thin films derived by sol-gel method, detect acetone at room temperature at the level of 2-8 ppm [19]. Yoon et. al.[20] first reported work function effect of plasma treated ZnO thin film for acetone sensing. They demonstrated that higher work function (5.51 eV) of plasma treated ZnO has better acetone sensing property than as deposited ZnO thin film. Firstly, positive effect of plasma treatment, secondly, the generation of the dangling bonds on ZnO surface and an increase in the binding energy of oxygen atoms are possibly responsible for the enhanced work function.

3.2 Nanorod Based Sensors

Amongst various nanoforms, nanorods has attracted immense attention owing to the relatively ease of fabrication and high surface to volume ratio. Though most of the nanorod based acetone sensors are ZnO based, the same with Co_3O_4 and SnO_2 has also been investigated. Different forms of nanorod structure (ladder, urchin and hierarchical shot rod) were grown by deposition of a seed layer followed by either hydrothermal, solvothermal or CVD method as shown in Fig 4 (a, b, c).

Low temperature hydrothermally grown vertically aligned/ along <001> directed ZnO nanorod with diameter in the range of 30-200 nm (as shown in Fig 5(a)) shows high dynamic range towards acetone with very first response/recovery of 5s/15s at optimum temperature about ~300-370ºC [21-23]. The response magnitude towards 100 ppm acetone and other analytes are shown in Fig 5(b). Almost at same temperature Nguyen et al.[24] reported reproducible and highly crystalline mesoporous Co_3O_4 nanorods for effective acetone detection. The higher surface area (83 m^2 gm^{-1}) of mesopore assembled Co_3O_4 nanorods (as shown in Fig 5 (c)) showed 7 folds higher sensitivity ~1850% (as shown in Fig 5 (d)) towards acetone than that of nanoparticle one. Barreca, and his coworkers [25] developed scalable catalyst free urchin like ZnO nanorod arrays on nanocrystalline Al_2O_3 substrate by plasma enhanced chemical vapor deposition to monitor acetone in breath and food. The sensor shows a sensitivity of 246 for 100 ppm acetone at an operating temperature of 300ºC.

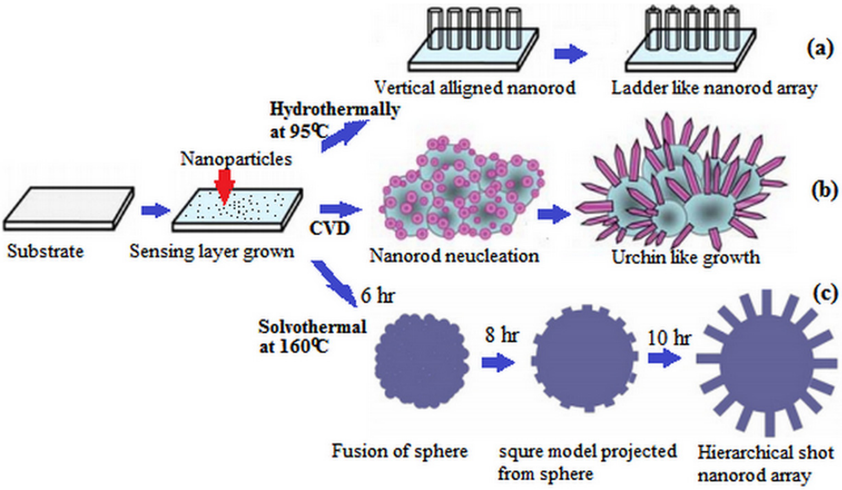


Fig. 4. Formation of (a) hierarchical ZnO nanorods (following reference [21]). (b) Urchin like ZnO nanorod [following reference 25] (c) shot rods from coralloid SnO_2 (following ref. [29]).

It was reported that effect of doping/modification with noble metal catalyst on nanorod surfaces improves the acetone sensing performance [6, 26-27]. The doped metal act as catalyst on nanorod surface which dramatically increases the adsorption and desorption kinetics on the metal oxide lattice site. Li et al. [27], reported that Sn doped on ZnO nanorod improve the response/recovery time about 66% and 177% than that with pure ZnO nanorod whereas La^{3+} doped ZnO nanorod [26] offer R_a/R_g value of 1826 for 1000 ppm acetone at 425ºC. Banerjee et al [6] reported the effect of Pd modification on hexagonal ZnO nanorod. It was found that Pd modified ZnO hexagonal nanorod reduces the optimum temperature from 350 to 300ºC with appreciable response magnitude of 99%. It has already been established that one dimensional nanostructure with porous form enhances sensors performance drastically owing to presence of pores (interaction sites) on the nanorod surface. Long term annealing/calcinations at proper temperature may form pore on nanorod surface. Annealing at 420ºC for two hour changes Fe_2O_3 nanorods [28] to a randomly packed structure with plenty of pores, into which the acetone vapor can penetrate easily at an operating temperature of 200ºC. Based on calcinations time at 160ºC, Fang et al. [29] reported hierarchical growth of SnO_2 nanorod array. The authors demonstrated the growth of non uniform sphere like structure after 30 min of calcinations. A number of short rods formed around the sphere after 6 hr and the products obtained after 8h were entirely a square model that are consisting of nanorods projected from the surface. Finally, they obtained a hierarchical structure that look entirely like a coral with short nanorods growing on the surface as shown in Fig 4 (c).

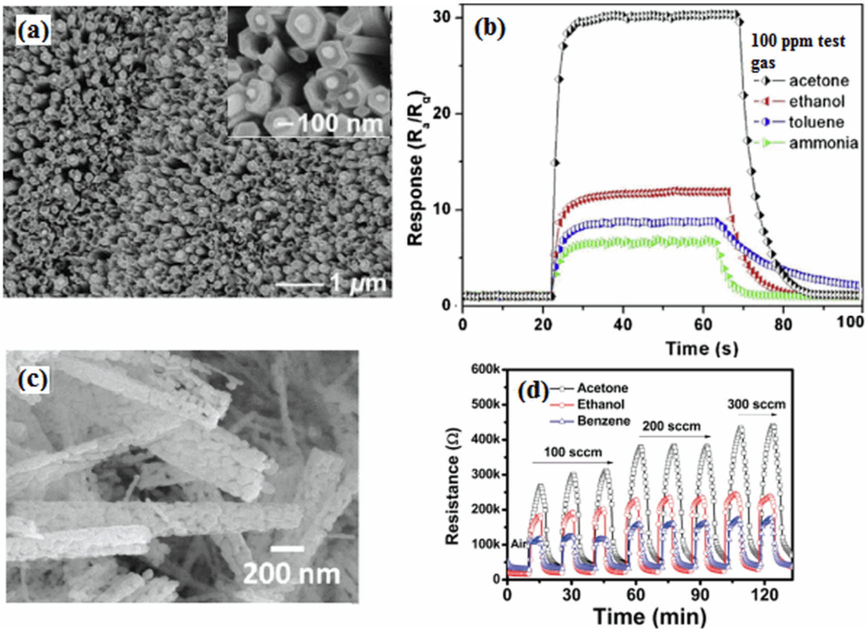


Fig. 5. (a) FESEM image of ZnO nanorods. [21] (b) Transient response of the ZnO nanorod for 100ppm various gases at 300ºC [21]. (c) FESEM image of calcined porous Co_3O_4 nanorod (Reproduced with permission [24]) (d) Dynamic response of porous Co_3O_4 nanorods to acetone, ethanol, and benzene measured at 300ºC (Reproduced with permission [24]).

3.3 Nanowire Based Sensor

Metal oxide nanowires (NWs) having the length of ~1-100nm [30-38] exhibit very special sensing characteristics e.g. (i) high sensitivity and fast response (ii) good selectivity and stability (iii) low operating temperature (iv) low power consumption [30-33]. Till date many nanowire based gas sensor have been investigated based on SnO_2, ZnO, In_2O_3, for acetone detection.

3.3.1 SnO_2 Based Nanowire Sensor

SnO_2 nanowire sensor is one of the potential oxide sensor investigated during the last couple of decades. It is well established [8] that when the diameter of the grain is close to or less than twice the thickness of the space-charge layer (shown in Fig 2), the gas sensing properties such as sensitivity, transient response, will be improved remarkably. Many researchers reported on precise control over grain size based on doping, annealing and composite material for enhancing gas sensing property [26-29]. In case of nanowire structure, the researcher concentrates on precise control over nanowires length and their orientation and distribution (as the synthesized nanowire remains in non uniform manner) rather than grain size. Therefore, for sensing application, it is required to maintain a continuous path (through NW) between two electrodes and a proper alignment of nanowires [30-33]. Also it is essential to reduce the angle between two nanowires which decreases the barrier height and hence increases the e^- transport phenomena.

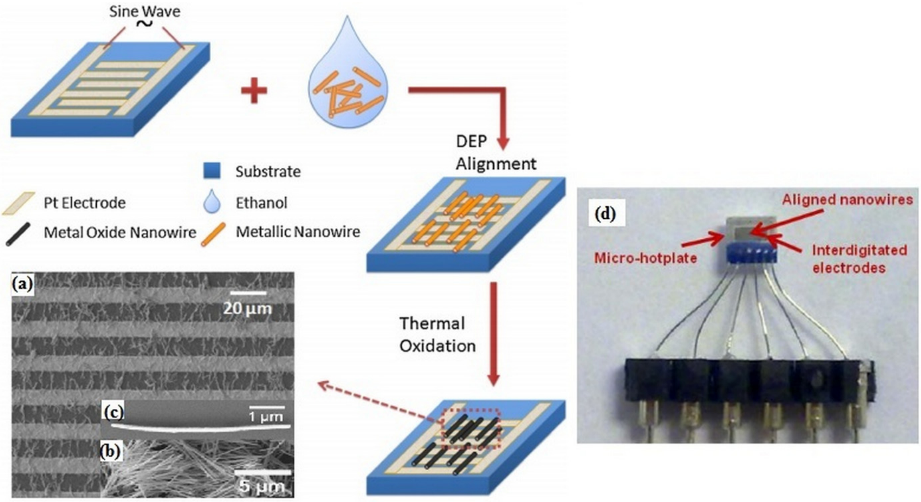


Fig. 6. Schematic diagram of nanowire sensor chip fabrication procedure and top-view of the sensor chip surface, (a) SEM image as-fabricated tin–copper metallic nanowires, (bottom left), (b) SEM images of as-fabricated tin–copper metallic nanowires (inset) (c) a zoom-in view of a single nanowire (inset) (reproduced with permission [31]), (d)) Overview of the integrated sensor chip on micro-hotplate (reproduced with permission [30]).

Dielectrophoretic (DEP) assembly technique is the most popular technique to grow aligned nanowires [30-32]. The DEP technique is diagrammatically shown in Fig 6. An integrated micro hotplate was used for temperature control up to 440ºC as shown in Fig 6 (d). The randomly oriented nanowires were pressurized to align them in vertical direction on the electrode fingers. The oxidation process transforms metal nanowires to metal oxide nanowires and subsequently, two step thermal treatments reduces the gaps between wires and electrodes. Fig 6 (a, b, c) shows the SEM images of the thermally treated nanowire. Though DEP assembly of SnO_2 nanowire [31-32] have not shown good response (0.7% and 0.6%) towards acetone but additional post assembly bonding process in ref [30], has found to be effective in joining of nanowires and hence improve the response magnitude (17%). Employing UV illumination Chen et al.[33] were able to reduce the contact angel of the thermally evaporated zigzag shaped Zn_2SnO_4 nanowires from 122° to 0°. They demonstrate that the zigzag Zn_2SnO_4 nanowire structure is capable of acetone sensing down to 1 ppm level possibly due to the zigzag nature of nanowires.

3.3.2 ZnO Nanowire Based Sensor

ZnO based nanowires, synthesized by RF magnetron sputtering [1], vapor phase transport deposition [34], electrochemical anodization [35] and hydrothermal [36] route have been investigated by many researchers. Recently Hsueh et al. [34] reported growth procedure of ladder like ZnO nanowire arrays. The authors suggested that apparent number of ladder can be increased by increasing the number of vapor phase transport deposition (VPTD) processes. The novel method allows precise control over the nanowire structure which is considered to be a significant achievement since last decade. ZnO nanowire (NW) array with high density and homogeneity by electrochemical anodization was studied by Galstyan et al. [35]. The develop nanowire diameter was in the range from the 20 to 110 nm with length more than 1 µm. They varied the thickness of NWs by tuning the anodization parameters. These anodized ZnO nanowire detects acetone down to 25 ppm at 500ºC.

Few research groups have tried to improve nanowire based sensor characteristics through injection of noble metal e.g. Pt, Au and Ag. For example Gua et al.[36] and Chang et al.[1] investigated the effect of Au nanoparticle injection on ZnO nanowire surfaces. Their result raveled that Au injection on the ZnO nanowire increases the sensitivity from 6.5 to 9.5 for 100 ppm acetone [36] and 18.5% to 82.5% at 300ºC for 200 ppm acetone [1]. The authors demonstrated that Schottky barrier contact between Au/ZnO interface and the catalytic effect of Au are the main reasons of fast acetone sensing.

3.3.3 Other Oxide Nanowire Sensor

Other oxide nanowire sensor such as In_2O_3, IrO_2 was developed by Vomiero et al.[37] and Hullavarad et al. [38]. In_2O_3 sensor through vapor solid (VS) or vapor liquid phase (VLS) process exhibited response about to 7 for 25 ppm acetone at an optimum temperature 400ºC, whereas IrO_2 nanowire sensor showed ultrafast response and recovery towards ethanol, acetone, and propanol, and are measured to be 3 (2.9 s), 4.7 (3.9 s), and 5.1 s (4 s), respectively.

3.4 Nanosheet Based Sensor

Two dimensional (2D) and three dimensional (3D) nanosheets originated through either exfoliation of layered compounds or assembly of the same has also been widely investigated during last few years. One of the easiest routes to create isolated single layer (oxide nanosheets) is chemical exfoliation where single layer is separated from the bunch as shown in Fig 7 (b) [39-42]. In 1996, Sasaki et al. [43] first reported on exfoliation of single titanate nanosheets from the layered titanates by such method. 2D building blocks (exfoliated nanosheets) can be utilized to create other nanostructure like sphere-like assemblies (3D nanosheets) as shown in Fig 7 (a). In the case of metal oxides, protonation followed by intercalation usually resulted in electrostatic repulsions that facilitated exfoliation [43]. Variation of crystallinity as well as porosity of the exfoliated nanosheets has also been investigated.

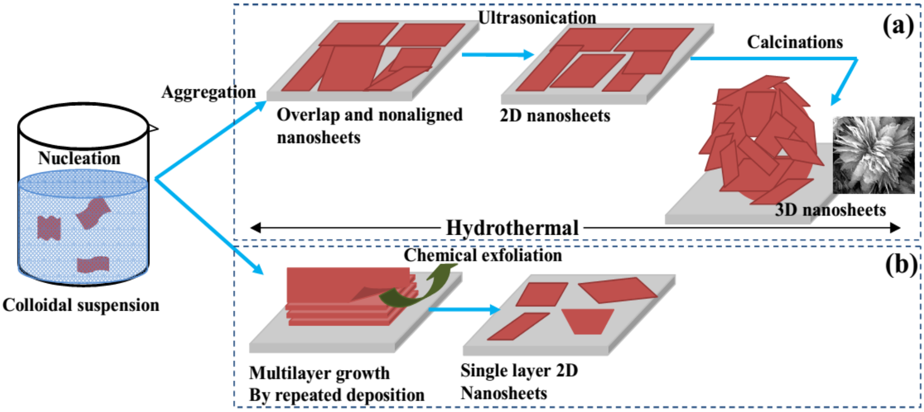


Fig. 7. (a) Schematic diagram of (a) 2D and 3D nanosheets synthesis though hydrothermal method (b) multilayer nanosheets growth by repeated deposition and chemical exfoliation for single layer nanosheets.

3.4.1 2D Nanosheets

Usually a bunch of nanosheets is first developed by material colloidal suspension followed by suitable deposition techniques [39] and subsequently single nanosheet is exfoliated from the bunch as shown in Fig 7 (b). Chang et al. [39] synthesized single crystalline WO_3 nanosheets by a very simple sonochemical method. At room temperature the sensors, made of such nanosheets, exhibited a response magnitude of 5.7 for 100ppm acetone. The leaf like potassium-doped thermally oxidized tungsten oxide nanosheets [40] exhibited higher response (~98) towards acetone at 300ºC. This is possibly attributed to the large number of potassium ions diffused into the tungsten oxide surface. Acetone sensing properties of ZnO nanosheets with porous structure was developed by the group of Guo[41] and XiaO [42]. Porous nanosheets increase the response magnitude than that of non porous one as reflected in the work of XiaO [42]. Pd modification of porous ZnO nanosheets enhanced the gas sensitivity by 3 fold than its unmodified counterpart at 500 ppm acetone.

3.4.2 3D Nanosheets

3D nanosheets, formed through assemblies of 2D nanosheets with different orientation, evolve different structures e.g., spherical, flowerlike, depending on the chemical synthesis procedure. Most of the researchers followed hydrothermal synthesis route for 3D nanosheets formation. For instance, Fan et al. [44] hydrothermally synthesized ZnO nanosheets with thickness 10 nm. The 3D flowerlike nanosheets offered high sensitivity (~30) towards 50 ppm acetone at 360ºC with a response/recovery time of 15s/30s respectively. Hierarchical porous nanosheets structures such as microsphere [45], multilayered microsphere [46], flowers [47-48], micro lotus [49], was investigated for acetone sensing application. It was found that all the reported hierarchal structure followed either hydrothermal or self assembly method. The probable growth mechanism might be like; the small solid particles/small clusters nucleate in the prepared colloidal solution to form nanoparticle which subsequently agglomerate to nanoparticles. Annealing of the nanosheet at an elevated temperature make a porous 2D nanosheet. Decomposition of the organic components during the annealing process is the main reason for the porous structure formation. Porous 2D nanosheets either form 3D nanosheets which on calcinations beyond 500ºC produces porous 3D nanosheets as shown in Fig 7 (a).

3.5 Hollow Sphere Based Sensor

In 1984, Kawalski et al. [50] first developed the hollow sphere structure. Further modification of basic hollowsphere like porous, hierarchical, rattle type hollow sphere have been developed by other researchers [51-53] employing various chemical and physiochemical processes such as sol-gel, template, hydrothermal etc. Hollow sphere are generated either by one step process like sol-gel, hydrothermal method, or by two step method employing templates (carbon template method, bacterial template method, polymer template method etc).

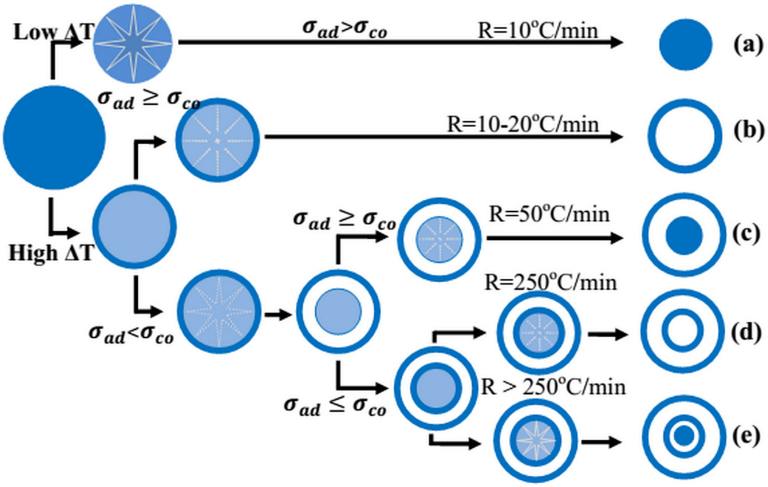


Fig. 8. The illustration for the formation mechanism of (a) solid (b) hollow (c) core-in-hollow-wall (d) double-wall hollow (e) and core-in-double-hollow-wall spheres (following ref. [54])

The wide variety of hollow micro/nanospheres with controllable interior structures varied from (a) solid (b) hollow (c) core-in-hollow-wall (d) double-wall hollow (e) and core-in-double-hollow-wall spheres as shown in Fig 8 (a-e). Heterogeneous contraction method based on cohesive force (σ_{co}), adhesive force (σ_{ad}) and non-equilibrium heat treatment [54] evolve the aforesaid varieties. As the detail formation mechanism of these hollowsphere structures has been discussed by Guan et.al. [54], in this chapter we avoid repetition of the same. At relatively low temperature with homogeneously heated ambient, minimization of surface energy with the mass loss of the organic component converts the 'gel sphere' into a solid one as shown in Fig 8 (a). However, at elevated temperature the outer surface of the gel sphere forms a dense shell whereas the interior still remains in gel form. After that, exterior of the gel sphere does not change much, but due to calcinations and heterogeneous contraction, the inner viscoelastic gel shrinks to different forms depicted in Fig 8 (b-e). Cohesive force (σ_{co}) and the adhesive force (σ_{ad}) at the interface layer between the inner gel and the dense shell acts in the opposite direction, as a result σ_{ad} resists the inward shrinkage of the inner gel while σ_{co} makes the inner gel further shrink inwards with occurrence of mass loss.

Recently Park et. al.[55] synthesized Co_3O_4 hollow spheres using solvothermal method to detect acetone. Morphology studies revealed that Co_3O_4 sphere has outer diameter of 200–300 nm and thickness of ~40 nm. Co_3O_4 hollow nanospheres detect acetone at 100ºC, with very fast response and recovery of 1s and 4s, respectively for 1000ppm acetone. Zhong et al. [56] reported an excellent acetone sensing using hollow core–shell η-Fe_2O_3 with the assistance of Polymer (PVP) as soft template. The η-Fe_2O_3 based hollowsphere showed sensitivity of 15.69 for 0.5 ppm acetone. Single core Ce doped SnO_2 hollow sphere prepared by polymer template method was also investigated by Song et al. [57]. The Ce doped SnO_2 hollowsphere showed 2.5 times higher sensitivity than pure SnO_2. Song and co workers [58] synthesize ZnO based multi core hollow sphere through single pot hydrothermal method. Hollow sphere having diameter of 1-2 μm with multi core detected acetone at 300ºC with response/recovery of 10s/10s for 500 ppm acetone.

Porous Hollow sphere based on sintered SnO_2 to detect acetone was studied by Ma et al.[59]. The authors demonstrated that proper amount of H_2O_2 make SnO_2 hollow sphere porous in nature. The average pore size was about to 6.2 nm. The contrast between the boundary and the center of the spheres confirms their hollow nature. SnO_2 nanosphere shows ultra fast response and recovery time (1s and 3s) towards acetone with high sensitivity about 18.5. Wang et. al. [60] synthesized porous α-Fe_2O_3 hollow sphere through carbon template method. Nanoparticles encapsulated porous α-Fe_2O_3 hollow sphere showed good dynamic range (10-1000ppm) for acetone at 350ºC operating temperature. Recently Zang et al.[51] reported ZnO hollow sphere, synthesized by environment friendly bacterial template method. They found that a small amount of Co_2O_3 incorporation on ZnO surface, shortened the response time from 10 s to 1.2 s.

There are very few reports on hierarchical hollow sphere for acetone detection. Hydrothermal synthesis of Au loaded NiO hierarchical hollowsphere was developed by Wang et.al. [52]. Such hollowsphere reduces response time from 21s to 1.7s. This is possibly attributed to easy diffusion of acetone on the hollow sphere surface encapsulated with nanosheets. Nanoparticles encapsulated in rattle type WO_3 hollow sphere were studied by Li et. al.[53]. Carbon template WO_3 hollow sphere offered fast recovery of 10s with high dynamic range (50-5000ppm) for acetone.

3.6 Nanotube Based Sensor

Unlike other nanoforms of oxides nanotube based acetone sensors are comparatively less investigated [61]. TiO_2, ZnO and Fe_2O_3 nanotube have been investigated for acetone sensing through either chemical anodization or Electrospinning route.

In 2008, Jianjun et.al. [61] first investigated the titania nanotube array for acetone sensing through a simple electrochemical anodization method. TiO_2 nanotubes lengths were 1.2μm, with diameter of 60nm and wall thickness of 30 nm were developed by 2 hours of anodization at 1M HNO_3, 1M NaOH and 0.5wt% HF electrolytes. The authors demonstrated that the thin wall thickness and long length of nanotubes might be the main reason of first acetone sensing at low temperature. TiO_2 nanotubes showed high sensitivity of 298 for 120 ppm acetone at 270ºC. But the most problematic issue regarding this sensor was their slow response and recovery time (>200s). Galstyan et al.[62] reported Nb doped TiO_2 nanotube prepared through anodization method. Their sensor brought down the optimum working temperature for acetone to 200ºC. The sensors were capable of detecting acetone down to 1 ppm.

Electrospinning is another method to grow nanotube structure. Recently, two groups Yu et al.[63] and Shan et al. [64] fabricated ZnO, Fe_2O_3 nanotubes through electrospinning method. ZnO nanotubes developed by Yu et al. [63] exhibited a high dynamic range of 1-3000 ppm with fast response and recovery of 5s and 10s respectively towards acetone. A faster response time of 3s for 50 ppm acetone was observed for La doped Fe_2O_3 nanotubes [64].

Table 1. Acetone sensing performance of different nanostructure based solid state sensors

Structure	Oxide	Method	^{a}Temp ^{o}C	^{b}Conc. range ppm	Sensitivity formula	S_{max}%	^{c}Res (s)	^{d}Rec (s)	Ref
Thin film	ZnO	Spray pyrolysis	300	25-100	(Ra-Rg)/Ra	96.8	15	48	[12]
	WO_3	Spray pyrolysis	400	20-600ppb	Ra/Rg	4.63	78	84	[14]
	WO_3	Sol-gel	300	0.05-5	Rg/Ra	204	<30	< 30	[17]
	SnO_2	Sol-gel	25	2-10	Rg/Ra	4	<30	< 60	[19]
	ZnO	Reactive sputtering	250	500-1500	(Ra-Rg)/Ra	100	--	--	[20]
Nanorod	ZnO	Hydrothermal	300	1-2000	Ra/Rg	30.4	5	15	[21]
	ZnO	CVD	300	1-1000	Ia/Ig	246	--	--	[25]
	ZnO	Hydrothermal	320	10-3000	Ra/Rg	25	1	3s	[22]
	ZnO	Solvothermal	385	0.01-1000	Ra/Rg	1826	16	3s	[26]
	ZnO	Deposition	300	400-4000	Ra/Rg	6.15	110	181	[27]
	ZnO	Hydrothermal	370	3000	Ra/Rg	3.2	3	5s	[23]
	Fe_2O_3	Hydrolysis	200	5-500	Va/Vg	4.75	<20	< 20	[28]
	Co_3O_4	Hydrothermal	--	7.5-22.5%	(Ra-Rg)/Rg	1850	24	180	[24]
	SnO_2	Solvothermal	25	10-50	Ra/Rg	4.25	15	20	[29]
Nanowire	SnO_2	Electrodeposition	200	100-1500	(Ra-Rg)/Ra	17.8	10	240	[30]
	SnO_2	Electrodeposition	380	500	(Ra-Rg)/Ra	0.7	--	---	[31]
	SnO_2	Electrodeposition	450	3.4-88.4	(Ra-Rg)/Ra	0.65	--	--	[32]
	Zn_2SnO_4	evaporation	--	1-200	Ra/Rg	27.5	49	18	[33]
	ZnO	Anodization	500	25-800	--	2	500	<500	[35]
	ZnO	Hydrothermal	320	200	Ra/Rg	9	4	15	[34]
	ZnO	RF sputtering	300	5-200	(Ra-Rg)/Ra	82.5	25	43	[1]
	In_2O_3	VS and VLS	400	25-100	--	7	120	<300	[37]
	IrO_2	Direct vapor phase	350	25-80	Ra/Rg	3.8	4.7	3.9	[38]
Nanosheet	WO_3	Sonochemical	25	100	Ra/Rg	5.7	7	5	[39]
	WO_3	Thermal	300	10-60	Ra/Rg	98	30	50	[40]
	ZnO	oxidation	420	10-500	Ra/Rg	70	9	6	[42]
	ZnO	Solvothermal	360	50	Ra/Rg	31	13	30	[44]
	ZnO	Hydrothermal	--	100-500	Ra/Rg	43	3	8	[49]
	ZnO	Hydrothermal	400	5-500	$^*1+AP^\beta$	32	5	28	[48]
	ZnO	Self assembly Hydrothermal	300	50-250	Ra/Rg	16	5	6	[47]
Hollowsphere	Co_3O_4	Solvothermal	100	10-1000	Ra/Rg	7.5	1-3	4-8	[55]
	Fe_2O_3	Calcinations	370	0.5-50	Ra/Rg	15.6	--	--	[56]
	SnO_2	Polymer template	250	50-1000	Ra/Rg	12.5	18	7	[57]
	ZnO	Hydrothermal	300	10-1000	Ra/Rg	12	10	<10	[58]
	ZnO	Bacterial	300	10-500	Ra/Rg	49	1.9	48	[51]
	SnO_2	template	250	1-500	Ra/Rg	18.5	1	2	[59]
	Fe_2O_3	Hydrothermal	350	10-1000	Rg/Ra	5.9	5	<30	[60]
	NiO	Carbon Template	350	5-100	Rg/Ra	15.3	1.7	11	[52]
	WO_3	Hydrothermal Carbon microsphere	240	50-5000	--	16	50	10	[53]
Nanotube	TiO_2	Anodization	220	80-160	Rg/Ra	298	200	200	[61]
	TiO_2	anodization	200	1-1000	Ia/Ig	--	>60	>60	[62]
	ZnO	Electro spinning	500	1-3000	Rg/Ra	12.3	5	10	[63]
	Fe_2O_3	Electro spinning	240	1-50000	Rg/Ra	26.2	3	10	[64]

^{a}Temp=temperature, ^{b}Conc= concentration, ^{c}Res=response time, ^{d}Rec=recovery tme

4 Conclusion

In this chapter concerns encapsulation the nanostructuring effect of semiconducting metal oxides on solid state vapor (viz. acetone) sensor performance. Nanodimensional semiconductor structures, owing to it unique structural and electronic transport related advantages, offer improved gas/vapor sensing characteristics in terms of pivotal sensing parameters like, sensitivity, operating temperatures, response time, recovery time and selectivity. It has already been established that gas/vapor sensing performance of the one dimensional structures (viz. nanorod, hollowsphere and nanotube) is the most attractive and efficient ones compared to its two (nanosheets) or three dimensional (bulk) counterparts. Future trends in sensor miniaturization and integration with electronics will require processing compatibility with silicon-based technologies. In view of this, the present article has discussed the underlying mechanisms governing the evolution of metal oxide nanoforms under different process routes and correlation of the structural attributes with the semiconducting behavior with special reference to the sensing of acetone vapor. In addition to processing nanostructured oxides, more fundamental work is needed to understand the role of nanostructured oxide materials on gas adsorption and subsequent electron transport mechanism. An area where greater improvements are necessary is gas/vapor sensor selectivity.

Acknowledgement. This book chapter was prepared while concurrently working on the full time basis on the grant given by Council of Scientific and Industrial Research (CSIR0, India (Sanction letter No.22 (0518)/10/EMR-II), Department of Science and Technology (DST) (Fast Track Scheme for Young Scientist: Sanction Letter No.SR/FTP/ ETA-041/2011), AICTE Career Award for Young Teachers (Letter No. 1-51/RIFD/CA/1/2011-12), Government of India and Indian National Science Academy (INSA). B. Bhowmik thankfully acknowledges COE, TEQIP –II, IIEST for his fellowship for pursuing PhD program. A. Hazra gratefully acknowledges Department of Science and Technology (DST), Govt. of India for his INSPIRE Fellowship.

References

1. Chang, S.J., Hsueh, T.-J., Chen, I.-C., Hsieh, S.F., Chang, S.P., Hsu, C.L., Lin, Y.R., Huang, B.R.: Highly Sensitive ZnO Nanowire Acetone Vapor Sensor With Au Adsorption. IEEE Transactions on Nanotechnology 7(6) (2008)
2. Flowers, L., Broder, M.W., Forsyth, C.: Toxicological review of acetone. EPA (May 2003)
3. Acetone: Health Information Summary, New Hampshire, Department of environmental service (2005)
4. Righettoni, M., Tricoli, A., Pratsinis, S.E.: Si: WO3Sensors for Noninvasive Diabetes Diagnosis by Breath Analysis. In: IEEE Sensors 2010 Conference (2010)
5. Bhowmik, B., Dutta, K., Banerjee, N., Hazra, A., Bhattacharyya, P.: Low Temperature Acetone Sensor Based on Sol-gel Grown Nano TiO2Thin Film. In: ICE-CCN, pp. 553–557 (2013)

6. Banerjee, N., Bhowmik, B., Sarkar, S., Sarkar, C.K., Bhattacharyya, P.: Anomalous recovery characteristics of Pd modified ZnO nanorod based acetone sensor. J. Nanosci. Nanotech. 13, 1–8 (2013)
7. Bhattacharyya, P., Basu, S.: Nanostructured Ceramic Materials for Chemical Sensors: Present Status and Future Prospects. Trans. Ind. Ceram. Soc. 69, 1–23 (2010)
8. Xu, C.N., Tamaki, J., Miura, N., Yamazoe, N.: N. Grain-size effects on gas sensitivity of porous SnO2-based elements. Sens. Actuat. B 3, 147–155 (1991)
9. Basu, S., Basu, P.K.: Nanocrystalline Metal Oxides for Methane Sensors: Role of Noble Metals. Journal of Sensors (2009), doi:10.1155/2009/861968
10. Zhang, H., Penn, R.L., Hamers, R.J., Banfield, J.F.: Enhanced adsorption of molecules on surfaces of nanocrystalline particles. J. Physical Chemistry B 103(22), 4656–4662 (1999)
11. Korotcenkov, G.: Metal oxides for solid-state gas sensors: What determines our choice? Materials Science and Engineering B 139, 1–23 (2007)
12. Khan, S.B., Faisal, M., Rahman, M.M., Jamal, A.: Low-temperature growth of ZnO nanoparticles: Photocatalyst and acetone sensor. Talanta 85, 943–949 (2011)
13. Prajapati, C.S., Sahay, P.P.: Influence of In doping on the structural, optical and acetone sensing properties of ZnO nanoparticulate thin films. Materials Science in Semiconductor Processing (2012), doi:10.1016/j.mssp.2012.04.015 (in press)
14. Righettoni, M., Tricoli, A., Pratsinis, S.E.: Si: WO3 Sensors for Highly Selective Detection of Acetone for Easy Diagnosis of Diabetes by Breath Analysis. Anal. Chem. 82, 3581–3587 (2010)
15. Ge, C., Xie, C., Cai, S.: Preparation and gas-sensing properties of Ce-doped ZnO thin-film sensors by dip-coating. Materials Science and Engineering B 137, 53–58 (2007)
16. Huang, G.G., Wang, C.T., Tang, H.T., Huang, Y.S., Yang, J.: ZnO Nanoparticle-Modified Infrared Internal Reflection Elements for Selective Detection of Volatile Organic Compounds. Anal. Chem. 78, 2397–2404 (2006)
17. Shi, J., Hua, G., Sun, Y., Gengb, M., Wua, J., Liu, Y., Gea, M., Tao, J., Cao, M., Dai, N.: WO3 nanocrystals: Synthesis and application in highly sensitive detection of acetone. Sensors and Actuators B 156, 820–824 (2011)
18. Hazra, A., Bhowmik, B., Dutta, K., Bhattacharyya, P.: Low Temperature Low ppm Acetone Detection by Pd/TiO2/p-Si Metal-Insulator-Semiconductor Devices. In: International Conference on Sensing Technology (ICST 2013), pp. 396–400 (2013), doi:10.1109/ICSensT.2013.6727683
19. Jie, Z., Hua, H.L., Shan, G., Hui, Z., Gui, Z.J.: Alcohols and acetone sensing properties of SnO2 thin films deposited by dip-coating. Sensors and Actuators B, 460–464 (2006)
20. Yoon, Y.S., Jee, S.H., Kakati, N., Maiti, J., Kim, D.J., Lee, S.H., Yoon, H.H.: Work function effects of ZnO thin film for acetone gas detection. Ceramics International 38S, S653–S656 (2012)
21. Zeng, Y., Zhang, T., Yuan, M., Kang, M., Lu, G., Wang, R., Fana, H., He, Y., Yang, H.: Growth and selective acetone detection based on ZnO nanorod arrays. Sensors and Actuators B 143, 93–98 (2009)
22. Lou, Z., Feng, Y., Liu, X., Wang, L., Zhang, T.: Acetone sensing properties of hierarchical ZnO urchinlike structures by hydrothermal process, Biomedical Engineering: Applications. Basis and Communications 24(2), 99–103 (2012)
23. Xu, H., Liu, X., Cui, D., Li, M., Jiang, M.: A novel method for improving the performance of ZnO gas sensors. Sensors and Actuators B 114, 301–307 (2006)
24. Nguyen, H., El-Safty, S.A.: Meso- and Macroporous Co3O4 Nanorods for Effective VOC Gas Sensors. J. Phys. Chem. 115, 8466–8474 (2011)

25. Barreca, D., Bekermann, D., Comini, E., Devi, A., Fischer, R.A., Gasparotto, A., Maccato, C., Sada, C., Sberveglieri, G., Tondello, E.: Urchin-like ZnO nanorod arrays for gas sensing applications. Cryst. Eng. Comm. 12, 3419–3421 (2010)
26. Chu, X., Zhu, X., Dong, Y., Ge, X., Zhang, S., Sun, W.: Acetone Sensors Based on La3+ Doped ZnO Nano-rods Prepared by Solvothermal Method. J. Mater. Sci. Technol. 28(3), 200–204 (2012)
27. Li, X., Chang, Y., Long, Y.: Influence of Sn doping on ZnO sensing properties for ethanol and acetone. Materials Science and Engineering C 32, 817–821 (2012)
28. Si, S., Li, C., Wang, X., Peng, Q., Li, Y.: Fe2O3/ZnO core–shell nanorods for gas sensors. Sensors and Actuators B 119, 52–56 (2006)
29. Fang, C., Wang, S., Wang, Q., Liu, J., Geng, B.: Coralloid SnO2 with hierarchical structure and their application as recoverable gas sensors for the detection of benzaldehyde/acetone. Materials Chemistry and Physics 122, 30–34 (2010)
30. Li, X., Chin, E., Sun, H., Kurup, P., Gu, Z.: Fabrication and integration of metal oxide nanowire sensors using dielectrophoretic assembly and improved post-assembly processing. Sensors and Actuators B 148, 404–412 (2010)
31. Li, X., Gua, Z., Cho, J.H., Sun, H., Kurup, P.: Tin–copper mixed metal oxide nanowires: Synthesis and sensor response to chemical vapors. Sensors and Actuators B 158, 199–207 (2011)
32. Cho, J., Li, X., Gu, Z., Kurup, P.: Recognition of Explosive Precursors Using Nanowire Sensor Array and Decision Tree Learning. IEEE Sensors Journal 12(7) (2012)
33. Chen, D., Xu, J., Liang, B., Wang, X., Chen, P.C., Zhou, C., Shen, G.: Electric transport, reversible wettability and chemical sensing of single-crystalline zigzag Zn2SnO4 nanowires. J. Mater. Chem. 21, 17236 (2011)
34. Hsueh, T.J., Chang, S.J., Lin, Y.R., Tsai, S.Y., Chen, I.C., Hsu, C.L.: A Novel Method for the Formation of Ladder-like ZnO Nanowires. Crystal Growth & Design 6(6), 1282–1284 (2006)
35. Galstyan, V., Comini, E., Baratto, C., Faglia, G., Brisotto, M., Bontempi, E., Sberveglieri, G.: Growth and Gas Sensing Properties of Rough ZnO Nanowires. In: IMCS 2012 – The 14th International Meeting on Chemical Sensors (2012)
36. Gua, C., Shanshan, L., Huanga, J., Shi, C., Liu, J.: Preferential growth of long ZnO nanowires and its application in gas sensor. Sensors and Actuators B 177, 453–459 (2013)
37. Vomiero, A., Bianchi, S., Comini, E., Faglia, G., Ferroni, M., Poli, N., Sberveglieri, G.: In2O3 nanowires for gas sensors: morphology and sensing characterization. Thin Solid Films 515, 8356–8359 (2007)
38. Hullavarad, N.V., Hullavarad, S.S.: Direct-Vapor-Phase Grown IrO2Micronanostructures for Ethanol, Acetone, and Propanol Gas Sensor. IEEE Transactions on Nanotechnology 9(5) (2010)
39. Chang, X., Sun, S., Yin, Y.: Green synthesis of tungsten trioxide monohydrate nanosheets as gas sensor. Materials Chemistry and Physics 126, 717–721 (2011)
40. Zhang, B., Liu, J., Guan, S., Wan, Y., Zhang, Y.N., Chen, R.: Synthesis of single-crystalline potassium-doped tungsten oxide nanosheets as high-sensitive gas sensors. J. Alloys and Compounds 439, 55–58 (2007)
41. Guo, W., Liu, T., Zhang, H., Sun, R., Chen, Y., Zeng, W., Wang, Z.: Gas-sensing performance enhancement in ZnO nanostructures by hierarchical morphology. Sensors and Actuators B 166, 492–499 (2012)

42. Xiao, Y., Lu, L., Zhang, A., Zhang, Y., Sun, L., Huo, L., Li, F.: Highly Enhanced Acetone Sensing Performances of Porous and Single Crystalline ZnO Nanosheets: High Percentage of exposed (100) facets working together with surface modification with Pd nanoparticles. ACS Appl. Mater. Interfaces 4, 3797–3804 (2012)
43. Sasaki, T., Watanabe, M., Hashizume, H., Yamada, H., Nakazawa, H.: Macromolecule-like Aspects for a Colloidal Suspension of an Exfoliated Titanate. Pairwise Association of Nanosheets and Dynamic Reassembling Process Initiated from It. J. Am. Chem. Soc. 118, 8329 (1996)
44. Fan, H., Jia, X.: Selective detection of acetone and gasoline by temperature modulation in zinc oxide nanosheets sensors. Solid State Ionics 192, 688–692 (2011)
45. Zhu, G., Xi, C., Xu, H., Zheng, D., Liu, Y., Xu, X., Shen, X.: Hierarchical NiO hollow microspheres assembled from nanosheet-stacked nanoparticles and their application in a gas sensor. RSC Advances 2, 4236–4241 (2012)
46. Xingfu, Z., Zhaolin, H., Yiqun, F., Su, C., Weiping, D., Nanping, X.: Microspheric Organization of Multilayered ZnO Nanosheets with Hierarchically Porous Structures. J. Phys. Chem. C 112, 11722–11728 (2008)
47. Shao, C., Tu, L., Yu, A., Li, B., Zhou, X.: From Zn4(CO3)(OH)6·H2O curling nanopetals to ZnO stretching porous nanosheets:Growth mechanism and gas sensing property. Thin Solid Films 525, 148–153 (2012)
48. Li, J., Fan, H., Jia, X.: Multilayered ZnO Nanosheets with 3D Porous Architectures: Synthesis and Gas Sensing Application. J. Phys. Chem. C 114, 14684–14691 (2010)
49. Yu, A., Qianb, J., Pan, H., Cui, Y., Xu, M., Tu, L., Chai, Q., Zhou, X.: Micro-lotus constructed by Fe-doped ZnO hierarchically porous nanosheets: Preparation, characterization and gas sensing property. Sensors and Actuators B 158, 9–16 (2011)
50. Kowalski, A., Vogel, M., Blankenship, R.M.: US Patent 4,427,836 (1884)
51. Zhang, H., Xu, C., Sheng, P., Chen, Y., Yu, L., Li, Q.: Synthesis of ZnO hollow spheres through a bacteria template method and their gas sensing properties. Sensors and Actuators B 181, 99–103 (2013)
52. Wang, L., Lou, Z., Fei, T., Zhang, T.: Enhanced acetone sensing performances of hierarchical hollow Au-loaded NiO hybrid structures. Sensors and Actuators B 161, 178–183 (2012)
53. Li, X.L., Lou, T.J., Sun, X.-M., Li, Y.D.: Highly Sensitive WO3 Hollow-Sphere Gas Sensors. Inorg. Chem. 43, 5442–5449 (2004)
54. Guan, J., Mou, F., Sun, Z., Shi, W.: Preparation of hollow spheres with controllable interior structures by heterogeneous contraction. Chem. Commun. 46, 6605–6607 (2010)
55. Park, J., Shen, X., Wang, G.: Solvothermal synthesis and gas-sensing performance of Co3O4 hollow nanospheres. Sensors and Actuators B 136, 494–498 (2009)
56. Zhong, J., Cao, C., Liu, Y., Li, Y., Khan, W.S.: Hollow core–shellg-Fe2O3 microspheres with excellent lithium-storage and gas-sensing properties. Chem. Commun. 46, 3869–3871 (2010)
57. Song, P., Wang, Q., Yang, Z.: Preparation, characterization and acetone sensing properties of Ce-doped SnO2 hollow spheres. Sensors and Actuators B 173, 839–846 (2012)
58. Song, P., Wang, Q., Yang, Z.: Acetone sensing characteristics of ZnO hollow spheres prepared by one-pot hydrothermal reaction. Materials Letters 86, 168–170 (2012)
59. Ma, X., Song, H., Guan, C.: Interfacial oxidation–dehydration induced formation of porous SnO2 hollow nanospheres and their gas sensing properties. Sensors and Actuators B 177, 196–204 (2013)
60. Wang, S., Wang, L., Yang, T., Liu, X., Zhang, J., Zhu, B., Zhang, S., Huang, W., Wu, S.: Porous α-Fe2O3 hollow microspheres and their application for acetone sensor. J. Solid State Chemistry 183, 2869–2876 (2010)

61. Jianjun, Y., Huiming, J.: Preparation and acetone sensitive property of titania nanotube arrays (OL) (June 3, 2008), `http://www.Paper.edu.cn/en_releasepaper/content/21925V`
62. Galstyan, V., Comini, E., Faglia, G., Vomiero, A., Borgese, L., Bontempi, E., Sberveglieri, G.: Fabrication and investigation of gas sensing properties of Nb-doped TiO2 nanotubular arrays. Nanotechnology 23, 235706 (2012)
63. Yu, X., Song, F., Zhai, B., Zheng, C., Wang, Y.: Electrospun ZnO nanotubes and its gas sensing applications. Physica E 52, 92–96 (2013)
64. Shan, H., Liu, C., Liu, L., Li, S., Wanga, L., Zhang, X., Bo, X., Chi, X.: Highly sensitive acetone sensors based on La-doped -Fe2O3nanotubes. Sensors and Actuators B 184, 243–247 (2013)

Design of Low-Cost Sensors for Industrial Processes Energy Consumption Measurement: Application to the Gas Flow Consumed by a Boiler

B. Hadid[1], R. Ouvrard[1], L. Le Brusquet[2], T. Poinot[1], E. Etien[1],
F. Sicard[3], and A. Grau[3]

[1] Laboratoire d'Informatique et d'Automatique pour les Systèmes – EA6315 (LIAS),
Université de Poitiers – ENSIP, France
`{baya.hadid,regis.ouvrard,thierry.poinot,erik.etien}@univ-poitiers.fr`
[2] Supélec Sciences des Systèmes – EA4454 (E3S), France
`Laurent.Lebrusquet@supelec.fr`
[3] EDF - R&D, EPI groupe E25, Renardières, France
`retd-epi-chic@edf.fr`

Abstract. The demand for energy is becoming increasingly important, and who says strong demands for energy says rising CO_2 emissions. Everyone agrees that a great part of the energy consumed by industry and households can be saved. The energy savings can take many forms. In addition to the necessity to build equipments more and more energy efficient, it is also necessary to get a clear view of how the energy is used. This obviously involves the implementation of an energy flow measuring system for long lasting optimization solutions. It is precisely in this context that the project CHIC (Low cost industry utilities monitoring systems for energy savings), funded by the French National Research Agency (ANR), emerged. The objective of this project is to develop and test low-cost non-intrusive sensors to monitor and analyze the energy consumption of major flows used in the manufacturing sector (electricity, gas, compressed air). With such sensors, it should be possible to tool up a factory, equipment by equipment, which is not feasible with intrusive sensors. The ultimate goal is the long term consumption monitoring and the detection of the consumption deviations rather than a precise measurement. The measurement accuracy is fixed to 5%. These developments are based on the recent approaches in system identification and parametric estimation. This project, concretely, involves the design of new low-cost sensors in the following areas: current sensors, voltage, power, and gas flow, relying on the international ISO 50001 standard for Energy Management Systems. The work presented in this chapter focuses on the modeling of the gas flow supplied to a boiler in order to implement a soft sensor. This implementation requires the estimation of a mathematical model that expresses the flow rate from the control signal of the solenoid valve and the gas pressure and temperature measurements. Two types of models are studied: LPV (Linear Parameter Varying) model with pressure and temperature as scheduling variables and a non-parametric model based on Gaussian processes.

A. Mason et al. (eds.), *Sensing Technology: Current Status and Future Trends IV*,
Smart Sensors, Measurement and Instrumentation 12, DOI: 10.1007/978-3-319-12898-6_2

Keywords: Soft sensors, Identification, Gaussian process modeling, LPV model, Flow measurement, Boilers, Energy Efficiency, Consumption monitoring.

1 Introduction

The concept of energy efficiency is becoming more and more important in the context of high energy demand. The international standard ISO 50001 represents the desire for saving energy. This standard is based on a preliminary energy audit and implementation of systems for measuring and monitoring to ensure that the objectives are achieved.

In the industrial sector, each investment is made in relation to the expected benefits. The cost of a program to improve energy efficiency must be offset by the gained benefits. Sometimes a project, though promising, is rejected on the basis of the amount of initial capital costs, the implementation requiring a production stop. To foster the acceptance of improved energy efficiency programs, production stops must be kept to a minimum and costs of measures need to be low.

It is in this context that the ANR CHIC project (Low cost industry utilities monitoring systems for energy savings, Fr: CHaînes de mesures Innovantes à bas Coût) was born. The objective is to develop and to test low-cost sensors to monitor and to analyze the energy consumption of the major fluids used in industrial sites (electricity, gas, compressed air). The studied sensors in the ANR CHIC project should allow monitoring of consumption and drift detection consumption. EDF R&D, the initiators of this project gave the objective of achieving a measurement accuracy of about 5%. The project is to develop new sensors (both physical and "soft") at low cost in the following areas: current sensors, voltage sensors, power sensors, gas flow meters. The work presented in this chapter only concerns the study of gas flow measurement.

The objective of this study presented in this chapter is the modeling of a boiler with the aim of developing a "soft" sensor. The concept of soft sensor is to combine measures available or easily achievable and mathematical models which link the measured quantities and the quantities to be determined. This concept is used in various fields and especially in chemical processes [5,7,19], or biological processes [3,6,15,22]. The implementation of the soft sensor is based on a simulation, on an observer or on an inverse method; modeling is then a key point for the measurement quality. Modeling can be based on physical principles, on empirical approaches, or on a combination of both.

The study focuses on installation of gas boiler with a power of $750kW$, located on the *Renardières* site of EDF R&D near Paris, France. For the sake of economy, it is desirable that the soft sensor can easily be installed. The development of a physical model dedicated to a plant is excluded because it would induce a too high cost of development. For this purpose, it is proposed to build black-box behavioral models. In the case of the gas flow measurement, the dynamic behavior of the signal to be modeled is very fast, consequently, the construction of static models is sufficient regarding the objectives of energy monitoring. Two modeling approaches are explored. The first one consists of a parametric model where the parameters depend on the pressure and on temperature, *i.e.* an LPV (*Linear Parameter Varying*) model is estimated [8,23]. The second approach is to estimate a non-parametric model [1]. Developed models allow representing the mass flow of gas in a boiler from the gas pressure, the gas temperature and the solenoid valve control signal.

This chapter is organized in the following way: Section 2 introduces the CHIC project and details its motivations and issue. The experimental bench and the data collection and selection are illustrated in Section 3. Section 4 presents and explains the choice of the exploited models and algorithms. Section 5 is devoted to the experimental results including estimation, validation, and implementation on site. Finally, some conclusions and prospects (Section 6) conclude the chapter.

2 ANR CHIC Project Motivation

2.1 Implementing Energy Efficiency Improvement Programs

There exists a strong potential for energy savings within the French manufacturing industry, probably equally in the european one. This potential is yet to be revealed and exploited.

Nowadays, more and more industries are willing to save energy and therefore are implementing Energy Efficiency improvement programs. Most of such programs rely on national or international standards.

The best tools available nowadays are: the international ISO 50001 “Energy Management Systems” standard and the International Performance Measurement and Verification Protocol (IPMVP). Both rely on the proper measurement of key energy efficiency indicators.

ISO 50001 Standard for Energy Management Systems. The ISO 50001 standard was published in June 2011. It is the result of a collaborative effort of 61 countries, including countries from the European Committee for Standardization.

The ISO 50001 standard specifies requirements for an energy management system that is based on a continuous improvement principle: Plan – Do – Check – Act and then Plan – Do – Check – Act, etc.

- Plan: determine the main energy consuming systems and establish performance targets for it,
- Do: install a metering and monitoring system,
- Check: compare the measured energy performance to the targeted one,
- Act: identify corrective actions and implement Energy Efficiency improvement programs.

The ISO 50001 standard relies on a preliminary energy audit to determine the main energy consuming systems of the plant, and then requires setting performance targets for those systems and implementing metering and monitoring devices to check that these performance targets are respected.

The International Performance Measurement and Verification Protocol (IPMVP). The IPMVP was first released in 1996 and has evolved ever since. It is free to download from the Efficiency Valuation Organization (EVO) web site (http://www.evo-world.org/). EVO is a non-profit organization "dedicated to creating measurement and verification tools to allow efficiency to flourish".

This protocol presents a framework and defines the terms that are to be used for determining the savings one should expect after implementing an Energy Efficiency improvement program. IPMVP focuses on three major issues which are: defining Performance, Performance Measurement and Performance Verification.

Defining Performance is a prerequisite. Performance can be defined at the plant level or at an intermediate level according to the Energy Efficiency improvement program that is to be implemented. For instance, if a compressed air system is to be refurbished, then the protocol can only focus on that specific compressed air system.

Performance Measurement requires installing measuring devices wherever needed, which depends on the Performance Verification protocol that will be used. Performance Verification is the trickiest part of the protocol, since it is impossible to measure energy savings per se. Only energy consumption can be measured. It has to be compared to forecasted energy consumption in order to estimate how much energy has been saved. According to the protocol, the forecasted energy consumption will be calculated using a baseline / reference energy consumption and several adjustment factors that have to be defined. Typical adjustment factors would be the production load factor, the outside temperature, etc.

Measuring Is the Key for Improving Energy Efficiency in the Manufacturing Sector. Whatever the industrial sector considered (food, cement, metal ...), the optimization of a manufacturing plant is a complex process that requires monitoring. To identify and evaluate energy savings, one must get a clear view of how the energy is used. As stated in the ISO 50001 standard, measuring is the first step towards energy consumption awareness and thereafter Energy Efficiency. The ability to measure, monitor and control energy consumption at several key locations in a manufacturing plant is a major prerequisite for any efficient energy management program.

Furthermore, all manufacturing plants are continuously evolving and what was optimized at one moment may not stay optimized for a long time. Once more, measuring is the key to maintain Energy Efficiency throughout time. Energy savings programs, when their impacts are not continuously measured, prove themselves inefficient in the long - or even short - term. Usually several months is a period of time long enough to get into a non optimized situation again. Therefore, continuously measuring energy flows is one of the necessary conditions for long lasting energy-efficient solutions.

In the manufacturing industry, two different types of energy consumption must be distinguished: the one related to the process itself and the one related to the systems that deliver compressed air, vapour, cold water, etc... through the plant. Whereas it is generally very difficult to modify the energy consumption related to a manufacturing process, because this might have a strong impact on production, it is most of the time much easier to optimize auxiliary energy consumption, as long as it is well known and understood and therefore well measured.

Cost-Benefit Analysis for Energy Efficiency Improvement Programs. Within the industrial sector, every investment program, and especially an Energy Efficiency improvement one, is or is not implemented according to its cost-benefit analysis. Unfortunately, most of the time, the implementation of an Energy Efficiency improvement program, because of its mandatory measurement phase, is seen as not acceptable.

Several values need to be measured within an Energy Efficiency improvement program. Some physical parameters, such as temperatures, are easy and not expensive to measure. On the contrary, power and flow rates are either rather expensive or totally impossible to measure, especially if the plant is already in operation.

To measure power for instance, one must cut the power off, for safety reasons, which generally disrupts production. To measure flow rates, one can use some regular flow meters which installation requires cutting through the pipes. This once again generally disrupts production. For many manufacturing plants, it is not acceptable to stop production to install measuring devices.

What penalizes measurements is not technology. It is costs. IPMVP suggests an additional cost for measuring of less than 10 to 15% of the program total energy savings. What penalizes measurements is the additional cost of disrupting production during the installation of the meters, which is most of the time way above the recommended and acceptable 10 to 15%.

2.2 The CHIC Research Project

This research project focuses on creating and experimenting new solutions that are:

- non intrusive,
- low cost,
- plug and play,
- low energy consumption systems,
- efficient and robust, even with noise and perturbations.

The following meters are developed during the CHIC project:

- a physical clamp-on power meter that could be installed around three-conductor electrical cables anywhere in the plant [4],
- a soft power sensor, for industrial electrical furnaces, that derives power from the furnace control signal [13],
- a soft compressed air flow sensor, that derives the air flow rate from the compressor consumed power [14],
- a soft gas flow sensor for boilers, that derives the gas flow rate from its inlet valve opening position.

Every soft sensor must be dedicated to a specific equipment because it relies on extra variables and on mathematical models that are strongly dependent on the physics involved.

The facilities used to test the prototypes are similar to those found in most French manufacturing plants. They may be a little less powerful, but they will allow testing the soft measuring devices in real and industry-like operating situations (with noise, perturbations, etc.).

2.3 A Comparison of Existing Devices

Methodologies and costs for flow rates measurement with actual commercial devices were investigated at the beginning of the project.

Measuring Flow Rates with Actual Commercial Devices. Two types of actual commercial flow meters were evaluated at the beginning of the CHIC project:

- a standard electromagnetic flow meter that is very common in manufacturing plants (Figure 1a), and that needs cutting the pipe to be installed (the same evaluation could have been done with other types of intrusive flow meters (Coriolis, Vortex, etc...) - whatever the technology used, the results would be similar),
- a non intrusive flow meter that is based on ultrasound technology, which installation does not need cutting the pipe (Figure 1b). Taking the meter off the pipe does not require cutting the pipe either.

The total costs for flow rates measurement with these two commercial devices have been evaluated for an exploitation period of 10 years. The assumptions for the pipe and operating conditions were as follows: the diameter of the pipe was of 80 mm, the fluid flowing in the pipe was water, its pressure was below 10 bars, its temperature was comprised between 60°C and 80°C, the speed of the water was of about 7 m/s.

The assumptions for the maintenance of the meters were as follows: the standard electromagnetic flow meter needs maintenance every year, which requires

(a) Standard commercial electromagnetic flow meter

(b) Measuring flow rates with an ultrasound flow meter

Fig. 1. Physical flow rates measuring devices

emptying the pipe and sending the meter for checking (the total labour costs for taking the meter off the pipe and putting it back again on the pipe is of at least 2h per year), the ultrasound meter needs maintenance every 5 years, which requires sending the meter for checking (the total labour costs for taking the meter off the pipe and putting it back again on the pipe is on average of 15 min per year). The results of this analysis are showed in Table 1.

Table 1. Total costs for one flow rate measurement with 2013 commercial products over a 10 years exploitation period (no interest rate)

	Standard electromagnetic flow meter	Ultra sound flow meter
Purchase cost for one meter	500 €	5 000 €
Total costs (sum of purchase, installation and maintenance costs)	6 540 €	5 845 €
Number of times for which the pipe must be emptied and the process stopped during the 10 years period.	10	0

Although intrusive flow meters purchase costs are very low (there is a factor of 10 between the purchase costs of the two flow meters presented above), their exploitation costs are, over a 10 years exploitation period, at the same level as those of non intrusive flow meters. Nevertheless, ultrasound flow meters are seen as very expensive and are mostly dedicated to time-limited energy audits.

Plant managers are usually reluctant to install flow rate meters on existing and operational pipes. Different reasons explain this attitude, according to the type of flow meter:

- non intrusive flow meters are seen as too expensive,
- all other flow meters, which are intrusive, require emptying and cutting the pipe to be installed.

Costs and Benefits of Soft Sensors. The flow soft sensors targeted purchase price has been set to 1 300€, which is an intermediate value between the purchase costs of the two commercial flow meters presented above. If flow soft sensor purchase price were too high, they will be rejected as are nowadays the ultrasound flow meters. It is foreseen that this sensor won't need maintenance, because they are basically software sensors.

Being non intrusive, they should be accepted fairly widely as long as their price remains within acceptable boundaries. Within the project a particular attention is devoted to decrease as much as possible the cost of these flow meters. For example, the targeted total exploitation costs for CHIC power meters over a 20 years period has been set to between 1 409€ and 1 726€, which is the sum of the actual purchase, installation and maintenance costs for existing commercial meters over a 20 years exploitation period. The main advantage of CHIC power sensors over the actual commercial sensors is that there is no need to cut power to install them. This would be a real technological breakthrough because, as we have seen on many manufacturing plants, roughly 30% of the installed power meters do not deliver correct values, as they are not installed properly. It takes time and efforts to make sure that the installed power meters are trustworthy.

It is also very difficult to know the gas consumption of a given industrial gas boiler, since most of the time very few gas meters are installed on industrial sites and the ones that are installed usually measure the site total gas consumption. Once a gas boiler is operational, it is very difficult to convince a plant manager to install a dedicated gas meter because this would cost money and this would imply shutting it down. As a consequence, there is a need for low cost and non intrusive gas flow meters for boilers. The following study focuses particularly on this system.

3 Gas Flow Soft Sensor of a Boiler

3.1 Boiler Description and Instrumentation

The experimental tests are done on a boiler plant which has a nominal power of 750 kW (see Figure 2). A schematic representation of the boiler installation and its instrumentation is shown in Figure 3. The mass flow rate of gas is measured before the gas pressure regulator, *i.e.* in the high pressure part. The gas pressure, temperature and volume flow are measured in the low pressure side. The gas pressure is a manually controlled value but for modeling purposes it is considered as an input. As for the gas temperature, it depends on meteorological conditions.

Fig. 2. Boiler side view

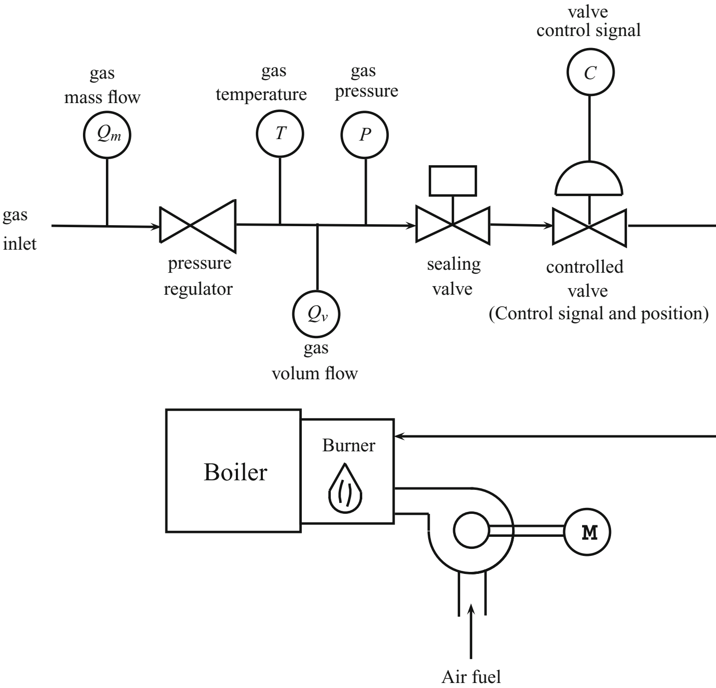


Fig. 3. Schematic view of the boiler

In the models that is presented in the following sections, the mass flow is modeled instead of volume flow, because of the dependency of the gas density to the absolute pressure, and thus, to the atmospheric pressure. It also depends on the fuel gas composition. The mass flow rate depends directly on the control signal fixed by the operator and sent to the electric valve. This control signal is converted into a position signal depending on the type of used regulating valve. In our case, it is a plug valve.

3.2 Experimental Protocol

The experiments realized on this installation consist of a stepwise increase of the electrical control signal C of the modulating gas plug valve. It is fairly easy to collect the valve control signal (which is a value that varies between 0 and 100%). These tests are repeated for different values of pressure p and temperature T of the gas. The pressure is an input data which can be practically controlled by a pressure regulator. It varies between 80 and 200 effective mbar. The gas having a long air routing system, its temperature is thus influenced by weather conditions. The temperature range is from 14 to 33 °C. These experiments are the same ones realized in operation with the aim of calibration of the proposed models.

Figure 4 shows a typical experiment. Figure 5 lists the operating points used in this study.

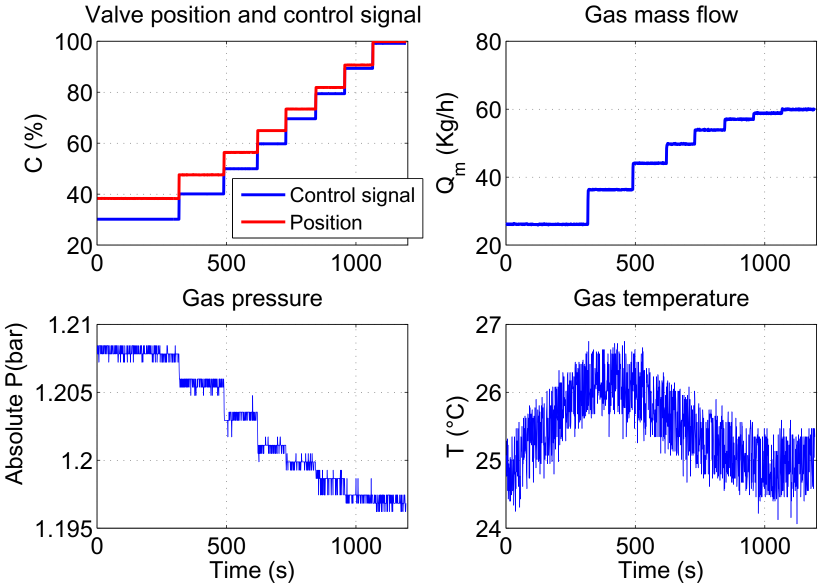


Fig. 4. Typical test at 200 mbar (effective pressure)

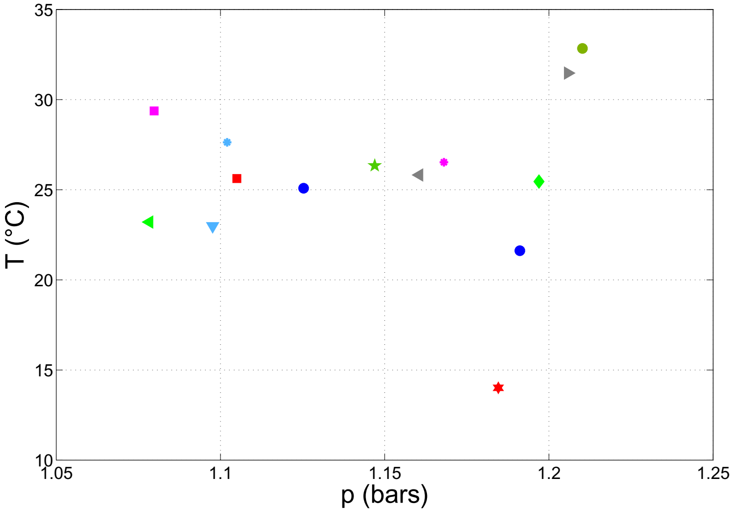


Fig. 5. Used operating points defined by constant values of the gas pressure and temperature

4 Modeling

Two different modeling approaches are studied: a parametric approach and a non-parametric approach.

One concept of engineering, today, is to model the signals and systems to facilitate the study, analysis and control. This model should be easy to estimate, but at the same time, it must be able to reproduce the main characteristics of the studied system. To circumvent the complexity of such models, we can use parametric models, which depend on a fixed number of parameters, and whose structure is pre-determined.

Generally, in the second approach which is the non-parametric theory, it is assumed that the number of parameters that describe the observations distribution is an increasing function of the observations dimension, or that the number of parameters is infinite. The non-parametric modeling studies the problems in which the parameterization is not considered as fixed, but there is a choice between multiple parameterization and the objective is to find one that leads to the most efficient procedures.

4.1 Parametric Modeling

In this modeling, the primary idea is to consider the mass flow Q_m as an output and C as an exogenous input. However, as can be seen in Figure 6, the pressure p influences the flow value and a simple law only based on the control signal cannot provide a good estimation of the flow. Therefore, we propose an LPV model with one scheduling variable p or two scheduling variables p and T.

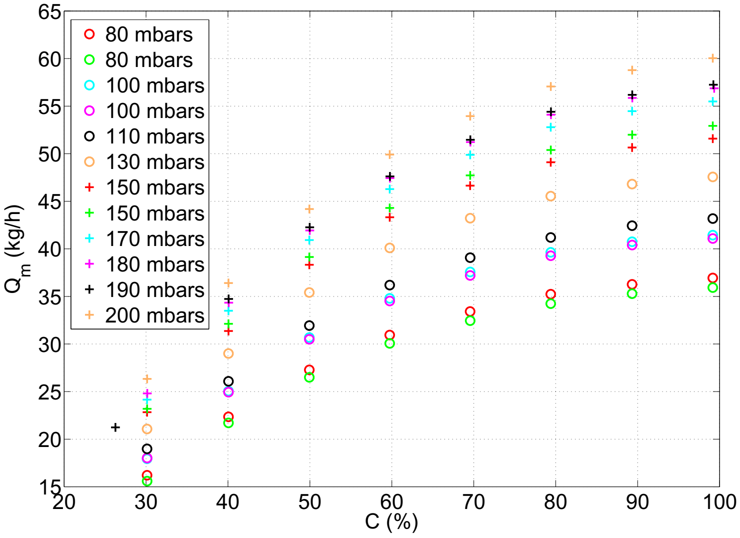


Fig. 6. Flow–control characteristic shape for different gas pressures – measured data

At first, knowing that the pressure has, physically, more influence than the temperature on the gas flow, it is proposed to model the mass flow by taking only the pressure as a scheduling variable. In a second step, we introduce the temperature as a second scheduling variable to see if it has also any influence on the gas flow.

The LPV model, like the other models described in this chapter, is static.

The estimated LPV model is obtained by a local approach [8,23] which consists of:

- estimating local models for different operating points of the scheduling variables,
- and calculate the global LPV model by a local models interpolation.

LPV Model with One Scheduling Variable. Local models are estimated at different operating points defined by a constant pressure and temperature. With regard to the evolution of the gas flow Q_m depending on the control signal C, the chosen model is presented in the following polynomial form:

$$Q_m(t) = \theta_1 C(t)^2 + \theta_2 C(t) + \theta_3 \tag{1}$$

The global LPV model as a function of pressure p is determined from the local models. A fixed pressure value is considered throughout the experiment, and is equal to the average of the level corresponding to the highest control signal value. The choice of a fixed value is justified because the pressure varies slightly around a value set by the user via the pressure regulator. It justifies again to consider the pressure as a scheduling variable to fit to different installations. Thus, the parameters θ_1, θ_2 and θ_3 variations depending on the pressure are represented by the following polynomials:

$$\begin{cases} \theta_1 = \sum\limits_{i=0}^{\text{degP}} \alpha_i p^i \\ \theta_2 = \sum\limits_{i=0}^{\text{degP}} \beta_i p^i \\ \theta_3 = \sum\limits_{i=0}^{\text{degP}} \delta_i p^i \end{cases} \tag{2}$$

where degP represents the polynomial degree of p. The global LPV model becomes:

$$Q_m(t) = \sum_{i=0}^{\text{degP}} \alpha_i p^i C(t)^2 + \sum_{i=0}^{\text{degP}} \beta_i p^i C(t) + \sum_{i=0}^{\text{degP}} \delta_i p^i \tag{3}$$

LPV Model with Two Scheduling Variables. The considered local models are the same as those given by 1. The global LPV model is still obtained by interpolating the evolution of θ_1, θ_2 and θ_3. The only differences are:

- the average test pressure p is replaced by the instantaneous pressure $p(t)$;
- instantaneous temperature $T(t)$ is also taken into account.

The general model is now given by:

$$\begin{aligned} Q_m = & \sum_{i=0}^{\text{degP}} \sum_{j=0}^{\text{degT}} \alpha_i p^i(t) T^j(t) C(t)^2 \\ & + \sum_{i=0}^{\text{degP}} \sum_{j=0}^{\text{degT}} \beta_i p^i(t) T^j(t) C(t) \\ & + \sum_{i=0}^{\text{degP}} \sum_{j=0}^{\text{degT}} \gamma_i p^i(t) T^j(t) \end{aligned} \tag{4}$$

where degP and degT represent the polynomials degrees of $p(t)$ and $T(t)$.

Taking into account the instantaneous measurements including the temperature, it is hoped that more accurate estimates than those provided by the first model will be obtained. Each coefficient θ_1, θ_2 and θ_3 is modeled by a polynomial $p(t)$ and $T(t)$.

4.2 Non-parametric Modeling

Modeling using Gaussian processes is also considered. It is a non parametric approximation method that aims to build an approximation $\hat{f}$ of the function $Q_m = f(C,p,T)$ from n observations $Q_{m_i} = f(C_i,p_i,T_i), 1 \leq i \leq n$ (observations may contain measurement errors), and from *a priori* about the speed variations of the searched function. To simplify notations, note $x_i = (C_i,p_i,T_i)$.

The *a priori* is expressed assuming that the searched function is the realization of a regular random process, in practice a Gaussian process determined by its mean and covariance function. The mean is here taken equal to zero to reflect the absence of *a priori* about a possible tendency of $f(x)$. The covariance function is chosen from a set of parameterized covariance functions family (also called kernels) whose parameters are estimated using the maximum likelihood criterion. We considered that the process was stationary and we chose to model its covariance by a Matérn covariance [21]. This family of covariance was chosen both for its ability to represent a wide range of processes, because its parameters are easily interpretable, and also because it avoids potential numerical problems.

To express the constraint that the searched function $\hat{f}(x)$ is close to the n observations, we search among all the Gaussian process realizations, those that explain the observed points: it is the principle of the modeling with Gaussian process that consists of conditioning of the process law with respect to the observations. The conditioned process is actually a new process with a law, including both the *a priori* (regularity, process variation speed) and the information provided by the observation of the process at some points, can be calculated. This principle is shown in Figure 7 on an example where the variable x is a scalar.

The estimate $\hat{f}(x)$ commonly used to estimate the function $f(x)$ at one point x is the mean of the process conditioned at this point. The covariance function of the conditioned process enable also to calculate confidence intervals for the function $f(x)$. Figure 8 includes the data of figure 7 (same function $f(x)$, the same abscissa x_i and same observed values) and gives the estimate $\hat{f}(x)$ and the associated confidence intervals.

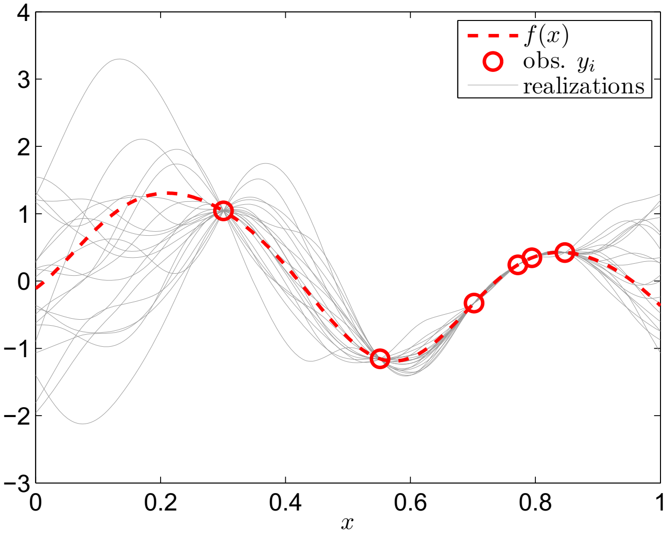


Fig. 7. Example of randomly generated realizations from a Gaussian process with Matérn covariance of parameters $(A = 1, h = 1, \nu = 2)$ conditioned to a set of $n = 6$ observed values (assumed here free-noise case)

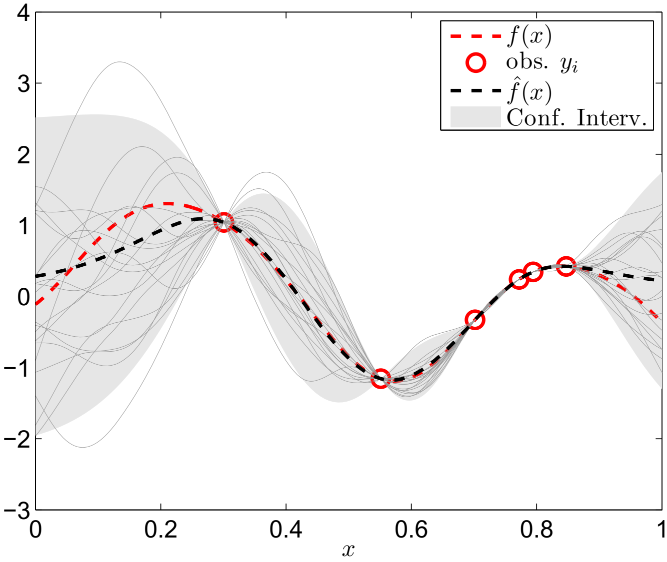


Fig. 8. Illustrative example of the Gaussian process modeling – Red dotted lines: function to estimate, red circles: observations, black dotted line: estimation, shaded area: confidence intervals 95 %

5 Experimental Results

5.1 Parametric Modeling

Local Models Identification. Figure 9 lists the local models, defined by (1), estimated with all 14 available experiments. The parameters have been estimated with a standard least-squares method.

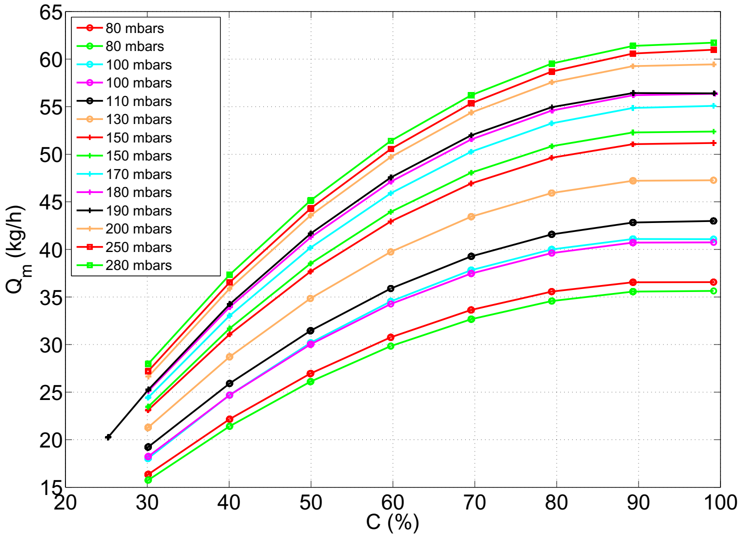


Fig. 9. Local models $Q_m(t) = \theta_1 C(t)^2 + \theta_2 C(t) + \theta_3$ for different pressure values p

Global Models Identification

1. LPV model with one scheduling variable

Figures 10 and 11 show the parameters of the different local models based on the operating point of the test. The polynomials, defined by (2), allow a good approximation of the estimated values of the parameters θ_1, θ_2 and θ_3, depending on the pressure as it can be seen in Figure 10. After several tests, the best results are obtained for $degP = 2$ for θ_1 and θ_2, and $degP = 3$ for θ_3. Figure 11 shows that it is more difficult to define a mathematical law that fits these points. Initially, it is proposed to use a global LPV model only function of the control signal and the pressure.

The global LPV model can then be used directly as a soft sensor; for a measured control signal C and a pressure setting, we simply simulate the equation (3) to estimate flow gas. The results of models simulation for an experiment are shown in Figure 12 and compared to the measured data. The maximum relative error (stepwise averaged) is shown in Figure 13 for all experiment set. Maximum relative error equal to 3.92 % is obtained for a $80\,mbar$ pressure.

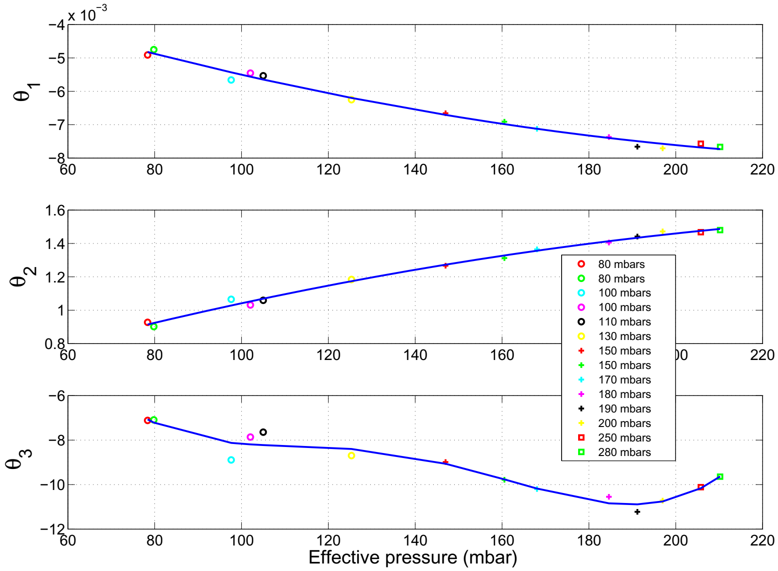


Fig. 10. Evolution of the local parameters models with respect to the pressure p (markers) and the polynomial models (2) (solid line)

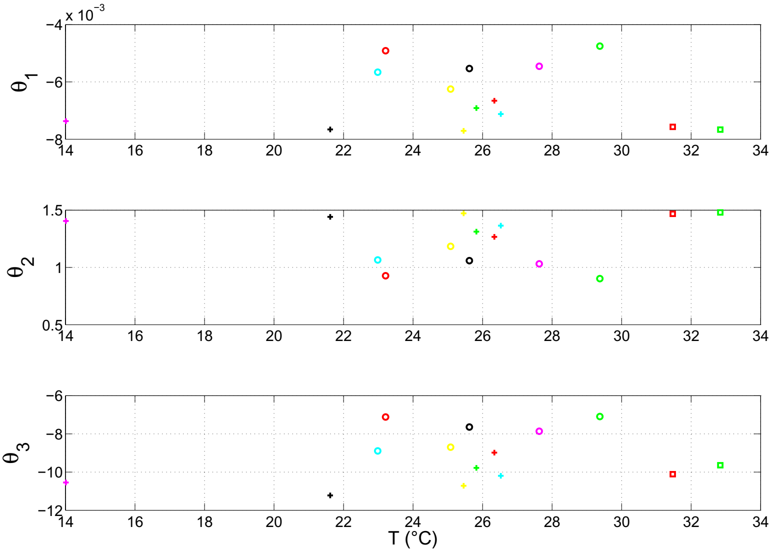


Fig. 11. Evolution of the local parameters models with respect to the temperature

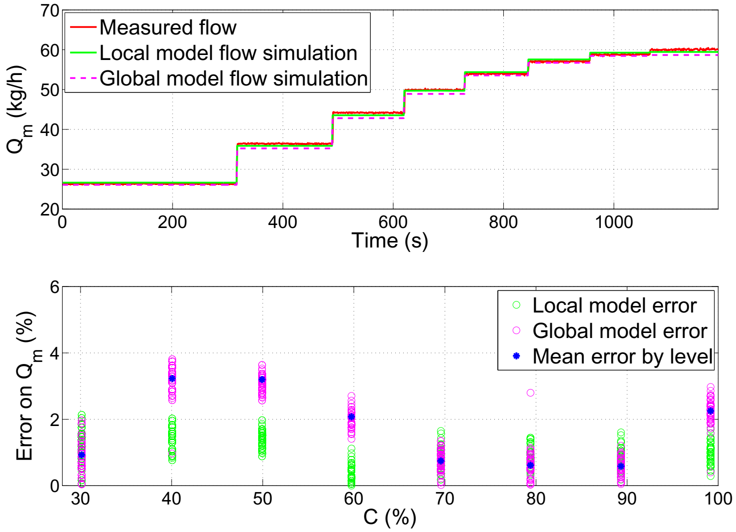


Fig. 12. Local model and one scheduling input global model simulations with 200 mbar experiment data

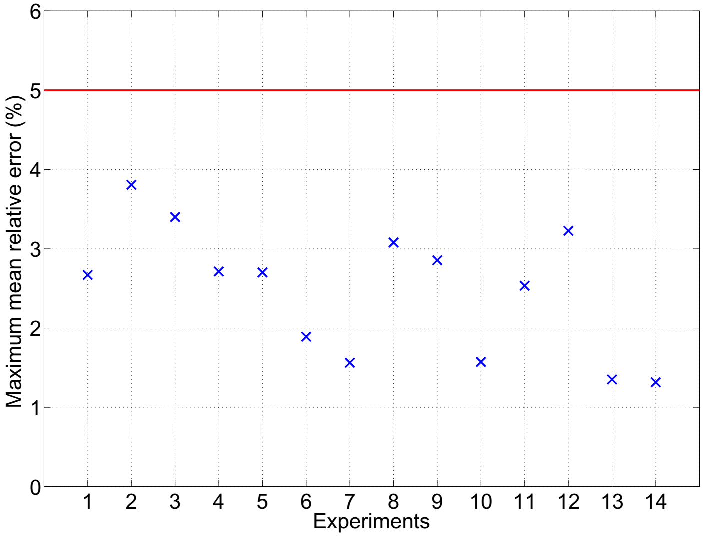


Fig. 13. Maximum relative errors (stepwise averaged) for 14 tests – LPV model with one scheduling variable

A cross-validation was performed to verify the behavior of the virtual sensor for all experimental conditions potentially faced in the operating phase. The number of experiments is relatively low (14 experiments); we chose to use a Leave–One–Out approach [16]. It consists to use 13 of the 14 tests for identification and one for validation, and repeat this operation so that each test is used as a validation.

The results are shown in Figure 14. The relative errors on the model simulation, estimated using 14 experiments are represented by crosses. The relative errors of the estimated models using 13 experiments and simulated on the validation test are represented by circles. A higher error is noted in validation. Nevertheless, it remains less than 5 % as shown in this figure.

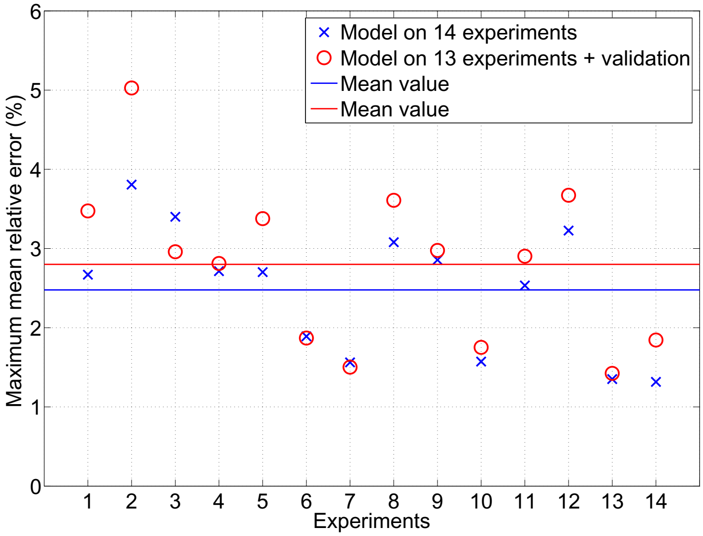


Fig. 14. Comparison between the estimated model using 14 experiments and the models from cross-validation – LPV model with one scheduling variable

2. LPV models with two scheduling variables

The global model is now based on the valve control signal $C(t)$ and the instantaneous measurements of pressure $p(t)$ and temperature $T(t)$. Figure 15 shows that, at the same pressure, the gas flow at two different temperatures is not really the same. Remains to be seen if it is really interesting to add a second scheduling variable, and thereby complexify the model.

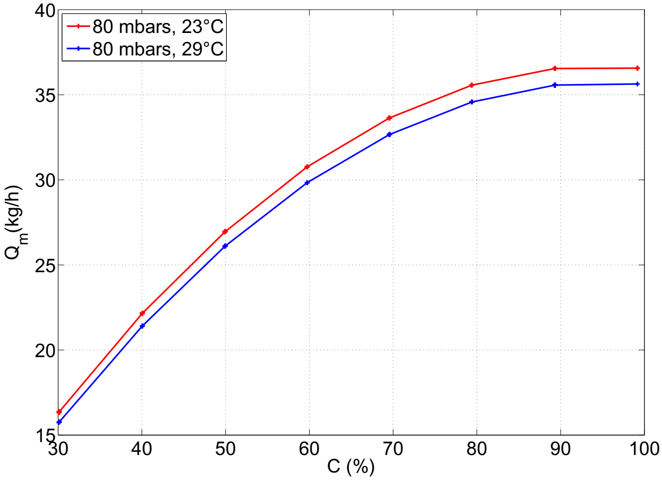


Fig. 15. Local models at the same pressure 80 mbar and two different temperatures (23 and 29°C)

After several tests, the best results are obtained for $degP = 2$ and $degT = 1$. The stepwise maximum relative errors in cross-validation are given in Figure 16. The maximum relative error is equal to 3.7 %, i.e. lower than those of the first model. However, it has a higher complexity. Figure 17 presents the simulation of the LPV model obtained for a value of $C = 50\,\%$ and varying pressures and temperatures. We can note that the influence of pressure on the flow variations is higher than the temperature, which may justify the use of a model taking into account only of the pressure.

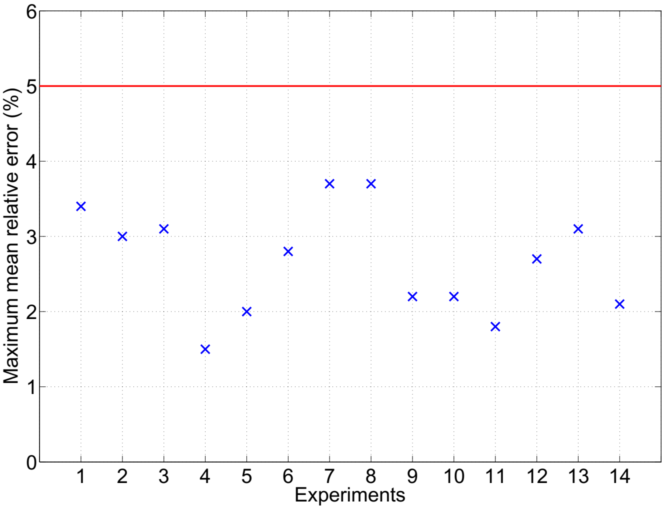


Fig. 16. Maximum relative errors (stepwise averaged) in cross-validation – LPV model with two scheduling variables

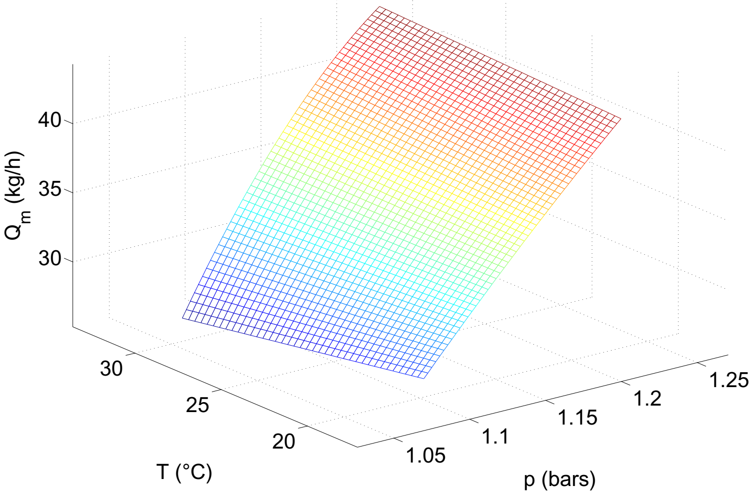


Fig. 17. Simulation of the Q_m estimated model defined by (4) for a fixed control signal at 50 %

5.2 Non-parametric Modeling

The implementation of this method on the 14 experiments realized on the boiler was performed using the Matlab toolbox STK (*Small Toolbox for Kriging*) [2]. As for the other simulations, Figure 18 gives the maximum stepwise relative errors obtained by cross-validation. We can note that the errors are lower than the parametric model errors. In addition, the non-parametric model provides reasonable errors without having to specify structure models.

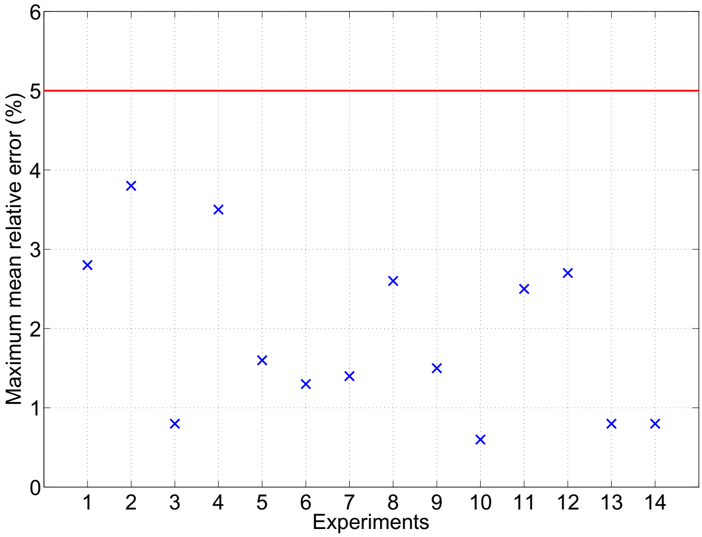


Fig. 18. Maximum relative errors (stepwise averaged) obtained in cross-validation – non-parametric model

Figure 19 shows the results obtained by cross-validation tests on 4 from the 14 tests. As suggested by the results shown in Figure 18, the predictions are close to the real values. The simulation of the obtained model at variable pressures and temperatures for a 50 % control signal, provides similar results to those presented in Figure 17 .

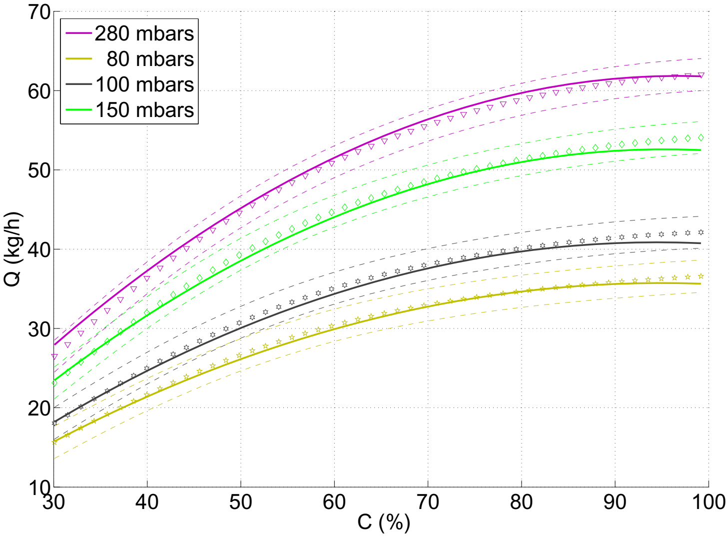


Fig. 19. Cross-validation results for $Q_m = f(C, p, T)$ models with respect to C – Solid lines: identified polynomials; markers: predictions; dotted lines: associated confidence intervals

5.3 Synthesis

The synthesis of the three different models results is illustrated in Table 2. Although the fact that the parametric model with two scheduling variables and the non-parametric model provide better results in terms of relative error in cross-validation, nevertheless, the parametric LPV with one scheduling variable model still the easiest to implement and calibrate to the real operating conditions. Indeed, it is less complicated to calibrate a system with one variable input than with two variable inputs.

5.4 Soft Sensor Implementation on the Industrial Boiler

The LPV model with one scheduling variable was experimented on site owing to its simplicity. The model was directly implemented on the PLC with C# programming language. The pressure setpoint has been tuned and fixed to the value read on the manometer during the pressure regulator setting. Figure 20 shows the flow measurement with the soft sensor. As we can see, the results are acceptable with a maximum mean relative error of 3.5 % for the LPV model with one scheduling variable.

Table 2. Summary table showing the results of the relative errors calculation in cross-validation of the three previous models

Model	Relative error in cross-validation
parametric LPV 1 scheduling variable	5 %
parametric LPV 2 scheduling variables	3.7%
non-parametric	3.8 %

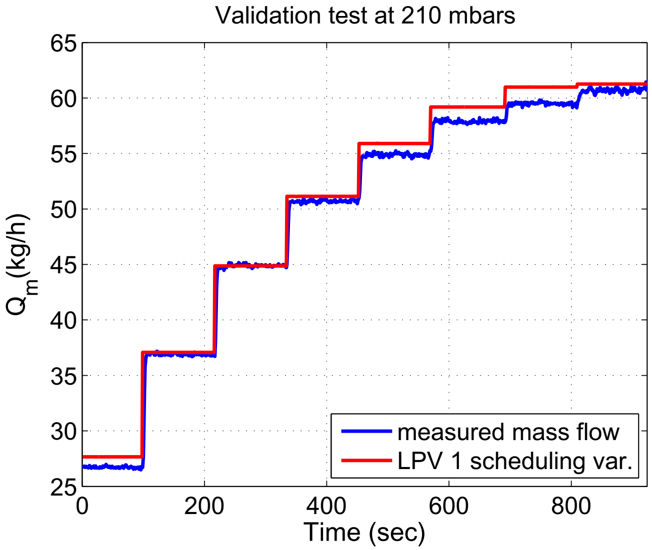


Fig. 20. Experimental simulation of the soft sensor based on the LPV model with one scheduling variable

6 Conclusion

This chapter reviewed the current technologies intended to flow and power measurements from a material and financial point of view. We have seen that measuring is the key for improving Energy Efficiency in the manufacturing sector, hence the necessity to look for smart low cost sensors.

In particular, the modeling of an industrial boiler was investigated towards a consumed gas flow measurement. Three static models were estimated: two LPV parametric models and a non-parametric model. A cross-validation showed that the simulation of these models gives a flow measurement error lower than 5 %. This value corresponds to the fixed objectives of low-cost sensors implementation that allow consumption monitoring and detection of possibles drifts.

Models degraded uses should also be considered. While the online temperature measurement could be considered low-cost, this is not the case of the pressure. But, in practice, the boiler engineer tunes the pressure with the pressure regulator and measures it with manometer.

Finally, the genericity of models to different installations and other kinds of valves, should be studied. It is then necessary to define a model parameters calibration methodology which is the least intrusive. To do this, a grey box model is being tested. Its purpose is to find generic models for all sorts of valves existing in industry for all types of boilers, with different pipe dimensions. There exists a valve opening law: $Q_{gas} = f(C\%)$, Q_{gas} being the gas flow, which should be different from a valve to another. The equation $Q_{gas} = K_v\sqrt{p_1 - p_2}$, with p_1 and p_2 the pressures before and after the valve, and K_v the flow coefficient, needs also to be considered. The objective here is to estimate Q_{gas} froms these two expressions, taking into account the valve manufacturer datasheet and the pipe dimensions.

Acknowledgment. This work has been supported by the French Research National Agency (ANR) through *Efficacité énergétique et réduction des émissions de CO_2 dans les systèmes industriels* program (project CHIC n° ANR-10-EESI-02).

References

1. Arkoun, O.: Estimation non paramétrique pour les modéles autorégressifs. Thesis, University of Rouen, France (2009)
2. Bect, J., Vasquez, E.: A small (Matlab/GNU Octave) Toolbox for Kriging, Supelec (2011), `http://sourceforge.net/projects/kriging/`
3. Bogaerts, P., Vande Wouwer, A.: Software sensors for bioprocesses. ISA Transactions 42, 547–558 (2003)
4. Bourkeb, M., Ondel, O., Joubert, C., Morel, L., Scorretti, R.: Méthodes numériques pour la mesure de courant dans un systéme polyphasé, Numélec 2012. In: European Conference on Numerical Methods and Electromagnetism, Marseille, France, July 3-5 (2012)
5. Cecil, D., Kozlowska, M.: Software sensors are a real alternative to true sensors. Environmental Modelling & Software 25, 622–625 (2010)
6. Chéruy, A.: Software sensors in bioprocess engineering. Journal of Biotechnology 52, 193–199 (1997)
7. Dochain, D.: State and parameter estimation in chemical and biochemical processes: a tutorial. J. Proc. Control 13, 801–818 (2003)
8. Dos Santos, P.L., Perdicoúlis, T.A., Novara, C., Ramos, J., Rivera, D.: Linear Parameter-Varying System Identification: new developments and trends Advanced Series in Electrical and Computer Engineering. World Scientific (2011)
9. Etien, E.: Modeling and simulation of soft sensor design for real-time speed estimation, measurement and control of induction motor. ISA Transactions 52, 358–364 (2013)
10. Gever, M.: Toward a joint design of identification and control? H.L. Trentelman and J.C. Willems Birkhäuser, Boston (1993)
11. Grau, A., et al.: Low cost power and flow rates measurements for manufacturing plants. In: 16th International Congress of Metrology, Paris, France (October 2013)
12. Grospeaud, O.: Contribution à l'identification en boucle fermée par erreur de sortie. Thesis, University of Poitiers, Poitiers, France (2000)

13. Hadid, B., Etien, E., Ouvrard, R., Poinot, T., Le Brusquet, L., Grau, A., Schmitt, G.: Soft Sensor Design for Power Measurement and Diagnosis in Electrical Furnace: a Parametric Estimation Approach. In: 39th Annual Conference of the IEEE Industrial Electronics Society IECON, Vienna, Austria (November 2013)
14. Hadid, B., Ouvrard, R., Le Brusquet, L., Etien, E., Poinot, T., Sicard, F.: Modeling for gas flow measurement consumed by a boiler. In: Towards a Low-cost Sensor for Energy Efficiency, Seventh International Conference on Sensing Technology ICST. Massey University, Wellington (2013)
15. James, S.C., Legge, R.L., Budman, H.: On-line estimation in bioreactors: a review. Rev. Chem. Eng. 16(4), 311–340 (2000)
16. Kohavi, R.: A study of cross-validation and bootstrap for accuracy estimation and model selection. In: Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence, vol. 2(12), pp. 1137–1143 (1995)
17. Le Mouel, A., et al.: Fostering Energy Efficiency in manufacturing plants through economical breakthroughs in power and flow rate measurement. In: ECEEE Summer Study on Industry, Arnhem, Germany (2012)
18. Ljung, L.: System identification. Theory for the user, 2nd edn. (1999)
19. Queinnec, I., Spérandio, M.: Simultaneous estimation of nitrification/denitrification kinetics and influent nitrogen load using orp and do dynamics. ECC, Cambridge, UK (2003)
20. Poinot, T.: Contribution à l'identification des systèmes par la méthode de surparamétrisation en traitement des eaux. Thesis, Université de Poitiers, France (1996)
21. Rasmussen, C.E., Williams, K.I.: Gaussian Processes for Machine Learning. The MIT Press (2006)
22. Sotomayor, O.A.Z., Won Park, S., Garcia, C.: Software sensor for on-line estimation of the microbial activity in activated sludge systems. ISA Transactions 41, 127–143 (2002)
23. Tóth, R.: Modeling and Identification of Linear Parameter-Varying Systems. LNCIS, vol. 403. Springer, Heidelberg (2010)
24. Van Donkelaar, E.T., Van den Hof, P.M.J.: Analysis of closed-loop identification with a tailor-made parametrization. In: European Control Conference, Brussels, Belgium, vol. 4 (1997)

Measurement of Ultrafine Particles in the Exhaust of Combustion Engines Using Light Scattering – Theory and Application

H. Axmann[1,*], A. Bergmann[2], and B. Eichberger[3]

[1] AVL DiTEST GmbH, Alte Poststraße 152, 8020 Graz, Austria
`harald.axmann@avl.com`
[2] AVL List GmbH, Hans List Platz 1, 8010 Graz, Austria
[3] Institute of Electronics, University of Technology,
Inffeldgasse 12, 8010 Graz, Austria

Abstract. The particle emissions of both diesel and direct injection gasoline vehicles are a major environmental problem and pose a risk to human health. As a countermeasure recent legislation has limited the allowed particle emissions of newly built vehicles. These low limits led to the introduction of new engine technologies and exhaust aftertreatment systems like e. g. particulate filters. In comparison to older vehicles the particle emissions have been decreased dramatically due to these measures. To ensure the compliance to the limits over the vehicles' lifetime, periodic emission tests are necessary. However, the devices currently used for this purpose lack the necessary sensitivity.

In this chapter a novel device based on light scattering, intended to fill this gap, is described. After an overview of current legislation and exhaust measurement, a thorough introduction into the theory of light scattering is given. Commonly used approaches to model the light scattering behavior of particles and exhaust particles in particular are discussed and compared. Subsequently a possible practical realization of a light scattering device is outlined with special emphasis on the electronic sensor design. Furthermore the issue of correlating to established measuring units is discussed and results of comparison measurements with reference devices are shown. Finally, possibilities how to calibrate light scattering devices are evaluated and an approach based on the use of a calibration gas is covered in more detail.

Keywords: light scattering, diesel particles, periodical technical inspection, optical methods, calibration, gases.

1 Introduction

The particle emissions produced by internal-combustion engines pose a considerable problem to our environment. For the most part those particles are generated by incomplete combustion of the fuel [1]. In former days only diesel

---

[*] Corresponding author.

A. Mason et al. (eds.), *Sensing Technology: Current Status and Future Trends IV*,
Smart Sensors, Measurement and Instrumentation 12, DOI: 10.1007/978-3-319-12898-6_3

engines where known to produce particles, whereas the exhaust of gasoline engines was formed essentially of gaseous components. However modern gasoline direct injection (GDI) engines emit as much particles as current diesel engines do [2, 3]. The generated particles effect global climate and human health [1, 4, 5]. Diesel soot was even proven to be carcinogenic in a resent study of the World Health Organization (WHO) [6]. Governments all over the world have reacted on this threat by introducing limits for the allowed particle concentration in the exhaust of newly built engines. In Europe for example the European Emission Standards have been created in 1992 with emission limits for several pollutants including particles. Starting with a relatively high limit value of 0.14 g/km for the allowed particle mass of diesel vehicles in EURO 1, the limits have been rigorously tightened since then [7]. In the current EURO 6 norm the mass limit is as low as 5 mg/km (see fig. 1). Additionally a number limit has been introduced in EURO 5 with a value of 6×10^{11}/km [8]. Comparison measurements show that this number limit is again lower than the mass limit by one order of magnitude [9]. With EURO 6 the same number limit value has also been adopted for GDI engines. Comparable regulations have been established in other regions all over world, e. g. the United States and Japan [3].

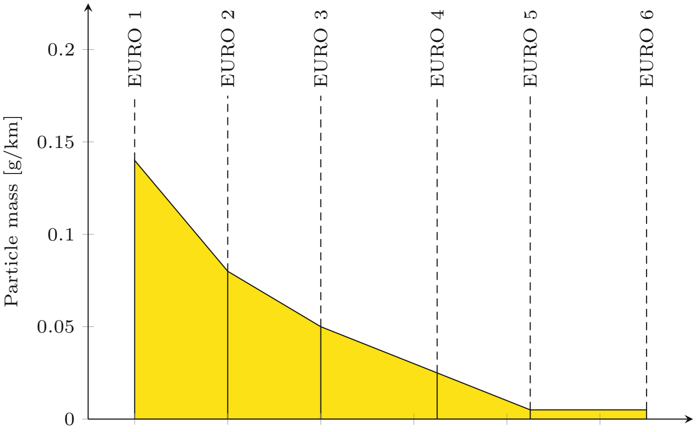


Fig. 1. Particle mass limits in the European Emission Standards EURO 1-6

In order to reduce the emissions and to conform to the regulations, new engines technologies and exhaust aftertreatment systems have been developed. The most important of these is the diesel particle filter (DPF). It reduces the particle output of the vehicle by collecting a big part of the particles downstream of the engine. With the reduction of emissions comes the need for improved measurement instruments, which quantify the particle amount. In the sector of engine development and homologation very sensitive, yet complex measurement devices like photoacoustic sensors and condensation particle counters (CPC) have become state-of-the-art [3]. Besides there is the need to check the particle emissions over the vehicles' lifetime, since aging or damages of the exhaust system may lead to elevated particle emissions. In Europe those measurements are performed during the periodical technical inspection (PTI), as prescribed by the European Commission Directive 2010/48/EU [10] and its amendments. The officially required measurement parameter in those roadworthiness tests is the turbidity of the exhaust gas, the opacity N. It is defined as the relative amount of light that is removed from an incident light ray due to the exhaust components [3, 4]:

$$N = \frac{I_0 - I}{I_0} = 1 - e^{kL}. \tag{1}$$

Herein I_0 is the incident light intensity, I the measured light after passing the sample, and L is the measurement length. k is the extinction coefficient, which is given in 1/m. It is also the typical threshold value defined for roadworthiness and commonly referred to as the "absorption coefficient". Though physically inexact, this nomenclature is extremely widespread. k and N are practically equivalent measures, however the extinction coefficient is preferred, because it is not depending on the measurement length.

During the PTI the opacity or the extinction coefficient respectively are determined in a free acceleration measurement. The engine is periodically accelerated at no load to cut-off speed. From each acceleration the maximum opacity value is taken. The average of a prescribed number of measurement cycles must be lower than the threshold value.

The device used to measure the smoke opacity is the opacity meter, also called opacimeter. It is a moderately priced, robust optical measurement device and usually consists of a light source on one side of a measurement chamber of length L and a light detector (e. g. a photodiode) on the other side. Typical implementations use a measurement length of 430 mm and a green light source. The green color was chosen for comparability to the perception of the human eye, since the original idea was to quantify the visual obstruction caused by the exhaust gases. For the black clouds of soot emitted by old diesel engines, this measurement method was by far satisfactory. However for modern diesel engines, whose emissions are practically invisible, typical opacity meters lack the necessary sensitivity. There is hardly any attenuation of the light over the short measurement distance, and the stability of the light source influences the detection limit directly. An improvement of the sensitivity is only possible up to a certain degree.

Consequently for an essential improvement of the sensitivity a new measurement technique is necessary. The PTI is performed in garages and vehicle inspection institutions. They cannot afford the expensive devices made for the engine developers and need robust devices with easy operability and low maintenance. Their requirements have to be taken into account for the selection process.

On the basis of a thorough evaluation of different contemplable methods the light scattering principle was selected. It is again an optical technique and appears very promising. The following sections deal with the theory behind light scattering and explain its advantages compared to the light transmission method.

2 Theory of Light Scattering

Light scattering is the deflection of light due to heterogeneities in the medium the light travels through. A typical example for such a heterogeneity is a particle. When the particle gets hit by a light ray, it will scatter the incoming light into all directions. The intensity of the scattered light can be measured by an optical detector placed at some distinct angle to the incident light ray. The more particles interact with the light ray, the more intensity will be scattered to the detector. Thus the signal on the detector can be used to determine the concentration of the particles (see fig. 2). The amount of light that is scattered is typically rather low compared to the incident one. However, if there are no particles, there will ideally be no scattered light at all. This is the biggest advantage over the light transmission technique, because it is typically easier to measure small, but absolute quantities than to measure small relative changes of big quantities.

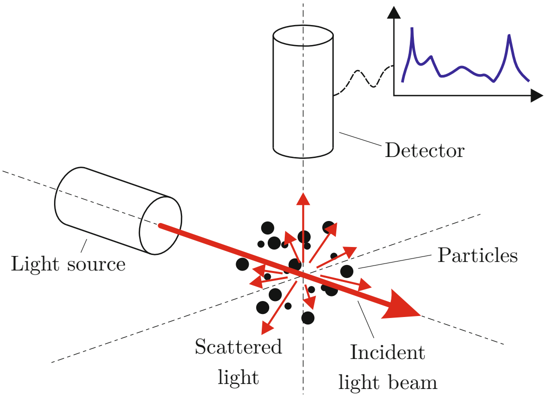


Fig. 2. Principle of light scattering

The distribution of the scattered light intensity is depending on numerous properties of both the particles and the incident light ray. The most important properties are the shape of the particles and their size in relation to the wavelength. In the following sections models for the most relevant particle shapes and sizes are described.

2.1 Spherical Particles

A sphere is the simplest geometric form, and so it is also the simplest case of scattering. The first models for the scattering of spherical particles date back to the end of the 19th century. At that time John William Strutt, better known as Lord Rayleigh, developed a model to describe the scattering of the sun light in the atmosphere. The Rayleigh model, named after him, can be used to describe the light scattering of particles very small in size compared to the incident wavelength [11–13]. For such small particles, the shape is irrelevant, so a spherical shape can be assumed. The scattering intensity of such a particle for unpolarized incident light at a distance r is equal to

$$I_s = \frac{8\pi^4 a^6}{\lambda^4 r^2} \left| \frac{m^2 - 1}{m^2 + 2} \right| \cdot \left(1 + \cos^2\theta\right) I_0. \tag{2}$$

I_s is the scattered intensity at an angle θ for a spherical particle with radius a and a refractive index m. I_0 is the irradiance and λ the wavelength of the incident light. The refractive index $m = n + jk$ is a complex number, composed of the real part n, which is a measure of the refraction in the particle medium, and the imaginary part k, which quantifies the light attenuation or absorption. Both numbers are positive, for the latter this is a matter of definition. The refractive index of the surrounding medium is assumed as $m_s \approx 1$ (e. g. air).

The Rayleigh scattering for two particles of different size is displayed in fig. 3. It is symmetric around the ordinate and the abscissa, with minima at 90° and 270°.

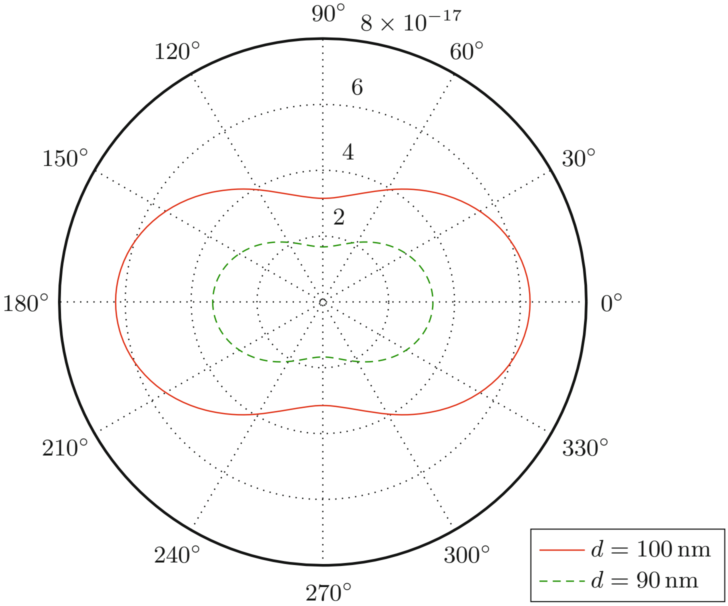


Fig. 3. Rayleigh scattering for particle sizes of $d = 90\,\text{nm}$ and $d = 100\,\text{nm}$

The upper threshold for the applicability of the Rayleigh model is between one sixth and one tenth of the wavelength. As the particles grow bigger, the scattering behavior starts to deviate from the Rayleigh model, even for the spherical case.

The forward scattering increases, and the scattered light waves from different positions on the particle's surface start to interfere with each other, creating additional maxima and minima. This scattering behavior is described by the Mie theory, which has been published by Gustav Mie in 1908 [14]. The formulas are much more complex and cannot be calculated in closed form.

The scattering intensity calculated by the Mie model is depicted in fig. 4. The increasing forward scattering lobe (in direction of 0°) and the interference patterns related to the bigger particle size are clearly visible. For very small particles the Mie theory will deliver results identical to the Rayleigh model.

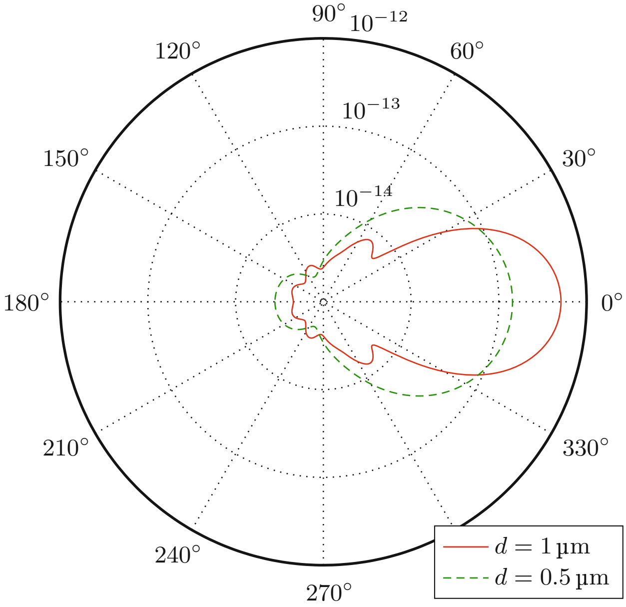


Fig. 4. Mie scattering for particle sizes of $d = 0.5\,\mu\text{m}$ and $d = 1\,\mu\text{m}$

The Mie theory is appropriate for calculating the scattering intensities for arbitrarily sized particles, however it is limited to spheres. There is a small number of models for some well-defined non-spherical objects like cubes or cylinders. For them the complexity of the calculation increases further, since the orientation of the particles has to be taken into account. Of course modern computer technology allows to simulate the scattering behavior of more or less any shape, but it is a time-consuming process. Furthermore a typical scattering experiment is not limited to a single particle. One has to deal with an ensemble of N arbitrarily positioned and oriented particles. For an exact calculation the scattering contribution of each particle needs to be determined and added with respect to its phase. Under the condition of randomly positioned particles, the phases add randomly[1], too [12]. It is therefore possible to sum up the single scattering intensities:

$$I_{\text{total}} = N I_s. \tag{3}$$

Still the calculation of the exact scattering of each particle in the ensemble has to be performed, rendering this approach practically impossible. It is therefore advantageous to seek further simplifications. Some authors approximate the particles by spheres [15–17]. Such approximations have to be used with care, as will be shown later. For the case of soot particles the fractal aggregate theory can be applied, which is described in the following section.

2.2 Fractal-Like Aggregates

In order to model the scattering behavior of a soot particle, it is necessary to model the soot particle first. A suitable mathematical description is not easily found. The relevant particles for mass, opacity and scattering measurement are the accumulation mode particles [18]. Those are sparse, chain-like aggregates formed of smaller, almost spherical primary particles called monomers (see fig. 5). The number of primary particles per aggregate varies greatly, resulting in aggregate sizes between 30 nm and 500 nm. For an incident wavelength in the visible range, e. g. red light with approximately 650 nm, most of those particles are well beyond the Rayleigh limit. Accordingly their shape and their structure need to be considered. This structure varies widely, even for particles with the same number of monomers.

It was Forrest and Witten who first found an easy mathematical description for such particles based on the work of Mandelbrot [20, 21]. They found out that the particles are self-similar and scale-invariant, at least to a certain degree, so their structure can be described by fractal theory. Hence the particles are called fractal-like aggregates. The use of fractal theory results in a relatively easy way to relate the particles' optical size R_g to the number of primary particles N:

$$N = k_0 \left(\frac{R_g}{a} \right)^{D_f}, \tag{4}$$

where k_0 is the fractal pre-factor of order unity, a is the common radius of the spherical primary particles, R_g is a root mean square radius called radius of gyration, and D_f is the fractal dimension. The latter is a measure of the particles' porosity. Particles with sparse chains have values of D_f near 1, whereas dense, sphere-like particles tend to values near 3. For soot particles typical values around 1.78 for diffusion-limited aggregation and 1.95 for ballistic aggregation processes have been reported [22–24]. Higher values published by some other authors seem highly doubtable [25].

---

[1] An exception to this rule is the forward direction (0°), where all phases add constructively. For a light scattering experiment, a measurement at 0° is impossible, and this case is therefore irrelevant.

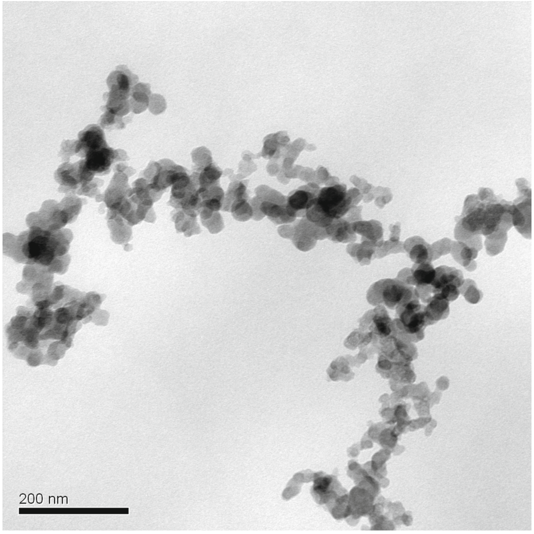


Fig. 5. Typical soot particle from propane combustion [19]

Based on this fundamental equation, formulas to calculate the scattering of such particles can be developed. It is assumed that all primary particles are spherical and equal in size and that they are all hit by the same incident light wave. The typical radius a of the primary particles is in the order of 20 nm. As they are relatively small compared to typical wavelengths used for light scattering, their scattering behavior can be calculated with the Rayleigh model. To determine the scattering at a scattering angle of 0°, i. e. in direction of the incoming light, all scattering contributions by the primary particles are in phase and can be summed up. For any other scattering angle θ, the phase of the scattering contribution must be incorporated. As the monomers are arranged in a static structure, the phases will not add randomly. Often it is advantageous to substitute the scattering angle θ by a different variable, the scattering vector q. It can be derived from the wave vectors of the incident and the scattered light and it is related to the scattering angle by

$$q = 4\pi/\lambda \cdot \sin(\theta/2) . \tag{5}$$

The structure factor S can then be defined depending on q to describe the angle dependent interference behavior. It is equal to 1 for the small q values in forward direction, where pure constructive interference occurs ("Rayleigh regime"). When going to higher q values it starts to diminish ("Guinier regime") and it finally decreases with $(qR_g)^{-D_f}$ ("Power law regime"). In a multi angle scattering experiment the three mentioned regimes can be used to determine different properties of the particle ensemble. For the experiment discussed here, the total scattering intensity $I_{agg}(q)$ on a detector at one distinct scattering angle is relevant. It can be calculated by summing the amplitudes, squaring them to get the intensities $I_m(q)$, and correcting for their phases by multiplication with the structure factor $S(q)$:

$$I_{agg}(q) = N^2 I_m(q) S(q). \tag{6}$$

This model is called Rayleigh-Debye-Gans Fractal-Aggregrate Theory (RDG-FA). It is to a huge amount based on the work of Sorensen, who has studied the scattering of those fractal-like aggregates thoroughly since the end of the 20th century [24]. For the relevant range of parameters its deviation from the exact results should be within 10 % [26].

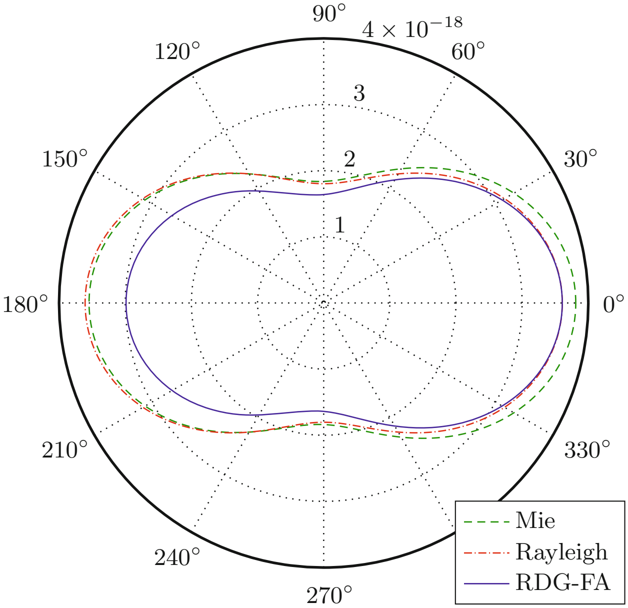


Fig. 6. Comparison of the RDG-FA scattering model with the Mie and Rayleigh model for volume equivalent particles

Fig. 6 shows the comparison of the fractal model with the Mie and the Rayleigh model. The optical size of the fractal aggregate is $2R_g = 80\,\mathrm{nm}$. For the Mie and the Rayleigh model spherical particles with the same volume and mass as the fractal aggregate have been chosen, since this will deliver the best results concerning comparability. This can be proved easily, when regarding the scattering at small q values, where $S(q) \approx 1$:

$$\frac{I_{agg}(q)}{I_s(q)} \approx \frac{N^2 I_m(q)}{I_s(q)} = \frac{N^2 a^6}{a_s^6}. \tag{7}$$

I_s is the Rayleigh scattering of a sphere with radius a_s. The volume of the aggregate is proportional to Na^3, for the sphere it is proportional to a_s^3. The volumes are equal, if $a_s = \sqrt[3]{N}a$, leading to a ratio of 1 in the above equation. Thus fractal model and Rayleigh model deliver exactly the same results for small q. The slightly higher values of the Mie model in forward direction result from the size of the sphere approaching the Rayleigh limit.

When going to higher q values the similarity between fractal and spherical models decreases. For bigger particles the deviations increase even more. Of course the Rayleigh model still delivers the same result as the fractal model at $q \approx 0$, but completely fails for the rest of the angles. The Mie model performs better at higher angles. What is completely missing in the fractal model are interference maxima and minima. The whole signal curve is rather smooth. This is not surprising, as the fractal model delivers an average value for all possible structures and orientations of particles with the same key features (N, k_0, R_g and a).

2.3 Scattering of Gas Molecules

Gas molecules are much smaller than typical exhaust particles, hence they are very small compared to the wavelength, when using visible light. Their scattering behavior can be calculated using the Rayleigh model. However, one cannot determine the "radius" of such gas molecules, as they are not necessarily spherical. For such cases the radius must be replaced by a characteristic length. Determining such a characteristic length can be a laborious task [11].

Instead it is possible to calculate the scattering of gas molecules by use of their polarizability α. This results in a very simple equation [12]:

$$I_s = \frac{8\pi^4\alpha^2}{\lambda^4 r^2}\left(1 + \cos^2\theta\right) I_0. \tag{8}$$

It can be seen that the scattering of gases is directly proportional to α^2. The value of the polarizability for common gases can be taken from the literature [27]. Tab. 1 shows a small selection. Obviously propane (C_3H_8) will scatter 4.4 times more than nitrogen dioxide (NO_2) and almost sixteen times more than oxygen. In relation to the scattering of particles, the effect is extremely small. High gas concentrations are necessary to achieve a measurable signal. Nevertheless it can have a practical application in a device optimized for particle scattering, as discussed in section 5.

Table 1. Polarizability α of some common gases

Name	Molecular Formula	Polarizability [10^{-24} cm^3]
Oxygen	O_2	1.5812
Nitrogen	N_2	1.7403
Carbon dioxide	CO_2	2.911
Nitrogen dioxide	NO_2	3.02
Propane	C_3H_8	6.29

3 Technical Realization

Light scattering allows the measurement of a number of particle properties. Depending on the desired properties the realization of the measurement setup may vary, e. g. in the number and power of the light sources and in the number and choice of the detectors. This device aims at measuring the integral particle concentration in the sample. Principally, as described before, this requires only one light source and one detector.

In the design of the measurement device the properties of the medium under test must also be taken into consideration. This is the raw exhaust gas taken directly from the tailpipe. In addition to particles it may contain oil droplets, sulfur compounds and water vapor. The device and especially the optical system need an adequate protection against soiling by these components.

Besides, there are some non-functional requirements. The device should replace the opacity meter in the garages and vehicle inspection institutions, which perform the periodical technical inspection. They need devices that are rugged, mobile, easy to handle and service, and quickly ready to measure. The dynamic range should allow measurements at both old high emitters and new vehicles with hardly any emissions. Last but not least they should feature a relatively low price.

There is already a number of light scattering devices that are commercially available. However, they do not fulfill the above requirements. Consequently a new device had to be developed. The measurement setup of this device is outlined below.

3.1 Optical Setup

Light scattering is a very weak effect. Only a tiny portion of the incident light is scattered, as could be seen in the preceding chapters. It is obvious that a high power light source in combination with a sensitive optical receiver circuit is necessary to achieve a good detection limit.

High optical output powers can nowadays be realized using a diode laser or a high power LED. The issue with LEDs is that they are designed for wide radiation angles. Their light is not easily collimated. Proper collimation of the incident light beam is one of the key factors for reducing background light in the measurement cell. Ideally there should be no detectable light at all, when there are no particles in the measurement volume. This would lead to the highest possible sensitivity. In practice the incident light is diffracted at the optical elements and reflected by the walls of the measurement cell, creating a significant background light. The better the light beam is focused and absorbed in a beam dump following the measurement volume, the lower is the background signal. The level of the background is irrelevant, as long as it is stable or compensated by other means. It depends on the output power of the light source though, and compensation is not always perfectly possible. Consequently one must aim to keep the background level low, in order not to run into the same issues as with the opacity meter.

Lasers have the advantage over LEDs that their radiation angles are smaller by default. Accordingly it is much easier to collimate their beams. We have therefore chosen a diode laser. It has a wavelength of 660 nm (deep red) and an optical power in the order of 10 mW. Our tests proved that this power suffices in combination with a good optical detection system. Naturally the wavelength needs some further discussion, since it has a considerable effect on the scattering intensity. To determine this effect, the Rayleigh model should be used for ease of calculation. In this case the scattering intensity declines with the forth power of the wavelength. Using a blue light source with e. g. 440 nm instead of the red one would improve the intensity by a factor of five. Unfortunately blue laser diodes are still very expensive, although their prices are falling due to their spreading in consumer electronics (BluRay players). A second drawback will be discussed when coming to the detection side. After all a red laser is still the preferred choice.

For detection of the scattered light some optoelectronic sensor must receive the light and convert it to an electrical signal. The options range from photomultipliers over avalanche photodiodes to silicon PIN photodiodes. The first two devices deliver the highest sensitivity [28]. It is even possible to detect single photons with them. They are costly though, and their electronic operation is rather complex. We therefore use a simple silicon photodiode. It is operated in short circuit for highest detection performance, best linearity and lowest noise [29]. To maximize the input light on the active area, a special lens system focuses the light from the measurement center onto the diode. This enables the use of a photodiode with a small active area in the order of 1 mm^2, resulting in smaller costs and lower noise. PIN photodiodes have their maximum spectral efficiency in the near infrared range. At 440 nm (blue) it usually drops to 20 – 30 % of its maximum. This is one more reason to choose a red light source.

In all optical systems soiling is an essential problem, as it leads to a decreased sensitivity due to the absorption of light. This must be detected by the system to prevent erroneous measurements. Still a reduction of the sensitivity is undesirable, therefore keeping the optical components clean is a must. This is typically done using a particle-free sheath air flow between the sample and the optics. The design of the measurement chamber needs to be optimized for a laminar sample flow, too. Otherwise in a turbulent flow particles might go through the curtain of particle free air and deposit on the optics.

3.2 Electrical Signal Processing

The photodiode converts the incoming light to a proportional current. This current is rather low, in the range of picoamperes or even smaller. The current can be measured with an analog-to-digital converter (ADC) after converting it to a voltage using a transimpedance amplifier. For these low currents a very high amplification (transimpedance) is necessary. Furthermore it must be taken care of electric noise, which would have a huge effect on these small signals.

Typical ADCs for precision applications feature a differential input structure. A normal transimpedance amplifier only creates a single-ended voltage.

An additional conversion to a differential signal adds noise and degrades the performance. It is advantageous to have a direct conversion of the photo current to a differential voltage. This is achieved by a sophisticated amplifier design consisting of two symmetric transimpedance amplifiers [30]. The patent issuance for this design is expected for summer 2014. After the conversion into a differential voltage signal, electrical disturbances occur as common-mode interferences and are filtered out in the amplifier input stage.

The target time resolution of the system is in the order of 100 Hz. This data rate suffices to resolve any transient process in the exhaust gas. The analog digital converter samples the input signal at a rate of 1 MHz. It performs a first downsampling step to 32 kHz, which increases the resolution. The microprocessor adds a second downsampling stage to the final 100 Hz update rate, resulting in an overall additional resolution of more than six bits.

All in all this special amplifier design provides a dynamic range of six orders of magnitude. The lowest detectable optical power is 500 fW, which is 2×10^{-10} times lower than the incident light power. Our test measurements indicate that this resolution is well adequate to perform credible measurements at vehicles that are equipped with a DPF.

3.3 Pneumatic System

In typical garage opacity meters the measurement cells are filled by the exhaust pressure. During idle speed the exhaust pressure is often too small to fill the measurement chamber completely. This might result in smaller opacity values at idle operation. In the scattering device a pump is used to extract the exhaust gas sample from the tailpipe and to maintain a constant flow of 2 L/min through the measurement chamber. This enables the device to measure constantly at different operating points.

The probe tube between the tailpipe and the device has a typical length of at least 2 m. In an unheated tube the vapor contained in the exhaust gas tends to condense, as the sample travels through the tube and cools down. Water droplets in the sample may cause spurious high peaks in the measurement signal, when they reach the measurement chamber. If they deposit on the walls of the measurement tube, they can also filter out parts of the particles. Neither of the two effects are tolerable. It is therefore very import to use a heated sampling tube to keep the exhaust gas always above the condensation temperature. For exhaust gases this temperature is around 45 °C to 55 °C. Therefore the heated probe should be set to a temperature above this value. A temperature setting of 80 °C delivers good results.

4 Correlation

Replacing the opacity meter in the garages by this new light scattering device requires backward compatibility to older systems. This means that the light scattering device must be able to cover more or less the range of vehicles supported by the opacity meters and extend it to the low concentration region of

DPF equipped vehicles. Furthermore the device needs to display the same measurement units: the opacity or the absorption coefficient respectively. Here one needs to rely on empirical formulas, determined by comparison measurements. This is due to the fact that there is no simple relation between those units and the scattering signal. It rather depends on the composition of the exhaust gas and on numerous particle properties. The most important example of composition dependent sensitivity is NO_2. This is a gaseous exhaust component with a brown color, which also attenuates light. It will lead to a considerable amount of opacity. At a measurement length of 430 mm a concentration of 100 ppm causes an opacity of approximately 2 % [31]. This is equal to an extinction coefficient of 0.047/m. Although negligible for older vehicles, for modern vehicles this amount may surpass the effect of the particles. The scattering effect of NO_2 however is extremely small at such concentrations and typically below the detection limit. Consequently, in contrast to light transmission, light scattering can be regarded as a pure particle measurement technique.

Despite the issues discussed above the correlation between the light scattering signal and the extinction coefficient is very good. The result of a comparison measurement campaign using a Combustion Aerosol Standard (CAST) particle generator is depicted in fig. 7. The CAST allows the production of soot particles with different sizes and concentrations [32]. The correlation between the absorption coefficient measured by a reference opacity meter and the absorption coefficient calculated from the scattering signal is good over the whole displayed range of data.

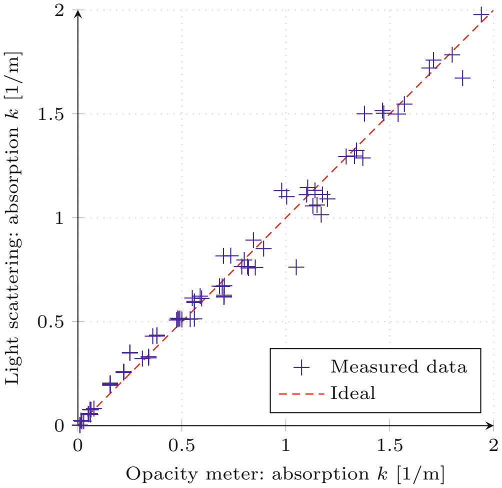


Fig. 7. Correlation between the light scattering device and an opacity meter for various CAST operating points

The extrapolation of the conversion formula permits the calculation of the resolution achievable by light scattering. It is in the order of $k = 0.0001/\mathrm{m}$. Compared to typical garage opacity meters, whose resolution is $k = 0.01/\mathrm{m}$, an improvement by a factor of 100 is attained [33]. Good results are also achieved at k values as high as 2/m. It is obvious that a device based on light scattering is well capable of replacing the opacity meter in the garages and to perform measurements at both current and future internal-combustion engines.

5 Calibration

One prerequisite for the usage of the light scattering device for roadworthiness evaluation of cars in the garages is the availability of a traceable calibration methodology. One of the possible ways to achieve this goal is the usage of a CAST as a well-defined exhaust source. This source is used for the calibration of the light scattering device against a so called "golden device" as a reference. The golden device might be a reference Opacimeter [34] or a gravimetric standard as used in light and heavy duty type approvals [35]. For type approval this might be a way to follow, but for the need of yearly calibration of the devices in the field this methodology lacks practicability.

Another way is the usage of a standardized scattering medium, which can be inserted into the laser light path of the light scattering device. The requirements for this medium are well known: stable optical properties and a rather uniform angular scattering intensity distribution. These requirements are fulfilled by a calibration gas showing rather strong Rayleigh scattering. As derived from equation (3),

$$I_R = I_0 N l_0 \Omega \left(\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega}\right)_{\mathrm{eff}} . \tag{9}$$

Consequently there is the possibility to calculate the scattering intensity per unit solid angle from first principle equations by knowing the effective Rayleigh cross section of the specific calibration gas. In the above equation I_R is the detected signal, I_0 the incident intensity, N the number concentration, l_0 the length of the probe volume, Ω the solid angle for detection and $(\mathrm{d}\sigma/\mathrm{d}\Omega)_{\mathrm{eff}}$ the Rayleigh

Table 2. Rayleigh cross sections for different gases at $\theta = 90°$, $\lambda = 532\,\mathrm{nm}$, $T = 273\,\mathrm{K}$ and $p = 1.013 \times 10^5\,\mathrm{Pa}$

Name	Formula	σ [$10^{-28}\,\mathrm{cm}^2/\mathrm{sr}$]	I_R/I_0 [$l_0 \times 10^{-8}/(\mathrm{sr\,m})$]
Oxygen	O_2	5.08	1.36
Nitrogen	N_2	6.13	1.64
Hydrogen	H_2	1.34	0.359
Carbon monoxide	CO	7.78	2.09
Carbon dioxide	CO_2	13.8	3.70
Water vapor	H_2O	4.43	1.19
Propane	C_3H_8	79.8	21.4

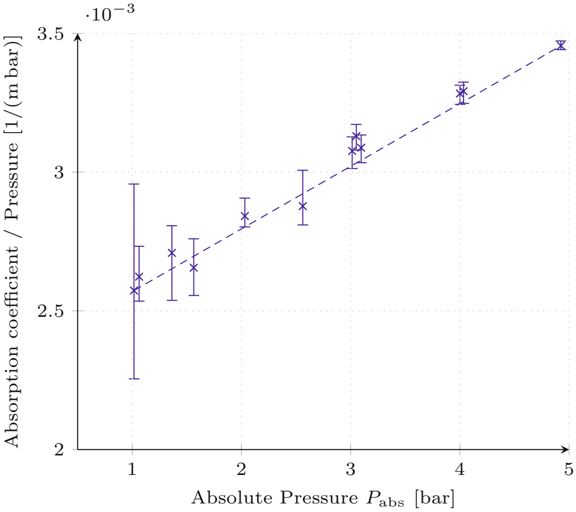


Fig. 8. Pressure dependent absorption coefficient measured in the scattering device with a propane-like gas

cross section. In tab. 2 effective Rayleigh cross sections of different gases are shown.

In order to calibrate scattering intensities comparable to modern engine exhaust, gases with scattering cross sections in the range of $80 \times 10^{-28}\,\mathrm{cm}^2/\mathrm{sr}$ have to be chosen, which result in extinction coefficients in the range of 0.001/m. For reasons of practicability the gases should be neither toxic nor inflammable. In fig. 8 the response of the light scattering device with a propane like calibration gas is shown. The error bars indicate the measured signal noise, which decreases significantly when going to higher gas pressures.

For higher scattering intensities a different approach has to be chosen. One possible way is the usage of a plug which is reproducibly brought into the laser light beam. The plug consists (see fig. 9) of a light absorbing glass tube filled with a medium, comprised of micrometer sized polystyrene spheres suspended in neutral-buoyancy-solution (water/glycerol), whereas the liquid phase of the matrix allows speckle averaging through Brownian motion.

Due to the spatial separation of absorbing layer and scattering spheres individual adjustment of the angular scattering distribution as well as the scattering intensity is possible. The scattering spheres provide isotropic light distribution over broad angular range via multiple Mie scattering and the absorbing glass allows independent adjustment of the overall intensity. By that means calibration in the respective extinction coefficient range from 0.5/m to 2/m is possible.

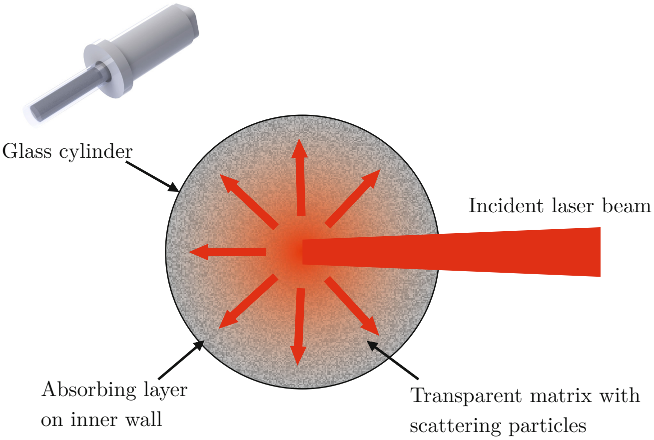


Fig. 9. Calibration plug consisting of an absorbing glass tube and a scattering medium. For visual clarity, individual components are not to scale.

6 Conclusion

The exhaust measurement during the periodical technical inspection (PTI) is an important tool to check the compliance of the vehicles to their emission limits. Due to the strong reduction of the particle emissions the opacity meter, which is currently in use for that purpose in the garages, lacks the necessary sensitivity for new engines. Devices based on light scattering can replace them. They are robust and highly suitable for operation in the garages. They deliver both a high temporal resolution ($f = 100\,\text{Hz}$) and a very good resolution of the measured value ($k = 0.0001/\text{m}$). The calculation of the extinction coefficient from the scattering intensity is very promising. Further series of measurements aim at validating the conversion formulas. The calibration of the devices can be performed using a calibration gas. It is especially applicable for low level calibration points. One issue is the inflammableness of the propane-like gas used in our calibration tests. For higher calibration points a calibration technique based on some type of a solid plug can be used.

References

[1] Ristovski, Z.D., Miljevic, B., Surawski, N.C., Morawska, L., Fong, K.M., Goh, F., Yang, I.A.: Respiratory health effects of diesel particulate matter. Respirology 17(2), 201–212 (2012)
[2] Giechaskiel, B., Mamakos, A., Andersson, J., Dilarab, P., Martini, G., Schindler, W., et al.: Measurement of automotive nonvolatile particle number emissions within the european legislative framework: A review. Aerosol Science and Technology 46(7), 719–749 (2012)

[3] Mollenhauer, K., Tschöke, H.: Handbook of Diesel Engines, 3rd edn. Springer, Heidelberg (2007)
[4] Eastwood, P.: Particulate Emissions from Vehicles. Wiley, Chichester (2008)
[5] Nickel, C., Kaminski, H., Hellack, B., Quass, U., John, A., Klemm, O., Kuhlbusch, T.A.: Size resolved particle number emission factors of motorway traffic differentiated between heavy and light duty vehicles. Aerosol and Air Quality Research 13(2), 450–461 (2013)
[6] International Agency for Research on Cancer, World Health Organization, Diesel engine exhaust carcinogenic. Press Release No. 213 (2012)
[7] Rakopoulos, C.D., Giakoumis, E.G.: Diesel Engine Transient Operation: Principles of Operation and Simulation Analysis. Springer, London (2009)
[8] Mamakos, A., Manfredi, U.: Physical characterization of exhaust particle emissions from late technology gasoline vehicles. JRC scientific and policy reports, Joint Research Center (2012)
[9] Kirchner, U., Gallus, J., Börensen, C., Vogt, R.: Particle number emission of light-duty vehicles during real-world driving. In: 17th ETH Conference on Combustion Generated Nanoparticles, Zürich (2013)
[10] European Commission, Commission Directive 2010/48/EU of 5 July 2010 adapting to technical progress Directive 2009/40/EC of the European Parliament and of the Council on roadworthiness tests for motor vehicles and their trailers. OJ L 173, 42–72 (2010)
[11] Bohren, C.F., Huffman, D.R.: Absorption and scattering of light by small particles. Wiley, New York (1998)
[12] van de Hulst, H.C.: Light scattering by small particles. Dover Publications, New York (1981)
[13] Young, A.T.: Rayleigh scattering. Phys. Today 35(1), 42–48 (1982)
[14] Mie, G.: Contributions to the optics of turbid media, particularly of colloidal metal solutions. Annals of Physics 25(3), 377–445 (1908)
[15] Chakrabarty, R.K., Moosmüller, H., Arnott, W.P., Garro, M.A., Slowik, J.G., Cross, E.S., Han, J.-H., et al.: Light scattering and absorption by fractal-like carbonaceous chain aggregates: comparison of theories and experiment. Applied Optics 46(28), 6990–7006 (2007)
[16] Görner, P., Bemer, D., Fabriés, J.F.: Photometer measurement of polydisperse aerosols. Aerosol Science 26(8), 1281–1302 (1995)
[17] Hull, P., Shepherd, I., Hunt, A.: Modeling light scattering from diesel soot particles. Applied Optics 43(17), 3433–3441 (2004)
[18] Kittelson, D.B.: Engines and nanoparticles: A review. Aerosol Science 29(5/6), 575–588 (1998)
[19] Reinisch, T.: Über die Morphologie von Rußpartikeln von Verbrennungskraftmotoren im Hinblick auf Partikelanzahlmessung, Bachelor Thesis, Graz, University of Technology (December 2009)
[20] Forrest, S.R., Witten, T.A.: Long-range correlations in smoke-particle aggregates. Journal of Physics A: Mathematical and General 12(5), L109–L117 (1979)
[21] Mandelbrot, B.B.: Fractals: Form, chance, and dimension. W. H. Freeman & Co., San Francisco (1977)
[22] Friedlander, S.K.: Smoke, Dust, and Haze: Fundamentals of Aerosol Dynamics, 2nd edn. Oxford University Press, New York (2000)
[23] Lapuerta, M., Ballesteros, R., Martos, F.J.: A method to determine the fractal dimension of diesel soot agglomerates. Journal of Colloid and Interface Science 303, 149–158 (2006)

[24] Sorensen, C.M.: Light scattering by fractal aggregates: A review. Aerosol Science and Technology 35(2), 648–687 (2001)

[25] Sorensen, C.M.: The mobility of fractal aggregates: A review. Aerosol Science and Technology 45(7), 765–779 (2011)

[26] Farias, T.L., Carvalho, M.G., Köylü, U.O., Faeth, G.M.: Computational evaluation of approximate rayleigh-debye-gans/fractal-aggregate theory for the absorption and scattering properties of soot. Journal of Heat Transfer 117(1), 152–159 (1995)

[27] Miller, T.M.: Atomic and molecular polarizabilities. In: CRC Handbook of Chemistry and Physics, vol. 71 (2000)

[28] Sobolewski, R., Verevkin, A., Gol'tsman, G.N., Lipatov, A., Wilsher, K.: Ultrafast superconducting single-photon optical detectors and their applications. IEEE Transactions on Applied Superconductivity 13(2), 1151–1157 (2003)

[29] Jung, W.G.: Op Amp Applications Handbook, 1st edn. Newness, Burlington (2005)

[30] Axmann, H., Eichberger, B.: Circuit configuration for evaluating photoelectric measurements, European Patent Application EP2501036A1 (September 2012)

[31] Pongratz, H., Schindler, W., Singer, W., Striok, S., Thaller, W.: Transient measurement of diesel engine emissions. MTZ Worldwide eMagazine 64(10), 13–16 (2003)

[32] Jing, L.: Standard combustion aerosol generator (SCAG) for calibration purposes. In: 3rd ETH Workshop "Nanoparticle Measurement", ETH Zürich (1999)

[33] Norris, J.O.W.: Low emission diesel research CP17/18/770. Phase 3-Report, Vehicle and Operator Services Agency (VOSA) (2005)

[34] AVL, AVL 439 Opacimeter – emission test instruments product description (2008), `https://www.avl.com/opacimeter`

[35] European Commission, Regulation No 83 of the Economic Commission for Europe of the United Nations (UN/ECE). OJ L 375, 223–495 (2006)

Noninvasive Estimation of Blood Glucose Concentration Using Near Infrared Optodes

Swathi Ramasahayam, K. Sri Haindavi, and Shubhajit Roy Chowdhury*

Center for VLSI and Embedded Sytems Technology,
IIIT Hyderabad, Gachibowli, Hyderabad, India
`src.vlsi@iiit.ac.in`

Abstract. Early detection of diabetes is indispensable for safer and healthier society as millions of people suffer from diabetes mellitus. Traditionally invasive techniques are used to estimate the concentration of blood glucose. Many efforts have been made to reduce the level of invasiveness of the glucose monitoring by decreasing the blood sample volume. The challenge for non invasive assays is to develop transducers with high sensitivity, capable of detecting weak blood signals that loose energy through intervening tissues (bone, fat, skin etc), and also to separate glucose information from other overlapping blood constituents with much higher concentration (water, haemoglobin, uric acid, urea, proteins). Many biomedical spectroscopy studies have been carried out in the mid infra-red region of the spectrum where the absorbance is high but the cost of the equipment becomes prohibitively high because of the high cost of LED. This chapter deals with the spectroscopic techniques carried out in near infrared (NIR) part of the spectrum to estimate the concentration of blood glucose. The spectroscopy has been performed at the second overtone of glucose which falls in the NIR region. The NIR spectroscopy has been performed based on NIR LED and photo detector constituting an optode pair, using transmission photoplethsymography (PPG). The analog front end system has been implemented to get the PPG signal at the near infra-red wavelengths of 1070nm, 950nm, 935nm. The PPG signal that has been obtained is processed and double regression analysis has been carried out with the artificial neural network for estimating the glucose levels. The root mean square error of the prediction was 5.84mg/dL.

Keywords :on invasive estimation of blood glucose concentration, NIR spectroscopy, photoplethsymography, near infra red optodes, artificial neural network.

1 Introduction

Diabetes mellitus is a wide spread disease effecting millions of people around the world. According to the world health organization (WHO), approximately 350 million people are suffering from diabetes worldwide. In the human body the pancreas is

---

* Corresponding author.

A. Mason et al. (eds.), *Sensing Technology: Current Status and Future Trends IV*,
Smart Sensors, Measurement and Instrumentation 12, DOI: 10.1007/978-3-319-12898-6_4

responsible for blood glucose control. By producing and releasing the counteracting hormones insulin and glucagon blood glucose concentration can be decreased or increased respectively and stabilized within the physiological range of 70 – 120mg/dl. Diabetes mellitus is characterized by a dysfunction of the pancreas often in combination with reduced insulin sensitivity. These metabolic disturbances are generally caused byreduced insulin sensitivity of the glucose consuming cells or deteriorated glucose sensing of the pancreas. Long term diabetes is dangerous as it can lead to blindness, damage nerves and kidneys. It also increases the risk of heart stroke and other heart diseases [1].

Currently invasive methods are being used to estimate the blood glucose concentration. These invasive approaches involve pricking the finger for blood sample which is painful and has risk of spreading infectious diseases. It does not provide instantaneous and continuous results as there is a delay between blood sample collection and the analysis. Many efforts have been done in order to reduce the level of invasiveness by decreasing the volume of blood being collected to few microliters. These methods result in decreased patient compliance and a failure to control the blood sugar levels. The need therefore becomes pertinent for a convenient noninvasive blood glucose monitor. Recently some significant progress toward, noninvasive optical glucose sensing have come about because of the availability of the new technologies. Optical glucose measurement techniques are particularly attractive for several reasons.They utilize nonionizing radiation to interrogate the sample, they do not generally require consumable reagents and they are fast. Cost of using reagent constitutes a substantial portion of the operating expenses in invasive method.

Different spectroscopic measurement techniques like raman spectroscopy, polarimetry, and fluroscence spectroscopy [2-4] have been employed for noninvasive measurement of glucose concentration in blood. Due to disadvantages like lack of selectivity and requirement of heavy, delicate and expensive instruments these methods are not widely used. Photoplethsmograpy (PPG) based near infrared spectroscopy is one of the noninvasive measurement technique which is most appreciated because of its easy methodology and inexpensive experimental setup[5]. Although optical approaches for glucose sensing are attractive, they are often plagued by a lack of sensitivity or specificity since variations in optical measurement depends on variations of many factors in addition to glucose concentration is a significant challenge in itself. Blood contains many components whose spectra overlap with each other in the near infra-red region (NIR) as the overtones of the many chromophores fall under this region. This requires developing transducers with high sensitivity, capable of detecting weak blood component signals that lose energy through intervening tissues like bone, fat, skin and to separate analyte information from other overlapping constituents with much higher concentration. Spectrophotometric measurements are done based on the structure of the analyte molecule. The measurement is done using PPG at the near infrared part of the spectrum ranging from 0.7μm to 1.5μm [6-8] which constitutes second overtone absorption bands primarily associated with C-H, O-H and N-H stretching vibrations.

2 Basics of NIR Spectroscopy

Near-infrared spectroscopy (NIRS) is a spectroscopic method that uses the near-infrared region of the electromagnetic spectrum starting from 780 nm to 2500 nm (12800cm^{-1} to 4000cm^{-1}). NIR spectroscopy is a sub-set of infrared spectroscopy.

Near infrared (NIR) spectroscopy has raised a lot of interest in the area of medical diagnosis because of its non-invasive analytical technique (Blanco et al., 1998). A great advantage of NIR spectroscopy is that there is no need for extensive sample preparation. The chemo-metric methods (i.e. the mathematical procedures for extracting the useful information) used are complicated and demand acquaintance with statistics (Lavine, 1998).NIR spectra contain information of both chemical and physical properties of the sample.

2.1 PhysicochemicalPrinciples of NIR Spectroscopy

Near infrared region lies between the visible and mid infrared (MIR) regions of an electromagnetic spectrum. It is defined as the range of electromagnetic radiation in the region of about 780-2500 nm (12821-4000 cm^{-1}).[9] Physics of NIR spectroscopy deals with the interaction of light with matter that occurs at the atomic and molecular levels. These interactions generally occur quickly and over small distances. As a result one must rely on chemical basis of NIR spectroscopy. In general molecules can be thought of having two types of properties, static and dynamic. The static properties include atomic composition, conformation, isomeric structure, stereochemistry and morphology. The dynamic properties such as molecular translation, rotation and vibration can also be considered as forms of molecular energy. From the law of conservation of energy, we can be certain that the specimen does not create light by itself. Instead it must produce the light form from selective absorption of light entering the sample or by emission of certain light forms as result of the sample losing some of its internal energy.

As mentioned, molecular energy can take many different forms, such as translational, rotational, electronic and vibrational. Vibrational energy refers to the oscillations of atoms through their bonds in a molecule much like two balls could oscillate if they were connected by a spring. The energy of the system consisting of two atomic balls of masses m_1 and m_2 attached by a bond spring is

$$E = \frac{h}{2\pi}\sqrt{\frac{k}{\mu}} \tag{1}$$

Where k is the force constant of the bond between the atoms and μ is the reduced mass

$$\mu = \frac{m_1 m_2}{m_1 + m_2} \tag{2}$$

Absorption of radiation of molecules in this region is primarily caused by overtones and combinations of fundamental vibration bands that appear in MIR region.

3 Theoretical Analysis

Spectroscopy is the study of the structure and dynamics of moleculesvia the interaction of light with matter. The different parts of the light spectrum have different effects on molecules and atoms.These interactions of light with matter are used to obtainchemical information about the samples. Near Infrared spectroscopy is a method based onvibrational transitions of molecular bonds.

3.1 Molecular Composition and Relative Absorptivities

The bonds in a molecule can be modeled with a spring having attached weights, where the whole system can vibrate.Hence vibrational frequencies are sensitive to the structure of the compound and each bond has different vibrational frequencies and the frequency varies slightly with molecule as neighboring atoms affect the bond energy , charge distribution.

When a photon is incident on a molecule, there will be bond deformations or bond vibrations at different energy levels related to different bonds, depending on the energy of incident photon. So, only the photon with energy that corresponds to the difference between two of its energy levels can be absorbed [10]. The frequency of the vibration is given by the

$$v = \frac{1}{2\pi}\sqrt{\frac{k}{\mu}} \tag{3}$$

Where ‘k’ is the bond strength and ‘μ’ is the reduced mass of the two bond atoms of masses m_1 and m_2.

Depending on the energy of incident light bond deformations or bond vibrations are possible. Generally bond deformations need less energy than bond vibrations. From the above equation fundamental frequency of vibration of a bond can be determined and the bond can vibrate only at frequencies multiples of fundamental frequency and these frequencies are called overtones.

So light of only those frequencies will be absorbed which can excite bond vibrations at fundamental or overtone frequencies.

From equation 3 fundamental frequencies of bond vibrations can be determined and from them overtone frequencies can be determined. Overtone frequencies are not exactly multiples but close to them. For a glucose molecule, the molecular structure is as shown in Fig1.

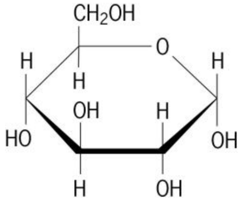


Fig. 1. Molecular structure of D-glucose

Table 1 shows the different bonds present in a glucose molecule and fundamental frequencies of vibration of those bonds.

Table 1. Fundamental frequencies of glucose bonds

Bond	Exciting wavelength range
C-O	(9345-8695)nm
C-C	(12,500-7692)nm
C-H	(3508-3333)nm
O-H	2941nm

Wavelength of light required for molecular vibration of C-O and C-C are very high. Hence only higher overtones will fall under below 1500 nm wavelength range. Intensity of light absorbed depends on probability of the photon being absorbed by the material which decreases with increasing overtones. Hence relative absorbencies rapidly decrease with increasing overtones.

The fundamental IR absorption bands of glucose have been reported in solid pellets and in solution (11-15). The strongest bands that can generate intense combinations and overtones are the broad OH stretch at 3550 cm^1 and the COH stretch vibrations at 2961 and 2947 cm^1. Table 2 shows the fundamental and overtone frequencies for C-H and O-H bonds.

Table 2. Fundamental and overtone frequencies of C-H and O-H bonds

Overtone	Bond	wavelength	bond	wavelength
Fundamental	vC-H	3377nm	vO-H	2817nm
First overtone	2vC-H	1688nm	2vO-H	1408nm
Second overtone	3vC-H	1126nm	3vO-H	939nm
Combination	C-H+O-H		1536nm	

Possible combination bands are a second OH overtone band at 939 nm (3vOH) and a second harmonic CH overtone band at 1126 nm (3vCH). A first OH overtone band can be assigned at 1408 nm (2vOH). The 1536 nm band can be assigned as an OH and CH combination band (vOH + vCH). The 1688 nm band is assigned as a CH overtone band (2vCH). Other bands at 2000 nm are possibly combinations of a CH stretch and a CCH, OCH deformation at 2261 nm (vCH + vCCH, OCH) and 2326 nm (vCH + v CCH, OCH). The presence of the CCH, OCH ring deformation component confers some glucose specificity on these bands.

From above data we know the energy required for C-H fundamental vibration and vC-H +v (CCH, OCH). So we can get the energy required for bond deformations of CCH and OCH.

As wavelength decreases, energy increases. In order to obtain a wavelength for noninvasive glucose monitoring below 1500nm the O-H bond vibrational energy is added to C-H stretch and CCH,OCH bond deformations giving rise to an bond vibrational energy of vC-H + v CCH,OCH + vO-H at 1254 nm.

From the quantitative analysis we can say that 1254nm can be assigned as a combination band (vOH + vCH) and the deformation band (v CCH, OCH), 1126nm can be assigned as a second overtone C-H band and 939nm as the second overtone O-H band. Hence 939nm,1126nm, 1254nm can be used for noninvasive glucose detection.

At a deeper level, absorption of light can be seen as dependent on the probability of absorbance of a photon by the molecule. For nth overtone final energy is (n+1)*E, where E is the fundamental energy. As n increases, probability of absorbance decrease rapidly and hence intensities of absorbance decrease as overtones increase. The absorption at fundamental frequency is calculated and from that the absorption at second overtone is calculated relatively [11]. Table 2 depicts the relative absorptivities of C-H bond at different overtones with respect to the fundamental. It is evident that the absorptivity decreases rapidly as one goes from fundamental to the overtones in sequence.

Table 3. Relativeabsorptivities of C-H bond

Overtone	Wave number(cm-1)	Relative Absorptivity
Fundamental	3019	1
First Overtone	5912	0.088
Second Overtone	8677	3.2 * 10-3

3.2 Selection of the Wavelength of Peak Absorption

The absorption spectrum of the glucose has been studied in order to choose the wavelengths for LEDs. For this purpose an IR absorption spectrum of 0.1M aqueous glucose solution has been collected and analyzed in second overtone region of the near-infra red spectra using the Bruker tensor 27 FTIR spectrophotometer. Fig.2 shows the absorption spectrum of the glucose over the second overtone region of the near infrared spectrum where the selected wavelengths are shown. We can observe that the absorption peaks in this region are very narrow typically of the order of the 2nm to 5nm but the LED emits the light over a range of wavelengths. The wavelengths are chosen insuch a way that the weighted average of the absorption over the spectral bandwidth of the LED is high. While calculating this weighted average the intensity of light emitted by the LED acts as weight for the absorption at that particular wavelength.

The expected value of the absorbance over a range has been calculated as

$$E[A] = \int_{-\infty}^{+\infty} A(\lambda) * f(\lambda) d\lambda \tag{4}$$

Where ‘A’ is the absorbance of glucose over a range of wavelengths and ‘f (λ)’ is the probability density function of intensity of light emitted.

By multiplying the relative intensity of the light emitted by the LED shown in Fig.3 and the absorbance value of glucose taken from the spectrum and averaging it over spectral distribution of LED we calculate the mean value of the absorbance which is shown in Fig.4.

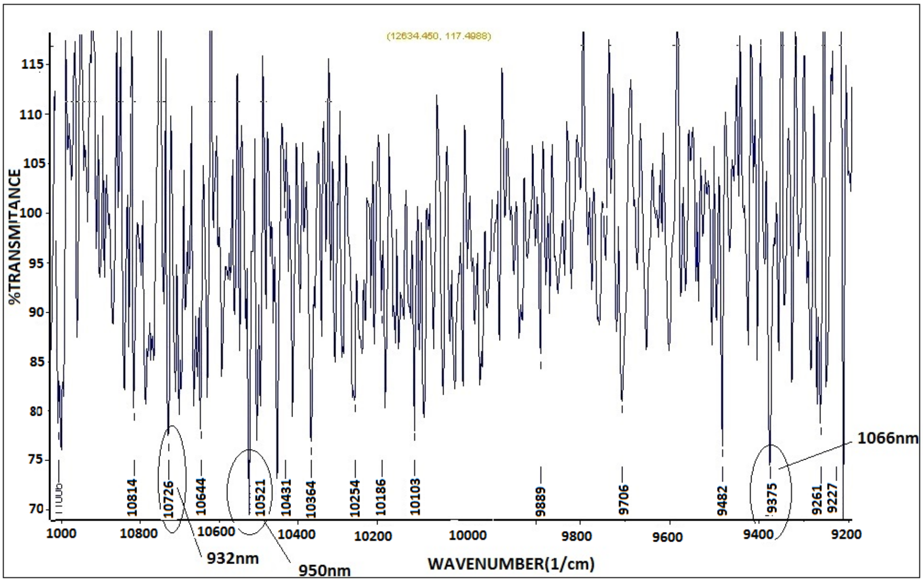


Fig. 2. Spectrum showing the transmitance of glucose

In order to improve the accuracy of the prediction of blood glucose level, the wavelengths have been chosen such that mean value of the absorbance of glucose over the frequency range of LED is high. For the current research, wavelengths of 935nm, 950nm, and 1070nm have been chosen based on the experimental analysis.

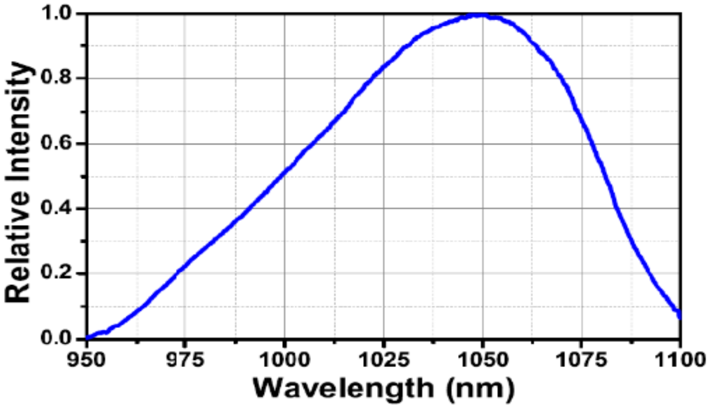


Fig. 3. Graph showing the relative senstivity of light emitted by 1070nm LED

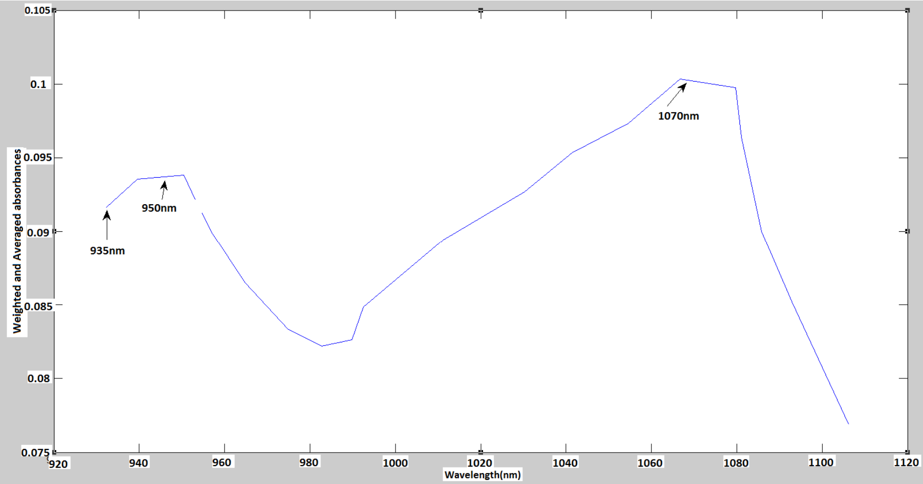


Fig. 4. Intensity weighted mean absorbances over 920nm to 1100nm

4 Photoplethysmography

Plethysmography is a technique of measuring the blood volume changes in any part of the body that result from the pulsation of the heart. These measurements are useful in the diagnosis of arterial obstructions and pulse wave velocity measurements. Photoplethysmography is the simple and low cost opto-electronic method for measuring the blood volume changes in the microvascular bed of tissue [5]. It is often used non-invasively to make the measurements at the skin surface. The PPG technology has been used in wide range of commercially available device like pulse oximetry which measures the oxygen saturation, blood pressure and cardiac pulse for assessing other cardiac related diseases.

4.1 Basic Principle of Operation

The interaction of light with tissue includes the optical processes of scattering, absorption, reflection, and transmission. The main constituent of tissue is water that absorbs light very strongly in the ultra violet and long infra-red wavelengths. The shorter wavelengths are also absorbed by melanin. There is a window where water does not show any peaks in the absorption spectrum which allows using the visible and near infra-red wavelengths for measurement of blood flow or volume. It uses a clip which contains a light source and a detector on the opposite sides to detect the cardio vascular pulse wave that propagates through the body. The PPG waveform consists of two components one is the AC component because of the pulsations of the heart with fundamental frequency typically around 1Hz. This AC component is superimposed on a large DC. This DC component is due to respiration, vasomotor activity and vasoconstrictor waves. By filtering and amplification both AC and DC can be extracted for further analysis. Fig.5 shows the PPG pulses and its various components.

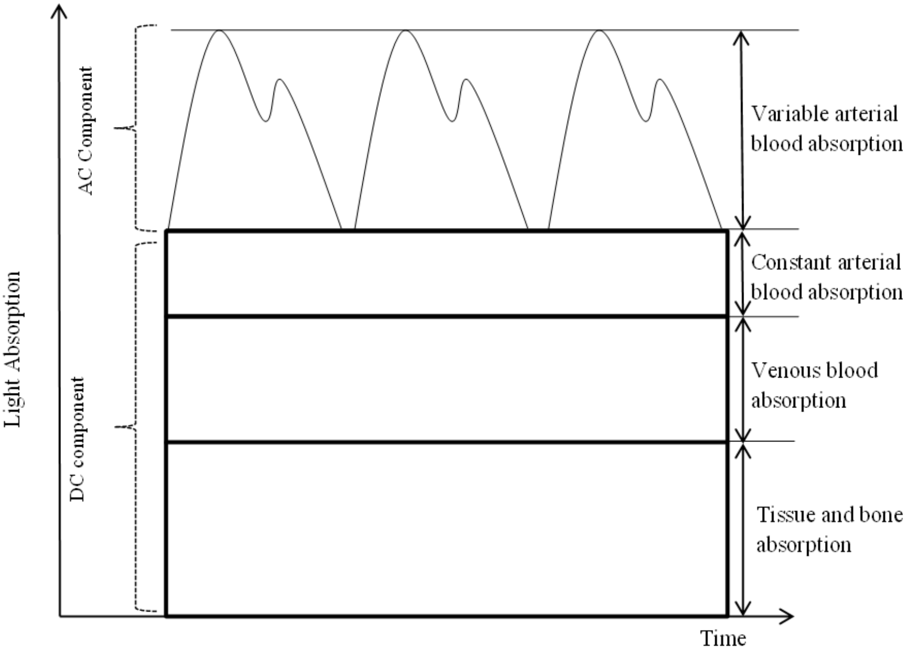


Fig. 5. PPG pulses with AC and DC components

According to Beer-lambert's law the absorbance of light by a liquid is related to the concentration of the material by

$$A = \in Cl \tag{5}$$

Where $\in$ the molar absorptivity of solute at a particular wavelength, C is is the concentration of the solute and l is the optical path length.

Hence for a specific wavelength i, equation (3) may be written as

$$A_i = \epsilon_i \, C_i l_i \tag{6}$$

5 Design Methodology

The PPG sensors are often made with the low cost LED and matched photo detector devices working at red or near infra-red wavelengths. LED's convert electrical energy to light energy, they typically have a narrow bandwidth of the order of 50nm. The photo detector must be chosen such that the spectral characteristics should match that of the light source. A photo detector converts light energy to electrical current. This electric current is converted to appropriate voltage by and low noise electronic circuitry which consists of trans-impedance amplifier and filtering circuitry. The high pass filter is used to remove the dominant DC component and the AC component is amplified and is low pass filtered for pulse wave analysis.

5.1 Analog Front End Sensing Circuit

As discussed in the previous section, three different wavelengths have been selected which have peak absorption of glucose in the near infrared region of the spectrum, viz 1070, 950 and 935 nm. The block diagram of the glucose sensing system is shown in the Fig.5.

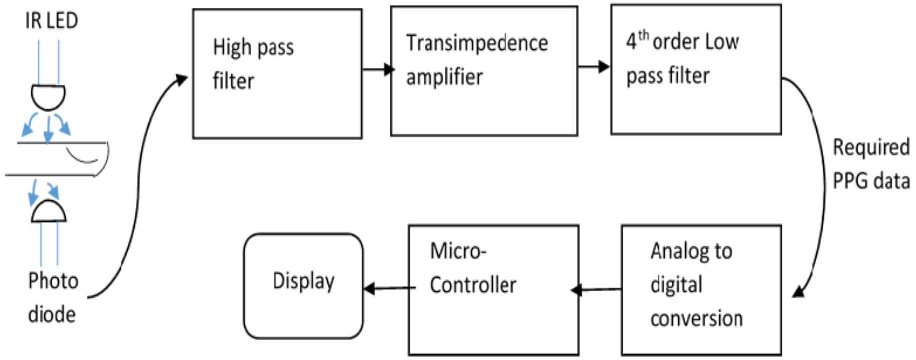


Fig. 6. Block Diagram of the glucose sensing system

It consists of the finger clip with LED acting as a light sensor and the photodiode as the detector to detect the small changes in the incident light as it passes through the finger. This light is converted in to an equivalent current by the detector and is high pass filtered with a cut off frequency of 0.8Hz. Then it is given to the trans-impedance amplifier for amplification of the signal. For this a low noise amplifier LMV796 has been used. The input referred noise of the amplifier is less than 17nv/√Hz. After necessary amplification it has to be low pass filtered in order to get PPG wave. A 4th order sallen-key topology has been used to build a low pass filter whose cut off

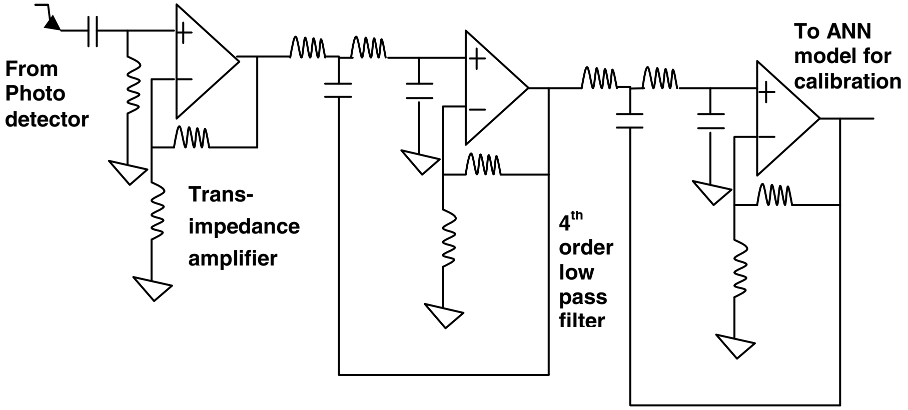


Fig. 7. Circuit diagram of analog front end of PPG system

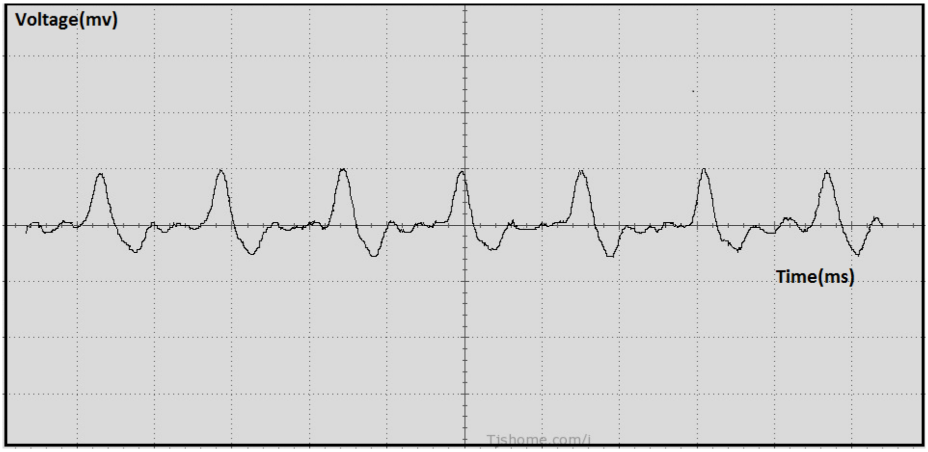


Fig. 8. PPG waveform for 950 nm

frequency is 10Hz. Fig. 6 shows the circuit level diagram of analog front end for PPG system and Fig.7 shows the PPG output waveform.

For getting glucose level in blood the absorbance which is mainly due to blood glucose is calculated from AC component of PPG wave as follows

$$\Delta OD_{\lambda} = \log\left[1 + \left(\frac{\Delta I_{\lambda}(t_i)}{I_{\lambda}(t_{i+1)}}\right)\right] \tag{7}$$

Where ΔOD_{λ} is difference between optical densities at time t_i and t_{i+1},$\Delta I_{\lambda}(t_i)$ is the pulsatile component at time t_i and $I_{\lambda}(t_{i+1)}$ is the intensity of light at time t_{i+1}.

This difference has the effect of removing the venous and the tissue contributions to yield only the change in intensity due to the pulsating arterial blood compartment [16,17].

The optical densities have been calculated at three wavelengths for 35 subjects. The actual glucose values have also been determined using standard invasive glucometer. Regression analysis has been done on these values using neural network toolbox in MATLAB.

5.2 Regression Analysis Using Artificial Neural Networks

Multivariate calibration models have come into wide use in quantitative spectroscopic analyses due to their ability to overcome deviations from the Beer Lambert's law caused by effects such as overlapping spectral bands and interactions between components [14].PLS and PCR are the most widely used chemo-metric techniques forquantitative analysis of complex multicomponent mixtures. These methods are notoptimal when the relationship between the IR absorbance's and the constituent concentrationdeviates from linearity. The theory and the application of Artificial Neural networks (ANN) in modeling chemical data have been widely presented in the literature [15].

Artificial neural networks can be used for function fitting to develop a model based on the PPG data and invasiveglucose measurements.

For calculating glucose concentration from beer-lambert's law we can say that

$$A[\{C_i\},\lambda] = \sum_{\mathrm{i}} (C_i A_i(\lambda)) \tag{8}$$

Where $A_i(\lambda)$ is the optical density at wavelength λ, of the i^{th} component, whose concentration is C_i, A(λ) the optical density of the mixture. If we have the spectra ofall the components $\{A_i(\lambda)\}$ and measure the spectrum of the mixture $\{A_i(\lambda)\}$, we canuse a mathematical tool to estimate the $\{C_i\}$.The three wavelengths have been chosen such that glucose has considerable amount of absorption when compared to other components. So we have given the matrix

$$\begin{bmatrix} A_{\lambda 1 i=1\ldots 25} \\ A_{\lambda 2 i=1\ldots 25} \\ A_{\lambda 3 i=1\ldots 25} \end{bmatrix} \tag{9}$$

as input and $[C_{i=1\ldots 25}]$ as output which are measured using standard invasive method. In this way we have three different optical densities as inputs and one output for a single sample.

MATLAB neural network toolbox has been used. A two layer feed forward network with sigmoid hidden neurons and linear output neurons has been chosen. Numbers of hidden layers have been chosen as

To minimize the deviations and increase the sensitivity of the device regression has been used twice i.e.., double regression. So we get two sets of regression coefficients for two different datasets.

Initially, the first set of regression coefficients are used to estimate Glucose levels. But, if it falls above 110 mg/dL second set of regression coefficients are used to estimate the glucose concentration accurately.

The performance of the regression analysis was evaluated in terms of root mean square error of prediction over the first data set which has glucose concentration below 110mg/dL is given by Eq. (10)

$$RMSE_1 = \sqrt{\frac{\sum_{i=1}^{N1}(x_i - y_i)^2}{N1}} \tag{10}$$

The test samples whose estimated glucose concentration fall above 110mg/dl are taken as second data set and regression analysis is performed whose root mean square error of prediction is given by

$$RMSE_2 = \sqrt{\frac{\sum_{i=1}^{N2}(x_i - y_i)^2}{N2}} \tag{11}$$

The overall root mean square error can be obtained as

$$RMSE_T = \sqrt{\frac{N1(RMSE_1)^2 + N2(RMSE_2)^2}{N1 + N2}} \tag{12}$$

Results of regression analysis on the first data set are shown in Fig.9 where the parameter R signifies the correlation between estimated glucose and the actual glucose levels. We can observe that R is nearly equal to 1 for training, validation and test. For this a root mean square error ($RMSE_1$) of 5.61mg/dL is obtained. Results of the second regression analysis performed are shown in Fig.10, the root mean square error ($RMSE_2$) is 6.32mg/dL.

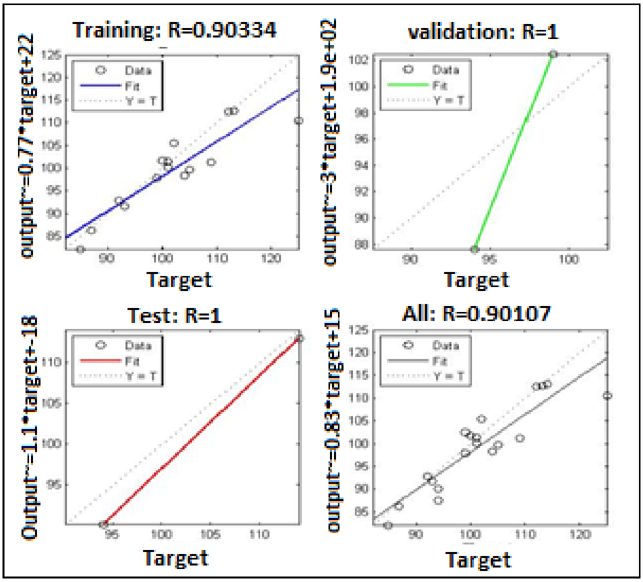


Fig. 9. Regression analysis on first data set

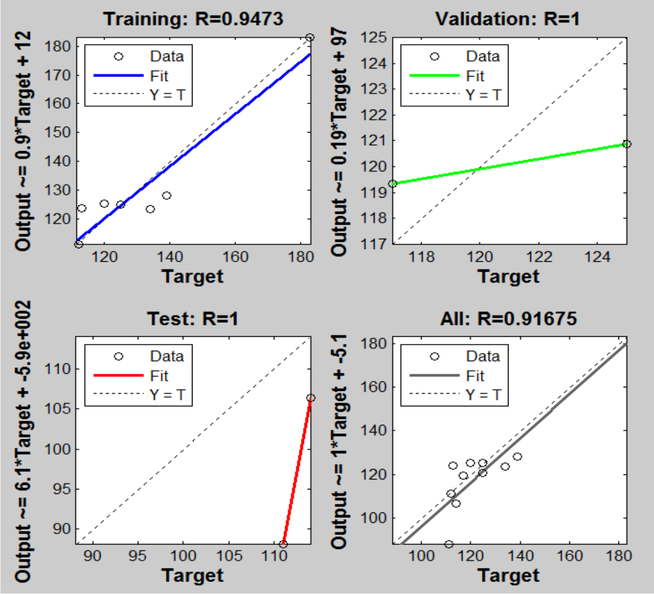


Fig. 10. Regression analysis on filtered data set (>110mg/dl)

6 Conclusion

It is observed that noninvasive monitoring of glucose is indispensible for safer and healthier society. There is need for the development of a convenient non-invasive sensor which can detect the trace levels of vital blood parameters like glucose as it holds the potential to significantly reduce the mortality associated with the diabetes. Use of opto- electronic based technique like photoplethysmography (PPG) combined with the near infra-red spectroscopy provides a fast and easy assessment of blood glucose while negating much of the cost,time delay and issues associated with traditional invasive methods.

This chapter reviewed the noninvasive estimation of blood glucose concentration based on the chemical and physical principles of the near infra-red spectroscopy of the analyte. It is observed that the choice of the light source plays the major role for getting better accuracy.A technique for implementation of measurement system based on the artificial neural networks is proposed.

References

[1] Heise, H.M., Meyers, R.A. (ed.) Encyclopedia of Analytical Chemistry, vol. 1, pp. 1–27. John Wiley Press, New York (2000)
[2] Maier, J.S., Walker, S.A., Fantini, S., Franceschini, M.A., Gratton, E.: Opt. Lett. 19(24), 2062–2064 (1994)

[3] Wicksted, J.P., Erckens, R.J., Motamedi, M., March, W.F.: Appl. Spectrosc. 49(7), 987–993 (1995)

[4] Cote, G.L., Fox, M.D., Northrop, R.B.: IEEE Trans. Bio-medical Engrg. 39(7), 752–756 (1992)

[5] Allen, J.: Photoplethysmography and its application in clinical physiological measurement. Physiol. Meas. 28, R1 (2007), doi:10.1088/0967-3334/28/3/R01

[6] Malin, S.F., Ruchiti, T.L., Blank, T.B., Thennadil, S.U., Monfre, S.L.: Noninvasive prediction of glucose by near-infrared diffuse reflectancespectroscopy. Clinical Chemistry 45, 1651–1658 (1999); Khalil, O.S.: Spectroscopic and clinical aspects of noninvasive glucosemeasurements. Clinical Chemistry 45(2), 165–177 (1999)

[7] Jeon, K.J., Hwang, I.D., Hahn, S., Yoon, G.: Comparison betweentransmittance and reflectance measurements in glucose determinationusing near infrared spectroscopy. Journal of Biomedical Optics (11), 014022 (2006)

[8] Amerov, A.K., Chen, J., Small, G.W., Arnold, M.A.: Scatteringand absorption effects in the determination of glucose in whole blood bynear-infrared spectroscopy. Anal. Chem. (77), 4587–4594 (2005)

[9] Miller, C.E.: Chemical principles of near infra red spectroscopy. In: Williams, P. (ed.) Near Infra Red Technology in Agriculture and Food Industries, 2nd edn. The American Association of Cerel Chemists, St. Paul (2001)

[10] Yorozu, Y., Hirano, M., Oka, K., Tagawa, Y.: Electron spectroscopy studies on magneto-optical media and plastic substrate interface. IEEE Transl. J. Magn. Japan 2, 740–741 (1987) (Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982)

[11] Wheeler, O.: Near-IR spectra of organic compounds. Chem. Rev. 59, 629–666 (1959)

[12] Vasko, P.D., Blackwell, J., Koenig, J.L.: Infrared and Raman spectros- copy of carbohydrates. I. Identification of OOH and COH vibra- tional modes for D-glucose, maltose, cellobiose, and dextran by deuterium substitution methods. Carbohydr. Res. 19, 297–310 (1971)

[13] Vasko, P.D., Blackwell, J., Koenig, J.: Infrared and Raman spectros- copy of carbohydrates. II. Normal coordinate analysis of a -D- glucose. Carbohydr. Res. 23, 407–417 (1972)

[14] Cael, J.J., Koenig, J.L., Blackwell, J.: Infrared and Raman spectroscopy of carbohydrates. IV. Identification of configuration and confor- mation sensitive modes for D-glucose by normal coordinate analysis. Carbohydr. Res. 32, 79–91 (1974)

[15] Low, M.J.D., Yang, R.T.: The measurement of infrared spectra of aqueous solutions using Fourier transform spectroscopy. Spectrochem. Acta 20A, 1761–1772 (1973)

[16] Hazen, K.H., Arnold, M.A., Small, G.W.: Measurement of Glucose in Water with First-Overtone Near-Infrared Spectra. Applied Spectroscopy

[17] Amerov, A.K., Chen, J., Arnold, A.M.A.: Molar Absorptivities of Glucose and Other Biological Molecules in Aqueous Solutions over the First Overtone and Combination Regions of the Near-InfraredSpectrum. Applied Spectroscopy

[18] Pavlyuchko, A.I., Vasilyev, E.V., Gribovb, L.A.: Calculations of molecular ir spectra in the overtone and combination frequency regions

[19] Yamakoshi, K., Ogawa, M., Yamakoshi, T., Tamura, T., Yamakoshi, K.: Multivariate regression and discreminant calibration models for a novel optical non-invasive blood glucose measurement method named pulse glucometry. Conf. Proc. IEEE Eng. Med. Biol. Soc. (2009)

[20] Yamakoshi, Y.: Pulse Glucometry: A New Approach for Non-invasive Blood Glucose Measurement Using Instantaneous Differential Near InfraredSpectrophotometry

[21] Bhandare, P., Mendelson, Y., Peura, R.A.: Multivariate Determination of Glucose in Whole Blood Using Partial Least-Squares and Artificial Neural Networks Based on Mid-Infrared Spectroscopy. J. Applied Spectroscopy 47, 1214–1221 (1993)

[22] Smits, J.R.M., Malssen, W.J., Buydens, L.M.C., Kateman, G.: Using Artificial Neural Net- works for Solving Chemical Problems. Part 1. Multi-layer Feed-forward Networks. J. Chemometrics and Intelligent Laboratory Systems 22, 165–189 (1994)

Optics and Radar Image Fusion Based on Image Quality Assessment

Ralf Reulke[1], Gianluca Giaquinto[2],
and Marcello Giovenco[2]

[1] Humboldt-Universität zu Berlin,
Department of Computer Science, Berlin, Germany
`reulke@informatik.hu-berlin.de`
[2] German Aerospace Center,
Institute of Sensor Systems, Berlin, Germany
{`Gianluca.Giaquinto`,`Marcello.Giovenco`}`@dlr.de`

1 Introduction

In many applications the image quality of remote sensing data is of fundamental importance and is usually assessed by visual analysis subjective to the interpreter. In many situations, a single sensor is not sufficient to provide an accurate perception of the real world. Data and image fusion is a common way to overcome this problem. The basic objective of data fusion is to derive a higher amount of information compared to that what would have from any individual input sensor. Therefore, multi-sensor data fusion has become an intense research area in recent years. The first part of this present work goes in this direction.

Image quality is a comprehensive view on a given data set from a certain application perspective. Several parameter of the imaging system, observation condition, etc. influences the image quality. For instance, the quality of the object detection depends strongly on the spatial resolution. Spatial resolution (for definitions of remote sensing terms see DIN 18716 [4]) is the ability of a sensor system to detect adjacent object structures separately and is not equal to ground sample distance (GSD). Resolution is also determined by a number of additional factors, e.g. the performance of the camera optics compared to pixel pitch, motion blur and the sensor noise. Image quality measures includes the determination of all these parameters, which are summarize in the Point Spread Function (PSF) or the Modulation Transfer Function (MTF) and the signal to noise ratio (SNR). But PSF and SNR are not always easy to evaluate in practice from real imagery. An alternative idea is to use particular spatial and radiometric targets to evaluate the edge response. Relative edge response (RER) is indeed easier to measure and is mathematically related to PSF and MTF.

The purpose of this study is to achieve a final number for image quality to assess and compare the quality of optical, radar and fused images. In terms of object detection or interpretability the NIIRS (National Imagery Interpretability Rating Scale) index (a number from 0 to 9) allows such a comparisons. In this

A. Mason et al. (eds.), *Sensing Technology: Current Status and Future Trends IV*,
Smart Sensors, Measurement and Instrumentation 12, DOI: 10.1007/978-3-319-12898-6_5

context NIIRS, GIQE (General Image-Quality Equation) and the calculation of its parameters will be performed.

Two relevant ways of carrying out the investigation should be put in evidence: Firstly find one or more measures for evaluating image fusion performance. On a quantitative side, this measure represents how much of the information in the final fused image was obtained from the input images. From a qualitative sense, it is interesting to evaluate the quality of fusion, whether it performed good or bad data integration, which is strongly algorithm-dependent. In other words, it is important to evaluate the goodness of fusion. The second aspect deals more with image quality and interpretability and the assessment of the image quality parameters. Furthermore this path will help face many other aspects in future projects regarding radar and optical data: When will it be possible to compare optical and radar data according to the same rating scale? What is a suitable rating scale for radar data? What is the information content conveyed from each of the pristine images into the fused image? Which image gives more contribution and under what conditions?

This paper is organized as follows. It starts with an overview of the background of this work. PSF, SNR and the influence on NIIRS will be explained. It follows a section about fusion. The third chapter describes data sets and methodology for this study. Results are presented in the following chapter. Finally some conclusions are presented.

2 Background

2.1 Image Quality and Interpretability Parameters

Photogrammetry and remote sensing techniques provide methods to obtain and process geometric and thematic information from image data. For capturing image data, a variety of aircraft and space-borne sensors are available. To ensure comparability of the sensors and image products, descriptive variables or metrics for image quality determination are necessary. The determination of the quality of remote sensing data can mostly distinguish in (spectral) radiometric and (spatial) geometric aspects. For the description, different metrics can be found relating to radiometric and geometric accuracy aspects (conversion of radiances into digital numbers (DN), geometric point accuracy, etc.) but also performance parameters like SNR and MTF). An overview about image quality and relevant metrics can be found in the paper from Hermiston [6] and Reulke [15]. In this presentation we focuses on NIIRS metric. The most important influences on NIIRS are GSD, PSF and SNR, but also image preprocessing aspects (e.g. image restoration). In the following sections, PSF, SNR and NIIRS are described more in detail.

2.2 PSF and SNR

The PSF can be calculated through the input signal $U(x, y)$ and the measurement of $V(x, y)$. If we assume a translation invariant PSF $H(x, y)$ the following equation is valid:

$$V(x, y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} dx' dy' \, H(x - x', y - y') \cdot U(x', y') \tag{1}$$

Signal-to-noise ratio compares the level of a desired signal to the level of background noise. It is defined as the ratio of signal power to the noise power. Noise consists of the Poisson distributed photon and dark noise. The third component is the Gauss distributed read noise. Read noise is defined as the temporal system noise of the detector in darkness. As an example the paper from Lukin [11] is mentioned here, which proposed a blind evaluation of noise type.

2.3 NIIRS

The National Imagery Interpretability Rating Scale (NIIRS) is a subjective scale used for rating the quality of imagery acquired from various types of imaging systems. It is related to object detection; in particular the scale is a 10-level scale, 0-to-9, defined by interpretation tasks or criteria. GIQE [10] is a heuristic approach for quantifying NIIRS and was developed to provide such predictions. The derivation of the fundamental parameters (RER and SNR) is explained in [15]. Gutchess [5] present an image quality metric and prediction model for SAR imagery. They derived two separate equations for a representation of GIQE, one to predict radar NIIRS and one to predict Automated Target Detection (ATD) performance.

2.4 Fusion

There are different applications, which use fusion of optic and radar data. Theil et al [17] describes fusion of radar and optic data for surveillance on land. When compared to camera only, a significant reduction of the number of false tracks is achieved. Tzeng [18] introduce multiple-classifiers approach for terrain cover classification. They take the complementary and supplementary information from different data sources into account. A new approach for classification of different areas in optical and SAR images was proposed by Kenduiywo [9]. Here conditional random fields for detection and support vector machine for classification are used. In Fan [20] a classifier was presented, which is based on integration of optical and radar data and the Maximum Likelihood classification for determination of soil moisture.

The proposed work was conceived as a theoretic study about image quality and radar-optical image fusion; hence no constraints on the chosen fusion algorithm were present in the proposed analysis. Indeed the choice of fusion algorithm is strongly dependent on the application or the objects observed. An overview of fusion algorithms can be found in Pohl [13]. Schultz [16] studied two different methods for fusing data from optical and radar sensors. They investigate fusion prior to feature extraction but also in a more traditional way, after feature extraction. In [8] image fusion method of SAR, Panchromatic (Pan) and multispectral (MS) data is described. Low pass SAR texture and high pass details extracted from the available Pan image is used for this approach. High-frequency information fusion is also described by Wang [19] in order to improve the resolution of the radar image and reserve more radar information.

Before fusion, basic pre-processing steps involve correction of the remote sensing images for system errors, and radiometric processing (speckle reduction in SAR images which can be performed at various stages). Optical imagery is influenced by the atmosphere during data acquisition and therefore needs correction and/or other radiometric enhancements. Following the radiometric processing the data are geometrically corrected (geocoding, co-registration). This is a major element of pixel level image fusion since these fusion techniques are very sensitive to bad registration. In Charlwood [2] a sensor alignment approach for optical and radar fusion systems is described. Key result of the experiment was, that spatial and time registration needs to be as accurately as possible. Jia [7] proposed an automatic registration algorithm based on different features. Here patches, edge and point features are used. Palubinskas and Reinartz [12] analyze disturbing effects in fused images, taken with sensors having different imaging geometries. For orthoimages they proposed an acquisition geometry which allows to minimize or compensate ground displacements in different orthoimages.

Further improvement can be achieved with model based methods based on high resolution DEM. With this approach the detection of Radar shadows is possible (see [14]). In this paper SAR data are used for target recognition. Anglberger [1] implemented an approach to transform optical images into the radar image space. Based on a digital surface model an optical image can project to the same slant-range image plane as the one from the radar image acquisition.

In some cases, especially if image data of very different spatial resolution are involved, the resampling of the low resolution data to the pixel size of the high resolution image might causes a blocky appearance of the data. Therefore a smoothing filter can be applied before actually fusing the images [3]. The resulting image map can be further evaluated and interpreted related to the desired application. In general, the techniques can be grouped into two classes:

- Color related techniques
- Statistical/numerical methods

The first comprises the color composition of three image channels in the RGB color space as well as more sophisticated color transformations, e.g., IHS and HSV (PAN-sharpening). Statistical approaches are developed on the basis of channel statistics including correlation and filters. Techniques like PCA and regression belong to this group. The numerical methods follow arithmetic operations such as image differencing and ratios but also adding of a channel to other image bands. A sophisticated numerical approach uses wavelets in a multiresolution environment. Some techniques only allow a limited number of input bands to be fused (e.g., RGB, IHS), whilst others can be performed with any number of selected input bands. One method that relies on statistics in order to select the data containing most of the variance is the selection method developed in [4] called Optimum Index Factor (OIF). As a case of study, the Principal Component Analysis (PCA) has been adopted. Nevertheless, there have been good reasons for supporting this preference. For instance radiometric pre-processing plays an important role in relation to the spectral content of the fused image. The appearance of SAR significantly influences the feature visibility in the fused optical image. As a consequence, features that are detectable on SAR data can be introduced to the optical data by image fusion to complement the data (e.g., soil moisture, urban area, oceanographic objects). The PCA technique enables the integration of more than three types of data [13]. The main goal of this work is to evaluate the quality and the interpretability of fused images and separately of optical and radar images and to compare each other. We will introduce the data set at first and then go through the methodology. The idea was to process the pristine images before fusing them in two different ways: on one hand at lower optical resolutions levels, and on the other hand at increasing multi-look levels of the radar images. This process led to disclose some break-even points of comparison between single and fused images, which are shown in the results and evaluated in the conclusion.

3 Data Set and Methodology

The selected data set is an area in the city of Dresden in Germany. The choice of this area has been driven by the need to take into account a great variety of classes that a real scene can contain, possibly with the largest range available. Hence, one can find the river Elbe and so water, buildings, roofs, jetties, fields, as well as vegetation and trees. As a matter of fact the scene is very complex and comprehensive of many classes. The data we used are:

- one radar image acquired by the Terra SAR-X satellite in Strip-Map mode on 02-07-2008 whose resolution is 2.75 meters
- four channel (blue, green, red and NIR) optical images acquired by the IKONOS satellite on 08-07-2007 whose resolution is 1 meter

In Figure 1 and Figure 2 the radar image and the blue channel of the optical image are shown. In these images are clearly visible the river area as well as the small vegetation zone on the top and the housing at the bottom part of it. Starting from these five images every other computation derived. Images were co-registered to one another through the use of ten Ground Control Points (GCPs) selected in the Image-to-Image mode. The GCPs were chosen in both open spaced and urban areas such as fields, buildings corners, streets and so on, using NNR (Nearest Neighbor Resampling) method to determine the digital pixel values of the corrected output image.

Fig. 1. Radar image from Terra SAR-X sensor of the selected area

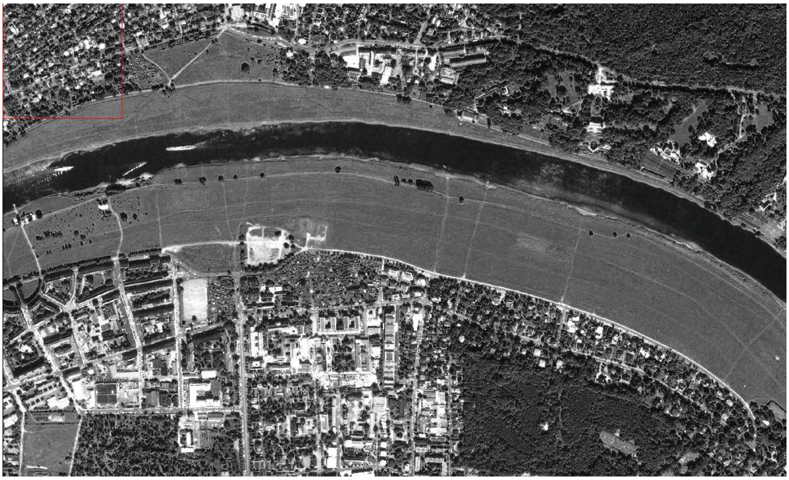


Fig. 2. Blue channel from IKONOS sensor of the selected area

An improvement can be expected, if bilinear or bicubic interpolation will be applied.

3.1 Processing Flows

Looking at the pristine images (optical and radar) separately, the idea was to work accordly with two processing flows, whose the main difference lies in the pre-processing level before fusion would be performed, changing the two resolutions: geometric resolution and radiometric resolution. As a matter of fact, they have indeed dominance within the GIQE terms.

- The first attempt (see Figure 3) was to work with lower spatial resolution, degrading the optical images by resampling and each time fusing them with the radar image at coarse resolution. Because this method works on a re-sampling of the original image, i.e. moving from 1 m resolution to lower resolutions, the first step has been the smoothing with a low-pass Gaussian filter of the optical images, as an anti-aliasing filter. Subsequently the images have been processed to 2, 5 and 10 meters resolution.
- The second workflow (Figure 4) was carried out, by fusing the high-resolution optical image with multi-look radar images.

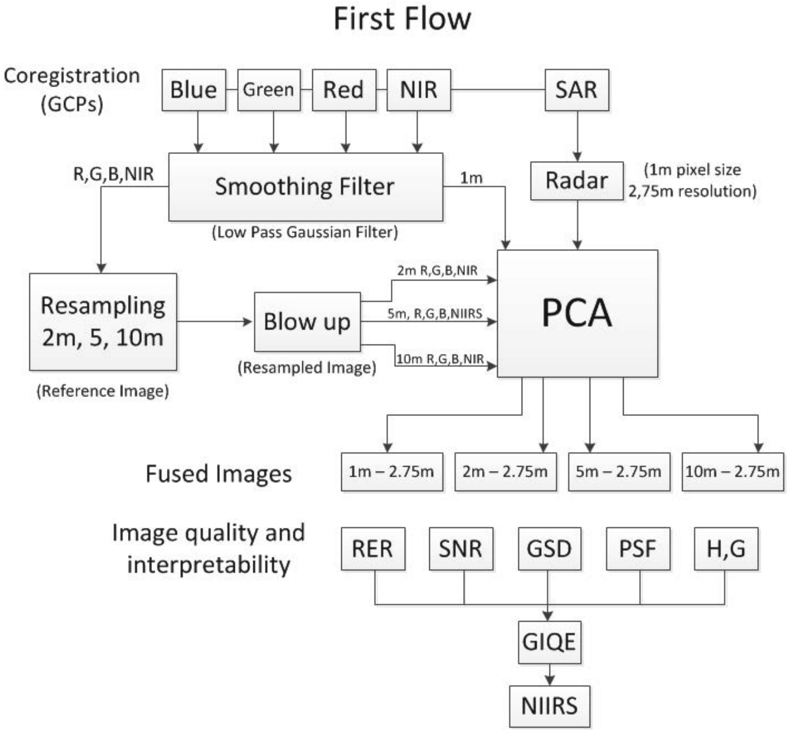


Fig. 3. First processing scheme

After the pre-processing part in both cases the fusion is applied. In Figure 5, 6 and Figure 7 the results of the fusion process are shown.

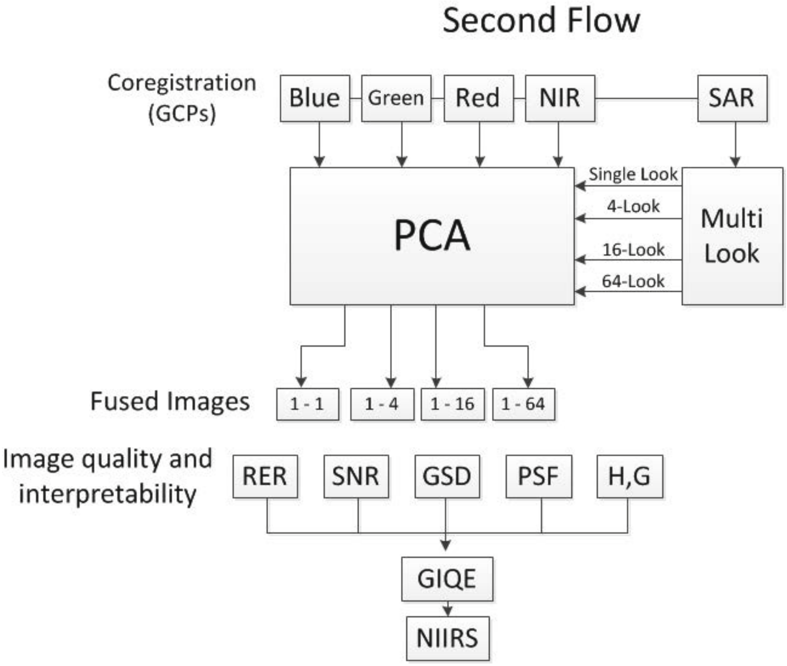


Fig. 4. Second processing scheme

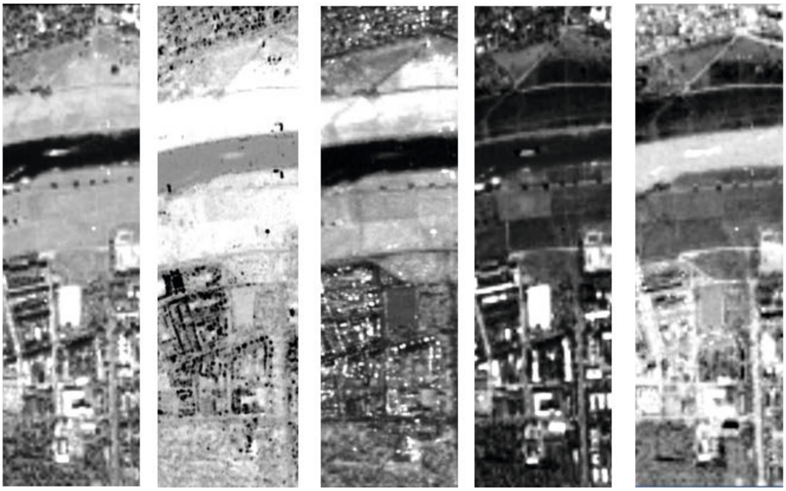


Fig. 5. PCA Radar/Optical (1 m/1 m). PCA from 1 to 5 (left to right)

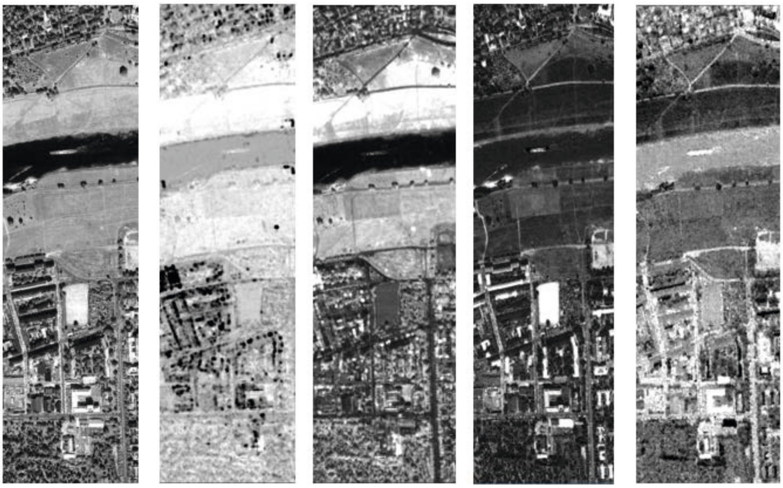


Fig. 6. PCA Radar/Optical (1 m/1 m). PC 1 to 5 (from left to right)

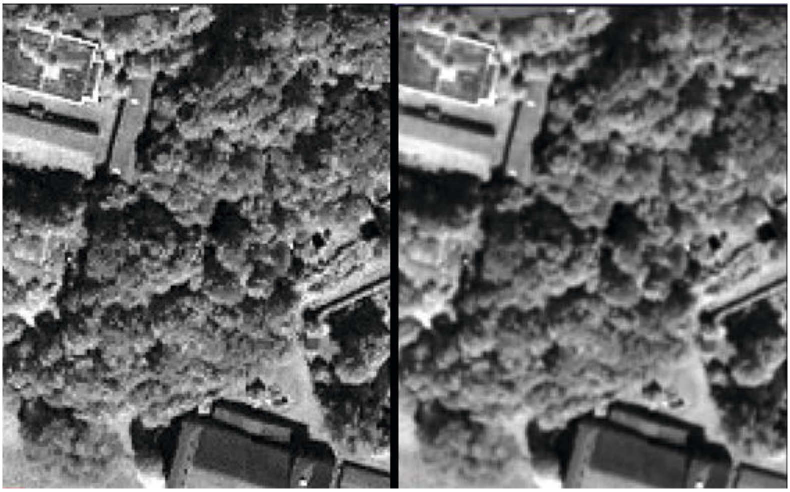


Fig. 7. Zoom Comparison: First PCs of the fused image optical/single-look (left) and optical/64-look (right)

3.2 NIIRS Calculation

Using the version 4.0 of GIQE (see Table 1), the NIIRS value was estimated:

$$NIIRS = c_0 + c_1 \cdot \log_{10}(GSD_{GM}) + c_2 \cdot \log_{10}(RER_{GM}) + c_3 \cdot \frac{G}{SNR} + c_4 \cdot H \quad (2)$$

Table 1. Versions 3.0 and 4.0 of GIQE and coefficients

	c0	c1	c2	c4	c5
GIQE 3.0	11.81	-3.32	3.32	-1	-1.48
GIQE 4.0 (for RER > 0.9)	10.251	-3.22	1.559	-0.334	-0.656
GIQE 4.0 (for RER < 0.9)	10.251	-3.16	2.817	-0.334	-0.656

4 Results and Discussion

In this section we will show the graphs regarding the NIIRS value for the two processing flows. The first graph we evaluated is shown in Figure 8. Here the comparison between radar and optical images prior to fusion is performed. The NIIRS is computed for the four channels of the optical image at each one of the resampling steps. Radar image NIIRS is also computed once in order to find a break-even point between the two curves. This is reached for a value of NIIRS equal to 2.2, which correspond to an optical resolution of 4.5 meters equal to a radar resolution of 2.75 meters.

Because there is no MTF improvement, neither noise gain (G=1), nor mean-height overshoot will be taken into account.

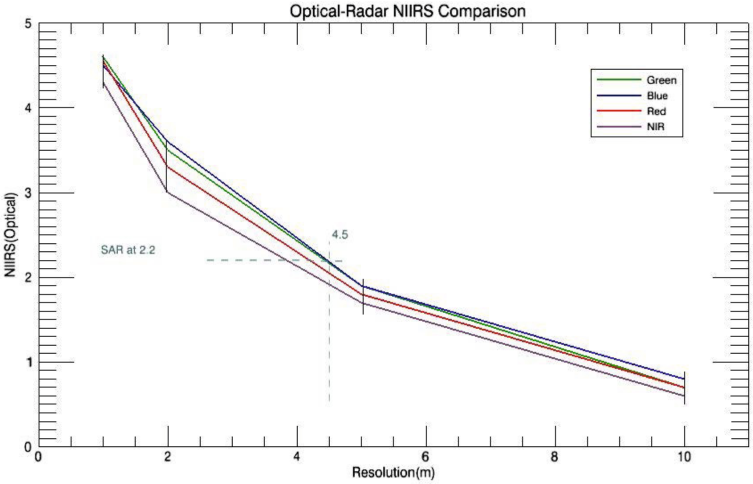


Fig. 8. Optical-Radar NIIRS comparison: the break-even point of radar and optical NIIRS is reached for a value of NIIRS equal to 2.2 corresponding to an optical resolution of 4.5 meters which is compared with the radar resolution set to 2.75 meters.

In other words to reach the same interpretability of a 2.75 meters geometric resolution radar image, an optical image of 1 meter geometric resolution should be brought to 4.5 meters resolution. This is reasonable if considering that the NIIRS value for optical images sensibly decreases with the resolution as we can also notice in Table 2. Since NIIRS is, in fact, a parameter for human interpretability of satellite imagery it is expectable that degrading the ground sample distance (e.g. becoming harder for a human eye to distinguish two separate points in the image) NIIRS values falls according to the entity of the degradation.

Table 2. NIIRS values for optical channel at different GSD

GSD [m]	1	2	5	10
Green	4.6	3.5	1.9	0.7
Blue	4.5	3.6	1.9	0.8
Red	4.5	3.3	1.8	0.7
NIR	4.3	3.0	1.7	0.6

In the following the results of fusing radar and optical images with PCA will be investigated. For the first processing flow, there is on one side the PCA obtained from the 1 meter geometric resolution optical image and the 2.75 meters geometric resolution radar image (keeping the pixel size always set to 1.0 for both images) and on the other side the PCA obtained by degrading the optical geometric resolution to 10 meters and keeping the radar resolution constant.

Table 3. Niirs value for fused images 1m optical bands/sar single-look and 10m optical bands/sar single-look

Bands	1	2	3	4	5
1m Opt - 1m Radar	4.5	3.0	4.1	3.8	3.3
10m Opt - 1m Radar	1.1	0.8	0.7	0.6	0.5

When computing an average of the four optical channels (1 meter resolution) NIIRS previously presented, the resulting NIIRS is equal to 4.4 which is slightly less than for the first principal component of PCA 1-1 which is equal to 4.5. That means that a small improvement from fusion is present in the first processing flow. From the analysis of RER for the first processing flow it results that the RER decreases with resolution being a measure of ability of the system to highlight sharpness. This would turn in a NIIRS decreasing trend according

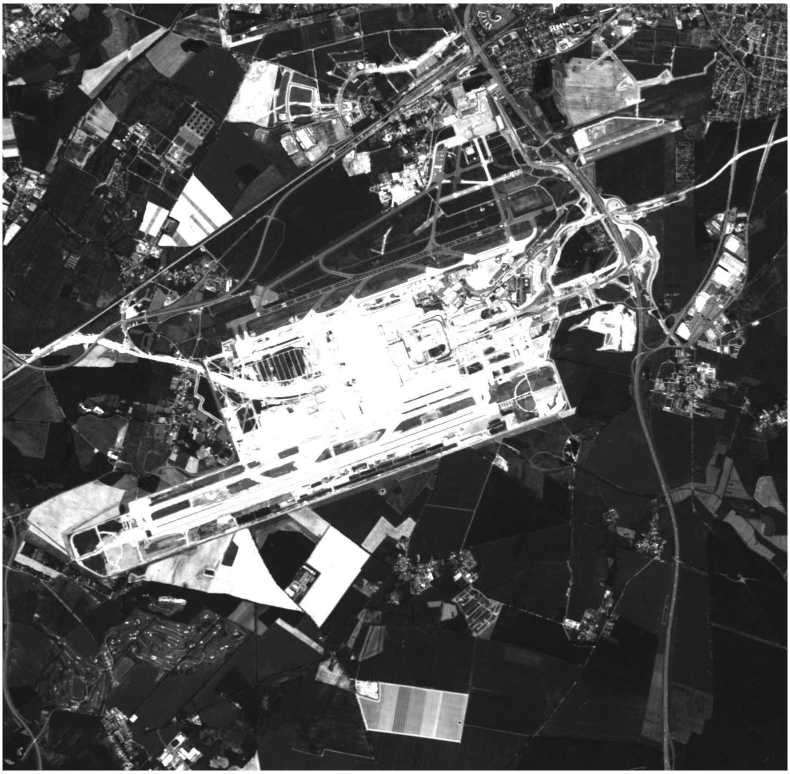


Fig. 9. Berlin Brandenburg Airport: RapidEye red channel

to the RERs one. One can notice, that NIIRS does not decrease as principal components decrease. That would be in contradiction with the definition of PCA according to which the first component would contain most of the information (for instance most of the variance) and the following components little by little would contain less. The problem arises when the complexity of the scene considered is high because of the presence of many small objects in it. These small objects have, in fact, a small variance that would end up in the last principal components (PCs) rather than in the first one causing the last PCs to also contain useful information and the first PC to contain noise. That explains why the second PC behaves quite badly in our case. Another consideration that has to be made is about the linearity of the PCAs trend in the second graph due mostly to the smaller value assumed by the second principal component in respect to the values assumed previously. That can be justified because degrading the geometric resolution of optical data from 1 to 10 meters can surely cause a flattening of the objects in the image and therefore the optical and the radar images are more comparable than before, hence the curve has a more linear trend. Same analysis was done for a set of images of Berlin Brandenburg International Airport in the southeast area of the German capital. The optical satellite involved is

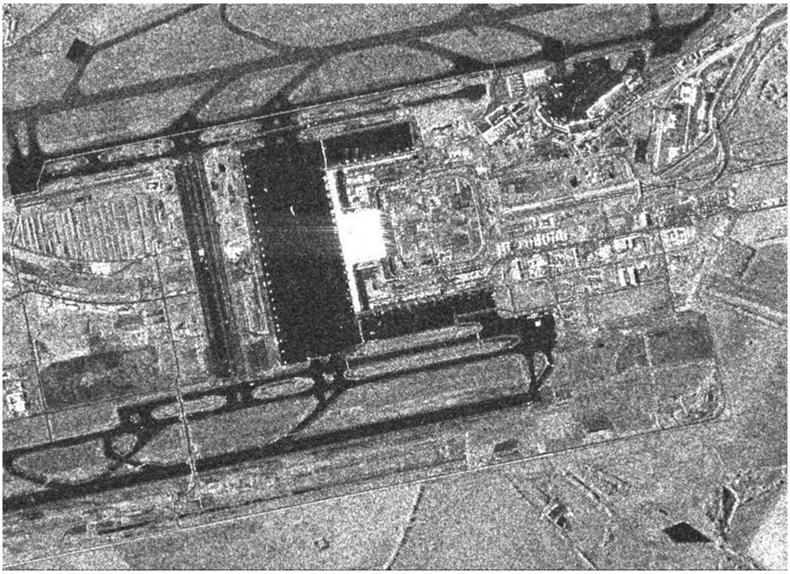


Fig. 10. Berlin Brandenburg Airport: Terra SAR-X image

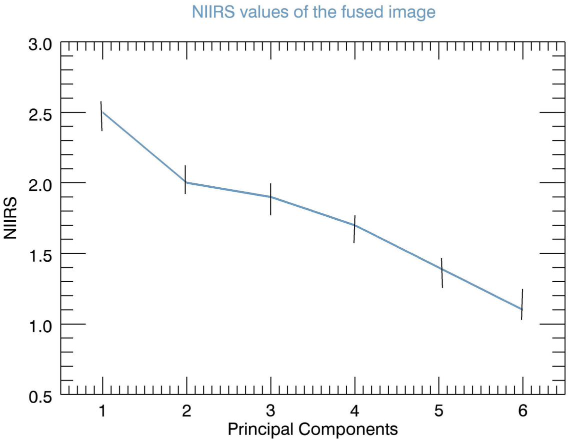


Fig. 11. NIIRS Level in Berlin Brandenburg Airport

RapidEye, a constellation of five satellites collecting 5 meters resolution images in 5 bands, blue, red, green, red-edge and near infrared. The images were captured on 24th September 2010. The radar satellite is Terra SAR-X, which provides 1-meter resolution images. Date of capture was 18th September 2010. Figure 9 and Figure 10 shows two images: red channel and radar.

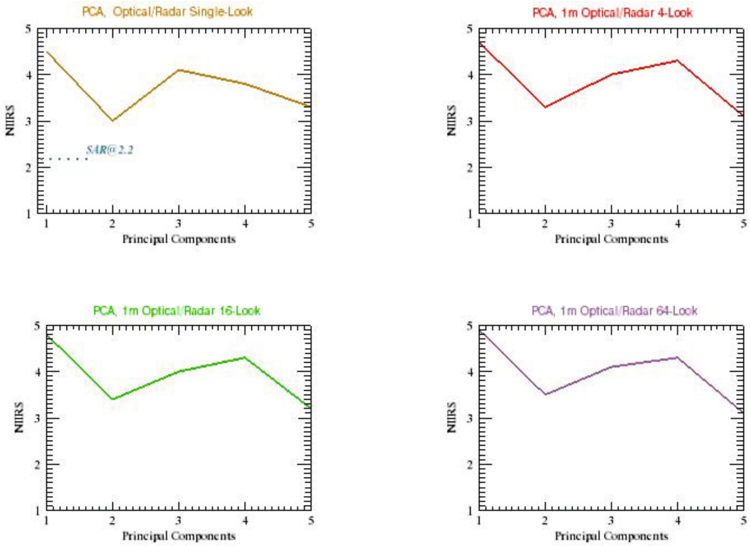


Fig. 12. NIIRS Comparison for PCA bands in the second processing flow

In this set of images the contribution of information is more comparable and thus the image quality is closer, although not similar (radar image has a NIIRS level of 1.1 and optical bands ca. 2.3). But the resulting NIIRS is slightly higher, as seen in Figure 11.

Now the NIIRS for the second processing flow shall be analyzed. As it is shown in Figure 12 and Table 4 NIIRS values increase as the number of multi-looks get higher. This means that the influence of the radar image in the PCA is visible even if its amount is not very comparable to the optical image, as it was expected.

Table 4. NIIRS values for the 5 PCA bands in the coarse 1 meter GSD optical image and the multi-looked radar images

Bands	1	2	3	4	5
1m-1 Look	4.5	3.0	4.1	3.8	3.3
1m-4 Look	4.7	3.3	4.0	4.3	3.1
1m-16 Look	4.8	3.4	4.0	4.3	3.2
1m-64 Look	4.9	3.5	4.1	4.3	3.1

The last graph we want to propose concerns the NIIRS values of the first principal components of the four multi-looked images analyzed and the maximum optical NIIRS value. As already mentioned, the first principal components are meant to contain the most information so it can be useful to observe how they behave in respect to the optical data alone. It is shown that optical NIIRS value is between the first component values with 4-Looks and 16-Looks radar image as in Figure 13. After that value, image fusion takes a slight advantage in respect to the optical image alone and it makes sense to fuse optical and N-multi-looked images.

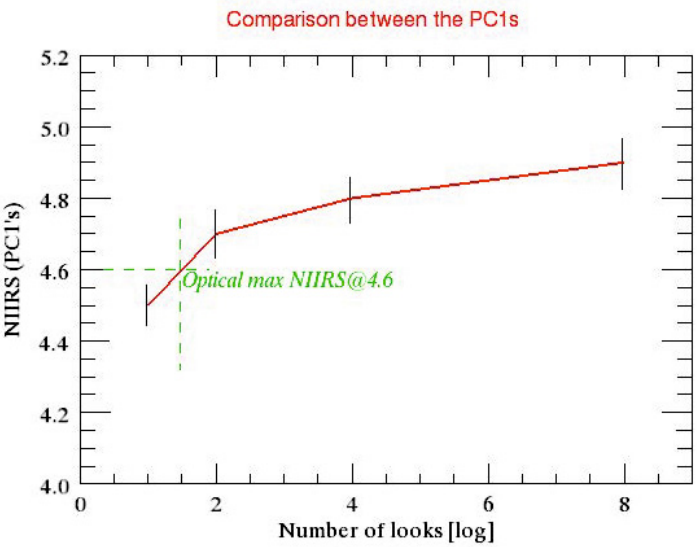


Fig. 13. Comparison between NIIRS values of the first principal components and the optical one

5 Conclusion and Outlook

The goal of image fusion techniques is to combine and preserve all of the important visual information present in multiple input images in a single output image. In many applications, the quality of the fused images is of fundamental importance. This contribution was carried out within multi-sensor image fusion, from optical and radar imagery. The main purpose was to assess the fused image quality and compare it with the pristine images in terms of visual interpretability. Relative edge response (RER) from natural edges and NIIRS value supported this measurement. Two evaluation procedures have been proposed, according to different pre-processing steps, separately on optical and radar images. The first method, by merging radar with progressive lower optical resolution, shows that optical images have to be decreased up to roughly 4 meters resolution to achieve the radar RER value. RER is a dominant term in the NIIRS equation, together with GSD. NIIRS values of fused images results show that RER confers the trend, while GSD the range of values. Fusing optical and radar images after PCA leads to a slight improvement in NIIRS on the first principal components, where most of information is conveyed. According to the first procedure optical and radar data are related in terms of NIIRS at roughly 4.5 times of the optical resolution, i.e. to achieve the radar interpretability the optical resolution has to lowered from 1 to 4.5 meters. This value can be considered as a possible interpretability-threshold of fusion. When optical resolution in the fused image is better than the threshold resolution, the trend of NIIRS after PCA fusion is unclear and objects cannot be well compacted into major PCs. In fact the radar image is still too noisy in terms of spatial resolution to be related with the same scale of contribution of information in the fusion. It ends up in the presence of a bad segregation of the noise in the last principal components that in our case showed up in the second principal component. Conversely, when the resolution exceeds the threshold value even with a lower range determined by a worse GSD, the trend is clear and regular decreasing with lower principal components. In the second procedure, the optical bands were kept at higher resolution, while the radar image was processed with progressive multi-looks. Results show that improving the radiometric resolution of the radar images performing multi-looks the NIIRS value had a small improvement of a 0.2-NIIRS. Increasing the number of multi-looks in a bad radar image involved in a fusion process, leads to a worse geometric resolution but removing the speckle noise to a better radiometric resolution. Results shows that this helps the interpretability of areas with a strong bright-dark transition. Indeed the edge response in this case ends up in being less sharp, but more regular and with lower variance of the pixels value in the edge transition. As a matter of fact, also the RER improves. Despite the bad resolution of the radar image, through the procedures proposed, the information originating from the input images was conveyed into the fused images, increasing the quality in terms of interpretability. Moreover, thanks to comparisons presented, an interpretability-threshold was found. It represents the value from which optical and Radar images can be compared in terms of interpretability and in order to perform a good fusion. New investigation using other fusion techniques can be assessed (Artificial Neural Network, Dempster-Shafer Theory, etc.). Noise-Adjustment Principal Component for instance, arranges the data according to SNR and not on the basis of variance, performing better data reduction and fusion when the input images are noisy and complex. Moreover, measurements of goodness of fusion and quantitative analysis of the contribution of each single pristine image in fusion can be evaluated. Quality metrics might be based on information theory. Information Theory based image fusion quality metrics can be taken into account with the investigation of subjective and objective metrics. The latter can be useful when no reference image is available and for real time applications, edge detection and change detection. As for the RER from natural edges, it is feasible but more tar-gets are required. The future work is based on optical RapidEye and high-resolution radar TerraSar-X imagery. In this sense the research team has obtained data sets of the new Berlin-Brandenburg airport at different stages of time. Here it might be possible to fuse optical and radar imagery to get products with higher information content. Indeed the spatial resolution of RapidEye and the high resolution of TerraSar-X by hypothesis will help the images to fuse more comparably. Hence, a further task is to find out what is meant by comparable, which term or parameter that will tell in advance that two or more images are comparable enough to be suitable for a good fusion. Based on this setup urban change detection and object detection will be studied. As for image quality, information theory can also be thought as an alternative quality metrics to GIQE and NIIRS because is direct, with no use of image performance middle-steps parameters (MTF, SNR).

References

1. Anglberger, H., Speck, R., Suess, H.: Transforming optical image data into a sar system's range-based image space, vol. 8714, pp. 871411–871411–8 (2013), doi:10.1117/12.2018395
2. Charlwood, E.C.E., Griffiths, R.J., Buttinger, M.R.: Sensor alignment in electro-optical/radar fusion systems. In: Acquisition, Tracking, and Pointing Xiv, vol. 4025, pp. 42–51 (2000)
3. Chavez, P.: Digital merging of landsat tm and digitized nhap data for 1: 24,000-scale image mapping (national high altitude program). Photogrammetric Engineering and Remote Sensing, 52 (1986)
4. DIN. Photogrammetry and remote sensing - terms and definitions (2012)
5. Gutchess, D., Irvine, J.M., Young, M., Snorrason, M.S.: Predicting the effectiveness of sar imagery for target detection, vol. 8051, p. 805110. SPIE (2011)
6. Hermiston, K.J., Booth, D.M.: Niirs and objective image quality measures. In: Solina, F., Leonardis, A. (eds.) CAIP 1999. LNCS, vol. 1689, pp. 385–394. Springer, Heidelberg (1999)
7. Jia, W.J., Zhang, J.X., Yang, J.H.: Automatic registration of sar and optics image based on multi-features on suburban areas. In: 2009 Joint Urban Remote Sensing Event, vol. 1-3, pp. 840–846 (2009)
8. Jia, Y., Blum, R.S., Li, F.: Fusion method of sar and optical images for urban object extraction, vol. 6790, p. 67900M–67906M (2007)
9. Kenduiywo, B.K., Tolpekin, V.A., Stein, A.: Detection of built-up area in optical and synthetic aperture radar images using conditional random fields. Journal of Applied Remote Sensing 8(1), 083672 (2014), doi:10.1117/1.JRS.8.083672
10. Leachtenauer, J.C., Malila, W., Irvine, J., Colburn, L., Salvaggio, N.: General image-quality equation: Giqe. Applied Optics 36(32), 8322–8328 (1997)
11. Lukin, V.V., Abramov, S.K., Ponomarenko, N.N., Uss, M.L., Zriakhov, M., Vozel, B., Chehdi, K., Astola, J.T.: Methods and automatic procedures for processing images based on blind evaluation of noise type and characteristics, vol. 5. SPIE (2011)
12. Palubinskas, G.: Peter Reinartz. Fusion of optical and radar remote sensing data: Munich city example. In: Wagner, W., Szkely, B. (eds.) ISPRS TC VII Symposium 100 Years ISPRS, vol. XXXVIII, Part 7A. IAPRS (2010)
13. Pohl, C., Van Genderen, J.L.: Review article multisensor image fusion in remote sensing: Concepts, methods and applications. International Journal of Remote Sensing 19(5), 823–854 (1998)
14. Surya Prasath, V.B.: Oussama Haddad. Radar shadow detection in synthetic aperture radar images using digital elevation model and projections. Journal of Applied Remote Sensing 8(1), 083628 (2014), doi:10.1117/1.JRS.8.083628
15. Reulke, R., Eckardt, A.: Image quality and image resolution. In: 2013 Seventh International Conference on Sensing Technology (ICST), pp. 682–685 (2013)

16. Schultz, J., Gustafsson, U., Crona, T.: Sensor data fusion of optical and active radar data, vol. 5429, pp. 490–500 (2004)
17. Theil, A., Kester, L., van den Broek, S.P., van Dorp, P., van Sweeden, R.: Fresnel program: fusion of radar and electro-optical signals for surveillance on land, vol. 4380, pp. 453–461
18. Tzeng, Y.-C., Chen, K.-S.: Image fusion of synthetic aperture radar and optical data for terrain classification with a variance reduction technique. Optical Engineering 44(10), 106202–106202–8 (2005)
19. Wang, Y., Qin, S.: Adaptive high-frequency information fusion algorithm of radar and optical images, vol. 8006, pp. 80060Z–80060Z–5 (2011), doi:10.1117/12.901986
20. Yu, F., Zhao, Y., Liu, Q.: A study on the extracting soil moisture information by fusion of optical and radar data, vol. 7498, pp. 749839–749839–8 (2009), doi:10.1117/12.832273

Development of Anal Position Detecting System for New-Toilet system

K. Tokoro*, K. Fujihira, and H. Kobayashi

Department of Mechanical Engineering, Graduate School of Engineering,
Tokyo University of Science, 6-3-1 Niijuku, Katsushika-ku, Tokyo, Japan
`tokoro@kobalab.com`

Abstract. Japan entered a super-aged society, and nursing care is a serious social problem. Especially, an excretion care causes physical and mental stress to both caregivers and recipients. In excretion care for bedridden elderly, diapers are used, however changing dispars and odor are problems. In this study, we have been developing an new nursing-care equipment for reducing the stress of the excretion care. The equipment consists of Support system which assist to transfer and New-Toilet system which assist of toilet. The New-Toilet system can capture feces and prevent odor dispersion by adhering tightly to buttocks. For attaching the device to buttocks, it is necessary to know the position of the anus. It is preferable that the device automatically moves to an anal position and adhere to buttocks without caregiver's help. Therefore we developed a sensing system of anal position using the infrared camera. This paper describes the position detecting system for the New-Toilet system and its estimation.

Keywords: Applications of Sensors, Infrared Temperature Sensors, Nursing Care, Toilet.

1 Introduction

Japan entered a super-aging society and the elderly population is increasing rapidly [1]. The population of the elderly people requiring care and support is also growing [2]. One of the problems facing in the nursing field is "excretion care." [3][4][5] In excretion care, not only caregivers but also recipients feel strains. Caregivers transfer recipients to the toilet and then have to take care of elimination. Whereas recipients have risk of fall while transferring and feel ashamed. Thus both sides feel physical and mental strain [6][7][8].

There are several assist devices and a toilet for an excretion care [9]. The assists device for transferring a user from the wheelchair to the toilet seat reduces the physical burden of a caregiver [10][11]. By moving the toilet-self autonomously, user does not have to move and then it becomes possible to avoid falling and incontinence [12][13][14]. In the case of a bedridden, disposable diaper are used mainly though,

---

* Corresponding author.

A. Mason et al. (eds.), *Sensing Technology: Current Status and Future Trends IV*,
Smart Sensors, Measurement and Instrumentation 12, DOI: 10.1007/978-3-319-12898-6_6

diaper exchange and odor are serious issues [15]. Then, the equipments which can be eliminated even lying on the bed without diaper have been developed [16][17][18][19]. Automatic disposal equipment of excretion automatically suctions, washing, and drying function are realized, though, leakage and skin rash were reported [19].

We have been developing a closely-contact-type toilet so that smell and excretion stay in the plastic bag mounted inside the toilet bowl [20]. We believe that this type of toilet is very useful for nursing care and using in the space. In order to realize the closely-contact-type toilet, it is necessary to find the position of anal and then the toilet bowl should fit to the anal automatically.

This paper describes an outline of the closely-contact-type toilet system and shows the anal position detecting system with its estimation.

2 Outline of a Closely-Contact-Type Toilet System

Configuration and mechanical structure of the closely-contact-type toilet system is shown in Fig. 1. It consists of a toilet seat which moves back and forth, a toilet bowl which collects excrement, and an infrared camera for detecting the anus. Aluminum angle pipe is used for the frame of the system.

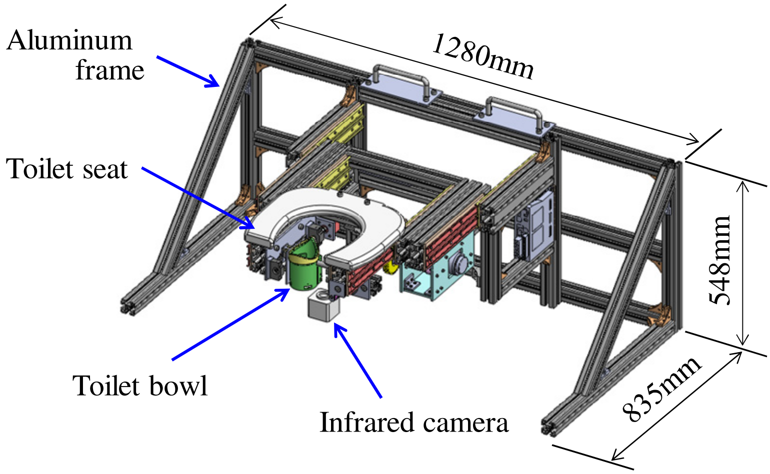


Fig. 1. Mechanical structure of the closely-contact-type toilet system

Fig. 2 describes the operation of the system. From the aesthetic point of view, the system is inserted in the wall. When using the toilet, the toilet seat moves forward and after user sits on it, the anal position is searched by using the infrared camera. The toilet bowl then moves to the anal position and contacts closely to the anus. The details of each structure are as follows.

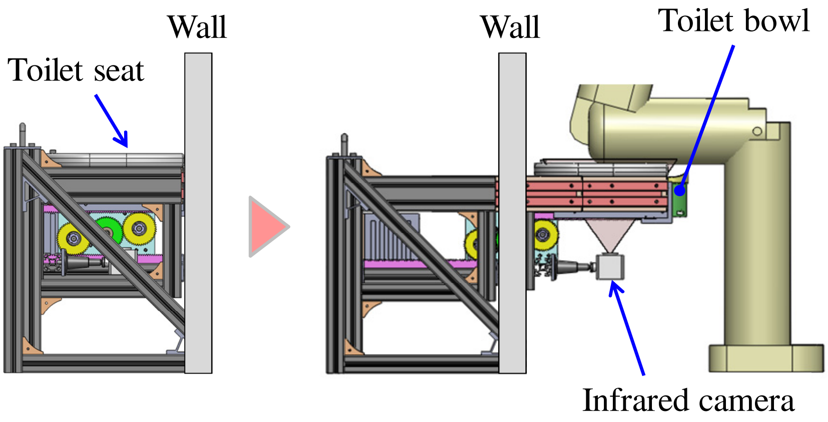
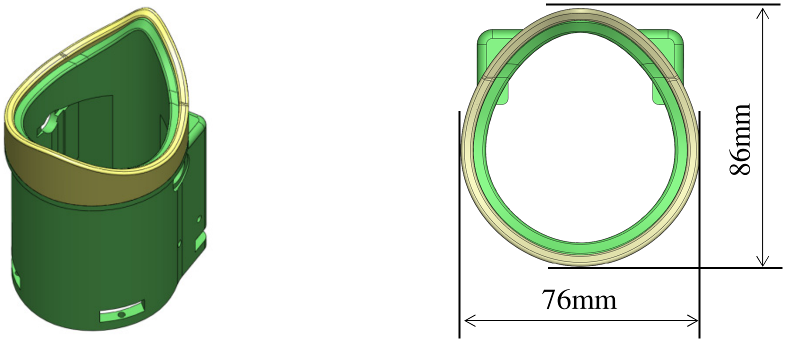


Fig. 2. Operation of the closely-contact-type toilet system

2.1 Toilet Bowl

The toilet bowl is a device which attaches closely to a buttock. By attaching to the buttock, the toilet bowl takes feces in the bag and prevents odor dispersion. The top portion of the toilet bowl is designed to conform to the shape of a buttock. The shape of the buttock is measured with a 3D scanner (Artec Group Inc., Artec MHT). The 3D scanner acquires three-dimensional shape data which can use in the 3D CAD. We made the shape of the toilet bowl based on the shape data of the buttock.

And, during a bowel movement, the rectum is pushed by abdominal pressure. Then the anal canal which is under the rectum is also pushed, so the shape of the buttock is changed [20]. With 3D printer, we make a comparison of shape of the buttock between before and during the bowel movement.



(a) External view (b) Top view

Fig. 3. Toilet bowl

2.2 Movement Mechanism of Toilet Bowl

The movement mechanism move the toilet bowl to the anal position where is detected by anal position detecting system. As shown in Fig. 4, the mechanism consists of stepping motor (MERCURY MOTOR, ST-42BYG020), ball screw, and migration stage. The coupling connects the stepping motor with the ball screw. The migration stage is moved by rotating the stepping motor and the ball screw. Lead of the ball screw is 4 mm and primary step angle of the stepping motor is 1.8 degree, so the migration stage can move at 0.02 mm intervals. Table 1 displays specification of stepping motor.

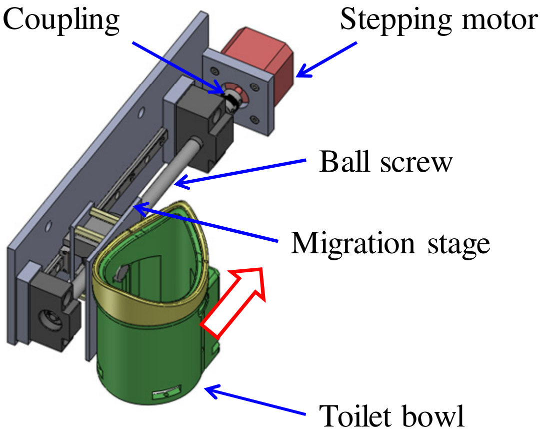


Fig. 4. Structure of the movement mechanism of toilet bowl

Table 1. Specification of stepping motor

Term	Specification
Step angle	1.8°±5%
Voltage	12V
Phases	4
Holding torque	1.6kgf.cm
Weight	0.21kg

2.3 Movement Mechanism of Toilet Seat

The movement mechanism of toilet seat consists of three sections; a toilet seat section, a motor box section, and a base section, as shown as Fig. 5. The toilet seat section has the toilet seat, the infrared camera, the toilet bowl and the movement mechanism of it (Fig. 5[A]). The motor box section is composed of an AC motor, gear wheels (Fig. 5[B]). The base section is made of aluminum frame (Fig. 5[C]). The mechanism uses AC motor, gear wheel and rack gear to move the toilet seat. The rack gear is attached to the toilet seat section and base section, and the gear wheel is attached to the motor box section. When the gear wheel is rotated by the AC motor, the motor box section and the toilet seat section is moved (Fig. 6).

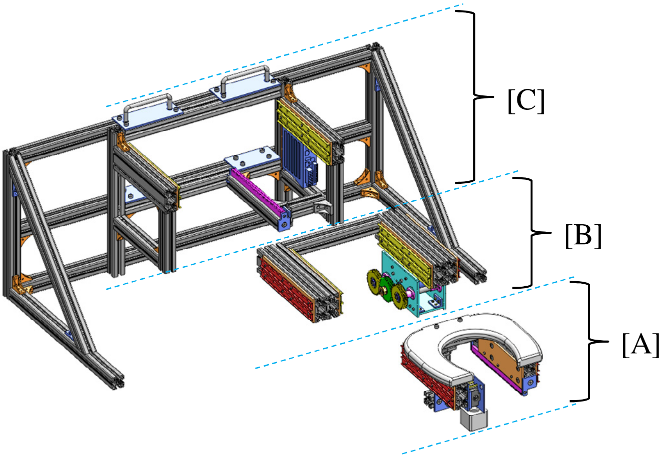
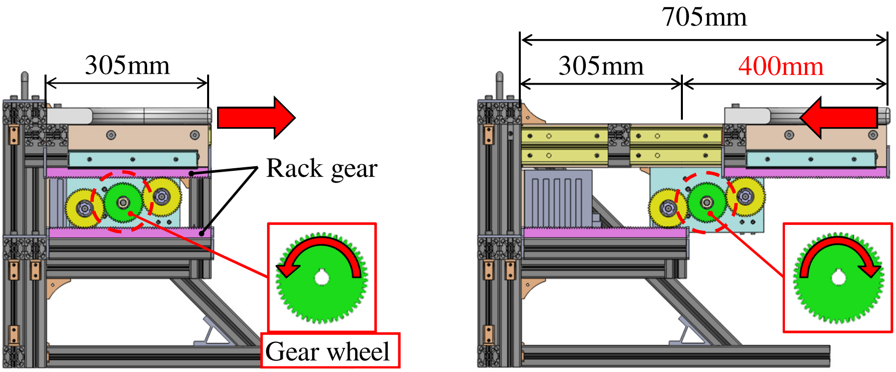


Fig. 5. Structure of the movement mechanism of toilet seat



(a)

(b)

Fig. 6. Operation of the movement mechanism of toilet seat

3 Anus Position Detection

Temperature of the internal human body, called “core temperature”, is higher than the temperature of a surface such as skin and stable. However, there are several points reflecting the core temperature exist. They are places to check the body temperature with a thermometer such as armpit, oral cavity, an ear or a rectum. Since the anus locates just after the rectum, we can expect that the temperature of the anus is the almost same as the internal body one and then higher than other places around the anus. In this system, therefore by using temperature differences, the position of the anus is extracted by using the infrared camera.

3.1 System Configuration

The system configuration is shown in Fig. 7. The infrared camera connects a PC with a USB cord, and data are sent or received through it. Infrared image and temperature date are sent to the PC from the infrared camera. The PC calculates an anal position based on the received data. The calculated result which is a moving distance of the migration stage and the toilet bowl is sent to the PIC via RS232 serial communication, the stepping motor is moved by microcomputer control. The toilet bowl is moved by rotating the stepping motor as described in 2.2.

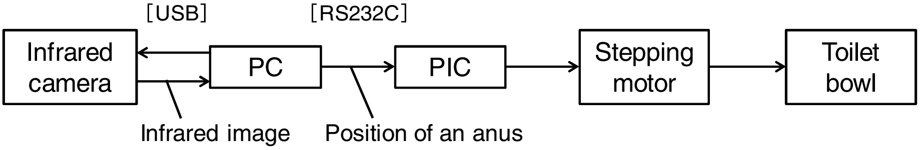


Fig. 7. System configuration

3.2 Infrared Camera

For detecting position of an anus, the infrared camera measures the temperature of whole buttock. TP-L0260UN small thermal imagery sensor (Chino Co., Tokyo, Japan) is utilized as an infrared camera. Fig. 8 describes the overview of the infrared camera and Table 2 shows the specification. The sensor is an infrared temperature sensor which has thermopile as detection element for infrared ray. It sends a measured temperature data to PC with USB connection. By dedicated software, the temperature data are created a picture on a monitor (Fig. 9). The resolution of the sensor is 47 × 48 = 2256 pixels, the view angle is 60° ×60°, and the space resolution is 21.8 mrad. When a measured distance to an object is 1.0 m, a measuring place is 1392mm×1363mm, and the width of view per pixel is 29mm as shown in Fig. 10. Equation (1) and (2) are equation relating a measuring place to measured distance.

$$R = 1.363\ D \tag{1}$$

$$W = \frac{R}{47} \tag{2}$$

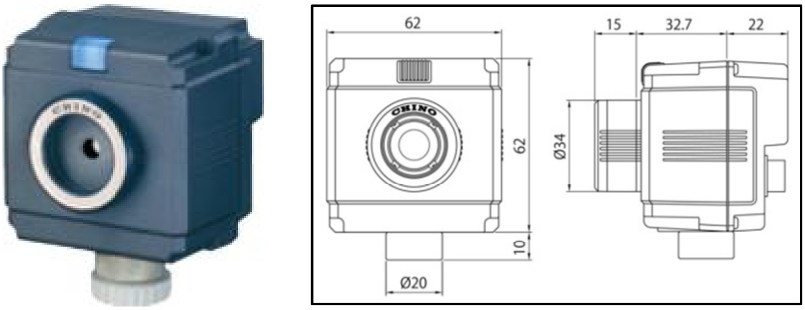


Fig. 8. Overview of the infrared camera

Table 2. Characteristics of the infrared camera

Term (and character)	Specification
Detection element	Thermopile element 47×48pixels
View angle	60°×60°
Filming range (R)	1.363m×1.392m in the case of 1m distance
Width of view per pixel (W)	29 mm in the case of 1m distance
Range of temperature	-20℃～300℃
Resolution	0.5℃ in the case of 100℃ black body

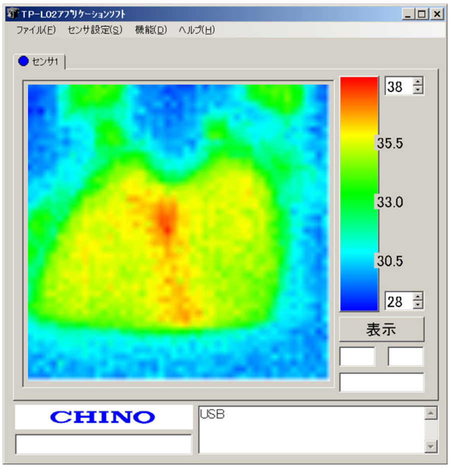


Fig. 9. Thermal image on the monitor

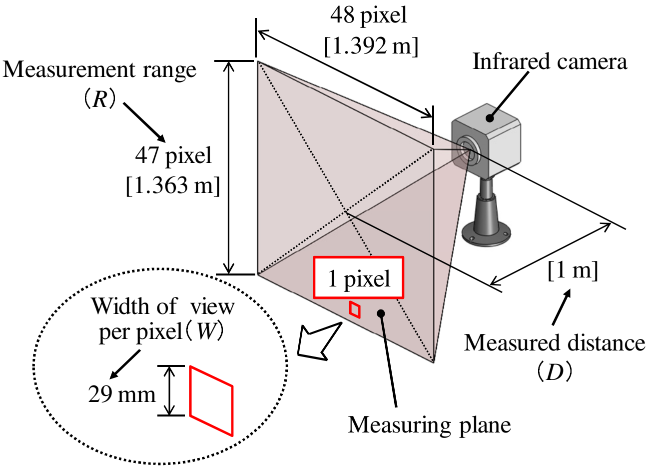


Fig. 10. View range of the infrared camera

3.3 Camera Configuration

Fig. 11 describe layout drawing of the infrared camera from a lateral view. The camera is located under the toilet seat, and the position which secure measured view for measuring whole buttock. The measuring place of the camera mach to the top of the toilet seat, and the front edge of the infrared image matches to the front end of the toilet seat. Then, the center of the toilet bowl is adjusted at the front end of the toilet seat as the default position. In this case, the distance between the top of the toilet seat and the infrared camera is 150mm. The measuring range comes to 204.45 mm×208.8mm and the width of one pixel becomes about 4.35 mm by calculating Equation (1) and (2), as shown in Fig. 12.

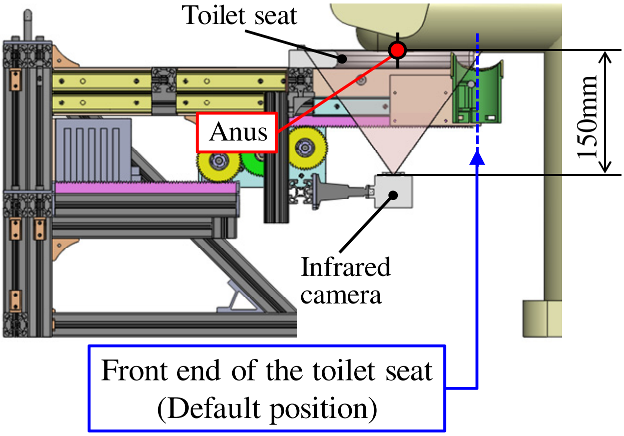


Fig. 11. Configuration of the infrared camera

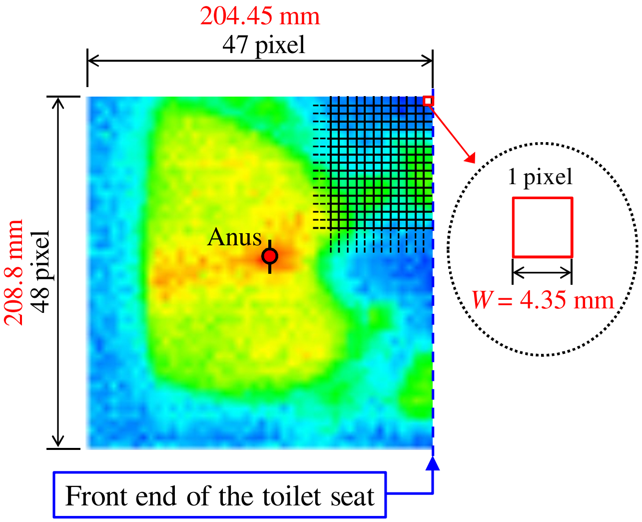


Fig. 12. Infrared image of buttocks

3.4 Anus Position Detection

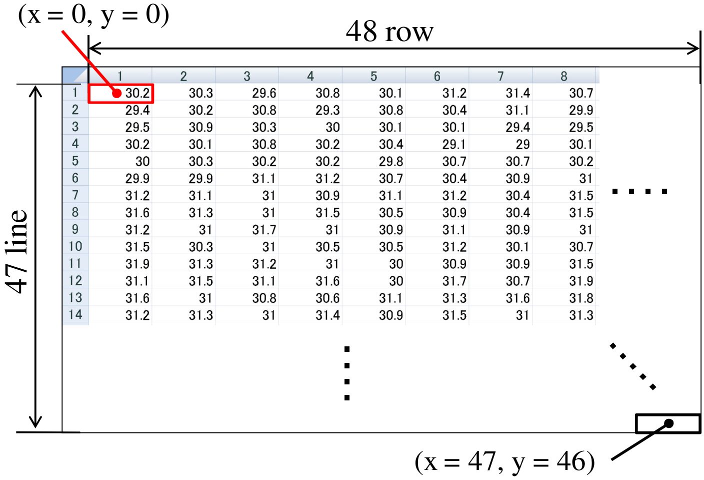
The system used the temperature data which are sent to the PC from the infrared camera via USB pinpoint the location of an anus. The temperature data are saved in the PC as a CSV file with dedicated software of the infrared camera. The CSV file displays the temperature value data of 2256 (= 47×48) pieces of the same number of pixels. The data is the table of 47 rows and 48 columns as shown in Fig. 13 on the CSV. When creating an infrared image, the image displays as shown in Fig. 14. The top of the image is a front direction of the person and the under is a rear side. Then, we assign the origin point (x=0, y=0) at the upper right and (x=47, y=46) at the lower left.

Since as we mentioned above, the anul position becomes high temperature, we detect it by following method (Fig. 15).

(1) From the temperature data in the CSV file, choose an area of 7×7 counting from the origin, and calculate the mean of temperature within that area (Fig. 15(a)).
(2) Then select an area of 7×7 shifted by one row, and calculate the mean of temperature as well (Fig. 15(b)).
(3) In the same way, calculating the average temperature in the area obtained by shifting one row to the extreme right. When complete, calculating by shifting one line in turn (Fig. 15(c)).
(4) As a result, the data of the average temperature of 42×41 pieces are obtained, and the area of the highest average temperature is detected from the data (Fig. 15(d)).
(5) Identifying the detected area of 7×7 as the place in which the anus is located, and outputting the coordinate data (Fig. 15(e)).

We select 7×7 area as the template size for the anus empirically by considering the distance between the anus and the infrared camera.

When moving the toilet bowl, matching the center of the bowl to the center point of the detected area. Calculating a moving distance from the center point coordinates in the detected area, and controls the position with the stepping motor.



	1	2	3	4	5	6	7	8
1	30.2	30.3	29.6	30.8	30.1	31.2	31.4	30.7
2	29.4	30.2	30.8	29.3	30.8	30.4	31.1	29.9
3	29.5	30.9	30.3	30	30.1	30.1	29.4	29.5
4	30.2	30.1	30.8	30.2	30.4	29.1	29	30.1
5	30	30.3	30.2	30.2	29.8	30.7	30.7	30.2
6	29.9	29.9	31.1	31.2	30.7	30.4	30.9	31
7	31.2	31.1	31	30.9	31.1	31.2	30.4	31.5
8	31.6	31.3	31	31.5	30.5	30.9	30.4	31.5
9	31.2	31	31.7	31	30.9	31.1	30.9	31
10	31.5	30.3	31	30.5	30.5	31.2	30.1	30.7
11	31.9	31.3	31.2	31	30	30.9	30.9	31.5
12	31.1	31.5	31.1	31.6	30	31.7	30.7	31.9
13	31.6	31	30.8	30.6	31.1	31.3	31.6	31.8
14	31.2	31.3	31	31.4	30.9	31.5	31	31.3

Fig. 13. CSV file

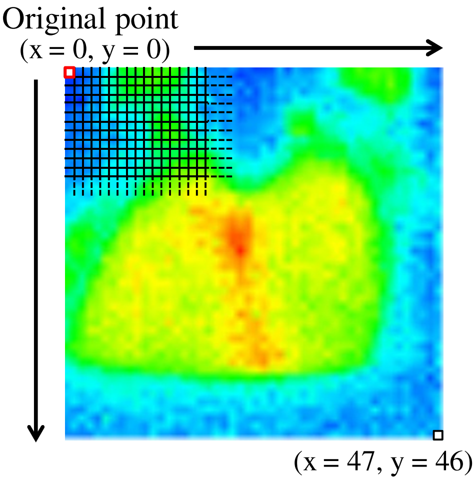


Fig. 14. Infrared image

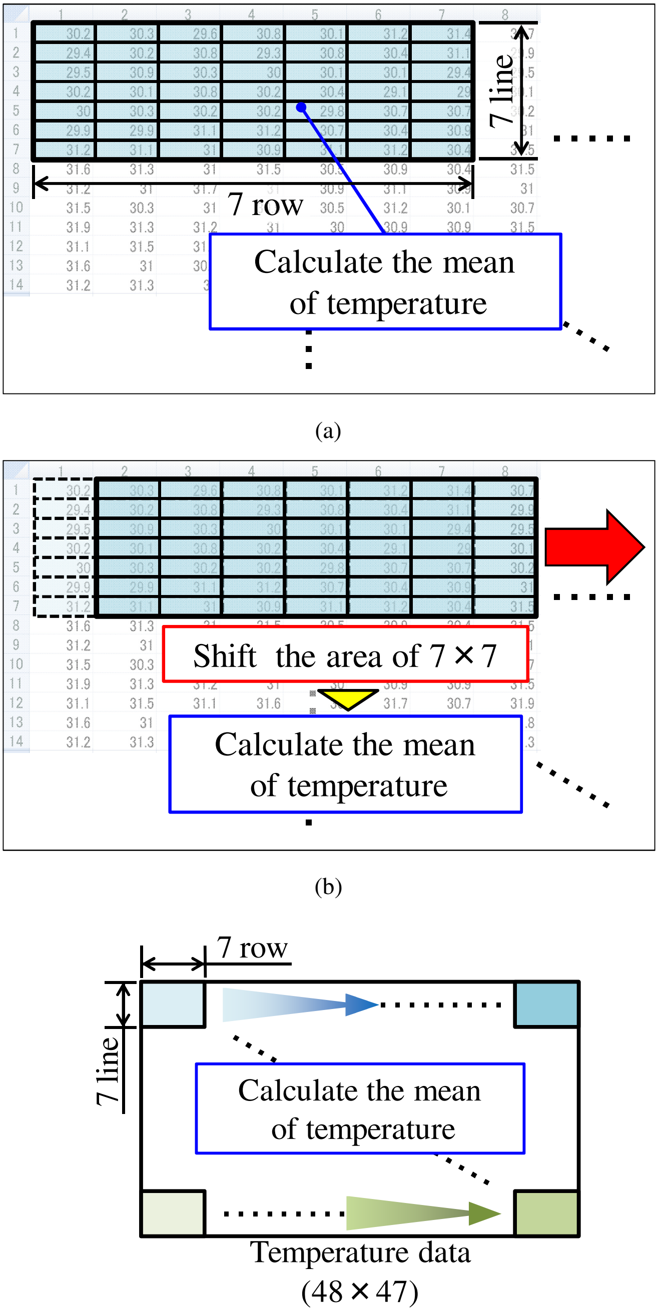
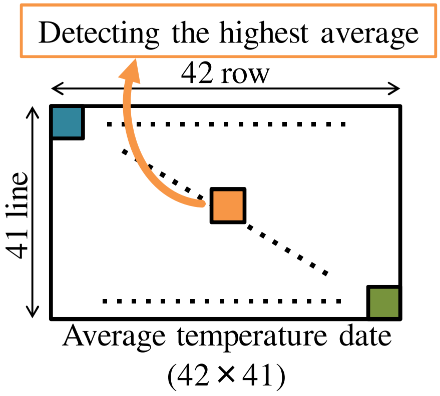
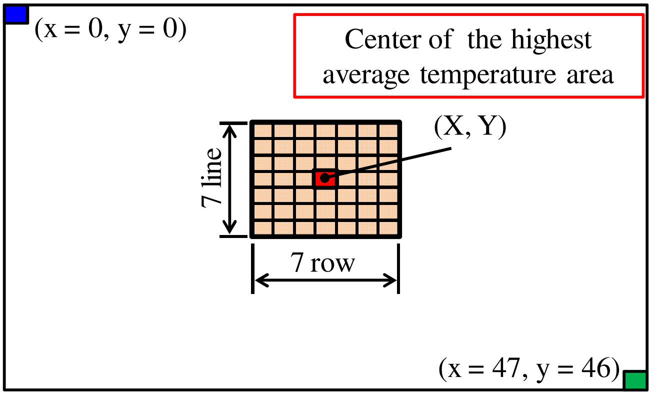


Fig. 15. Anal position detection process



(d)



(e)

Fig. 15. (*Continued)*

4 Estimation Experiment

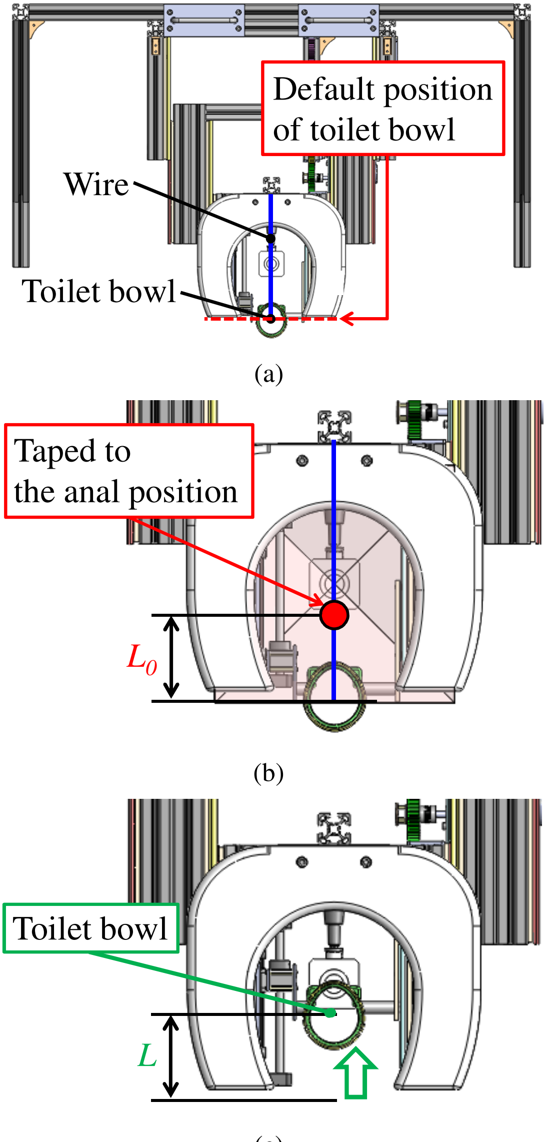
In order to check the precision of the anus position detection and the toilet bowl control, we investigate the following experiment.

4.1 Experimental Set up

As shown in Fig. 16(a), a wire is attached on the central line of the toilet seat. After subject sits on the toilet seat, the image of buttock is taken by the infrared camera and calculate the anal position (Fig. 16(b)). Then subject detects the position of the own anus by touching it and puts the thin tape on the wire at the anus position. We measure the distance (L_0) between the front end of the toilet seat and the tape position as the anus position, and compare L_0 to the position extracted by our method.

We apply PID control method for the toilet bowl control. We assign L_0 as the target position and compare the position of the toilet bowl controlled (L) (Fig. 16(c)) to L_0.

The experiment mentioned above is repeated 10 times and we investigate the precision of the method.



(a)

(b)

(c)

Fig. 16. Motion experiment of toilet bowl

4.2 Results and Discussion

Table 3 shows the extracted anal position and the error comparing to L_0. We find that the range of the error distributes from -5.2 mm to +5.1 mm. Since 1 pixel corresponds to 4.35 mm, 5.2 mm equivalents to 1.2 pixel. We are not sure, for such a system, how much accuracy is expected though, we may say that if the resolution of image becomes high, the error will be small. From this point of view, depending on the accuracy required, the resolution of the image should be selected. For our system, since the diameter of the toilet bowl is 70 mm (Fig. 17.), 5.2 mm error at maximum is acceptable.

Table 4 depicts the target position (*L*), controlled position of the toilet bowl and the error. We find that the range of the error distributes from -5.2 mm to +5.0mm. Since the diameter of the toilet bowl is 70 mm as mentioned before and also it is said the maximum diameter of the anus when it opens is 40 mm, 5.2 mm error at maximum is also acceptable.

Thus we confirmed that our proposed method is acceptable for the closely-contact-type toilet we have developed.

Table 3. The anal position extraction result

Position of anus (L0) [mm]	Extracted position [mm]	Error [mm]
74.0	78.3	+4.3
80.0	78.3	-1.7
95.0	100.1	+5.1
98.0	95.7	-2.3
104.0	108.8	+4.8
109.0	108.8	-0.2
114.0	108.8	-5.2
122.0	126.2	+4.2
126.0	130.5	+4.5
130.5	130.5	+0.0

Table 4. Toilet bowl control result

Target position (L_0) [mm]	Controlled position [mm]	Error [mm]
74.0	78.2	+4.2
80.0	78.2	-1.8
95.0	100.0	+5.0
98.0	95.7	-2.3
104.0	108.8	+4.8
109.0	108.8	-0.2
114.0	108.8	-5.2
122.0	126.1	+4.2
126.0	130.5	+4.5
130.5	130.5	+0.0

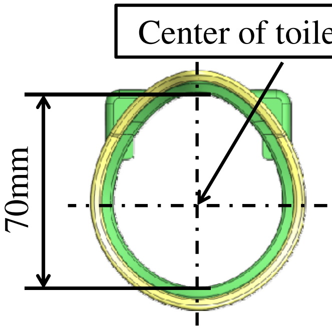


Fig. 17. Diameter of the toilet bowl

5 Conclusion

We developed the anal position detecting system for the closely-contact-type toilet we have developed. Since the temperature of the anus is higher than other places around it, the system extracts the position of the anus by using temperature differences. Infrared camera acquired the temperature data, PC calculated the position, and toilet bowl was moved by the PID control method. In order to check the precision of the anus position detection and the toilet bowl control, we conducted experiments using the system. The result of the error comparing to the actual anus position, the error range of position detection was from -5mm to 7mm, and the error range of toilet bowl control was from -5mm to 6mm. We thought the result was acceptable for the size of the toilet bowl which the diameter is 70mm, and size of the anus which the diameter when it opens is 40mm.

The development of an anal position detecting system enables to pinpoint the location of the anus. It is then considered that the system enables to clean the anus automatically and exactly. Therefore, it will be possible that the elderly and the disabled defecate independently without the help of caregivers.

We would like to create a cleaning system of the anus and develop more practical New-Toilet system.

References

[1] Ministry of Health, Labour and Welfare. Annual Health, Labour and Welfare Report 2011-2012, p. 2 (2013)
[2] Ministry of Health, Labour and Welfare. Annual report on the aging society: 2012, p. 2 (2013)
[3] Murakami, M., Yamaguchi, M., Suzuki, J., Hattori, K., Iwanaga, H.: A Study of Knowledge Supporting Basic Nursing Arts: A Philological Research on Elimination. The Journal of Juntendo Medical College of Nursing 4, 1–15 (1993)
[4] Hattori, K., Yamaguchi, M., Murakami, M., Suzuki, J., Omoto, M., Nagano, M.: A Study of Knowledge Supporting Basic Nursing Arts (2): Philological Research on Elimination. The Journal of Juntendo Medical College of Nursing 11, 72–87 (2000)
[5] Kumagai, S., Tainaka, H., Miyajima, K., Miyano, N., Kosaka, J., Tabuchi, T., Akasaka, S., Kosakai, H., Yoshida, J., Tomioka, K.: Load on the Low Back of Care Workers in Nursing Homes for the Elderly. Journal of Occupational Health 47(4), 131–138 (2005)
[6] Kikuchi, Y., Minai, J., Shimanouchi, S.: Factors of caregiver burden related with incontinence of the dependent elderly at home. Annals of the International University of Health and Welfare 15(2), 13–23 (2011)
[7] Yoshimoto, K.: Experience of Elderly People Receiving Excretion Assistance in the Care Facilities. Journal of Japan Academy of Gerontological Nursing 13(1), 57–64 (2008)
[8] Murakami, K., Fujisawa, H.: Questionnaire survey of care workers and examination of time required for toileting care prior to development of standing-position support equipment. Rehabilitation Science: Memoirs of the Tohoku Bunka Gakuen University Faculty of Medical Science & Welfare, Department of Rehabilitation 3(1), 41–48 (2007)
[9] Kojima, K., Abe, N., Oota, M., Kurihaara, N.: Market research report on excretion aids suitable for an aging society. The Bulletin of Reseach Institute for Interdisciplinary Science, Hachinohe Institute of Technology 6, 35–41 (2008)
[10] Ishikawa, N., Takahashi, Y.: Transfer Assist Robot System for Independence Excretion Support. Transactions of the Japan Society of Mechanical Engineers Kanto Branch 2003(9), 135–136 (2003)
[11] Yukawa, T., Nakata, N., Obinata, G., Makino, T.: Assistance System for Bedridden Patients to Reduce the Burden of Nursing Care (First report – Development of a multifunctional electric wheelchair, portable bath, lift, and mobile robot with portable toilet). In: 2010 IEEE/SICE International Symposium on System Integration (SII 2010), pp. 132–139 (2010)

[12] Homma, K., Yamada, Y., Matsumoto, O., Ono, E., Lee, S., Horimoto, M., Suzuki, T., Kanehira, N., Suzuki, T., Shiozawa, S.: A proposal of a method to reduce burden of excretion care using robot technology. In: Proc. 2009 IEEE Int. Conf. Rehabil. Robot. (ICORR), pp. 621–625 (2009)

[13] Homma, K., Yamada, Y., Matsumoto, O., Lee, S., Ono, E.: Ethical Review Process and Subject Protection in Demonstration Experiments of Assistive Robots for Care-Giving-Case Report of the Excretion Care Robot "Toilet-Assist"-Case Report of the Excretion Care Robot "Toilet-Assist". Journal of the Robotics Society of Japan 28(2), 181–190 (2010)

[14] Ueno, S., Imai, Y., Hayasaka, T., Okubo, M., Ishikawa, T., Yamaguchi, T.: Development of a mobile toilet system. In: The JSME Symposium on Welfare Engineering 2008, pp. 176–177 (2008)

[15] Itakura, T., Mitsuda, M., Tanamura, T.: Reseach on characteristics of the odors from excrement at the adult diaper exchange. J. Environ. Eng., AIJ 73(625), 335–341 (2008)

[16] Nagaya, K., Wang, U., Kojima, T.: Auto-Washing Panty Toilet for Care. Transactions of the Japan Society of Mechanical Engineers. C 70(692), 1141–1148 (2004)

[17] Kojima, T., Nagaya, K., Eda, K., Ando, Y.: Measurement of Urine, Moisture and Grip Pressure Control of a New Type Panty Toilet for Care. Transactions of the Japan Society of Mechanical Engineers. C 72(722), 3286–3293 (2006)

[18] Nakamura, K., Yamaguchi, K., Takakura, N.: Development of a bed taking care of excretion. In: The JSME Symposium on Welfare Engineering, vol. 2003(3), pp. 247–250 (2003)

[19] Fujimoto, T., Hashimoto, T., Sakaki, H., Higashi, Y., Tamura, T., Tsuji, T.: Automated Handling System for Excretion. In: Proceedings of the 20th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, vol. 20(4), pp. 1973–1974 (1998)

[20] Tokoro, K., Hashimoto, T., Kobayashi, H.: Development of New Toilet System. In: Proceedings of the 1st Annual IEEE Healthcare Innovation Conference of the IEEE EMBS, pp. 105–108 (2012)

Experimental Confirmation of Cylindrical Electromagnetic Sensor Design for Liquid Detection Application

K. Tashiro[1], H. Wakiwaka[1], T. Mori[1], R. Nakano[1], N.H. Harun[2], and N. Misron[2]

[1] Spin Device Technology Center (SDTC), Shinshu University, Wakasato 4-17-1,
Nagano, Japan
[2] Faculty of Engineering, University Putra Malaysia, Kuala Lumpur, Malaysia
tashiro@shinshu-u.ac.jp

Abstract. Design of cylindrical electromagnetic sensor is discussed. The purpose of this sensor is to detect a change in the conductivity or permittivity of the liquid. If the evaluation frequency is less than 100 kHz, skin effects would be negligible. The proposed sensor consists of two cylindrical sensors, a solenoid coil and a cylindrical capacitor. For an ideal shape condition, estimation methods for inductance and capacitance have been already proposed. However, the practical issues need to be clarified. It starts with fabrication of several coils and capacitors, to confirm the validity of the estimation methods. From experimental results, it was found that, the estimation error and existence of the parasitic element could not be neglected. Liquid detection demonstrations with fabricated sensors are also demonstrated.

Keywords: Liquid detection sensor, Electromagnetic field, Low-frequency, Cylindrical capacitor, Solenoid coil.

1 Introduction

Nondestructive electromagnetic inspections are widely used in liquid inspection applications. For distinguishing between a hazardous liquid, such as gasoline, ethanol, etc., and a drink having water, devices for judging types of liquid inside a plastic bottle [1] are installed in the most of Japanese airports. The sensor of this device is based on the capacitance measurement. When a sample is inserted between the electrodes of a capacitor, the value of capacitance changes in accordance with the permittivity of the sample. For an agricultural application, an inductive sensor for grading of oil palm fruit is a unique example [2]. Another example, micro coils which in a sample water can investigate bacteria concentration is reported [3]. They sensors are based on the inductance measurement. When a sample is placed on an air-core coil, the value of resonance frequency changes in accordance with the electromagnetic properties of the sample. Recently, many researchers focus on the a capacitive sensor, or an interdigital sensor, to detect contaminants in liquid, such as several chemicals

A. Mason et al. (eds.), *Sensing Technology: Current Status and Future Trends IV*,
Smart Sensors, Measurement and Instrumentation 12, DOI: 10.1007/978-3-319-12898-6_7

[4-9], and bacterial endotoxin [10]. In this measurement, the differences in frequency responses of the impedance are focused on.

In practical use, there are three remarkable points. The first is the impedance design of the sensor. In order to discuss the frequency dependence, a simple equivalent circuit model should be considered. Therefore, the estimation accuracy of the own inductance or capacitance is important issue. The second is the evaluation frequency of the electromagnetic field. If the frequency is large, the skin effect should be considered which limits the inspection area to the surface of the sample. The third is that conventional liquid inspections are usually microscopic or sampling measurements because the sensor should be touch to the liquid or its container. If the sensing area is relatively wide whose electromagnetic field affects the impedance of the sensor, macroscopic or total measurements are possible.

For liquid detection applications, this chapter presents the design of two cylindrical sensors, a solenoid coil and cylindrical capacitor. When the liquid sample is set inside the cylinder, the impedance profile of the sensor changes in accordance with the electromagnetic properties of the liquid. Most of liquids are non-magnetic material whose values of conductivity are less than 100 S/m [11]. The evaluation frequency is limited to below 100 kHz. It makes free from the skin effects because the corresponding skin depth is over 50 m. These sensor can be used for macroscopic or total measurements, because the sensing area is the inside of the cylinder and the liquid is required to touch the sensor. The theoretical estimation of both the inductance and capacitance have been already proposed from 1900's. However, the acceptable estimation error and the parasitic element should be considered for practical use. From several experimental confirmations mentioned below, several remarks are summarized.

2 Complete Elliptic Integral

To estimate both the inductance and capacitance, complete elliptic integral of first kind K and second kind E should be calculated. The theoretical backgrounds mentioned below referred to a famous, well-organized and compact Japanese mathematic formulae book (S. Moriguchi, K. Udagawa and S. Hitomatsu, "IWANAMI SUUGAKU KOUSHIKI", Iwanami publishing, 22th edition, 2010). The definitions of both K and E are described by following equations.

$$K = K(k) = \int_0^{\pi/2} \frac{1}{\sqrt{1-k^2\sin^2\theta}} d\theta \tag{1}$$

$$E = E(k) = \int_0^{\pi/2} \sqrt{1-k^2\sin^2\theta} \quad d\theta \tag{2}$$

Where k is the elliptic modulus. It should be noted that the Legendre's relation which represents the relationship between K and E.

$$EK' + E'K - KK' = \frac{\pi}{2} \tag{3}$$

$$K' = K(k') = K\left(\sqrt{1-k^2}\right) \tag{4}$$

$$E' = E(k') = K\left(\sqrt{1-k^2}\right) \tag{5}$$

$$k'^2 = 1-k^2 \tag{6}$$

Where k' is the complementary elliptic module, K' and E' are complementary complete elliptic integral of the first and second kind. The theoretical calculation of inductance for the solenoid coil needs K and E, and that of capacitance for the cylindrical capacitor needs K and K'. Because $K(k)$ and $E(k)$ could not be solved directly, the acceptable approximation should be considered. A famous example of the approximations which can be expressed as a power series.

$$\begin{aligned} K(k) &= \frac{\pi}{2}\sum_{n=0}^{\infty}\left(\frac{(2n-1)!!}{(2n)!!}\right)k^{2n} \\ &= \frac{\pi}{2}\left(1+\left(\frac{1}{2}\right)^2 k^2+\left(\frac{1\cdot 3}{2\cdot 4}\right)^2 k^4+\left(\frac{1\cdot 3\cdot 5}{2\cdot 4\cdot 6}\right)^2 k^6+\cdots\right) \\ &= \frac{\pi}{2}\left(1+\frac{1}{4}k^2+\frac{9}{64}k^4+\frac{25}{256}k^6+\frac{1225}{16384}k^8+\frac{3969}{65536}k^{10}\cdots\right) \end{aligned} \tag{7}$$

$$\begin{aligned} E(k) &= \frac{\pi}{2}\sum_{n=0}^{\infty}\left(\frac{(2n-1)!!}{(2n)!!}\right)\frac{k^{2n}}{1-2n} \\ &= \frac{\pi}{2}\left(1-\left(\frac{1}{2}\right)^2 k^2+\left(\frac{1\cdot 3}{2\cdot 4}\right)^2 \frac{k^4}{3}-\left(\frac{1\cdot 3\cdot 5}{2\cdot 4\cdot 6}\right)^2 \frac{k^6}{5}+\cdots\right) \\ &= \frac{\pi}{2}\left(1-\frac{1}{4}k^2+\frac{3}{64}k^4-\frac{5}{256}k^6+\frac{175}{16384}k^8-\frac{441}{65536}k^{10}+\cdots\right) \end{aligned} \tag{8}$$

Where $n!!$ denotes the double factorial. In 1955, another famous approximations were proposed by C. Hastings [12], which was aimed to be applied for computer calculations.

$$\begin{aligned} K(k) &= \left(1.3862944+0.1119723\,k'^2+0.0725296\,k'^4\right) \\ &+\left(\frac{1}{2}+0.1213478k'^2+0.0288729\,k'^4\right)\ln(1/k'^2) \end{aligned} \tag{9}$$

$$\begin{aligned} E(k) &= \left(1+0.4630151\,k'^2+0.1077812\,k'^4\right) \\ &+\left(0.2452727k'^2+0.0412496\,k'^4\right)\ln(1/k'^2) \end{aligned} \tag{10}$$

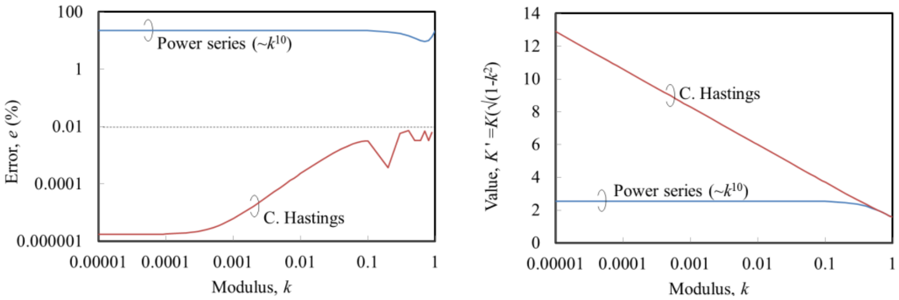
Both of them can calculate with a simple calculator as Excel spreadsheet program. To investigate the estimation error, the Legendre's relation can be used. Here, the estimation error e is defined by following equation.

$$e = \frac{EK' + E'K + KK' - \frac{\pi}{2}}{\frac{\pi}{2}} \times 100 \quad (\%) \tag{11}$$

Fig. 1(a) shows the comparison of the estimation error as a function of modulus. The estimation error of the approximations with power series was almost 20 % even though the modulus was taken in account up to the 10th order's, t. In contrast, the values of estimation error with the approximations proposed by C. Hastings were lower than 0.01 %.

Fig. 1.(b) shows the comparison of *K*' value as a function of the modulus *k*'. When the modulus *k*' is less than 0.1, the *K*' value converges to 2.5506. However this value should be divergent if the modulus *k*' is close to 0. Namely, the approximations with power series cannot use when the modulus *k* is close to 1.

Since the approximations with power series are derived from mathematic basis, it is not easy to obtain the good convergence in practical use. Hereafter, the approximations proposed by C. Hastings were used for the estimation of complete elliptic integral.



(a) Error, *e*

(b)Value, *K*'

Fig. 1. Calculated values as a function of modulus *k*. When the approximations with power series were used, the modulus were calculated up to the 10th order's.

3 Electromagnetic Sensor Design

Fig. 2. shows the schematic design of cylindrical electromagnetic sensors: 1) solenoid coil and 2) cylindrical capacitor. Where *l* is the length of the sensor, D_o and D_O are the inner and outer diameter of the non-magnetic core for the sensor, *t* is the diameter of coil windings of the solenoid coil and *θ* is the gap angle of the cylindrical capacitor, respectively.

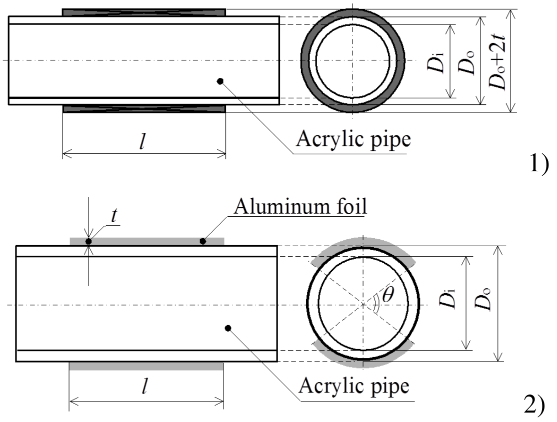


Fig. 2. Schematic design of cylindrical electromagnetic sensors: 1) Solenoid coil, 2) Cylindrical capacitor

3.1 Estimation of Inductance

The inductance of an ideal solenoid coil is described by following equation.

$$L_0 = \frac{\mu_0 \pi r^2 n^2}{l} \qquad \text{(H)} \tag{12}$$

Where μ_0 (H/m) is the permeability in vacuum, r (m) is radius of the coil, n is number of the coil windings, and l (m) is length of the solenoid coil. If the diameter of coil windings is not negligible compared with the that of the solenoid coil, the radius should be defined as follows.

$$r = \frac{D_o + t}{2} \qquad \text{(m)} \tag{13}$$

In practical case, the existence of open ends should be taken in account even if the solenoid coil has relatively long length. The inductance of a solenoid coil with finite length is expressed by following equation.

$$L = C_{nagaoka} L_0 \qquad \text{(H)} \tag{14}$$

Where C_{nagaoka} is Nagaoka coefficient, and L_0 (H) is the inductance of the ideal solenoid coil. The Nagaoka coefficient is expressed by following equation.

$$C_{nagaoka} = \frac{4}{3\pi}\frac{1}{k'}\left[\frac{k'^2}{k^2}(K - E) + E - k\right] \tag{15}$$

$$k = \sqrt{\frac{4r^2}{4r^2 + l^2}} \tag{16}$$

Where k is the elliptic modulus, k' is the complementary elliptic module, K is the complete elliptic integral of the first kind, and E is the complete elliptic integral of the second kind. Because the calculation of this coefficient requires the estimation of the complete elliptic integral, several researchers had tried to provide suitable approximations. In 1909 [13], H. Nagaoka proposed approximations with q-series. It is known that the complete elliptic integral of first kind can be estimated with q-series by following equations.

$$K = \frac{\pi}{2}\left[1+\sum_{n=1}^{\infty} q^{n^2}\right]^2 \tag{17}$$

$$q = a + 2a^5 + 15a^9 + 150a^{13} + 1707a^{17} + \cdots \tag{18}$$

$$a = \frac{1-\sqrt{k'}}{2\left(1+\sqrt{k'}\right)} \tag{19}$$

Because the use of q-series makes the convergence rapid, the following approximation was proposed.

$$C_{nagaoka} = \frac{2}{3(1-2q+2q^4)^2} \times \left\{1+8\left(q^2-q^4\right)+\frac{8(q-4q^4)}{1-2q+2q^4}\frac{k'^2}{k^2}\right\} - \frac{4}{3\pi}\frac{k}{k'} \tag{20}$$

It was noted that it is superfluous in practical use because it give up to six decimal points result. However, the q value is also expressed as a power series. From the practical point of view, C. Hastings' approximations may have advantages of both the simplicity and accuracy.

3.2 Estimation of Capacitance

The capacitance of an ideal parallel-plate capacitor is described by following equation.

$$C_0 = \varepsilon_0 \frac{w}{d} l \quad \text{(H)} \tag{21}$$

Where ε_0 (F/m) is the permittivity in vacuum, w (m) is width of the plate, d (m) is distance between the two plate, and l (m) is length of the plate. In 1940 [14-15], Y. Omoto and K. Morita proposed an approximation for the capacitance of a cylindrical capacitor by using conformal transformation.

$$C = \frac{10}{4\pi \times 9}\frac{K'}{K} \qquad \text{(pF/cm)} \tag{22}$$

$$k = \tan^2\left(\frac{\theta}{4}\right) \tag{23}$$

Where θ (rad) is the gap angle of two cylindrical plate, k is the elliptic modulus, k' is the complementary elliptic module, K is the complete elliptic integral of the first kind, and K' is the complementary complete elliptic integral of the first kind. Since the definition of k is described as $\tan^2(\theta / 2)$ in this original paper, it may be a mistake. Because if $k = \tan2(\theta / 2)$ were right and θ were more than 90 rad, module k would be more than 1. However module k must be $0 < k < 1$. This approximation can be expressed with SI unit as follows.

$$C = \frac{1}{4\pi \times 9} \times 10^{-9} \times \frac{K'}{K} \text{ (F/m)} \tag{24}$$

Where the value of constant is $8.8419... \times 10^{-12}$ whose value is regarded the permittivity in vacuum. This approximation can be expressed by following equation.

$$C = \varepsilon_0 \frac{K'}{K} l \quad \text{(F)} \tag{25}$$

On the analogy of the ideal parallel-plate capacitor, the values of K' and K are regarded as the width and distance of the two parallel-plates. It should be noted that the capacitance does not depend on the diameter of the cylinder. This estimation also needs the calculation of the complete elliptic integral. Although the approximation of the cylindrical capacitor was also presented as following equation, the estimation error will be large for large angle.

$$C = 0.129612 \times \left(0.60206 + 2\log_{10} \cot\frac{\theta}{4}\right) \qquad \text{(pF/cm)} \tag{26}$$

3.3 Design of Electromagnetic Sensor

Table 1 shows the specifications of solenoid coils, (a) and (b). The values of inductance were calculated from equation (14) with the equations (6), (9), (10), (12), (13), (15) and (16). Fig. 3(a). shows the elliptic modulus as a function of the length of the solenoid coil. Because the values of elliptic modulus were 0.316512 and 0.158375, the corresponding values of the estimation error calculated from equation (11) were 0.00643 % and 0.00179 %, for (a) and (b), respectively. Fig. 3(b). shows the estimated inductance as a function of the length of the solenoid coil. From the experimental results, the used values of the number of coil windings per 1 mm were 4.453 and 4.500 for (a) and (b), respectively.

Table 2 shows the specifications of cylindrical capacitors, (a), (b), (c) and (d). The values of capacitance were calculated from equation (25) with equation (4), (6), (9) and (23). To confirm the estimation error of the complete elliptic integral, values of the elliptic modulus were calculated. Fig. 4(a). shows the elliptic modulus as a function of the gap angle of the cylindrical capacitor. The values of elliptic modulus were 7.61582×10^{-5}, 0.0717967, 0.171572 and 0.446462, so that the corresponding values of the estimation error calculated from equation (11) were -0.00001 %, -0.002842 %, -0.001172 % and 0.005943 %, respectively. Fig. 4(b). shows the estimated capacitance as a function of the gap angle of the cylindrical capacitor.

If a liquid is set in the solenoid coil or the cylindrical capacitor, the changes in conductivity, permeability and permittivity of the liquid sample produce the changes in inductance or capacitance. It may behave as the changes in relative permeability or permittivity. If they have frequency dependence, their profile corresponds to the liquid property. The measured value of inductance and capacitance can be expressed by following equations.

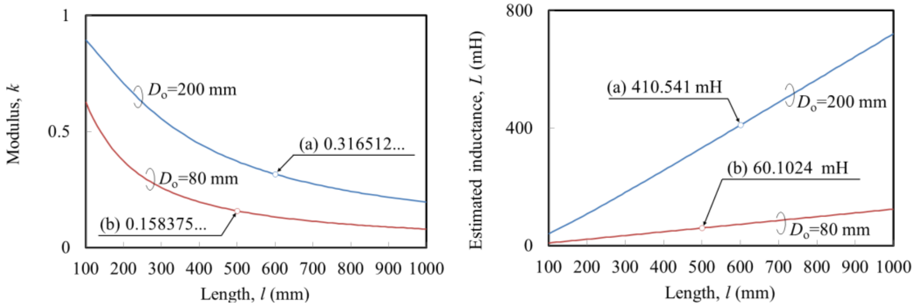
$$Ls(f) = \mu_r(f) \times L \qquad \text{(H)} \tag{27}$$

$$Cs(f) = \varepsilon_r(f) \times C \qquad \text{(F)} \tag{28}$$

In ideal, the values of L and C does not depend on the frequency, and can be defined by equation (14) and (25). In practice, the estimation error may occur due to the approximations and tolerance in fabrication. It should be also considered that these values may have frequency dependence because of parasitic elements.

Table 1. Specifications of solenoid coils

Property	(a)	(b)
Diameter of copper wire, *t* (mm)	0.2	
Number of turn, *N* (turn)	2672	2250
Length, *l* (mm)	600	500
Inner diameter, D_O (mm)	200	80
Pipe's inner diameter, *D*i (mm)	194	74
Estimated inductance, *L* (mH)	410.541	60.1024

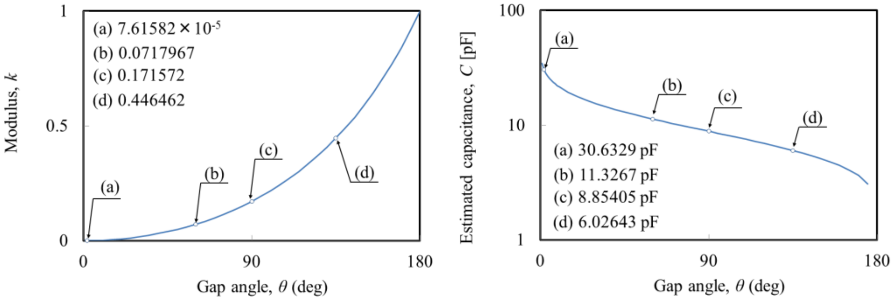


(a)Modulus, *k* (b) Estimated inductance, *L*

Fig. 3. Calculated values as a function of the length of the solenoid coil *l*

Table 2. Specifications of cylindrical capacitors

Property	(a)	(b)	(c)	(d)
Length, *l* (mm)	500			
Outer diameter, D_o (mm)	80			
Pipe's inner diameter, D_i (mm)	74			
Thickness of electrode, *t* (μm)	11			
Gap angle, *θ* (deg)	2	60	90	135
Estimated capacitance, *C* (pF)	30.6329	11.3267	8.85405	6.02643



(a)Modulus, *k* (b) Estimated capacitance, *C*

Fig. 4. Calculated values as a function of the gap angle of the cylindrical capacitor

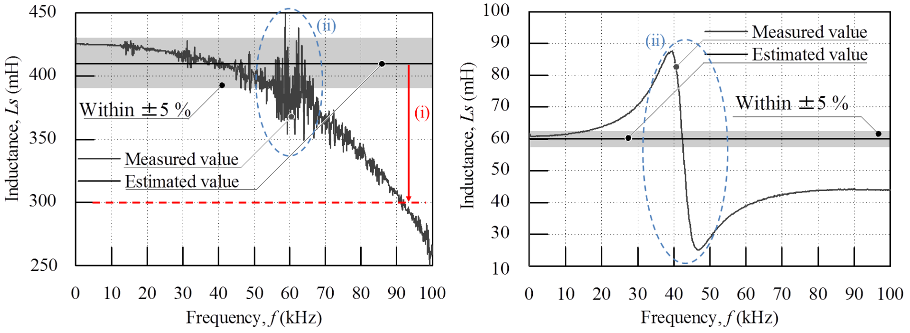
4 Experimental Confirmation

Basis and design of cylindrical electromagnetic sensors were discussed above. Through the measurement of the impedance or capacitance for both the solenoid coils and cylindrical capacitors, the validity of the design was confirmed. These were measured using an impedance analyzer (Agilent, 4294A). In order to demonstrate the liquid detection, three kinds of liquid were tested: city water, saturated salt water, and salad oil. First of all the values of conductivity of those liquid were measured with a conductivity tester (FUSO Model-7200). The measured values were 133 μS/cm for a city water, >20.0 mS/cm for the saturated salt water, and 11.0 μS/cm for a salad oil, respectively. The values of permittivity were not measured because a suitable permittivity tester was not found in commercial products. From a reference, it is found that the values of the relative permittivity at room temperature are known as 80 for a water, and 2~3 for an oil. Because they are not magnetic material, the values of relative permeability are 1 for the all.

4.1 Estimation Error

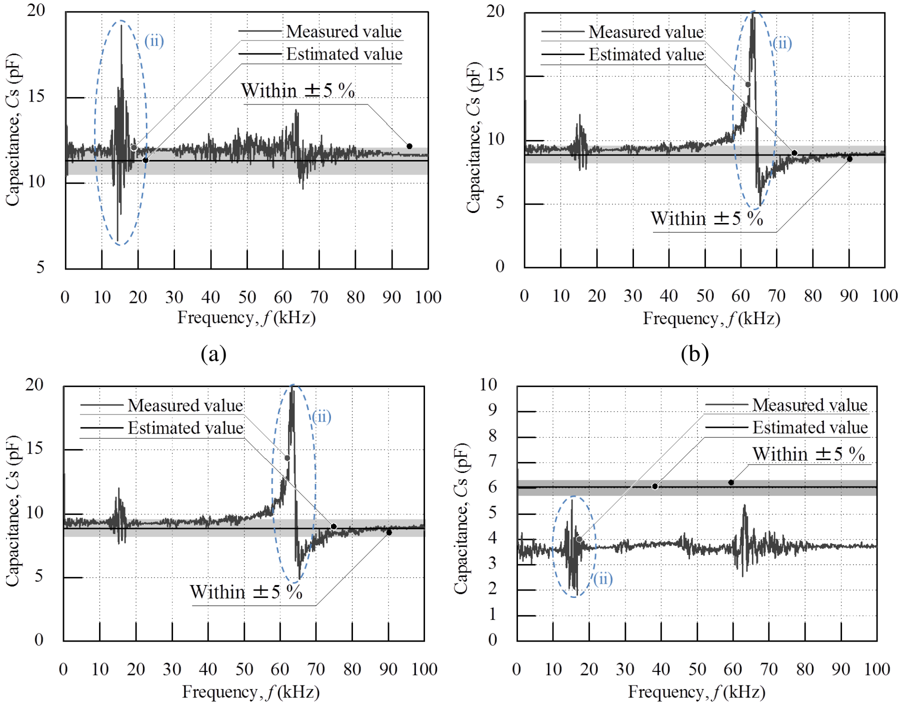
Fig. 5. shows the inductance profiles of the solenoid coil as a function of the frequency. The values of estimated inductance were 410.541mH and 60.1024 mH for (a) and (b), respectively. Fig. 6. shows the capacitance profiles of the cylindrical capacitor as a function of the frequency. The values of estimated capacitance were

30.6329 pF, 11.3267 pF, 8.85405 pF and 6.02643 pF, for (a), (b), (c) and (d), respectively. Excepting the cylindrical capacitor (d), the estimated values agreed with the experimental results at low frequency range. When the frequency was less than 10 kHz, the measured values were within ± 5 %.



(a) (b)

Fig. 5. Inductance profiles of the solenoid coil as a function of the frequency



(a) (b)

(c) (d)

Fig. 6. Capacitance profiles of the cylindrical capacitor as a function of the frequency

In ideal, the estimated values of both the inductance and capacitance should not depend on the frequency. Experimental confirmations revealed two undesired phenomena:

(i) The measured inductance decreased as the frequency was increased. It may be caused by the parasitic capacitance.

(ii) Perturbation of measured values were found. It may be caused by the environmental electromagnetic noise or condition of the impedance analyzer.

4.2 Demonstration of Liquid Detection

Fig. 7. shows the schematic design of liquid detection with the solenoid coil (a). Sample liquid sealed in a 1500 ml plastic bottle was set in this center. First of all, it was confirmed that the power source conditions affected the profile of the measured inductance. If the constant voltage mode of 1 V was used, perturbation of measured values as mentioned before were negligible. Fig. 8. shows the measured inductance as a function of the frequency. When the city water or saturated salt water was set, the differences in profile of the inductance were clearly confirmed. However, clear differences in the profile were not observed when the sample was the salad oil.

Fig. 9. shows the schematic design of liquid detection with the cylindrical capacitor (c). Sample liquid sealed in a 555 ml plastic bottle was set in this center. Fig. 10. shows the measured capacitance as a function of the frequency. The differences in profile of the inductance were clearly confirmed. Although some perturbations of the values were observed, hereafter the power source condition were fixed as the constant voltage mode of 1 V.

5 Discussion

The validity of approximations for both the inductance and capacitance were confirmed at low frequency range. From experimental confirmation, two undesired phenomena were observed. One is the decrease in measured inductance when the frequency was increased. Because the gaps between of the coil windings are short, parasitic capacitance may exist. Another is the perturbation of the measured values. It was found that a suitable source condition of the impedance analyzer made it negligible. It should be also considered that the requirement of a cable between the coil and impedance analyzer because it may affects both of them. In order to confirm the effect of the cable condition on the profile of both the inductance and capacitance, several experiments were conducted. In practical use, this matter is one of the most important problem which needs several cut-and-try. Some of reasonable results were summarized as follows.

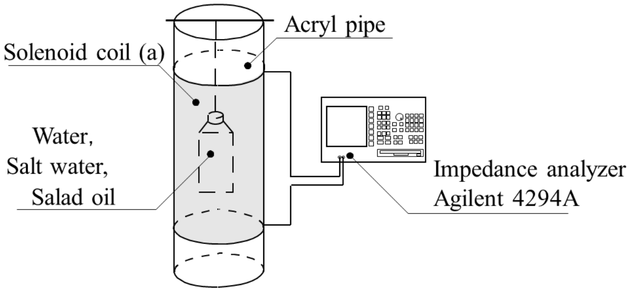
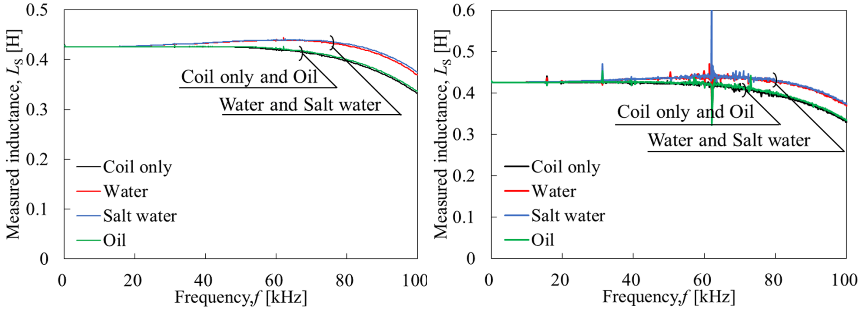
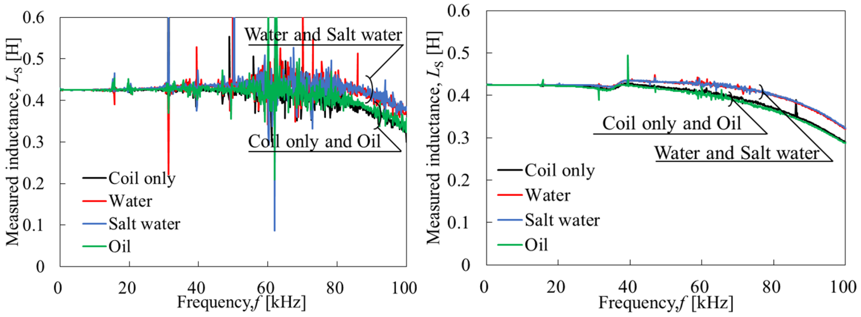


Fig. 7. Schematic design of liquid detection with the solenoid coil (a), whose estimated inductance L was 410.541 mH



(a) 1 V (b) 100 mV



(c) 10 mV (d) 20 mA

Fig. 8. Demonstration of liquid detection with the solenoid coil (a) when the power source condition was (a) 1V, (b) 100 mV, (c) 10 mV and (d) 20 mA

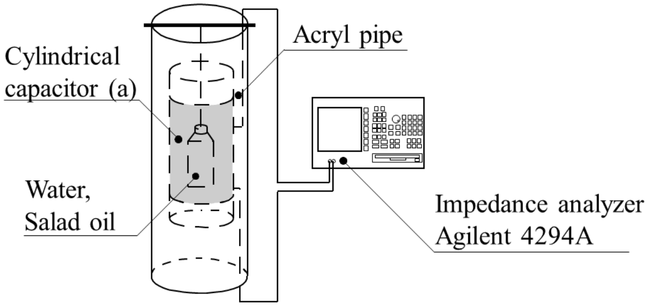
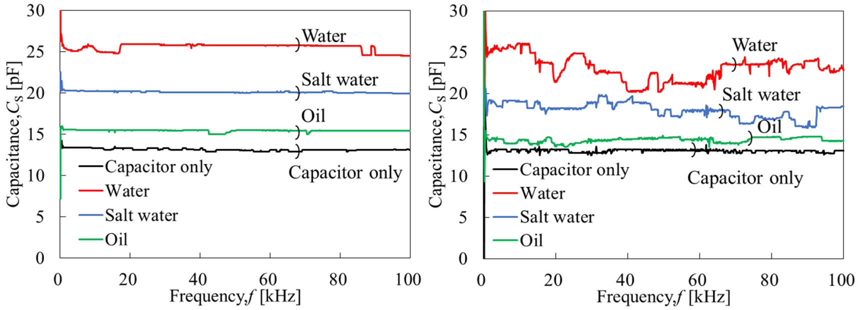
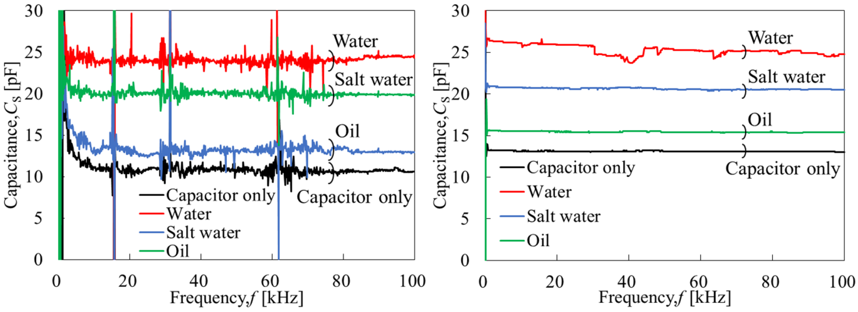


Fig. 9. Schematic design of liquid detection with the cylindrical capacitor (c), whose estimated capacitance C was 8.85405



(a) 1 V

(b) 100 mV



(c) 10 mV

(d) 20 mA

Fig. 10. Demonstration of liquid detection with the cylindrical capacitor (a) when the power source condition was (a) 1V, (b) 100 mV, (c) 10 mV and (d) 20 mA

A twist cable can be reduced the own inductance and induced noise from environmental magnetic field. In ideal, the external inductance of the cable can be negligible, and the value is close to the internal inductance, L_i.

$$L_i = \frac{\mu}{8\pi} 2l = 10^{-7} \times l \text{ (H)} \qquad (29)$$

Where μ (H/m) is the permeability of the wire ($=4\pi\times10^{-7}$), l is the length of the twist cable. Compared with the inductance of the solenoid coil (a), this value can be negligible. Although the approximation of the inductance is possible, that of the parasitic capacitance is not easy.

Fig. 11. shows the two kinds of cable condition for the solenoid coil (a). The terminals of the solenoid coil were both open ends, the length was defined by the lower side of the open end. Fig. 12. shows the profiles of measured inductance when the samples were set to the center. It was found that the parasitic capacitance caused the resonant phenomenon. The resonant frequency range was around 30 kHz for the use of twist cable, 80 kHz ~ 90 kHz for the use of the short cable. The resonant frequency f_r can be defined by the following equation.

$$f_r = \frac{1}{2\pi\sqrt{L_{total}C_{total}}} \quad \text{(Hz)} \qquad (30)$$

Where L_{total} (H) and C_{total} (F) are the synthesis inductance and capacitance, respectively. At the measurement of the inductance discussed before, the cable was relatively long and the two wires of the cable had relatively large distance. Because the parasitic capacitance of that were negligible, the values of resonant frequency were larger than 100 kHz. It was found that the distinction between the city water, saturated salt water and salad oil was possible with this solenoid coil if the parasitic capacitance was relatively small. From the values of resonant frequency, the values of parasitic capacitance were 67.6 pF and 8.13 pF for the cable (1) and (2), respectively.

The relationship of the resonant frequency was:

f_r (Solenoid coil(a)): Salt water < Water < Oil < Coil only

These experimental results show that an increase in the conductivity of the sample produces a decrease in the resonant frequency. The measured conductivity of the salt water was larger than 10 times compared with the salad oil. The difference in the resonant frequency was a few kHz when the short cable was connected. Since the difference between the city water and salt water was less than 1 kHz while the differences in conductivity were more than 10 times.

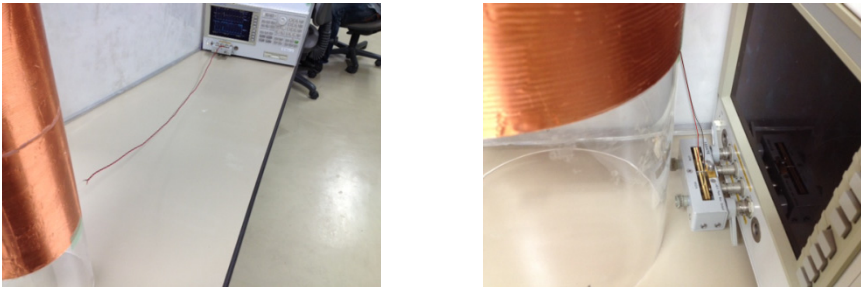
Fig. 13. shows the two kinds of cable condition for the cylindrical capacitor (c). The terminals of the cylindrical capacitor were the lower side of the open end, the both length of the wire of the cable were the same. Fig. 14. shows the profiles of measured capacitance when the samples were set to the center. The values of parasitic capacitance in both the cables increased the values of the measured capacitance. The values of estimated parasitic capacitance in both cables agreed with the results discussed before.

The values of measured capacitance:[16-18]

C_s (Cylindrical capacitor): Water > Salt water > Oil > Capacitor only

Compared with the solenoid coil, the order of results was different. These experimental results may exhibit that an increase in the permittivity of the sample produces an increase in the capacitance. In this liquid detection, the measurement of difference in the resonant frequency is not required. Because the measured values may contain the frequency dependency in the permittivity of the sample, it is an advantage for the inspection of the liquid in detail. Since the capacitance of the cylindrical capacitor does not depend on the diameter of that, the available value of the capacitance per length was less than 100 pF/m. Because the value of parasitic capacitance in the cable is considerable, a suitable connection between the cylindrical capacitor and impedance analyzer should be selected for a practical use.

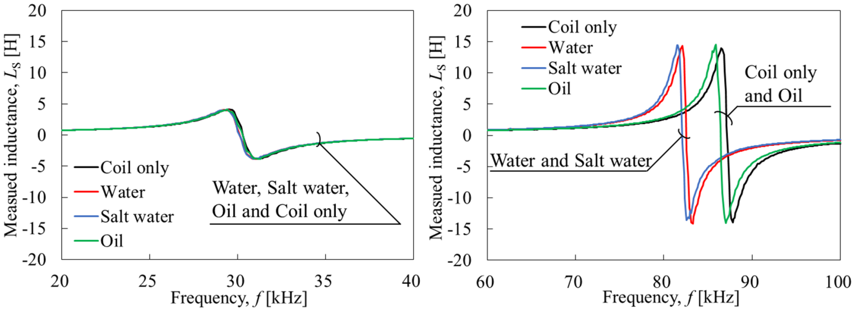
A suitable application with these electromagnetic sensors is a macroscopic inspection, or total inspection of flowing liquid inside a pipe. Compared with conventional microscopic or sampling inspections, this advantage is very attractive for practical use even though the accuracy is not high. Contamination detection of waste water, pureness evaluation of purified water, and salinity monitoring system are possible applications. Combination of the two sensors and other sensors, the inspection accuracy could be improved. The sensitivity dependence of the sample position is important in the design for these applications. Fig. 15 shows the definition of the city water sample positions for both sensors. The edge effects of both sensors on the measured impedance were investigated. Fig. 16 shows the summary of the experimental results. The difference of the measured value was low when the sample was away from the center of both sensor. In contrast, it was confirmed that the symmetry of the structure made the symmetry of the measured values of the impedance. The symmetry properties can be useful for the design of differential structured sensors.



(1) with a twist cable of 1m

(2) with a short cable of 0.1 m

Fig. 11. Two kinds of cable condition for the solenoid coil (a), whose estimated inductance L is 410.541 mH



(1) with a twist cable of 1 m

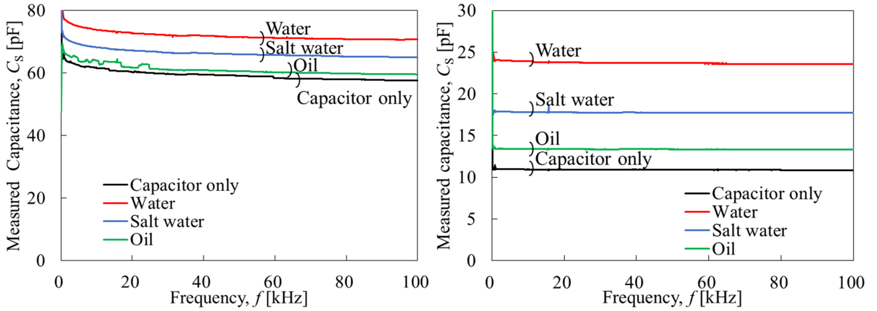
(2) with a short cable of 0.1 m

Fig. 12. Demonstration of liquid detection with the solenoid coil (a)

(1) with a twist cable of 1m

(2) with a short cable of 0.1 m

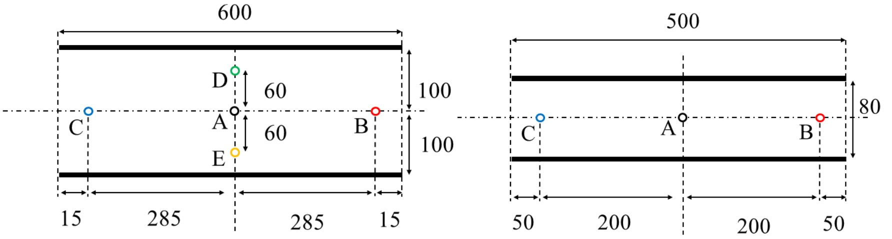
Fig. 13. Two kinds of cable condition for the cylindrical capacitor (c), whose estimated capacitance C is 8.85405 pF



(1) with a twist cable of 1 m

(2) with a short cable of 0.1 m

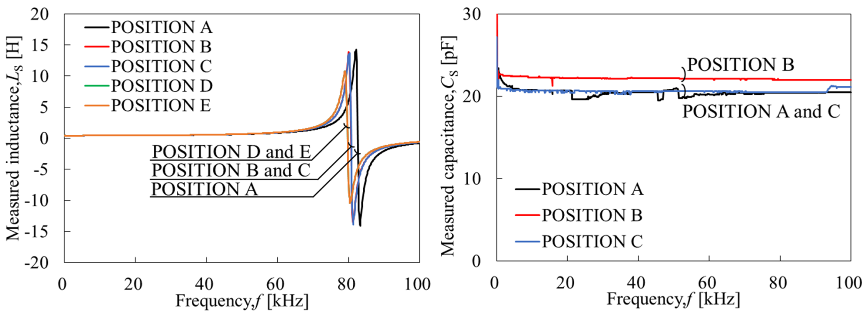
Fig. 14. Demonstration of liquid detection with the cylindrical capacitor (c)



(1) Solenoid coil (a)

(2) Cylindrical capacitor (c)

Fig. 15. Definition of the city water sample positions



(1) Solenoid coil (a)

(2) Cylindrical capacitor (c)

Fig. 16. Measured values of inductance or capacitance as a function of the frequency, when the city water sample was set to the corresponding positions

6 Conclusion

Experimental confirmation of cylindrical electromagnetic sensor design for liquid detection application was presented. The important tips were summarized as follows:

(1) Introduced approximations of the complete elliptic integral has enough accuracy to design the both sensors, the solenoid coil and cylindrical capacitor. Because of the simplicity and accuracy, they are well acceptable in practice.

(2) Both sensors has ability to distinguish three kinds of liquid: city water, saturated salt water, and salad oil. The solenoid coil can detect the difference in the conductivity from the measured value of the resonance frequency. The cylindrical capacitor can detect the difference in the permittivity from the measured capacitance.

(3) The experimental confirmation reveals two undesirable phenomena in practical use. One is the perturbation of the measured values. Another is the effect of parasitic capacitance.

(4) The measuring condition of the impedance analyzer makes the perturbation of the measured values. When the power source is constant voltage mode with a enough amplitude, it could be suppressed. However, a cable between the sensor and analyzer may also cause the perturbation of the measured values because of inducing the electromagnetic noise.

(5) A twist cable could be suppressed the electromagnetic noise, and own inductance could be negligible. However, the detection abilities of both sensors become low because the parasitic capacitance becomes high compared with a short cable.

(6) The advantage of the use of low-frequency electromagnetic field is discussed. Compared with conventional microscopic or sampling inspections, this macroscopic inspection is very attractive in practice.

Acknowledgements. The author's would like to thank Mr. S. Aoki who studied at our laboratory until 2014 March. Some of experimental data were based on his bachelor's thesis.

References

[1] Abe, T., Shinozawa, Y., Machida, T., Yamada, K.: Device for judging types of liquid in container and control method therefor, US Patent, US7466236B2 (2008)

[2] Misron, N., Harun, N.H., Lee, Y.K., Sidek, R.M., Aris, I., Wakiwaka, H., Tashiro, K.: Improvement in sensitivity of an Inductive oil palm fruit sensor. Sensors 14, 2431–2448 (2014)

[3] MartineZ, S., Munoz-Pascual, F.-X., Baldrich, E.: detection of sample conductivity and bacterial presence using inductive microcoils. Effect of device size and geometry. Sensors and Actuators B: Chemical 181, 816–822 (2013)

[4] Zia, M.I., Rahman, M.S.A., Mukhopadhyay, S.C., Yu, P., Al-Bahadly, I.H., Goicoechea, J., Gooneratne, C.P., Kosel, J., Liao, T.: Technique for rapid detection of phthalates in water and beverages. Journal of Food Engineering 116, 515–523 (2013)

[5] Korostynska, O., Mason, A., Al-Shamma'a, A.I.: Proof-of-concept microwave sensor on flexible substrate for real-time water composition analysis. In: 2012 Sixth International Conference on Sensing Technology (ICST 2012), pp. 547–550 (2012)

[6] Yunus, M.A.M., Mukhopadhyay, S.C.: Development of planar electromagnetic sensors for measurement and monitoring of environmental parameters. Meas. Sci. Technol. 22, 025107(9p.) (2011)

[7] Yunus, M.A.M., Mukhopadhyay, S.C.: Novel planar electromagnetic sensors for detection of nitrates and contamination in natural water sources. IEEE Sensors Journal 11, 1440–1447 (2011)

[8] Yunus, M.A.M., Mukhopadhyay, S.C.: A low-cost sensing system for quality monitoring of dairy products. IEEE Transactions on Instrumentatuin and Measurement 55, 1331–1338 (2006)

[9] Wylie, S.R., Shaw, A., Al-Shamma'a, A.I.: RF sensor for multiphase flowmeasurement through an oil pipeline. Measurement Science & Technology 17(8), 2141–2149 (2006)

[10] Rahman, M.S.A., Mukhopadhyay, S.C., Yu, P., Goicoechea, J., Matias, I.R., Gooneratne, C.P., Kosel, J.: Detection of bacterial endotoxin in food: New planar interdigital sensors based approach. Journal of Food Engineering 116, 515–523 (2013)

[11] National Astronomical Observatory of Japan. Chronological science tables. Maruzen Publishing Co. Ltd. (2005) (in Japanese)

[12] Hastings, H.: Approximations for digital computers (Sheet No. 46 and 49), Princeton (1955) (This information referred to a Japanese book: Moriguchi, S., Udagawa, K., Hitomatsu, S.: IWANAMI SUUGAKU KOUSHIKI, 22nd edn., vol. III, pp. 79–81. Iwanami Publishing (2010)

[13] Nagaoka, H.: The inductance coefficients of solenoids. Journal of the College of Science XXVII, Article 6 (1909), `http://repository.dl.itc.u-tokyo.ac.jp/dspace/bitstream/2261/32855/1/jcs027006.pdf`

[14] Omoto, Y., Morita, K.: On the electrode capacity of magnetron tube. Journal of IEEJ 60(619), 61–62 (1940), `http://dx.doi.org/10.11526/ieejjournal1888.60.61` (in Japanese)

[15] Kogo, H.: Electrode Capacity of Split-Coaxial Cylinder. Research Reports of Faculty of Technology 6(10), 54–67 (1955) (in Japanese)
[16] Oka, S.: Theory of dielectric, pp. 191–195. Gendaikougakusha Co. Ltd. (1977) (in Japanese)
[17] Ichijo, B., Arai, T.: A new method of measuring dielectric property of very high loss material at high frequency. IEEE Trans. Instrum. Means. IM-19, 73–77 (1970)
[18] Lattey, R.T.: Dielectric constants of electrolytic solutions. Philosophical Magazine and Journal of science 41(246), 829–848 (1921)

An Ultra-Low Power Miniaturised Wireless Mote for Ubiquitous Data Acquisition

A. Ivoghlian, K.I-K. Wang*, Z. Salcic, and S.A. Catapang

Department of Electrical and Computer Engineering, The University of Auckland,
38 Princes St., Auckland, New Zealand
kevin.wang@auckland.ac.nz

Abstract. The advancement of embedded sensing and wireless communication technologies facilitate the possibility of continuous monitoring of our activities and surrounding environment. This leads towards many practical applications such as environmental monitoring, structural health monitoring, wearable and context aware computing which could not be easily done in the past. However, design of wireless sensor nodes and networks for such applications need to follow stringent application requirements to ensure precise measurement, long operating life, reliability under varying thermal and impact conditions, and reasonable communication range. In this paper, a low power miniaturised wireless sensor node (or mote) is designed targeting ubiquitous data acquisition applications. The mote has a modular architecture to meet different application requirements. It is built using ultra low power miniaturised components with a novel power regulation method for achieving long operating life with the overall physical dimension of 30mm x 16mm. It works at sub-1GHz frequency range to allow better obstacle penetration and communication range of up to 1.7km. The mote is planned to be deployed in various applications including earthquake simulation experiments for measuring soil movement and wearable body area networks for vital signs monitoring.

Keywords: iniaturised mote, power efficient mote, AWSAM, environmental and structural health monitoring, pervasive healthcare.

1 Introduction

Accurate data acquisition from distributed information sources is a challenge that continues to hinder both discovery and practical applications of scientific knowledge. However, the advancement of embedded and wireless technologies has enabled new ways of acquiring information to realise applications which were once infeasible. Particular examples of such applications include, large scale environmental/ geographical sensing [1, 2], structural monitoring [3] and pervasive health monitoring [4]. Within this context, wireless sensor nodes have been instrumental in overcoming the barriers of ubiquitous data acquisition. Though, current research has demonstrated

---

[*] Corresponding author.

A. Mason et al. (eds.), *Sensing Technology: Current Status and Future Trends IV*,
Smart Sensors, Measurement and Instrumentation 12, DOI: 10.1007/978-3-319-12898-6_8

successful deployment of sensor nodes in numerous applications, a number of challenges remain. Of particular interest are the challenges introduced by requirements of ubiquitous wireless sensing applications such as, miniaturisation, flexibility, low power consumption and achieving adequate transmission range.

The primary aim of this research is to produce a node which meets the requirements of ubiquitous sensing applications with more generalised requirements as introduced in the previous paragraph. This is in contrast to application specific designs where nodes are only expected to fulfil requirements for specific circumstances. With this objective in mind, a new mote, AWSAM-3 is proposed which possesses a modular layered hardware architecture. Each layer is an independent PCB with components necessary for its particular application. For example, the core processing board contains power conditioning circuits, a CC430 MCU, flash memory and a radio section. Layers are unified via stacked connectors which allow different boards to be swapped and connected according to application requirements.

As sensor nodes aim to acquire information from their direct surroundings, physical disturbance to the sensing environment due to deployment should be minimised. It is then, desirable to shrink the physical footprint of nodes. Miniaturisation of nodes will allow seamless embedding into their target environment. AWSAM-3 possesses a 30mm x 16mm footprint which was achieved through selection of components in small packages and multi-layer PCB design.

Implementations of ubiquitous sensing systems also possess different requirements within and between applications. Use of nodes with rigid capabilities may lead to cost inefficiency and limited use cases as reuse for other applications is unlikely even between similar scenarios. Versatility to adapt to specific needs through flexible hardware configuration thus becomes vital. Due to its modular board stack, AWSAM-3 can be adapted to multiple applications by swapping modules. An example configuration presented in this Chapter includes a core processing board, antenna, inertial sensor board and power board.

Many ubiquitous sensing applications will require that nodes operate as autonomous data acquisition devices. However, limited battery life of nodes will negatively impact the effectiveness of continuous sensing applications. It follows that designing nodes with low power consumption is necessary for successful deployment of continuous sensing systems. Components used in the AWSAM-3 boards were all carefully selected with low power consumption in mind. Characteristics such as quiescent current draw, maximum power draw and low power modes were considered in the selection of sensor ICs, memory, power conditioning components and MCU.

Long communication range is desirable in sensor nodes to enable various deployment schemes and network architectures. This is especially important for applications which cover geographically large areas. Unlike many existing nodes AWSAM-3 employs a sub-1GHz radio which achieves a longer communication range, at the same power level, in comparison to 2.4GHz radios frequently used in sensor nodes.

A further aim of this research is to present an evaluation of the AWSAM-3 platform with regards to the criteria previously mentioned. Power consumption, communication range and component power profiling tests were performed to evaluate general performance. The power consumption evaluation results show the advantage of using a switching DC-DC voltage regulator against the common linear regulator approach used in most existing mote designs. The flash memory and radio communication power profiling also provide invaluable information in designing power efficient software and protocol stacks. Indoor and outdoor line-of-sight experiments were conducted to evaluate maximum communication range of the AWSAM-3. Using a quarter wave Linx Splatch antenna, AWSAM-3 is able to achieve up to 1.7 kilometers communication range at the data rate of 1.2kbps.

The rest of the Chapter is as follows. Section 2 presents the targeted mote design requirements, followed by the proposed hardware architecture in section 3. Section 4 and 5 illustrates and discusses the implemented mote performance in terms of power consumption and communication range. Section 6 gives a brief review of some existing mote designs with various processing and interface capabilities and compares them with the AWSAM-3. Section 7 draws conclusions of the Chapter with future perspectives presented in section 8.

2 Mote Design Requirements

As will later be discussed in section 6, most mote designs are adequate platforms on which WSN research can be carried out. However, most mote designs generally do not target in-field deployment and consequently, are not optimised for it. This can be seen in the choice of form factor and power supply. To achieve a balance between performance and flexibility, AWSAM-3 was designed with the requirements discussed in the following subsections.

2.1 Modularity

The AWSAM-3 adopts a modular architecture. Essentially, AWSAM-3 is designed with modular blocks that can be integrated in different ways for various applications. For wireless network prototyping purposes, the AWSAM-3 may simply consist of a radio transceiver, processor, and a large capacity power source. In other cases, such as in environmental monitoring applications, the power source may be minimised to achieve small physical size and a group of inertial sensors added to the mote. This modular architecture also provides the opportunity for new hardware to be incorporated into the AWSAM-3 well after the initial manufacturing run, extending mote capabilities to previously unplanned applications.

2.2 Power Consumption

The importance of mote battery life varies with application. For example, battery life may not be of great concern in lab experiments, but in the case of structural health monitoring, it is a key performance metric. As a flexible mote platform, the AWSAM-3 hardware is intended to be optimised for minimal power consumption without sacrificing vital functionality or performance. The design aims to remain flexible so as to allow the mote to perform well across a range of power profiles. In the research presented, particular focus was given to power consumption reduction during hardware design. Both the active and sleep mode power consumption was evaluated as either one can be severely detrimental to battery life depending on the application. At the hardware level, it is important to choose components that have an ultra-low sleep current consumption. In applications where the AWSAM-3 sleeps for the majority of the time, sleep current consumption becomes the limiting factor on mote lifetime.

2.3 Communication Range

It is desirable that the range of the AWSAM-3 be maximised in order to increase the feasibility of and reliability in existing applications. This also opens the door to new applications that require long range communications (e.g. agricultural applications). To achieve this aim, expansion of support to include other radio frequencies may be needed.

2.4 Miniaturisation

It is important for the mote to be physically small and light such that it will not affect or contribute to the sensor measurements. It follows then that this should be one of the design goals of the AWSAM-3. Miniaturisation can be applied to every aspect of the hardware design but especially so to components which occupy significant PCB real estate. The battery and the antenna front end often dictate the overall dimensions of the mote, as seen with many existing mote platforms. Some typical mote designs reviewed in section 6 are good examples of demonstrating this limitation. However, it must be noted that reduction in battery size causes a tradeoff with operating lifetime. Therefore, the requirement of miniaturisation further emphasises the importance of power efficiency of the mote design.

3 AWSAM-3 Hardware Architecture

As discussed in section 2, the goal of achieving a flexible design is realised through the modular partitioning of the mote elements. There are many ways that the entire system can be divided up. One logical way of decomposing the mote functionality, is to group components based on their dependence on each other and whether they are likely to be used together in most foreseeable applications. The general principal that has been followed is that any single module can be modified without requiring modification of any other module. Also, there should be no restriction on how modules can be modified, provided that prescribed interface rules are followed.

Each individual module can be stacked on top of each other, forming a layered architecture. This approach takes advantage of the often unused vertical space surrounding a mote and seeks to reduce the horizontal dimensions. The connection between each layer is accomplished through three sets of connectors. The connectors are all ultra-miniature, low profile, 20 pin header-receptacle sets. One set is used exclusively for the ADC inputs on the microcontroller in addition to analog VDD and GND levels. The other two sets are used for power lines, serial interfaces (i.e. I^2C, SPI and UART), and digital IO. Each header-receptacle set sit parallel to each other on two sides of the PCB with matching signals electrically connected through vias. This arrangement allows the stacking order, or even the presence of the different mote layers to be changed with maximum flexibility. The placement of the connectors on the PCB also enforces a polarity so that the PCBs can only be connected in one way. Excluding the connectors, each PCB should be populated with components only on the top side. This allows complete flexibility of component placement without considering the boards above or below. There may be exceptions to this rule on the uppermost and lowermost layers. In keeping with the modularity principle, each board has its own power source regulation to remove dependency on any other PCB layer in the stack. This does not rule out the passing of power supply outputs between the boards so that the IO interfaces can be matched.

Fig. 1. AWSAM-3 featuring antenna, core, and power, beside a Li-Po battery and NZ50c coin

The AWSAM-3 accentuates the multi-layered mote architecture. As a result, its form factor is somewhat different from the motes reviewed in section 6. By utilising vertical space, the 2D footprint is reduced to 30mm x 16mm (PCB only). The simplest AWSAM-3 node can be formed with a minimum of four layers from top to bottom, namely Antenna, Core, Sensors, Power. A battery can also be used and can be added to the stack or placed beside it depending on the enclosure used and any critical dimensions. Fig. 1 shows this arrangement with the battery detached, next to an NZ50c coin.

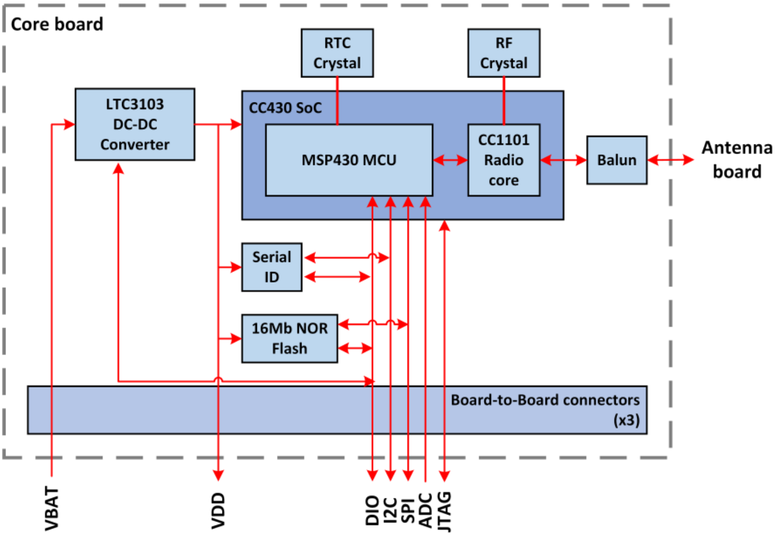


Fig. 2. Core board architecture

3.1 Core Board

As the name suggests, the Core board is the centrepiece of the AWSAM-3. It contains most of the components that typically define a mote. Fig. 2 illustrates the overall architecture of the Core board. The heart of the AWSAM-3 sensor node is the CC430 SoC from Texas Instruments. This IC is comprised of a microcontroller and a radio transceiver in a single package. The microcontroller is based on the well-known and popular MSP430 architecture and the transceiver is based on, and is very similar to, the CC1101 sub-1GHz radio IC. The 16-bit CC430 offers great interfacing capabilities including an 8-channel 12-bit ADC, two 16-bit timers, a hardware Real-Time Clock (RTC), SPI, I^2C, UART and 44 GPIO pins. The other important feature offered by CC430 is its ultra-low power consumption. It operates with just 160μA/MHz in active mode, and 2.0μA in standby mode (RTC running) at a cost of limited memory capacity of 4KB of RAM and 32KB of Flash.

The radio transceiver on the CC430 is based on the CC1101 radio IC from Texas Instruments. It supports four frequency bands below 1GHz (315MHz, 433MHz, 868MHz, & 915MHz), a maximum symbol rate of 500 kbaud (at higher frequencies), and dynamically adjustable output power exceeding 10 dBm. In contrast to higher frequency radio (i.e. 2.4GHz in many existing motes), lower frequency radio (such as in the sub-1GHz ISM bands) can, under the right conditions, provide greater communication range than higher frequency at the same output power levels. This is in line with the design requirements of longer communication range and lower power consumption. One pitfall however is the relationship between radio frequency and antenna size. The inherent range gains that can be obtained with lower frequency communication can be cancelled out by the loss of efficiency when the antenna is shrunk below a quarter wavelength in order to achieve the miniaturisation requirement. Because of the wide range of applications for the AWSAM-3, and the possibility of attaching different antennas for different applications, a sub-1GHz radio was settled on to provide flexibility. While it is useful to evaluate the microcontroller and radio individually, there are also advantages that the SoC structure offers. Because the size of the die is often much smaller than the package (package size being determined by the number of pins), combining two IC's inside a single package can save a fair bit of space. This is especially true in regard to the communication channel between the ICs. SPI pins, power pins, track routing all add to overall PCB footprint. This is in line with the goal of miniaturisation.

It is common these days to find that ICs operate around 3V. They often support a voltage up to 3.6V and some can operate down to well below 2V. When using AA or lithium coin cells batteries, most components can be connected directly to the battery. As one of the primary power sources, AWSAM-3 is powered from a Li-Po battery which offers larger capacity and discharge rate. However, the cell voltage on this type of battery is too high for a direct connection and so a voltage regulator needs to be used. Linear regulators are common in research motes even when low power consumption is being presented as a feature. With the AWSAM-3, the idea is to reduce the power consumption as much as possible. The losses associated with a linear regulator at high currents really undermine the low power goal, whereas a switching buck converter is more suitable for this purpose. Also, some applications of the AWSAM-3 involve very low current draw (below 10μA) for well over 99% of its process cycle, a device with low quiescent current is vital. There are some devices that compete with the linear regulators in this regard, but they introduce a lot of noise into the system due to the duty cycled nature of the circuit. The voltage regulator on the core board supplies all the components on this layer and so any noise on this rail will affect the radio communication. Even when communication can be established successfully, noise can cause packet loss as a result of signal degradation. This results in greater average transmit time and hence, greater power consumption. To address this, the DC-DC converter used must support a low-noise mode. It is essential that the low-noise feature be offered as a mode becauselow noise operations often draw considerably higher quiescent currents which would have severe impact on the AWSAM-3's lifetime. The LTC3103 DC-DC converter from Linear Technology is well suited to regulating the Core board voltage. It optimises the power consumption of the board while still satisfying the low noise requirements of the radio transceiver. The DC-DC converter supports two operating modes called "Automatic Burst Mode" and "Forced Continuous Operation". These modes refer to high efficiency and low noise respectively and can be controlled with a digital signal line. On the Core board, this signal is connected to a digital IO pin on the CC430.

Storing sensor data inside external memory can be considered application dependent but this is not the only reason for including it on a mote. External memory can be used for network routing tables, or as a buffer for radio links that are not always available. Wireless communication is an elemental part of the AWSAM-3 and so it follows that some additional storage be included on the Core board.

Despite the superior power consumption of NAND flash memory, the typically parallel interface poses some obstacles to implementation. The number of pins on the IC will be relatively high due to the parallel nature of the interface, resulting in a physically large IC. This hinders the miniaturisation objective of the AWSAM-3. In addition to this, the parallel interface consumes more valuable IO pins on the CC430 and complicates the software drivers. NOR based Flash memory overcomes these obstacles by way of its serial interface. Communication with these IC's is usually done via a standard SPI bus which simplifies interfacing with MCU's like the one inside the CC430. Program, erase, read, and device control are all achieved over a minimal number of signals and hence IC pins. NOR based Flash does however come at a cost of significantly increased power consumption.

The M25PX16 memory IC from Micron Technology is a 16Mb NOR based Flash IC with an SPI interface. It draws more current than similar serial NOR memory readily available in New Zealand but its redeeming quality is its low operating voltage. The M25PX16 can operate from voltages as low as 2.3V. Factoring in a buffer for power supply fluctuations, the Flash memory ICs can operate from 2.4V and 2.8V respectively. As mentioned previously, the entire Core board is supplied through a single voltage regulator. A second DC-DC converter cannot be added without significantly increasing the board size. Therefore, the minimum operating voltage of the Flash IC dictates the minimum input voltage of the other components on the Core board, including the radio transceiver in the CC430. Minimising the supply voltage to digital IC's is an essential step in reducing overall power consumption for the AWSAM-3. The radio is an essential feature of mote design and also one of the largest current consumer and so minimising its supply voltage is of the highest priority.

3.2 Antenna

As mentioned in section 3.1, the radio front end on the Core board has been made flexible so that the AWSAM-3 can operate at a range of frequencies. The 915MHz and 433MHz bands have initially been targeted and suitable antennas chosen accordingly. At sub-1GHz frequencies, antenna efficiency becomes an issue when the size is constrained. In order to achieve better communication range and more flexible RF design, the antenna on the AWSAM-3 does not reside on the same PCB as the radio transceiver. AWSAM-3 consists of a dedicated layer for the antenna. This is in line with the modularity principle surrounding many of the AWSAM-3 design decisions but also provides size advantages. Many common PCB antennas (Chip, helical, PCB trace, etc.) require that there be no copper or components above or below them. For a stacked design like the AWSAM-3, this creates an expanded horizontal dimension that no other layer above or below the Core board can utilise. By selecting antennas that can be placed on top of the AWSAM-3 layer stack, the horizontal dimensions can be minimised. It is expected that future antenna designs for the AWSAM-3 will follow the same practice.

The Splatch antennas from Linx Technologies are embedded quarter wave antennas for use in compact designs. The miniature antenna is a multilayer PCB that measures 27.94mm x 13.7mm x 1.5mm and can be soldered directly onto the underlying board. The Splatch antennas that have been chosen for the initial testing are tuned to 434MHz and 916MHz. The Splatch antenna board connects with the core board via a UMC-UMC cable. The Splatch antennas fit into the AWSAM-3 horizontal dimension (30mm x 16mm) on their own but leave very little room for a ground plane extending away

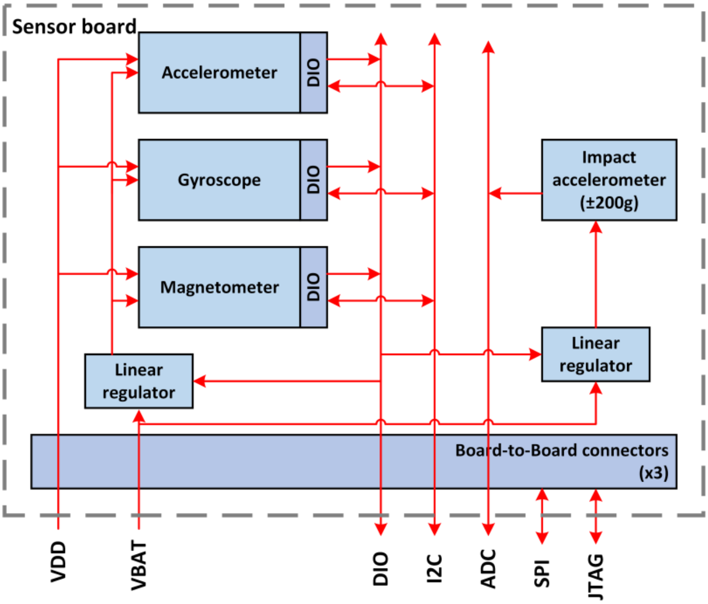


Fig. 3. Inertial sensor board architecture

from the antenna. This ground plane, or 'counterpoise' is suggested by the reference design given by Linx to realise the full potential of the antenna. Reducing ground plane size below the reference will shift the centre frequency higher and cause sub-optimal communication performance. Despite this, the antennas still perform reasonably well as demonstrated in the evaluation discussed in section 5.

3.3 Inertial Sensor Board

The AWSAM-3 offers a platform on which application specific sensor boards can be easily developed and deployed. To demonstrate the capabilities of the AWSAM-3 and validate its architecture, an inertial sensor board has been designed and manufactured for various orientation and motion sensing applications. Referring to Fig. 3, this board contains an accelerometer, a gyroscope, a magnetometer and an analog impact accelerometer which are introduced in this section.

The MMA8451Q is a three-axis capacitive accelerometer from Freescale Semiconductor. It supports dynamic ranges of ±2g, ±4g, and ±8g on each axis with 14-bits of resolution and a range of sampling rates up to 800Hz. It offers a 32 sample FIFO buffer which can be used for holding raw sensor data or for on-chip high-pass filtering. Communication with a microcontroller is achieved via an I^2C interface. Two programmable interrupt lines are also available. The MMA8451Q is positioned as a low power device which is part of what makes it attractive for use on the mote. Its current consumption while active ranges from 6μA to 165μA depending on sample rate and device power mode. Even more attractive is its low standby mode current which is typically 1.8μA. This fittingly supports architectures where the mote is asleep for long periods of time. Furthermore, the programmable interrupts can allow the rest of the system to sleep while the accelerometer is sampling, only waking the MCU when the FIFO is full.

Past experience has revealed situations where the accelerometer data alone was not enough to accurately describe the movement of the mote. When stationary, acceleration due to gravity can be used to determine the orientation of the mote in 2 dimensions (roll and pitch) using Pythagoras's theorem. However, when the mote is moving, changes in orientation cannot be positively determined. Furthermore, orientation in the horizontal plane parallel to earth (yaw) cannot be determined from acceleration, stationary or not. Therefore, a gyroscope was added to address this issue. The L3G4200D is a three axis angular rate sensor from STMicroelectronics. It supports ranges of 250, 500, or 2000 degrees-per-second (dps) with 16-bits of resolution and a range of sampling rates up to 800Hz. Like the accelerometer it is interfaced with the MCU via I^2C and two interrupt lines. Also like the accelerometer, it offers a FIFO with the ability to high-pass filter the on the chip. However, unlike the accelerometer, the gyroscope can apply a low-pass filter to the sensed data as well. In applications where smooth or low rate motion is expected, a low pass filter can further improve the accuracy of the sensor data by reducing high frequency noise.

The current consumption of the L3G4200D stands out from the rest of the ICs on the inertial sensor board. Unlike the other sensors which consume in the hundreds of microamperes when active, the L3G4200D typically consumes 6.1mA. This relatively high current consumption is not specific to the L3G4200D, rather it is a common characteristic of MEMS gyroscopes and as such is unavoidable in this design. It is up to the software developer to be aware of this and factor the power consumption in accordingly. Fortunately, in power down mode, the L3G4200D typically consumes only 5µA.

The L3G4200D provides angular rate and so to determine the angular displacement, the raw data must be integrated with respect to time. As with all dead reckoning approaches, this method is susceptible to accumulated errors. Electrical or mechanical noise or even the discrete time nature of digital signals will introduce inaccuracies in measurement that will result in a growing difference between the reported and actual angular position of the mote. In some cases it is desirable that the orientation or displacement data be in reference to an absolute reference position. Neither the L3G4200D gyroscope nor the MMA8451Q accelerometer can provide this information in three dimensions and thus a magnetometer is introduced for this purpose.

The LSM303DLHC is an e-compass MEMS module featuring a 3-axis accelerometer and a 3-axis magnetometer in a single package. Each sensor is individually addressable and can be operated independently of the other. This is particularly useful for leveraging the low power properties of the individual sensors. The accelerometer can be powered down while the magnetometer is active and vice-versa as the application demands. The entire module typically consumes 1µA of current when powered down. Like the MMA8451Q and L3G4200D, the LSM303DLHC communicates with the MCU over I^2C and provides two interrupt lines.

The 3-axis magnetometer can be configured for full scale ranges from ±1.3 to ±8.1 gauss with 16-bits of resolution and an output data rate of 30 Hz. The 3-axis accelerometer supports full scale ranges up to ±16g with 16-bits of resolution and an output data rate of over 5 kHz. These three specifications are all better than the MMA8451Q 3-axis accelerometer also on this board, however the accelerometer on the LSM303DLHC suffers from greater acceleration noise density and so it does not completely replace the MMA8451Q.

The ADXL377 is a 3-axis linear accelerator with a typical full scale range of ±200g, which allows impact acceleration to be monitored. Unlike the other sensors on the inertial sensor board, the ADXL377 does not provide a digital interface and acceleration values must be read from a set out output voltages using an analog to digital converter (ADC). The MCU in the CC430F6137 contains an 8-channel, 12-bit ADC which can be used for this purpose. The ADXL377 typically draws 300µA when powered up. Because of the lack of a digital interface, shutting down this IC is not as straight forward as it is with the other sensors on this board. 300µA, while comparable to other sensors and acceptable when active, is far higher than the sleep currents of other components. To address this, the LDO supplying the ADXL377 can be switched off, effectively shutting down the sensor. This reduces the current drawn by the LDO and ADXL3XX to 40nA.

In order to keep this sensor board modular, independent power source regulation has been implemented as part of the circuit. This particular sensor board contains three digital sensors and one analog sensor. None of the digital sensors are tolerant of voltages exceeding 3.6V so powering them directly from the battery is not an option. The supply voltage ranges of the digital sensors vary but overlap for the greater portion. As such, a single voltage regulator has been used. Similar to the radio, keeping power supply voltage ripple to the minimum is vital to get the most accurate readings possible out of the sensors. To achieve this, the DC-DC converter must be put into a low-noise mode, which increases the quiescent current by three orders of magnitude. The necessity of operating under low noise mode during continuous sensing consumes the advantage of using the DC-DC converter and has made the linear regulator a more favourable option. All three digital sensors on this board support a separate supply for the digital interface. This feature is what allows the use of a 2.5V low-noise linear regulator while interfacing with the MCU at 2.4V. The digital supplies on this board are connected to the output of the DC-DC converter on the Core board.

The ADXL3XX outputs acceleration values as analog voltages and its accuracy is therefore very dependent on the external analog to digital conversion. A separate low-dropout (LDO) linear regulator is used exclusively for this sensor to avoid noise from the other sensors on this board from coupling back into the positive supply rail. A 1.9V LDO has been chosen to match the 2V internal reference voltage of the MCU. This voltage allows the full range of the accelerometer to be sensed with the greatest resolution without clipping at the upper end. To further increase acquisition accuracy, the regulator output voltage has been connected to an ADC input channel to allow compensation for regulator voltage drifting. This particular LDO also has a high power supply rejection ratio (PSRR) of 68 dB at 1 kHz, to keep the supply voltage to the ADXL3XX as clean as possible.

4 Power Consumption Evaluation

Carrying out a practical experiment is vital for measuring the real life power consumption and estimating lifetime of the mote. Simplistic models based solely on datasheet specifications don't capture the true behaviour of the circuit and often overestimate the performance. One example of the oversimplification of a power consumption model is assuming a linear relationship between active/sleep mode duty cycling and average power consumed. Practical experiments on the AWSAM-3 seek to take into account often overlooked factors and produce an accurate power profile from which energy aware software can be developed.

Application of the AWSAM-3 may vary so to capture all possible use cases it is important that the power profile be formatted into a model that can be fit to the requirements. The modular architecture of the AWSAM-3 allows flexible and modular power profile to be developed as well. This section presents important models for the core module. A similar methodology should be applied to any boards developed for specific applications.

4.1 Core Board Regulator

The Core board in combination with an antenna and power represent the simplest form that the AWSAM-3 can take. In this arrangement, the LTC3103 is the sole voltage regulator and so a model of this regulator is indispensable. In order to accurately control the load, the LTC3103 and battery have been isolated from the rest of the circuit. A resistance has been chosen as an appropriate representation of the Core board load. By varying this resistive load (R_L), a model for power consumption versus current draw can be obtained. A shunt resistor R_S in series with the battery provides a way to measure the current drawn by the circuit. In order to reduce the effect of R_S on the measurement, it should be minimised as much as practical. A nominal resistance of 1Ω has been chosen as a balance between minimisation, availability, and dominance over PCB track impedances. More specifically, 1Ω is low enough to dissipate less than 2% of the total power used at the highest Core board current levels, while still being large enough to measure accurately on a PCB. 1Ω is the nominal impedance, however due to variances between parts and soldering quality, the impedance of the physical resistor is measured using a 4-wire measurement process on the multi-meter. This approach involves accurately pumping current through the resistor with and measuring the voltage across using a separate set of leads. This technique is particularly applicable to small resistances as it minimises the effect of multi-meter leads on the reading. R_S has been measured to be 1.110Ω for the conducted experiment.

The Core board's total current draw from its on-board regulator (the LTC3103) can range from as little as 7μA up to a maximum of 50mA (albeit in an irrational state). R_L is stepped up in decades from 1μA up to 100mA. As the output voltage of the LTC3103 is programmed to 2.4V (following the minimum IC voltage requirement of the FLASH), this can be achieved using resistors between 24Ω and 2.4MΩ. As the battery voltage can vary anywhere between 2.75V and 4.2V the experiment will need to take this into account as a variable. This experiment is carried out at five discrete voltages between 2.8V and 4.2V. One of the features of the LTC3103 which is utilised with the AWSAM-3 is the option of two different operating modes. These modes allow a trade-off between power consumption and output voltage ripple (noise). Because both modes are used practically, the experiment will have to test each of them.

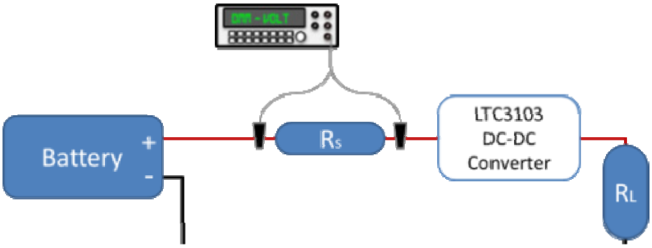


Fig. 4. Core board regulator (LTC3103) experiment set-up

Table 1. LTC3103 regulator input current versus output current in Low Power Mode

Load	**Input Current (mA)**				
V_{IN}	**2.8V**	**3.0V**	**3.3V**	**3.7V**	**4.2V**
1µA	0.006306	0.009009	0.006757	0.005583	0.004623
10µA	0.014414	0.014414	0.010811	0.008933	0.007397
100µA	0.098198	0.092793	0.082883	0.072092	0.064185
1mA	0.938739	0.88018	0.802703	0.698198	0.621622
10mA	9.354955	8.648649	8.302703	6.905405	6.136036
100mA	89.92171	85.13514	81.08108	70.21622	62.24324

Table 2. LTC3103regulator input current versus output current in Low Noise Mode

Load	**Input Current (mA)**				
V_{IN}	**2.8V**	**3.0V**	**3.3V**	**3.7V**	**4.2V**
1µA	1.262162	1.469369	1.696396	2.042342	2.362162
10µA	1.269369	1.473874	1.701802	2.048649	2.384685
100µA	1.345045	1.544144	1.766667	2.10991	2.418018
1mA	2.102703	2.243243	2.414414	2.681982	2.926126
10mA	9.763063	9.504505	9.459459	8.435135	8.024324
100mA	89.88288	85.58559	81.08108	70.27027	62.31532

The test is carried out by setting the battery voltage at a desired level and changing the load resistance R_L while measuring the voltage across the shunt resistor R_S. This measurement is taken using the same multi-meter used for the 4-wire measurement described earlier. It is the Agilent 34410A digital multi-meter. A diagram of the experiment set-up is shown in Fig. 4. Table 1 and Table 2 show the input current flowing into the Core board regulator at different input voltages for low power mode and low noise mode respectively. The current is calculated by taking the voltage measured across R_S and dividing it by 1.11Ω. The load is represented as the current flowing through the load resistor, R_L, stepped up in decades from 1µA up to 100mA.

For an ideal linear regulator with zero quiescent current, the efficiency would be 86%, 80%, 73%, 65%, and 57% for input voltages of 2.8V, 3.0V, 3.3V, 3.7V, and 4.2V respectively. A linear regulator is logically most efficient at the lowest input voltage of 2.8V. A simplified benchmark of performance would be to say that, so long as the Core board regulator's efficiency exceeds 86% with an input of 2.8V, the decision to use a DC-DC is valid on the grounds of power consumption.

Fig. 5 and 6 are plots of efficiency versus load. Input power is calculated by multiplying the input current into the Core board regulator by the input voltage (shown as V_{IN} on Table 1 and 2). Output power is measured by multiplying the load current by the fixed output voltage of the Core board regulator (2.4V).

It can be seen from Fig. 6 that the efficiency in low noise mode exceeds 86% for loads of 8.2mA or more (with an input voltage of 2.8V). As mentioned in section 3, the low noise mode of the DC-DC converter will primarily be used when operating the radio so that voltage ripple does not interfere with communications. The lowest active current that will be drawn by the radio transceiver is 15mA which occurs when it is in receive mode. At this point, the efficiency of the low noise mode is 92%.

When the radio transceiver is not being operated, the Core board regulator will, in most circumstances, be operating in low power mode. In this mode, the ideal linear regulator efficiency of 86% at 2.8V will be surpassed when the load exceeds 78μA. When the CC430 is carrying out some data processing, or when the Flash memory is being accessed, etc., this efficiency of the Core board regulator will be well above 86%. However, when the Core board is in its lowest power sleep state, the load is closer to 10μA. At this load the regulator is only 59% efficient at 2.8V. Despite this undesirable result, the effect on overall Core board power consumption will still be minimal. Firstly, 2.8V is not the nominal voltage of the battery. Secondly, the current drawn when the Core board is active and communicating (i.e. under low noise mode) is three orders of magnitude greater than the sleep current and hence has more significant impact on the overall power consumption.

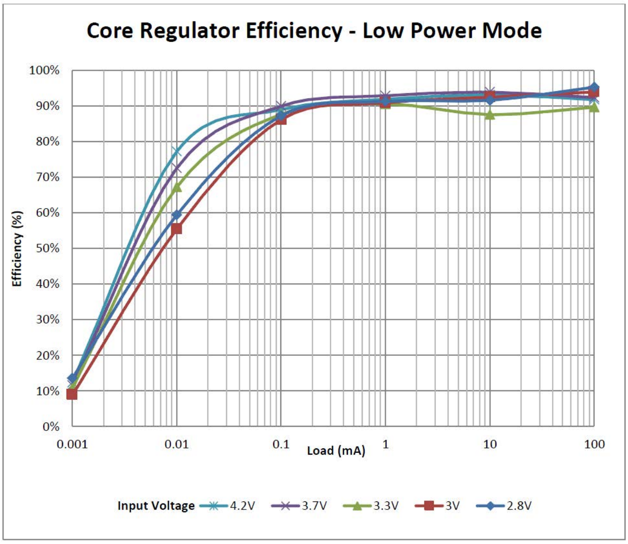


Fig. 5. LTC3103 regulator efficiency plot for Low Power Mode

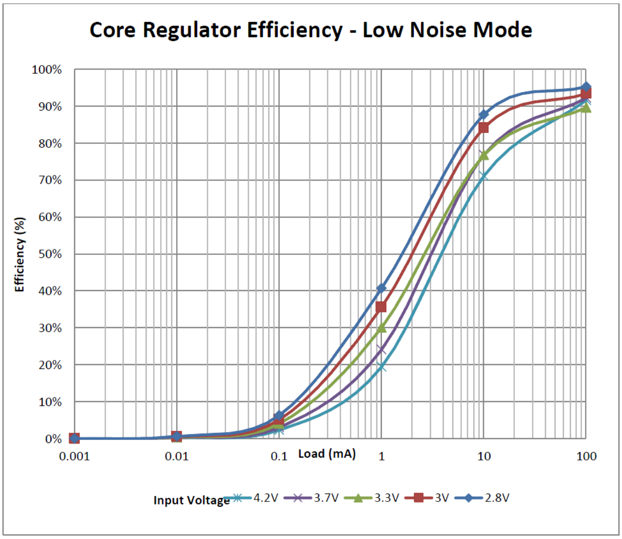


Fig. 6. LTC3103 regulator efficiency plot for Low Noise Mode

This simplified analysis is carried out at the best case scenario for a linear regulator in order to verify that the DC-DC converter will be more efficient across the whole range of input voltages. In truth, the input voltage is typically higher, sitting around the nominal battery voltage of 3.7V. At this point, the ideal linear regulator can only be 65% efficient. Fig. 5 and 6 show clearly how much the DC-DC converter outshines the ideal linear regulator. In low power sleep mode for example, where the Core board draws around 10μA, the DC-DC converter is 77% efficient at 3.7V while the ideal linear regulator is only 65% efficient. This is a particularly powerful figure because linear regulators are traditionally valued for their superior efficiency under light loads.

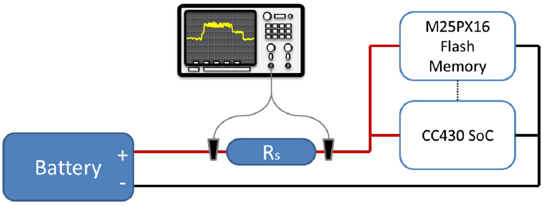


Fig. 7. Flash memory (M25PX16) experiment set-up

4.2 Flash Memory

The question of whether to store or to send has been repeatedly investigated and researched with different conclusions drawn. The hardware and applications in these investigations vary and it is important that practical experiments be undertaken on the AWSAM-3's specific hardware and use cases to determine the value of external storage and when it should be used.

The process for measuring the current drawn by the M25PX16 Flash and CC430 SoC are similar to the methods used with the LTC3103 voltage regulator discussed in the previous section. As such, the test set-up will not be discussed in the same level of depth here. Unlike the experiment carried out in the previous section, storing data to flash or transmitting data wirelessly are not continuous actions. They occur for fixed periods of time which in-turn determine how much energy is used for the action. As such, the results of this experiment will be expressed as energy to provide a more meaningful comparison. With this model the developer can decide when to store data to flash first and when to transmit it immediately.

The datasheets for the M25PX16 and the CC430 provide specifications for the amount of current drawn in any particular mode. They do not however provide any detail regarding start-up, shut-down, and other transitional behaviour. If these factors are not taken into consideration when calculating the energy cost of storing or transmitting, any advantage that external memory could provide may be masked in actual practices. Like the previous experiment with the LTC3103 regulator, the battery voltage is measured at its terminals and the current draw is measured through a shunt resistor R_S in series with the battery, as shown in Fig. 7. In this experiment, an oscilloscope (specifically the Agilent InfiniiVision DSO5012A) is used for measuring the current drawn from the battery. This is required as the limited sampling rate of the multi-meter is not appropriate for measuring the transitional behaviour mentioned above. The core board voltage regulator (LTC3103) is bypassed to provide a more flexible set of results from which algorithms can be developed, independent of any other components. The results from the core board regulator experiment can be used to transform these results into real-life values.

Different operations and power modes are simulated with a simple program executing on the CC430. The current consumption measured also includes the current consumed by the MCU as this represents an actual use case most accurately.

Table 3. Flash memory (M25PX16) current consumption

Operation / Mode	Current Drawn	Time Taken	Data Size	Energy Consumption	Energy Per Byte (nJ)	Power Consumption
Page Program	8.88 mA	0.969 ms	50B	21.97 µJ	439.3 nJ	22.67 mW
	9.19 mA	2.43 ms	100B	57.01 µJ	570.1 nJ	23.46 mW
	9.19 mA	5.16 ms	200B	121.05 µJ	605.3 nJ	23.46 mW
	9.5 mA	6.43 ms	250B	155.99 µJ	624.0 nJ	24.26 mW
Read Data	5.19 mA	-	-			13.25 mW
Bulk Erase	9.63 mA	-	-			24.59 mW
Stand By Mode	13.4 µA	-	-			34.21 µW
Deep Power Down Mode	4.7 µA	-	-			12 µW

For the page program operation, the notion of time comes into play and varies with the number of bytes that are written to the flash. This allows us to measure the task of page programming in terms of energy. To do this, the current drawn (Table 3) is cross referenced with the results of the Core board regulator experiment in section 4.1. The results of these conversions are shown in the last three columns of Table 3. Note that battery voltage is assumed to be 3.7V (the nominal voltage of the Li-Po battery), and the LTC3103 DC-DC converter is in low power mode. The values that haven't been measured against time are also converted to a power representation. The Read Data and Bulk Erase operations are also discrete tasks with notions of time, however measuring the time as accurately as the Page Program operation has proven difficult, and as such, is left out of this experiment.

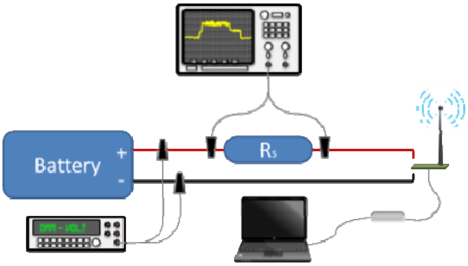


Fig. 8. Radio power consumption test set-up

4.3 Radio

The CC430 offers many adjustable radio transceiver parameters, each with a power consumption trade-off. Carrying out an experimental analysis of different radio modes is necessary for the power and link budget calculations. The CC430F6137 datasheet provides comprehensive figures for the power consumption in any of the radio modes, however like the transmission versus storage experiments, the results need to be analysed on an energy-per-task basis. This exposes the overheads of radio mode transitions and allows different data rates to be compared directly.

The procedure for measuring radio power consumption in these tests is much like the ones in Flash memory power consumption tests. The supply voltage is measured at the battery and the current is acquired by measuring the voltage of the shunt resistor R_S. The power consumption is transformed into energy by measuring the amount of time taken to complete the transmission. In order to capture time with millisecond accuracy, an Agilent InfiniiVision DSO5012A oscilloscope is again used to measure the current through R_S. R_S is fixed at approximately 10Ω. At a maximum radio current draw of 35mA this corresponds to 350mV voltage drop providing an acceptable trade-off between dynamic range on the oscilloscope and voltage drop from the supply. Since the supply voltage does not affect the current draw and as such, the 350mV voltage drop should not affect the results. This has been validated with the use of different supply voltages. The same 4-wire resistance measurement used in the other power consumption tests has been used here and R_S is found to be 9.955Ω. As with the Flash memory power consumption tests, the core board voltage regulator is bypassed during the experiment. SmartRF studio from Texas Instruments is used to control the radio core and initiate data transmission. A diagram of the complete test set-up is shown in Fig. 8.

The radio symbol rate has been set to 1.2 kbaud. This has been done to prolong the transmission of a packet, effectively "stretching" the waveform horizontally while keeping the packet payload size to a minimum. The payload size is set to 1 byte to that the effect of packet overhead can be seen clearly.

Table 4. Radio power consumption test packet format

Preamble	Sync Word	Length Byte	Payload	CRC	Total
4 Bytes	4 Bytes	1 Byte	1 Byte	2 Bytes	12 Bytes

As mentioned in the experiment set-up, the intention with this test is to calculate the amount of energy consumed when transmitting a certain amount of data. Table 4 outlines the format of the packet used for this test, which takes 80ms transmission time at 1.2kbaud. It should be noted at this point that, as there hasn't been any encoding done on the transmitted data, 1.2kbaud is effectively 1.2kbps.

Now, Table 5 lists the results of the experiment at different output power levels and transmission frequencies. Considering that we have an accurate notion of time, we can easily convert these figures into energy. To do this, the input current is first translated into power. This is achieved by cross referencing the current with the results of the Core board regulator experiment in section 4.1. The results of these conversions are shown in the last two columns of Table 5. Note that battery voltage is assumed to be 3.7V (the nominal voltage of the Li-Po battery), and the LTC3103 DC-DC converter is in low noise mode.

Table 5. Current consumption and energy per byte at different power levels and transmission frequencies

TX Frequency	TX Output Power (dBm)	R_s Shunt Voltage (mV)	R_s Shunt Current (mA)	Power From Battery	Energy Per Byte
915 MHz	10	370	37.17	27.71 mW	184.73 μJ
	0	191	19.19	14.64 mW	97.60 μJ
	-6	184	18.48	14.10 mW	94.00 μJ
	-12	153	15.37	12.01 mW	80.07 μJ
433 MHz	10	291	29.23	21.92 mW	146.13 μJ
	0	179	17.98	13.72 mW	91.47 μJ
	-6	181	18.18	13.87 mW	92.47 μJ
	-12	142	14.26	11.15 mW	74.33 μJ

In order to calculate the amount of energy required to transmit an arbitrary amount of data, Eqn (1) could be used.

$$E = (Energy\ Per\ Byet) \times (11 + payload\_size) \tag{1}$$

This experiment was carried out at a low data rate in an effort to elongate the waveforms. In truth, 1.2kbps is a very inefficient data rate to use when low power consumption is desired. The AWSAM-3 supports a maximum data rate of 250kbps which is logically the most power efficient and should be used when long range communication is not required. The consumed energy per byte ad the maximum data rate (250kbps) can be calculated using Eqn. (2). The results presented in Table 6 is computed based by Eqn. (2).

$$E_{250} = E_{1.2} \times \frac{1.2}{250} \tag{2}$$

Table 6. Radio power consumption test energy per byte at 1.2kbps and 250kbps

TX Frequency	TX Output Power	Energy Per Byte, 1.2kbps	Energy Per Byte, 250kbps
915 MHz	10 dBm	184.73 µJ	886.72 nJ
	0 dBm	97.60 µJ	468.48 nJ
	-6 dBm	94.00 µJ	451.2 nJ
	-12 dBm	80.07 µJ	384.32 nJ
433 MHz	10 dBm	146.13 µJ	701.44 nJ
	0 dBm	91.47 µJ	439.04 nJ
	-6 dBm	92.47 µJ	443.84 nJ
	-12 dBm	74.33 µJ	356.8 nJ

It is worth noting that for output power levels below 0dBm, the energy savings are not substantial and even negative as seen at -6dB, 433MHz, despite the significant reduction in output power. It may not be worthwhile to use these modes unless a very limited communication range is required and a very limited amount of energy is available.

By comparing the energy per byte between writing to flash and transmitting over the radio interface (Table 3 and 6 respectively), it can be seen that, in the case of the AWSAM-3, writing to flash may possibly consume more energy than transmitting over the radio (at 250kbps). For packet sizes large enough to make the 11-byte overhead negligible, radio transmission at 0dBm or lower consumes less power than Page Program operations on the flash. This is also not taking into account the Erase and Read Data operations which must logically accompany flash programming, and would certainly drive up the energy cost of using flash even further.

5 Communication Range Evaluation

The applications for which the AWSAM-3 has been developed may have different needs in terms of wireless communication range and so a range of tests would be more appropriate for effectively evaluating the maximum communication distance. These experiments evaluate the radio performance at 915MHz.

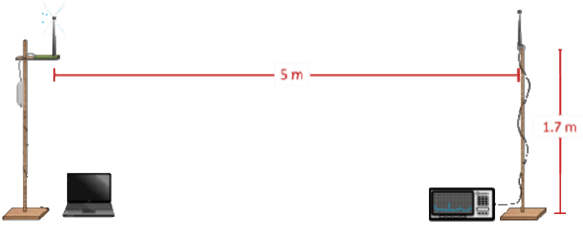


Fig. 9. Spectrum analyser test set-up

5.1 Spectrum Analyser Experiment

The first and simplest communication range experiment carried out on the AWSAM-3 can be done in a lab environment. It involves measuring the transmit power of a mote at a known distance with light of sight (LOS) and projecting it forward using the free-space-path-loss (FSPL) calculations [5]. This coarse experiment applies to LOS situations which are typical of outdoor applications. It serves as a starting point to compare to more realistic experiments involving the maximum communication range in the next subsection.

This experiment is carried out in a lab containing workbenches, workstations, and electronic lab equipment. Referring to Fig. 9, the AWSAM-3 is placed in free space, 1.7m above the floor, 2m from the nearest unrelated object in any direction. It is held in place with a wooden post and attached to a nearby computer using a MSP-FET430UIF debugging tool from Texas Instruments. This is the transmitting end of the test rig. The receiving end of the test rig is composed of a whip antenna attached to a spectrum analyser. A S463AM-915 half-wave whip from Nearson is used. It has a voltage-standing-wave-ratio (VSWR) of 1.9. The antenna is held at 1.7m above the floor to match the height of the transmitter. The cable between the antenna and the spectrum analyser is 0.5m in length.

To control the transmitting side, SmartRF Studio from Texas Instruments is used. This software tool provides the developer direct access to the radio control and data registers with some built-in pre-sets for testing radio communication. In order to get a clear reading on the spectrum analyser, the CC430 is set to transmit an un-modulated carrier instead of a stream of data packets. The spectrum analyser is set to 5kHz bandwidth. This attribute is minimised in order to reduce the noise floor. After setting the transceiver into transmit mode, the average signal strength is measured on the spectrum analyser. These setting are set on the CC430 radio registers through the SmartRF Studio user interface. The test is repeated at different radio output power levels and the results are discussed in this section.

Table 7. Spectrum analyser test results on the received power and RF front-end gain

Nominal Transmitter Output Power	ANT-916-SP Splatch Received Power (dB)	ANT-916-SP Splatch RF Front-End Gain (dB)
10	-56	-20
0	-63	-17
-6	-69	-17
-12	-75	-17

At 5m of separation, the FSPL should be ~46dB for 915MHz calculated using the Friis Transmission Equation [5] shown in Eqn. (3).

$$P_r = P_t + G_t + G_r + 20\,log_{10}\left(\frac{\lambda}{4\pi R}\right) \tag{3}$$

$$FSPL = -20\,log_{10}\left(\frac{\lambda}{4\pi R}\right) = P_r + G_t + G_r - P_t$$

$$FSPL = 45.6\,dB$$

where P_r and P_t is the received power and transmitter power respectively,
G_r and G_t is the receiver gain and transmitter gain respectively,
S_r is the receiver sensitivity, R is the distance between transmitter and receiver,
d is the maximum communication range and λ is the wavelength of carrier

Applying this value to the received power results gathered from the spectrum analyser experiment in the middle column of Table 7 yields the RF front-end gain values presented in the last column. They are obtained by taking the received power value from the test, subtracting the nominal output power, and adding the FSPL (46dB). These values represent the total losses of the RF front end of the transmitter and the receiving antenna, expressed as negative gain.

Taking the following rearrangement of the Friis Transmission Equation (Eqn. (4)), the theoretical LOS range can be projected out by using the following equation

$$d = \lambda \times \frac{1}{4\pi 10^{\frac{S_r - P_t - G_t - G_r}{20}}} \tag{4}$$

The values for RF front-end gain are taken from Table 7. The values for receive sensitivity are taken from the CC430 datasheet [6]. Table 8 lists the theoretical LOS ranges at signalling rates of 250kbaud and 1.2kbaud respectively.

Even with the assumedly low efficiency Splatch antenna (as mentioned in section 3), LOS ranges are predicted to be greater than 1km when the output power is ≥ 0dBm. The radio power consumption experiments in section 4.3 suggest that power output levels below 0dBm are not worth the trade-off in range unless strictly necessary.

Table 8. Projected LOS range based on spectrum analyser test results at 250kbaud

Nominal Transmitter Output Power	**ANT-916-SP Splatch** Theoretical LOS Range at 250kbaud (meters), $S_r = -92$	**ANT-916-SP Splatch** Theoretical LOS Range at 1.2kbaud (metres), $S_r = -109$
10	329	2,326
0	147	1,039
-6	74	521
-12	37	261

5.2 Line-Of-Sight Experiment

The outdoor LOS experiments are similar to the ones carried out using a spectrum analyser in that the effects of multipath signals are insignificant compared to the strength of the direct signal. The justification of these experiments is to verify the accuracy of the spectrum analyser experiments and to provide a benchmark on which further tests can be based. A practical difficulty is locating an unobstructed space large enough to push the AWSAM-3 to its maximum communication range. Fortunately, the city of Auckland has many large parks and towering volcanoes that can be seen for kilometers.

Similar to previous experiments, the Splatch antenna is used to represent a typical mote to base station link. The CC430 offers many configurable radio parameters. Two particular arrangements have been chosen that are applicable to two different but equally valid use cases. One maximises data rate, the other maximises range. In both tests the base station (½ wave whip antenna) is placed in a fixed position while the mote (Splatch antenna) is moved away in a straight line, maintaining a LOS. A PVC pipe is used to elevate the mote 30cm above the human carrier to minimise the effects of the body. The radio link is considered "broken" when the bit-error-rate (BER) exceeds 10%. Two laptop computers running SmartRF Studio are used to facilitate the test.

The experiment has been carried out in clear weather, in free and open air. Under these conditions, GPS readings are typically accurate to within 10 metres. A smartphone based GPS receiver is used to record positions for these tests. The latitude, longitude, and altitude are used later to calculate the approximate range of the radio communications. Eqn. (5) calculates the great circle distance between two points on globe based on the geographical coordinates provided in Table 9. Note that to use this equation, the latitude and longitude values must first be converted to radian representation.

$$d = r \times 2\, arcsin\left(\sqrt{sin^2\left(\frac{\Delta_{Lat}}{2}\right) + cos(Lat_f)\, cos(Lat_m)\, sin^2\left(\frac{\Delta_{Lon}}{2}\right)}\right) \quad (5)$$

where d is the distance, r is the radius of Earth,

Lat_f and Lon_f is the latitude and longitude of the fixed node,

Lat_m and Lon_m isthe latitude and longitude of the mobile node,

Δ_{Lat} and Δ_{Lon} is the absolute difference between latitudes and longitudes

Table 9. LOS test results at 250 and 1.2k baud shown in geographic coordinates of nodes

High Data-Rate - 250kbaud			
	Latitude	**Longitude**	**Altitude**
Fixed Node	36° 51.674' S	174° 46.604' E	61m
Mobile Node	36° 51.734' S	174° 46.489' E	60m
Long Range - 1.2kbaud			
Fixed Node	36° 52.690' S	174° 45.832' E	182m
Mobile Node	36° 53.605' S	174° 45.524' E	62m

This provides values for the distance between the fixed and mobile nodes. Pythagoras' theorem can be used to account for the difference in altitude between the fixed and mobile nodes and results in the following absolute ranges. The achieved communication range at 250kbaud and 1.2kbaud is summarised in Table 10.

Table 10. Range achieved corrected for altitude

Test	**Range**
High Data-Rate (250kbaud)	204m
Long Range (1.2kbaud)	1760m

These values exceed the ranges predicted by the spectrum analyser experiments considerably. There can be many reasons for this result, but one explanation immediately stands out. The ranges predicted in the spectrum analyser experiment are based on the typical receive sensitivities provided by the CC430 datasheet [6]. These values would most likely attempt to capture a range of conditions and would intentionally be conservative.

Examining the test procedure for this experiment, one would notice that the maximum bit-error-rate (BER) in the CC430 datasheet is 1% while this experiment uses 10%. You could come to the conclusion that this is why the receive sensitivity varies, however visual monitoring of the BER while conducting the experiment revealed that the shift from 0% BER to 10% BER was only a few meters.

The results of this experiment are quite promising, providing plenty of headroom for many applications. However, LOS conditions are really very ideal and the range can be devastatingly lower when LOS cannot be achieved. The actual values will vary greatly based on environment and must be evaluated per application.

Table 11. Processing specifications of surveyed motes

	Processor	Maximum Frequency	Bus Width	FLASH Memory	RAM	External Storage
TELOSB	MSP430F1611	8 MHz	16-bits	48 kB	10 kB	1 MB
MICAz	Atmega 128L	16 MHz	8-bits	128 kB	4 kB	512 kB
IRIS	ATmega1281	16 MHz	8-bits	128 kB	8 kB	512 kB
LOTUS	Cortex M3	100 MHz	32-bits	512 kB	64 kB	64 MB
Imote2	PXA271 XScale	416 MHz	32-bits	32 MB	32 MB	N/A
TinyNode	MSP430F1611	8 MHz	16-bits	48 kB	10 kB	512 kB
AWSAM-1/2	CC430	12 MHz	16-bits	32 kB	4 kB	1/2 MB
AWSAM-3	CC430	12 MHz	16-bits	32 kB	4 kB	2 MB

Table 12. Processing specifications of surveyed motes

	Freq.	Data Rate	TX Output Power	RX Sensitivity (max data rate)	TX Current (at 0dBm)	RX Current
TELOSB	2.4GHz	250 kbps	0 dBm	-94 dBm	17 mA	23 mA
MICAz	2.4GHz	250 kbps	0 dBm	-94 dBm	17 mA	20 mA
IRIS	2.4GHz	250 kbps	3 dBm	-101 dBm	15 mA	16 mA
LOTUS	2.4GHz	250 kbps	3 dBm	-101 dBm	15 mA	16 mA
Imote2	2.4GHz	250 kbps	0 dBm	-94 dBm	17 mA	20 mA
TinyNode	Varies	152 kbps	12 dBm	-101 dBm	23 mA	14 mA
AWSAM-1/2	915MHz	250 kbps	10 dBm	-92 dBm	18 mA	15 mA
AWSAM-3	Varies	250 kbps	10 dBm	-92 dBm	18 mA	15 mA

Table 13. Processing specifications of surveyed motes

	Physical Size (2D Area)	Battery Type	Supply Voltage	Active Current (Processing)	Sleep Current
TELOSB	20.2 cm2	AA (x2)	2.1V - 3.6V	4 mA	5.1 µA
MICAz	18.6 cm2	AA (x2)	2.7V - 3.3V	8 mA	15 µA
IRIS	18.6 cm2	AA (x2)	2.7V - 3.3V	8 mA	8 µA
LOTUS	25.8 cm2	AA (x2)	2.7V - 3.3V	50 mA	10 µA
Imote2	17.3 cm2	AAA (x3)	3.2V - 4.5V	66 mA	390 µA
TinyNode	12 cm2	Li-Ion	2.4V - 3.6V	4 mA	4 µA
AWSAM-1/2	9.8 cm2	CR2032.Li-Po	2.0V - 3.6V	4.6 mA	20.7 µA
AWSAM-3	4.8 cm2	Li-Po	2.5V - 15V	4.6 mA	4.2 µA

6 Existing Wireless Motes – Review and Comparison with AWSAM-3

Motes are the fundamental element in forming wireless sensor networks (WSNs) and various sensing and control applications. They are typically used as end nodes in a network although they can also be used as base stations or gateways in WSNs. A vast amount of research is being carried out in wireless sensor networks and motes design both at universities and in industries. Many motes are designed for research and development purposes allowing prototypes to be built before committing to actual applications. The section provides a brief review of some typical motes. The motes discussed in this section is summarised and compared with the AWSAM-3 in terms of Processing, Radio and Electromechanical specifications in Table 11, 12, and 13.

6.1 TELOSB

The TELOSB mote platform was initially developed at UC Berkley [7] but is currently being manufactured and distributed by MEMSIC as a research and experimentation tool for the wider community. It has an IEEE 802.15.4 compliant radio operating in the 2.4GHz ISM band. Like other radios operating within IEEE 802.15.4 specifications, it supports a maximum data rate of 250kbps. At its core the TELOSB features an MSP430 microcontroller positioning it in the low power, low performance category. It provides a range of interfaces including: DIO, ADC, I^2C, and SPI [7, 8]. The TELOSB is powered by a pair of AA batteries. The size of the battery pack defines the overall 2D dimensions (65mm x 31mm) which includes an embedded PCB antenna. While the physical size is not tiny, it can fit into the palm of a hand and is suitable for the purpose of research and experimentation.

One particular variant of the TELOSB has been fitted with temperature, humidity, and light sensors and also provides two expansion connectors for interfacing with additional peripheral devices. 1MB of external flash memory also adds data logging capacity to the TELOSB. This data can be accessed over the air or via an on-board USB connector. The USB connector is also used for programing the TELOSB and can power it directly when attached to a computer. The TELOSB is capable of running TinyOS, also developed at UC Berkley. TinyOS is a lightweight, embedded operating system targeting WSNs. It is open source and is now under the TinyOS Alliance.

6.2 MICAz/IRIS

The MICAz wireless measurement system is currently being manufactured and distributed by MEMSIC. It has an IEEE 802.15.4 compliant radio operating in the 2.4GHz ISM band, with a maximum data rate of 250kbps. The MICA series mote (including MICAz) use the 8-bit ATmega128L low power MCU to handle the processing and provide analog and digital interfaces for peripherals such as sensors [9]. The MICAz has a similar form factor to the TELOSB and is also powered by a pair of AA batteries and hence have very similar dimensions (58mm x 32mm). One distinguishing feature of the MICAz is its lack of on-board USB connector and the use of an upright whip antenna instead of an embedded PCB antenna. The MICAz has a flexible peripheral architecture. The main processor and radio board has no sensing capability. Instead, it provides an expansion connector to which a variety of ready-made peripheral boards can be attached. The MICAz supports the MoteWorks platform [10] and is capable of running TinyOS.

The IRIS wireless measurement system is a direct descendant of the MICAz featuring two primary improvements. The ATmega1281 MCU on the IRIS provides twice the SRAM that the MICAz has (based on ATmega128L). Also, the CC2420 based radio has been replaced with an AT86RF230, yielding over three times the communication range of the MICAz. This is achieved from the superior output power and receive sensitivity of the AT86RF230. Outdoor LOS tests have achieved ranges of up to 500m [11]. Beyond these improvements, the IRIS is very similar to the MICAz. It is compatible with the same peripheral boards and has exactly the same physical dimensions. The IRIS supports the MoteWorks platform and is capable of running TinyOS.

6.3 LOTUS

The LOTUS WSN platform definitely sits in the high-performance category. It features an ARM Cortex M3 32-bit processor with an operating frequency of up to 100MHz. It contains 512kB of data memory, 64kB of RAM and a massive 64MB of external serial flash [12]. The radio transceiver is IEEE 802.15.4 compliant and is similar to the other radios used on the other motes in this section. It achieves similar performance and operates in the 2.4GHz band. The LOTUS also has the standard IO interfaces (DIO, ADC, I^2C, and SPI) allowing a range of sensors to be attached. Like the MICA and the IRIS, the LOTUS features a stacked design. The main board does not have any sensors on it, but instead allows plugging-in of the standard set of MEMSIC sensor boards.

Measuring 76mm x 34mm, the LOTUS is larger than the motes discussed so far. It features an integrated antenna on the PCB and sit on top of a double AA battery pack. Like the MICAz and IRIS, it contains a USB connector directly on the main PCB in the form of a mini-B socket. The high performance of the LOTUS processor in combination with its massive data storage capacity opens up the mote to applications requiring substantial and complex on-board processing.

6.4 Imote2

The primary motivation of the Imote2 is to move the data processing in a WSN onto the edge of the network (i.e. the individual motes). Some applications that would benefit from this approach are: predictive maintenance, public key cryptography, and data compression [13, 14]. The Imote2 platform is well suited to these usage cases due to the advanced nature of its processor, the Intel XScale CPU. Integrated DSPs and complex ALUs (features not found on low power MCUs) allow the Imote2 to carry out computationally intensive applications with ease.

By compressing data before sending across the radio interface, power consumption can be reduced. However, the power consumption benefits gained by performing data compression on the mote must be considered in the context of increased power consumption by a relatively powerful processor. Furthermore, the sleep mode power consumption is also higher on the Imote2 than on low-power/low-performance motes mentioned previously. This potentially makes the Imote2 unsuitable for applications which do not utilise its superior processing power.

The Imote 2 uses an Intel XScale CPU which operates at adjustable frequencies between 13MHz and 416MHz. It has 32MB of flash memory, 32MB of SDRAM, and 256kB or SRAM [13]. Altogether these specifications make the Imote2 a very powerful platform. The radio transceiver is similar to that used on other motes reviewed in this section. It is 802.15.4 compliant, and operates at 2.4GHz. Quite unique for the Imote2 however is that Bluetooth and GPRS functionality can be added easily using expansion boards. Some interfaces that are quite unique in mote technology that the Imote2 supports are I^2S, AC97, and a Camera Chip Interface.

6.5 TinyNode

The TinyNode is a mtoe designed both as a research tool and as a final product for in-field deployment [15, 16]. Of the devices reviewed in this research, the TinyNode is unquestionably the most optimised in terms of the characteristics required for practical, in-field use. Its primary focuses and achievements are: long communication range, low power consumption, and miniaturisation.

The most recognisable characteristic of the TinyNode that supports its long communication range is the use of a XE1205 radio transceiver that is capable of 12dB of output power [16]. It must be noted however that the specified communication range (600m, outdoor, 76.8kbps) is not directly comparable to the specifications of other WSN platforms reviewed, which are often quoted at a data-rate of 250kbps. However, with a 12dB output power level, the TinyNode can surpass the communication ranges of other motes. The radio can operate at 868MHz and 915MHz with a maximum data-rate of 152kbps.

The power consumption of the TinyNode is in the same ballpark as other low power motes, with consumption naturally increasing when operating at a high radio output power. In sleep mode it draws between 4μA and 7μA of current from a choice of flexible battery options. The TinyNode is reasonably small, measuring 30mm x 40mm. The main module does not include any sensors or a battery but provides a connector through which these can be attached. Current support includes: expanded memory, solar energy harvesting, and sensors.

6.6 AWSAM-1 and AWSAM-2

AWSAM-1 is the first generation of mote developed at the University of Auckland [17]. This mote has been used to evaluate soil response in a shake table excited laminar box for earthquake simulation [18]. Physical size is of distinct importance to this particular application, and as such the hardware has been designed with miniaturisation as a focus. Having a physically small module allows the node to more closely resemble a natural particle. At the heart of the AWSAM-1 lies the CC430F5137 SoC from Texas Instruments which is made up of a CC1101 sub-1 GHz radio transceiver and an MSP430 microcontroller. It consumes very little power and measures just 8mm x 8mm. The CC430 has 32KB of program memory and has been configured to operate at 12MHz and offers abundant peripheral interfaces including a 6-channel 12-bit ADC, two 16-bit timers, hardware Real-Time Clock (RTC), SPI, I^2C, and 10 digital I/O lines. The radio front end of the AWSAM-1 is configured to operate at 915MHz.

The AWSAM-1 is driven directly from a 3V CR2032 lithium coin-cell battery. The battery is mounted on the back side of the node in an enclosure, resulting in a total node thickness of 10mm. The particular variety of CR2032 battery that is being used with this node has a nominal capacity of 225mAh with a cut-off voltage of 2.0V (the lowest voltage required to operate the radio on the CC430). The battery is often the largest component on a mote and constrains the miniaturisation of the mote. The size and form factor of the CR2032 allows the base configuration of the AWSAM-1 to be just 35mm x 28mm x 10mm.

The most noticeable differences between the AWSAM-1 and AWSAM-2 [18] are the two additional inertial measurement sensors, the doubling of the external serial flash capacity, and support for a Li-Po battery pack. To support the sensing of the node's angular velocity, the AWSAM-2 includes the L3G4200D 3-axis gyroscope from STMicroelectronics. It can measure up to ±2000 degrees per second with 16-bit resolution. The LSM303DLHC 3-axis magnetometer and accelerometer combo from STMicroelectronics has been chosen for the purpose of providing absolute orientation. The LSM303DLHC provides 3-axis' of magnetic field sensing at up to ±8.1gauss. In addition to this, 3-axis' of acceleration sensing are available at up to ±16g. Instead of using the CR2032 battery, the AWSAM-2 implements a Li-Po battery which offers high energy density, high maximum discharge rate, and its ease of recharging with off the shelf chargers.

7 Conclusions

Distributed and precise ubiquitous data acquisition is of crucial importance for future embedded monitoring and situational aware applications. While many research efforts have been devoted to sensor networks and motes designs, most existing motes still suffer from being power hungry (i.e. short operating lifetime) and being physically large. In this research, a new mote design, AWSAM-3, featuring modularity/flexibility, physical size miniaturisation, power efficiency, and long communication range has been presented.

The principle of the AWSAM-3 architecture is modularity and it has been followed throughout the hardware architecture and performance profiling. Modularity was identified as a goal of the AWSAM-3 and the layered architecture allows the

AWSAM-3 to be miniaturised while meeting application specific requirements easily. This is achieved through smart component partitioning and a standardised framework on which further hardware and software can be developed. As the result, AWSAM-3 is 2-dimensionally smaller than the earlier AWSAM generations and is also smaller than other popular mote platforms reviewed by this research. The AWSAM-3 measures just 30mm x 16mm in two dimensions, with the vertical height varying based on application.

The AWSAM-3 has targeted low power consumption and practical experiments have demonstrated just how well it performs in this regard. Motes commonly use linear regulators to regulate their supply voltage however the AWSAM-3 utilises a DC-DC converter to regulate the voltage to its core components. An experimental analysis has confirmed that by using a DC-DC converter, power consumption is noticeably smaller when operating the relatively power hungry radio transceiver, compared to an ideal linear regulator. Furthermore, the experiment demonstrated that under typical conditions, the DC-DC converter was more efficient than an ideal linear regulator even when the AWSAM-3 is in its lowest power sleep state. This result verifies the low power standing of the AWSAM-3.

The power consumption of operating the radio transceiver and reading/writing to serial flash memory have been evaluated as a foundation model from which power profiles can be built. An interesting finding for the AWSAM-3's specific hardware shows that, in some cases, transmitting data over the radio can consume less energy per byte than writing to flash memory. This invalidates the use of accumulate-and-send techniques on the AWSAM-3 for the purpose of reducing power consumption. However, other beneficial uses for the flash exist and the results of these experiments do not undermine them.

Superior communication range has also been identified as a goal of this research. Tests with a spectrum analyser have shown promising results. A series of practical line-of-sight tests seeking to validate the spectrum analyser results have surpassed expectations. The communication range has been shown to be over 200m at 250kbps and over 1.7km at 1.2kbps. These results correspond to an output power level of 0dBm while the AWSAM-3 is capable of 10dBm.

8 Further Perspectives and Challenges

Modularity has been a goal of the AWSAM-3 platform, which has been applied in the hardware design and facilitates modular power profiling of individual components. The same objective should be targeted in software development as well. Drivers for the hardware and peripherals must be developed in a modular and hierarchical way, allowing code to be reused and drivers for future hardware to be implemented easily. This software should abstract away the hardware, allowing the developer to focus efforts on application specific development. This also extends to the protocol stack design. For example, the common physical layer radio core driver provided in the Texas Instruments SimpliciTI protocol stack can be reused in developing more efficient application specific protocols. The same radio core driver has already been utilised to develop a bare-metal IPv6 compatible 6LoWPAN protocol stack for AWSAM motes to support large scale embedded interoperable applications.

The experiments carried out so far were meant to reveal the potential of the AWSAM-3 hardware. However there is a need to carry out further testing in order to get a complete overview and model of the AWSAM-3 performance metrics. The power consumption of the radio has not been evaluated with the depth required to model its start-up and shut-down behaviour. This will require a more complex testing procedure. The line-of-sight radio range tests highlight the notable performance of the AWSAM-3 but they do not capture all usage scenarios. Indoor environments and those involving visible obstructions need to be tested as they represent typical conditions for practical applications.

There are some metrics that have already been planned for this research. First, a model of the 3-dimensional radiation patterns is yet to be done. This needs to be carries out in an anechoic chamber with specialised equipment. A major component that is yet to be evaluated is the battery. Self-discharge and battery aging can have significant effects on applications where the mote is expected to stay alive without maintenance for extended periods of time. A long term set of experiments are planned to seek to precisely model the behaviour and expected lifetime of the battery for the whole scope of possible use cases.

At the hardware level, there are many more sensor and/or actuator boards that can be developed to open up the AWSAM-3 to more applications. The modular architecture of the AWSAM-3 will simplify this process greatly. Currently, two main application fields, environmental monitoring and pervasive healthcare are targeted. More sensing and actuating boards are planned to strengthen environment and human vital signs monitoring abilities of the AWSAM-3.

References

[1] Mao, X., Miao, X., He, Y., Li, X., Liu, Y.: CitySee: Urban CO2 monitoring with sensors. In: Proceedings of INFOCOM 2012, pp. 1611–1619 (2012)

[2] Somov, A., Dupont, C., Giaffreda, R.: Supporting Smart-city Mobility with Cognitive Internet of Things. In: Future Network & Mobile Summit (2013)

[3] Jiang, X., Tang, Y., Lei, Y.: Wireless sensor networks in Structural Health Monitoring based on ZigBee technology. In: Proceedings of the 3rd International Conference on Anti-Counterfeiting, Security, and Identification in Communication (ASID 2009), pp. 449–452 (2009)

[4] Li, X., Jiang, L., Li, J.: Home healthcare platform based on wireless sensor networks. In: Proceedings of the International Conference on Information Technology and Applications in Biomedicine (ITAB 2008), pp. 263–266 (2008)

[5] IEEE Standard for Information technology– Local and metropolitan area networks– Specific requirements– Part 15.4: Wireless Medium Access Control (MAC) and Physical Layer (PHY) Specifications for Low Rate Wireless Personal Area Networks (WPANs), IEEE Std 802.15.4-2006 (Revision of IEEE Std 802.15.4-2003), pp. 1–320 (2006)

[6] Texas Instruments, CC430F613x, CC430F612x, CC430F513x MSP430 SoC With RF Core (Rev. G), ed (2013)

[7] Polastre, J., Szewczyk, R., Culler, D.: Telos: enabling ultra-low power wireless research. In: Fourth International Symposium on Information Processing in Sensor Networks, IPSN 2005, pp. 364–369 (2005)

[8] MEMSIC, TELOSB Datasheet 6020-0094-04 Rev B, ed.

[9] MEMSIC, MICAz Datasheet 6020-0065-05 Rev A, ed.

[10] Crossbow Technology, MoteWorks Brochure 6030-0001-02 Rev A, ed.

[11] MEMSIC, IRIS Datasheet 6020-0124-02 Rev A, ed.

[12] MEMSIC, LOTUS Datasheet 6020-0705-01 Rev A, ed.

[13] Crossbow Technology, Imote2 Datasheet 6020-0117-01 Rev A, ed.

[14] Nachman, L., Huang, J., Shahabdeen, J., Adler, R., Kling, R.: IMOTE2: Serious Computation at the Edge. In: International Wireless Communications and Mobile Computing Conference, pp. 1118–1123 (2008)

[15] TinyNode, TinyNode 584 Fact Sheet V.1.1, ed. (2011)

[16] Dubois-Ferriere, H., Meier, R., Fabre, L., Metrailler, P.: TinyNode: a comprehensive platform for wireless sensor network applications. In: Proceedings of the 5th International Conference on Information Processing in Sensor Networks, pp. 358–365 (2006)

[17] Wang, K.I.K., Salcic, Z., Wilson, M.R., Brook, K.M.: Miniaturized wireless sensor node for earthquake monitoring applications. In: 2012 7th IEEE International Symposium on Industrial Embedded Systems (SIES), pp. 323–326 (2012)

[18] Ivoghlian, A., Wang, K.I.K., Salcic, Z.: AWSAM - Miniaturised Wireless Nodes for Structural Health and Environmental Monitoring. In: The 19th Electronics New Zealand Conference (ENZCon), Dunedin, New Zealand, pp. 43–48 (2012)

Intellectus: Multi-Hop Fault Detection Methodology

Tiziana Campana and Gregory M.P. O'Hare

School of Computer Science and Informatics University College Dublin
`tiziana.campana@gmail.com, gregory.ohare@ucd.ie`

Abstract. Wireless Sensor Networks (WSNs) can experience problems (anomalies) during deployment, due to dynamic environmental factors or node hardware and software failures. These anomalies demand reliable detection strategies for supporting long term and/or large scale WSN deployments. Several strategies have been proposed for detecting specific WSN anomaly, yet there is still a need for more comprehensive anomaly detection strategies that jointly address network and node level anomalies. Intellectus methodology build a tool that detected a new limited set of faults: sensor nodes may dynamically **fail**, be **isolate** and **reboot** and **local topology control**. These bugs are difficult to diagnose because the only externally visible characteristic is that no data is seen at the sink, from one or more nodes. This chapter evaluate Intellectus methodology by different experiment in a Testbed network. In fact, Intellectus is be able to detect the injected fault and assess different scenarios of topology change.

1 Introduction

The main problems and difficulties of Wireless Sensor Network monitoring and debugging are:

- **Node side**: errors and faults that happen in a node, including.
  - **Node reboot**: the sensor node perceives reboot events together with their associated frequency and advices the user-side when the frequency is greater then zero (self-detection).
  - **Node crashes**: errors in low-level hardware, which render the node inoperative. The neighbouring nodes monitor each other allowing the detection of crash failures (group-detection).
- **Network side**: errors and faults that happen in a network, related to connectivity, environment and so on.
- **Connectivity issues**: resulting from topology changes and dynamics of the network. The Intellectus methodology provides connectivity monitoring and topology change control. Intellectus seeks to develop sensor nodes that are both self and environment aware. The sensor node relies on local information in order to monitor itself and that of its neighbourhood, by adding an intelligent approach based upon perceived events and their associated frequency (self and group-detection).

A. Mason et al. (eds.), *Sensing Technology: Current Status and Future Trends IV*,
Smart Sensors, Measurement and Instrumentation 12, DOI: 10.1007/978-3-319-12898-6_9

Furthermore, all of the above fault scenarios are worsened by the multi-hop communication nature of sensor networks. It often takes several hops to deliver data from a node to the sink; therefore, failure of a single node or link may lead to missing reports from the entire region of the sensor network.

A diverse range of faults and errors can occur within a wireless sensor network (WSN), and it is difficult to predict and classify them, especially post-deployment within the environment. Current monitoring and debugging techniques prove deficient for systems which contain bugs characteristic of both distributed and embedded systems. The challenge that faces researchers is how to comprehensively address network, node and data level anomalies within WSNs through the creation, collection and aggregation of local state information while minimizing additional network traffic and node energy expenditure.

Within Intellectus [12, 13], a sensor node relies on local information in order to monitor itself and that of its neighbourhood, by adding a deductive approach based upon perceived events and their associated frequency. In Intellectus nodes maintain state information about self and their neighbourhood. This is compiled by recording the occurrence of successful events. The challenge is to maintain local information results without an increase in energy consumption and network overhead associated with sending debug information on the channel.

The sensor node can detected it own faults like software bugs, energy, local topology change and advice the sink in a second moment without extra communication traffic. Furthermore, self -detection save battery-consumance but can consume more memory. The code size of Intellectus on a single sensor node is only 446 bytes. Therefore, Intellectus is easy to implement and not consumes too much data and program memory, during the monitoring and debugging process on the sensor node. This chapter focus on the evaluation of Intellectus in a single sensor node on Indriya [14] testbed, whereby a node controls its own actions and takes decision on its local states in different scenarios of topology change.

2 Related Work

Significant work on conventional network management and debugging tools exists [2, 7]. One of the main challenges is determining how to incorporate network intelligence for detecting and localizing anomalies. The literature can be segmented into two broad approaches: *centralized* and *distributed*. The centralized approach typically necessitates access to more comprehensive network state information available at a back-end and are thus simpler to implement. The *distributed approach* provides more scalable and responsive anomaly detection. Generally, all monitoring techniques can be divided into either *active* or *passive* monitoring. *Active* monitoring injects probes or event reports from the managed nodes, and network-internal performance can thereafter be inferred from the measurement parameters. In contrast, *passive* monitoring observes the traffic already present and then infers the network condition.

A number of systems can be characterized as centralized and passive approaches [1, 2]. Sympathy [2] is based upon a central premise that in a data-collect network there is a direct relationship between the amount of data collected at the sink and the existence of failures in the system. In general in centralized approaches nodes send additional information and it is consequently very expensive. The central node can easily become a single point of data traffic concentration in the network, as it is responsible for all the fault detection and fault management. This causes a high volume of message traffic and resulting energy depletion.

The *Distributed approach* [5] fosters the concept of local decision-making, permitting a local node to make certain levels of decision before communicating with the central node. This approach is predicated on the belief that the more decisions an individual sensor can make, the less information needs to be delivered to the central node [3, 5, 9, 10, 16]. A total *distributed approach* includes *node self-detection* and *neighbour coordination*. The anomalies typically detected are binary data of abnormal sensor reading [3]. *Node self-detection* schemes [6, 8, 11] identify possible failures by performing a self-diagnosis. In [8] the authors propose an approach whereby a node would perform self-diagnosis based upon the measurements of accelerometers to determine if the node suffers from an impact that could lead to hardware malfunctions. Faults caused by battery exhaustion can also be predicted when the hardware allows the measurement of the battery voltage [11]. Failure detection via *neighbour coordination* [7, 15] represents another example of fault management distribution. Nodes coordinate with their neighbours to detect and identify the network faults before consulting with the central node. This approach may be slow, and is error-prone and consequently may end up routing into a new neighbour that has also failed. In [7] nodes in the network coordinate with the neighbouring nodes to detect faulty nodes before contacting the central point. In this schema, the reading of a sensor is compared with its neighbouring' median reading. While this algorithm is scaleable the probability of sensor faults needs to be small. Intellectus [12] seeks to design and build sensor nodes that can recognize patterns. Accurate pattern recognition by sensor nodes would be immensely useful in monitoring its local states. In Intellectus, a sensor recognizes an event, understands its internal and external behaviour, identifies its local states and advices the user-side of this local state.

3 Intellectus

Significant changes occur throughout the node life cycle which demand provision for controlling and state monitoring for sensor nodes (see Fig. 2). Monitoring and control are iterative and ongoing processes undertaken within each sensor node at each point of time (see Section 3.2) in order to develop the course of action required to attain control of the sensor node and their associated state changes. Intellectus generates information by which to improve the management of both a single sensor node and that of the entire network. The Monitoring and control process groups tracks, reviews, and regulates the progress and performance of a sensor node (see Section 3.4).

The Intellectus methodology adds to the event-action paradigm in the sensor mote the logic component necessary to observe its behaviour. Sensor nodes report performance about local status by reporting their local decision In Intellectus, the *Current and Past local states* are comprised of current and past roles of the node in the network with an index represent the frequency of events.

$$CurrentState \rightarrow \langle RoleInNetwork \rangle \langle Evidence \rangle$$
$$PastState \rightarrow \langle RoleInNetwork \rangle \langle Experience \rangle$$

In Intellectus, a sensor recognizes an event, understands its internal and external behaviour, identifies its local states, and advices the user-side of its internal local decision by augmented standard message traffic with such classification. The follow sub- sections describe the basic structure of Intellectus as well as other important considerations including how Intellectus is implemented, managed and executed on operational sensor nodes.

3.1 Event Definition

Define events is the automatic process of identifying events that are performed by the application within the sensor node. The application source code is typically decomposed into smaller components called events-handler that will call events that represent the work necessary to complete a given sensor node activities. Decomposition, as applied to defining events, involves subdiving the application source code into smaller, more manageable components called *events*. Events represent those things necessary to complete a given action. The outputs from the define events process is an event list.

```
enum {
  FSM_BOOT = 0x00,
  FSM_RADIO_CONTROL  = 0x01,
     ...
  FSM_INTERCEPT = 0x09,
 };
```

Fig. 1. Event List data encoded with *IntellectusEventCode*

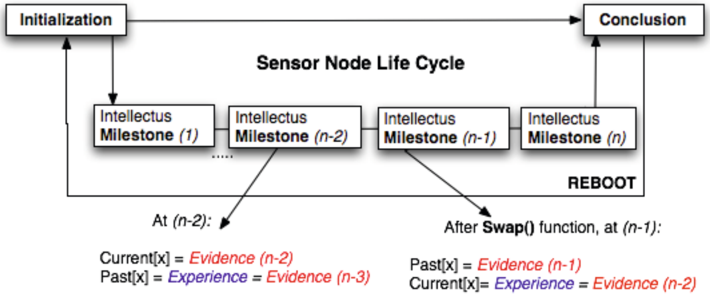


Fig. 2. Overview of the Sensor Node Life Cycle in Intellectus System

The event list includes the event identifier and a group description for each event in sufficient detail to ensure that the user-side understands what work is required to be completed. The event identifier is the *IntellectusEventCode*, encode each discrete event. Fig. (1) illustrates the event list with *IntellectusEventCode* for the *MultihopOscilloscope* application. Each event is comprised of a number of attributes: Event name (i.e. FSM BOOT), Event ID (i.e. FSM RADIO CONTROL = 0x01), Intellectus group classification (i.e. Reboot, Internal and Network). In Intellectus, each event is categorized as one of three Intellectus group classification, explained as follows:

- **Reboot Events**: are all events and states concerning the reboot of sensor node.
- **Internal Events**: are all events and states concerning the internal activity of the sensor motes, like an internal timer that expires every few seconds.
- **Network Events**: are all events and states concerning the network or external activity of the sensor motes, an example is the 3-tuple (Receive, Snoop Receive, Intercept Receive) for the Collection Tree Protocol (CTP) [4].

Marked events are collect and analyzed in the monitoring and controlling process where logical relationships and pattern of events are highlighted (see Section 3.4 and Fig. 6).

3.2 Intellectus Milestones and Observations

A sensor node life cycle is in essence a collection of generally sequential milestones whose number are determined by the reboot event or turn off event. The life cycle of a sensor node is divided into scheduled milestones (see Fig. 6 and Fig. 2). In each milestone, the monitoring and controlling processes are executed. Those processes are required to track, review, and regulate the progress and performance of a sensor node are initiated. The outcome of each milestone is a local decision, observations about current events and variance analysis between current and past observations milestones. Observations progressively assemble measurements about events and their associated frequencies collected in the current and previous milestones. One *observation* is comprised of (*Evidence, Experience*), see Fig. (6). An *evidence* is a current frequency of events in the current milestones and *experience* is a past frequency of events (in the previous milestones). Fig.(2) shows how the *evidence* and *experience* are created in each intellectus milestone. A *swap* function exists where *evidence* becomes *experience. Experience* will be updated with current events and current frequency. Fig. (3) illustrates how this *swap* function is implemented.

```
    // Evidence
    int current[11];
    // Experience
    int past[11];

    int (*puntNew)=current;
    int (*puntLast)=past;
    int (*tmp)=NULL;

// Swap Function: Evidence becomes Experience
// Experience will be update with new current frequency
    void swap(){
       tmp= puntNew;
       puntNew = puntLast;
       puntLast = tmp;
     }
```

Fig. 3. Calculating Evidence and Experience Observations Source Code

3.3 Intellectus Role Structure

The role of sensor node within the network is a factor, which can affect its local state and furthermore influence how behaviour (evidence and experience) is conducted. In the Intellectus methodology sensor nodes can perform three different roles, namely: root, router and leaf (see Fig. 4). The Intellectus roles are strongly related to the CTP interfaces [4]. In Intellectus, the sensor node is responsible for identifying its role and responsibility within the network. A root node receives data from messages delivered by the collection tree and uses the interface *Receive.* The root node role is the only one to have the *Receive* interface active and therefore a node knows about its role in the network and behaves with

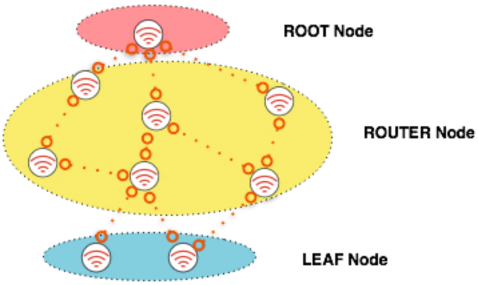


Fig. 4. The Intellectus Roles

appropriate monitoring activities in order to accomplish this role. Fig. (5) shows the *Receive* interface source code. When this interface is active a local variable (*local.fieldOne*) will be updated with the role of the node that will be used in its local decision and thereafter will be communicated to the user-side.

```
event message_t*
Receive.receive(message_t* msg, void* payload, uint8_t len){
      radio_msg_t* in = (radio_msg_t*)payload;
      local.fieldOne='R';
       .......
   return msg;
}
```

Fig. 5. Root node with active *Receive* interface

The router node is an internal node (or inner node) of the tree; it is any node of a tree that has child nodes. The router node role is the only to have the interface *Intercept* active. A collection service signals an Intercept event when it receives a packet for forwarding. When the *Intercept* interface is active a local variable (*local.fieldOne*) will be updated with the role of node that will be used in its local decision and subsequently communicated to the user-side.

The leaf node is an external node (or outer node); it is any node that does not have child nodes. The leaf node is a node that is not a root or router role. The leaf node can use the interface *snoop* to listen to neighbourhood nodes. Within the Intellectus implementation a leaf node can be characterized as having neither the *Intercept* nor *Receive* interfaces active. Within the Intellectus, each sensor node has one clear role based upon the nature of the active interface and organization events (see Fig. 4). During the network deployment a sensor node may change

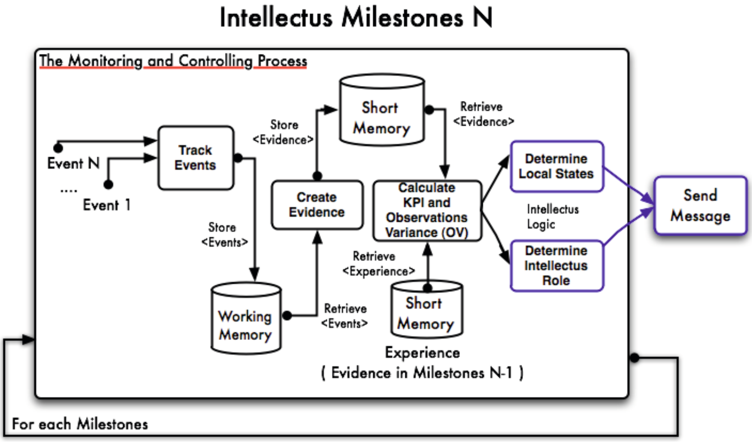


Fig. 6. Overview of the Architecture of the entire Intellectus System

its role in the network since the network topology is often subject to change. The role of a sensor node is distinct, a sensor node can only have one role at any given instance.

3.4 Intellectus: Monitoring and Controlling Process

The monitoring and controlling process is invoked as appropriate to effectively drive the activities of each sensor node in a controlled manner. A sensor node will expect to change during its lifetime and it is necessary to deliver processes that manage and control such change. The monitoring and controlling process group generates information to improve the management of sensor node. Events marked with *IntellectusEventCode* are collected. The track events process refers to a collection of events that are grouped together in order to create observations (*evidence* and *experience*) which facilitate effective management of the monitoring and controlling process group in meeting strategic monitoring and debugging objectives. Create *observations* is the process of analyzing event sequences, occurrences, and logical relationships to create the *evidence* for the current milestone. Observations continue throughout the life cycle as work progresses, the monitoring and controlling process group checks for change, and the nature of such change.

Baseline Components are used to review performance over milestones, see Fig. (6). The performance deviation from the past milestone baseline will be monitored and the following information is determinated:

- *Observations Variance (OV)* [12]: is a measure of observations performance on two consecutive milestones. It is equal to the *evidence* minus the current *experience*. Variance analysis is used in Intellectus to compare experience to evidence performance, (see Fig. 2 and Fig. 6). Equation: *OV = Evidence - Experience.*

- *KPIs* [12]: are a trend analysis that examines sensor node performance over past milestones to determine if performance is improving or deteriorating. KPI techniques are in valuable in understanding past performance and comparing it to future performance goals [12].
- *Group Control Variance*: is a measure of neighbour activity within a milestone. The ID and frequency of messages received (or *snoop* or *intercept*) are measured, (see Section 3.5).
- *Intellectus Local Decisions* [12]: combine the observations variance with KPI analysis with group control variance to determine local decisions about sensor node local states and update the Intellectus baseline [12, 13].

Only the group control variance is described in the subsequent sections.

3.5 Group Control Variance

The group control variance metric measures the neighbour activity within a milestone. As a metric it will advice the user-side when a new node is within a neighbourhood or not (it is not *received* or not *intercepted* or not messages). Every time that a sensor node *receive* (or *intercept*) a message, two local arrays (i.e. *IDnetwork[numberNeighbour]* and *network[numberNeighbour]*) will be updated.

```
if(local.fieldOne=='I'){
    if(evidence[9]-experience[9]<0){
       if(MNID<=0){
          for(g=0;g<numberNeighbour;g++){
              if(IDnetwork[g] != -1){
                 if(network[1][g]==0){
                    local.fieldTwo |= FSM_NOT_INTERCEPT_ID;
                    local.fieldFour = IDnetwork[g];
                  }
                }
           }
       }
  }
```

Fig. 7. Group Control in the source code: No Messages Intercepted

The first array (i.e. *IDnetwork[numberNeighbour]*) contains the neighbour node ID and the second array (i.e. *network[numberNeighbour]*) contains the frequency of messages received (or intercepted). When a message from a new neighbour is intercepted, the neighbour ID array and the local decision will be

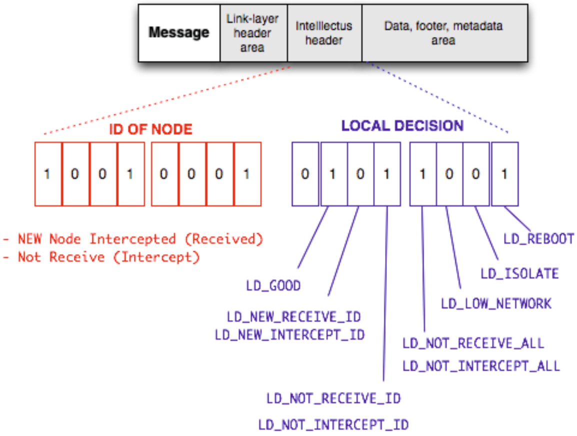


Fig. 8. Intellectus Standard Message Header

updated with the new ID and status. The sensor node registers the frequency of neighbour messages received (or intercepted) then checks in each milestone the value of those frequencies. A value of frequency message less than zero indicates an issue in the current milestone. The sensor node will send to the user-side the node ID that fails to send the messages anymore, a milestone counter and local status are updated (see Fig. 7). A frequency of message greater than zero indicates normal activity within the current milestone and not alert are communicate to the user-side, (see Fig. 7). Two data structures are implemented into Group Control, one to update the neighbour IDs and one to update the frequency of messages; both have the same size equal to the active neighbour nodes. In each milestone a cycle *for* will check the frequency of message and control the neighbour activity. The Group control variance realizes a Group-detection; the neighbouring nodes monitor each other allowing the detection of crash, isolate failures. Node knows the existence of its neighbourhood from the frequency of network events that happen in it, there is no extra traffic to the coordination of nodes. Therefore it is **not** produce **extra network traffic** and not consume **extra energy**. The sensor mote with its local analysis can give to the user-side the correct information about its local states. The aim is to create knowledge inside the network; therefore the computation is more **decentralized**.

3.6 Closing Process

When the closing process is initiated, knowledge about sensor node local decisions are transferred between the monitoring and controlling process to the header of the message for implementation of the delivered work, see Fig. (6). This process documents sensor node local decisions, specifying the local states in the header of the message. The Closing process is the process of finalizing all activities across the entire monitoring and controlling process group to formally complete the milestone. The Closing process is the process of incorporating local decisions into the message header and then sending the message to the user-side for further analysis.

```
enum {
  LOCAL_DECISION_Reboot= 0x01,
  LOCAL_DECISION_Low_Network= 0x04,
  LOCAL_DECISION_Isolate = 0x02,
  LOCAL_DECISION_Not_Receive_All = 0x08,
  LOCAL_DECISION_Not_Intercept_All = 0x08,
  LOCAL_DECISION_Not_Receive_ID= 0x10,
  LOCAL_DECISION_Not_Intercept_ID=0x10,
  LOCAL_DECISION_New_Intercept_ID=0x20,
  LOCAL_DECISION_New_Receive_ID=0x20,
  LOCAL_DECISION_Good= 0x40,
 };
```

Fig. 9. The Local Decisions in the Header of Message

The local decision attributes are: Decision Name (i.e. *LOCAL DECISION Reboot*), Decision ID (i.e. *LOCAL DECISION Reboot= 0x01*), Intellectus group description (i.e. *Reboot, Internal and Network*). Fig. (9) shows the decision list coded with the *IntellectusDecisionCode* for the first implementation of the *Intellectus* approach. The first column is the name of the local decision and the second column is the *IntellectusDecisionCode*. At runtime, the end of milestones triggers a local decision about the local state which is marked with an *IntellectusDecisionCode* and then communicated to user-side for subsequent analysis. The *IntellectusDecisionCode* is a hexadecimal code which encodes the local decision. The Intellectus header comprises two bytes of the normal message; the local decision is a bit of the header. Every local decision corresponds to a specific bit within the Intellectus header, this is the reason why *IntellectusDecisionCode* is encoded as an hexadecimal code is not a sequential code. If the bit is set to one the corresponding local decision is taken by the sensor node otherwise that local decision is not taken (see Fig. 8).

Fig. (10) documents how the local decision is taken. If the *evidence* (i.e. *curren[numerEvents]*) has the occurrence of reboot event (i.e. *FSM BOOT* or *FSM RADIO CONTROL* or *FSM SERIAL CONTROL*) greater then zero, the Intellectus algorithm will update the local header of the message (i.e. *local.decision*) with the local decision. The bit of header corresponding to the local decision Reboot (i.e. *LOCAL DECISION Reboot = 0x01*) will be set to one, (see Fig. 8 and Fig. 10), and then communicated to the user-side. The sensor node is primarily responsible for communicating its state with other sensor nodes and

```
 // Current local message
 message_t  local;

 void Intellectus(){
     ....
  void RebootEvent_Group(){
// Current Observation
// Reboot group have frequency greater of one
   if(current[0] > 1 && current[1] >1 && current[2] >1){
       local.decision = (nx_uint16_t )LOCAL_DECISION_Reboot;
       local.milestone = milestone;
   }
  }
```

Fig. 10. The Local Decision REBOOT

the user-side. When the user-side receives the message, it parses the Intellectus header and will establish that the sensor node rebooted from the value of the Reboot bit. The process described above is valid for all states within the decision list. This chapter focuses on Intellectus in a single sensor node. The Intellectus approach involves micro to macro model, where the entire network is considered the macro level and the micro level concerns the single sensor node. In the next section, Intellectus performance will be examined and evaluated through a series of experiments in a real WSN testbed.

4 Experiment Setup

Indriya [14] is a three-dimensional wireless sensor network deployed across three floors of the School of Computing, at the National University of Singapore. The testbed hardware comprises of 139 TelosB sensor "nodes". Nodes 1-39 are mounted in the basement, 40-86 are deployed across floor one, and 87-139 are in the floor two. Nodes run the TinyOS operating system and are programmed in the NesC programming language. The network has several inter-floor links providing 3D connectivity. Interestingly, most of these inter-floor links have packet reception ratio (PRR) close to 1.0 at 0dBm transmission power. The 3D connectivity remains even at lower transmit power levels. By artificially injecting faults of varying intensity into Indriya [14] testbed, the detection performance of Intellectus are studied. Two parameters specifically are changed to create instability into network: **transmit power** and the **radio channel**. Intellectus is able to detect the injected fault and assess different scenarios of topology change (link connectivity and node crashes). By artificially injecting faults of varying intensity into the Indriya [14] testbed, it is possible to assess the efficiency of Intellectus to recognize the occurrence of such faults. The experiments can be

Table 1. Experiment I: Node and Link Failure Parameter Settings

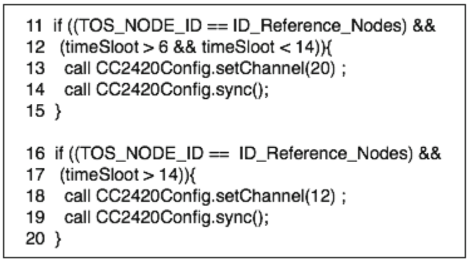
Parameter	Value
PHY Layer	2.4GHz IEEE 802.15.4
MAC Layer	Default CSMA
Routing Layer	CTP
Microcontroller	TI-MSP430
Memory	10KB (RAM) 48KB-1MB (Flash)
Transmit power	Variable (31 to 1)
Radio channel	Variable (12 to 20)
Network Size	139
Indoor range	Variable (20 to 30 *meter*)
No of BS/Sink	1 (ID=20)
Internal Timer	1024 *ms*
Intellectus Milestones	200 length
Testbed Duration	40 *min*
No. of iterations	2
Reference nodes	5,15,30,37

achieved through the manipulation of the parameters described in the Table (1). The description of the analysis of Intellectus results are discussed in the subsequent sections.

4.1 Local Topology Control - Channel Results

In the first experiment the parameter *radio channel* is changed at run time. The valid channels for the CC2420 are 11 to 26, with the default value of channel 11. *Experimental Hypothesis*: In the Indriya topology, faults are injected into the reference nodes, see Table (1). At run time, the reference nodes change their radio channel from 12 to 20 between and including milestones six to fourteen, while the rest of the network have their radio channel set to 12, see Fig. (11). Between milestone interval [6,14] the reference nodes have a different radio channel to the rest of the network, and consequently are not be able to communicate with the other sensor nodes.

By changing the *radio channel* after after a certain milestone up to milestone fourteen included, the information of Intellectus fault detection are reported and and analyzed. The change in the local connectivity results in a change of network activity. Neighbour nodes should be able to observe. The neighbourhood of reference nodes which ought no longer to be able intercept messages from nodes 5, 15, 30, 37 and the corresponding local states should be reported. Furthermore, after the milestone fourteen, the reference nodes will reset the radio channel to 12, the same value of the remainder of the network. Thereafter the neighbourhood of reference nodes will intercept again messages from the nodes 5, 15, 30, 37. Therefore, the reference nodes resume communication with the rest of the network and its neighbourhood should have a low intensity of ”No Messages



```
if ((TOS_NODE_ID == ID_Reference_Nodes) &&
 (timeSloot > 6 && timeSloot < 14)){
  call CC2420Config.setChannel(20) ;
  call CC2420Config.sync();
}

if ((TOS_NODE_ID ==  ID_Reference_Nodes) &&
 (timeSloot > 14)){
  call CC2420Config.setChannel(12) ;
  call CC2420Config.sync();
}
```

Fig. 11. The Channel Parameter

Intercepted from node with ID state" and a high occurrence of "New Messages Intercepted state" local decisions. The reference nodes and its neighbour nodes should report the local states associated with the change injected during the milestone interval [6,14]. During this experiment, only the local decision of node 5 is reported. The local decision of nodes 15, 30 and 37 are not shown but in keeping with the performance of other reference nodes.

The Fig. (12) shows the local decision states of node 5 and its neighbour nodes. The local decision of node 5 and its neighbour nodes are analyzed and compared with the initial hypothesis. Node 5 reports that its role is that of a leaf [12] an external node (or outer node) that does not have child nodes. In the first milestone [0,6], node 5 reports the local state "Reboot", the reboot local decision bit advices that the sensor node rebooted in a specific *milestone* counter, see Fig. (12), first graph. In fact, the neighbour nodes 16 and 20 report the local state "New Messages Intercepted state" with ID 5 during their first milestone [0,6], see Fig. (12), second graph. During the milestone [6,14], node 5 changes its radio channel and no local decisions are reported, see Fig. (12), first graph. The neighbour nodes 10 and 18 report the local state "New Messages Intercepted state" with ID 5 during the milestone interval [19,25] when node 5 communicate again with the rest of the network. During the milestone interval [6,14], the node 5 does not communicate with the rest of the network, only the node 10 reports the local decision "No Messages Intercepted from node with ID state" with ID 5 in milestone interval [31,37], not before because the node 10 had a previous local state "New Messages Intercepted state" with ID 5 only in milestone 25, see Fig. (12), second and third graph.

The node 18 change its role often in the network flipping between router and leaf role. The node 18 never reports the local decision "No Messages Intercepted from node with ID 5 state" because it has a leaf role in the network during the milestone interval [6,14]. During the network deployment a sensor node may change its role in the network because the network topology changes continuously. The role of a sensor node is distinct, a sensor node can only have one role at any given time. The neighbour node 16 has a router role only in the first milestone interval [3,7] and became a leaf node in the rest of its life. This can be explained because the neighbour node 16 reports the local state "New Messages Intercepted state" with ID 5 during the milestone interval [0,6] but never

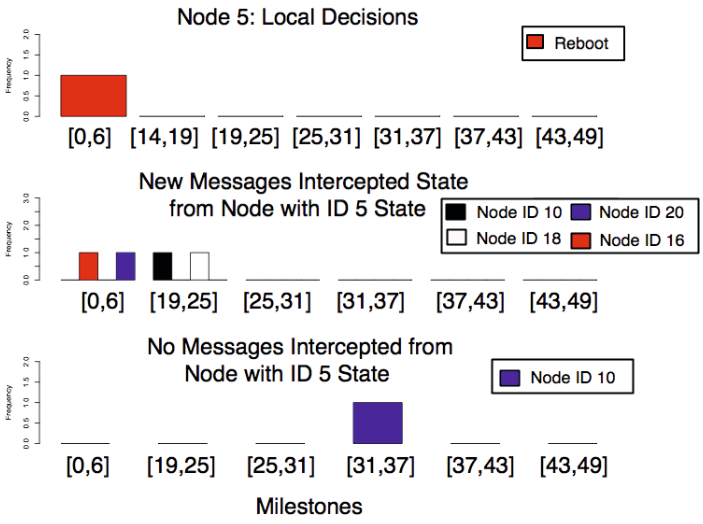


Fig. 12. Node 5 with Variable Radio Channel

the local decision "No Messages Intercepted from node with ID state" during or after the milestone interval [4,16].

Node 20 in contrast has a root role; the root node receives data from the collection tree and the local state "No Messages Received from node with ID state" is not implemented during this experiment. During the experiment the node 5 is a leaf node and Intellectus local decision for a leaf node is based upon snoop activity. After the the milestone interval [6,14], node 5 has no snoop activity and it did not report local decisions about its local states. Furthermore, the user-side received messages from node 5 after the milestone interval [6,14].

The local decisions made by the sensor nodes concour with the fault injected and the new topology created within the network. Through the aggregation of each node's local state determinations the user-side is able to debug and monitor the network as a whole. The temporal and spatial correlation of local decisions collected by neighbouring nodes are used in overseeing the entire network [12, 13].

The accuracy of Intellectus during the experiment is summarized in Table (2). The overall accuracy achieved was 100% for the local decisions "Reboot", "Isolated" and " New Messages Intercepted" states. Node Rebooted and Isolated faults are hard to detect in real deployment network. The accuracy of Intellectus was measured and found to be good for those type of faults. In Case T, the accuracy of 53.4% was due to some false positives and negatives detected at the user-side. False Negatives were caused by continuously changes of neighbour roles in the network (network topology change); while false positives were caused by milestones size parameter [13] and because a recognition system was not implemented where a feedback from user-side could be sent to stop the alert.

Table 2. Transmit Radio Channel Detection Accuracy

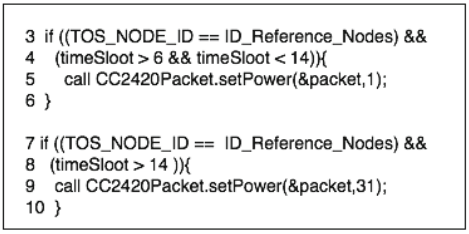
Local Decision	ID	Role	Number Inserted	Number Detected	False Negative	False Positive	Accuracy *(%)*
Isolated (X)	15	L	1	1	0	0	100
New Msg Inter. (W)	15	L	6	6	0	0	100
No Msg Inter. (T)	15	L	6	2	4	0	33
New Msg Inter. (W)	37	R	5	5	0	0	100
No Msg Inter. (T)	37	R	13	8	3	2	61.5
Reboot (D)	5	L	1	1	0	0	100
New Msg Inter. (W)	5	L	4	4	0	0	100
No Msg Inter. (T)	5	L	4	1	3	0	25
New Msg Inter. (W)	30	R	3	3	0	0	100
No Msg Inter. (T)	30	R	20	12	4	4	60
Total (W)				18	0		100
Total (T)				23	20		53.4
Total (X)				1	0		100
Total (D)				1	0		100

* **New Msg Inter. (A)** = New Messages Intercepted state.
* **No Msg Inter. (B)** = No Messages Intercepted from node with ID state.
* **R** = Router, * **L** = Leaf.

4.2 Local Topology Control - Transmit Power Results

The parameter *transmit power* is changed at run time. The valid values are 1 through 31 with power of 1 equal to -25dBm and 31 equal to max power (0dBm). When transmitting packets, the reference nodes, after a certain milestone set the value of transmitting power to 1, the min power, see Fig. (13). When end of milestone fourteen is reached the transmit power reverts back to 31, the max value, see Fig. (13). This experiment creates instability in the wireless links around the reference nodes. Between milestone interval [6, 4] the reference nodes have the transmit power to the minimal value, see Fig. (13). This results in link asymmetry around the reference nodes. The reference nodes can become isolated as a consequence of connectivity issues or different routing paths can be created in the network. Spatial and temporal correlation between local decisions of sensor nodes within the same neighbourhood advice the user-side about the reference nodes instability. By changing the *transmit power* of the referencer nodes, the effects of *group-detection* within Intellectus are reported and analyzed.

Experimental Hypothesis: During the experiment, the reference nodes in the milestone interval [6,14] changed their transmission power to the minimum value. After milestone 14 their transmission power reverted again to the maximum value. The transmit links are therefore not stable and a local connectivity change exists around the reference nodes. Therefore, the corresponding local states should not be reported in the milestone interval [6,14]; their transmit power is too low to reach the user-side. The local decision of their neighbour nodes are analyzed.



```
if ((TOS_NODE_ID == ID_Reference_Nodes) &&
 (timeSloot > 6 && timeSloot < 14)){
   call CC2420Packet.setPower(&packet,1);
}

if ((TOS_NODE_ID ==  ID_Reference_Nodes) &&
 (timeSloot > 14 )){
  call CC2420Packet.setPower(&packet,31);
}
```

Fig. 13. The Transmit Power Parameter

The neighbour nodes should report a local state "No Messages Intercepted from node with ID state" for nodes 5, 15, 30, 37, the occurrence of this local state should be high and compared with the initial hypothesis. Furthermore, after the milestone 14 the neighbour nodes should report a local state "New Messages Intercepted state" for nodes 5, 15, 30, 37 and report activities about the reference nodes.

During this experiment, only the local decision of node 15 is reported. The local decision of nod 5, 30 and 37 are not shown but are concordant with the performance of other reference nodes. Intellectus will concentrate on the local topology change and the local decision of the reference nodes and its neighbourhood will capture those changes.

The Fig. (14) shows the local decision states of node 15 and its neighbour nodes. Node 15 has a router role. During the milestone interval [6,14] it does not report activities to the user-side in accord with the transmission power change, see Fig. (14) first graph. The local state "New Messages Intercepted state" with ID 15 is reported in the milestone intervals [0,6], [6,14] and [14,19]. The high frequency of this local state in the milestone interval [14,19] shows that the transmission power of node 15 has returned to the maxim value and its neighbourhood intercepted the node again, see Fig. (14) second graph. The local decision of neighbour nodes of reference nodes correspondes with the change injected in the topology and are reported as follow.

- The neighbour node 9 reports the local decision "New Messages Intercepted state" with ID 15 in the milestone interval [6,14] when node 15 does not communicate with the rest of network, such false positive can be due to the condition of network or the milestones size parameter [12], [13]. Furthermore, the correlation of local decisions of other neighbour nodes creates an alert to the user-side about nodes 15 in the milestone interval [6,14]. The neighbour node 9 become a leaf node after the milestone 14 for that reason never reports the alert "No Messages Intercepted from node with ID state" with ID 15 in the successive milestone intervals.
- The neighbour node 3 is a leaf role until the milestone interval [0,13] and became a router role in the successive milestones after the fault was injected in the topology during the experiment. Therefore, its local decisions cannot give information about the fault injected in the topology.

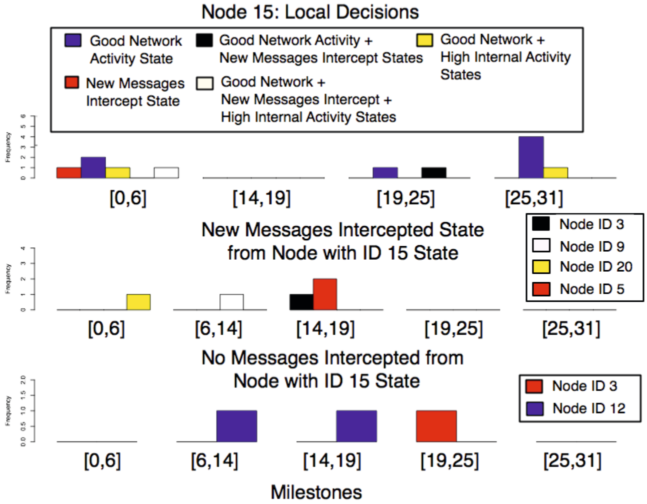


Fig. 14. Node 15 with Variable Transmit Power

- The node 20 has a root role. The root node reported the local state "New Messages Received state" with ID 15 in the first milestone interval [0,6] but never reported the local state "No Messages Received from node with ID state" with ID 15; because this alert is not implemented during this experiment, Fig. (14) third graph.
- The neighbour node 12 reports the alert "No Messages Intercepted from node with ID state" with ID 15 in milestone interval [6,14] in keeping with the fault injected in the topology. Furthermore, node 12 never report the local state "New Messages Intercepted state" with ID 15, this false negative can be due to the condition of network or milestones size parameter [12, 13] see Fig. (14) second graph.

The false positives and false negatives that Intellectus introduces in the network are likely to be caused due to either a change of the Intellectus role [12] during the sensor node life cycle or the condition of network or milestone size parameter [12]. Furthermore, the high frequency of alert "New Messages Intercepted state" with ID 15 in milestone interval [14,19] by neighbour nodes 5 and 3 correspond with the topology change injected in the network, see Fig. (14) second and third graph. Therefore, temporal and spatial correlation of local decision states creates an alert to the user-side regarding node 15 during milestone interval [6,14] in accord with the fault injected into the network. The overall accuracy achieved was 100% for the local decisions "New Messages Intercepted" states. In Case "No Messages Intercepted from node with ID state", the accuracy of 57% was due to some false positives and negatives detected at the user-side. False Negatives were caused by continuous changes of neighbour roles in the network; while false positives were caused by the milestone size parameter and because a recognition system was not implemented where a feedback from user-side could be sent to stop the alert. The accuracy of Intellectus was measured and found to be good in the testbed experiments for the injected faults. Intellectus is able to detect the injected different scenarios of topology change, in fact the local decisions of neighbour nodes of reference nodes corresponds with the change injected in the topology.

5 Conclusions

Intellectus delivers a tool, which leverages existing debugging and monitoring techniques to recreate the global state of wireless sensor networks through the aggregation of local states of network nodes. This research sought to develop a novel monitoring and debugging system for a real deployed network. This chapter has offered an evaluative perspective of the Intellectus methodology in a single sensor node and was conducted in testbed environment. During the experiments two parameters, namely **transmit power** and the **radio channel** are changed at run time in order to create controlled faults and managed topology change. It has been shown how Intellectus is capable of building an overview of a network topology as a bi-product of such experiments which assists in the better understanding of the real WSN (link connectivity, network topology and sensor node roles).

References

[1] Liu, K., Li, M., Li, M., Guo, Z., Liu, Y.: Passive Diagnosis for Wireless Sensor Networks. In: SenSys (2008)

[2] Ramanathan, N., Chang, K., Kapur, R., Girod, L., Kohler, E., Estrin, D.: Sympathy for the sensor network debugger. In: SenSys (2005)

[3] Sharma, A.B., Golubchik, L., Govindan, R.: Sensor Faults: Detection Methods and Prevalence in Real-World Datasets. ACM Transactions on Sensor Networks, TOSN (2010)

[4] Gnawali, O., Fonseca, R., Jamieson, K., Moss, D., Levis, P.: Collection Tree Protocol (CTP). In: Proceedings of the 7th ACM Conference on Embedded Networked Sensor Systems, SenSys 2009, Berkeley, California, pp. 1–14. ACM (2009)

[5] Sa de Souza, L.M., Vogt, H., Beigl, M.: A Survey on Fault Tolerance in Wireless Sensor Networks. PhD Dissertation (2009)

[6] Asim, M., Mokhtar, H., Merabti, M.: A Self-Managing Fault Management Mechanism for Wireless Sensor Networks. International Journal of Wireless and Mobile Networks (IJWMN) 2(4) (November 2010)

[7] Ding, M., Chen, D., Xing, K., Chen, X.: Localized fault-tolerant event boundary detection in sensor networks. In: Proceedings IEEE 24th Annual Joint Conference of the IEEE Computer and Communications Societies, INFOCOM 2005, vol. 2, pp. 902–913 (2005)

[8] Harte, S., Rahman, A., Razeeb, K.M.: Fault tolerance in sensor networks using self-diagnosing sensor nodes. In: IEE International Workshop on Intelligent Environments (Ref. no. 2005/11059) (2005)

[9] Tanenbaum, A.S., Van Steen, M.: Distributed Systems: Principles and Paradigms, 2nd edn. Prentice Hall (2002)

[10] Krishnamachari, B., Iyengar, S.: Distributed Bayesian algorithms for fault-tolerant event region detection in wireless sensor networks. IEEE Transactions on Computers 53, 241–250 (2004)

[11] Rakhmatov, D., Vrudhula, S.B.K.: Time-to-failure estimation for batteries in portable electronic systems. In: Proceedings of 2001 International Symposium on Low Power Electronics and Design, pp. 88–91 (2001)

[12] Campana, T., O'Hare, G.M.P.: Intellectus: Intelligent Sensor Motes in Wireless Sensor Network. In: Proceedings of the Second International Conference on Sensor Networks, SENSORNETS 2013, Barcelona, Spain (2013)

[13] Campana, T., O'Hare, G.M.P.: Intellectus: Multi-Hop Fault Detection Methodology. In: International Conference on Selected Topics in Mobile and Wireless Networking, MoWNet 2013, Montreal, Canada (2013)

[14] Doddavenkatappa, M., Chan, M.C., Ananda, A.L.: Indriya: A Low-Cost, 3D Wireless Sensor Network Testbed. In: Korakis, T., Li, H., Tran-Gia, P., Park, H.-S. (eds.) TridentCom 2011. LNICST, vol. 90, pp. 302–316. Springer, Heidelberg (2012)

[15] Hsina, C., Liub, M.: Self-monitoring of wireless sensor networks. Computer Communications 29, 462–478 (2005)

[16] Mehranbod, N., Soroush, M., Panjapornpon, C.: A method of sensor fault detection and identification. Journal of Process Control 15, 321–339 (2004)

New Approaches for Singularization in Logistic Applications Using Low Cost 3D Sensors

C. Prasse[1], J. Stenzel[1], A. Böckenkamp[2], B. Rudak[2], K. Lorenz[2],
F. Weichert[2,*], H. Müller[2], and M. ten Hompel[3]

[1] Fraunhofer Institute for Material Flow and Logistics, Dortmund, Germany
[2] Department of Computer Science VII, Technical University Dortmund, Germany
[3] Chair for Materials Handling and Warehousing,
Technical University Dortmund, Germany
`frank.weichert@tu-dortmund.de`

Abstract. In this chapter, novel approaches for the detection of logistical objects (loading units) in the field of material flow applications are comparative presented, focusing on solutions using low cost 3D sensors. These approaches realize substantial changes in comparison to traditional system design of logistic processes. Complex 3D-vision systems, costly laser scanners or throughput decreasing local sensor solutions integrated in grippers are substituted by low cost Photonic Mixing Device (PMD) cameras or structured light sensors (like Asus Xtion or Microsoft Kinect). By using low cost sensors and modern point cloud processing algorithms for detection and classification in logistic applications like depalletizing, automation of usually manual processes will be economically feasible. Besides the description of different basic solution concepts for 2.5D and 3D, two practical applications are presented.

Combining measurements of the PMD sensor and a predetermined model of loading situations, stored during the assembly of the pallet, is the first practical application for contour checking in the automated depalletizing process. This approach can compensate for the drawbacks of the comparatively low resolution of the PMD camera. Thus, it is possible to detect the deviation between the nominal and the actual loading positions and–if necessary–an automated correction of the packaging scheme may be initiated.

A 3D scanning approach (with dynamic sensor positioning) to acquire a full, registered 3D model of the pallet load is explained within the second example.

An essential part of both approaches are computer-graphics methods specific to the given problem. As a trans-applicable function, (auto) calibration techniques for 2.5 and 3D sensor applications will be presented. From an economic point of view, these approaches could decrease the costs of automated facility logistic processes. Within the evaluation the critical requirements to reach this aim are discussed on the application layer.

Keywords: PMD sensor, structured light sensor, point cloud, low cost, contour check, palettizing, de-palettizing.

A. Mason et al. (eds.), *Sensing Technology: Current Status and Future Trends IV*,
Smart Sensors, Measurement and Instrumentation 12, DOI: 10.1007/978-3-319-12898-6_10

1 Introduction

Driven by the internationalization of supply chains and global competition, new trade corridors, and privatization in emerging countries [11], typical processes in logistics became significantly more complex and dynamic during the last years. Among other things, shorter product life cycles, mass customization and stricter quality requirements [3, 14] demand for an increasing efficiency in this field of industry. Palletizing and especially de-palletizing are two of numerous logistic processes which contain, if properly dimensioned and highly utilized, a large potential for automation. To keep up in the global competition, especially in high wage countries, an increasing automation in facility logistics is unavoidable. But the use of modern technologies is not only limited to developed countries. For example german companies producing intralogistic equipment registered the highest rates of growth in export trade in 2010 for so called BRIC-states (Brazil, Russia, India and China) [19]. Concerning some industrial sectors like food, brewery, chemistry, plastic and wood, automated palletizing is established as a quasi-standard for producing subsequent packaging of goods [4].

Humanization of labor is a main objective, besides the need to make de-palletizing more efficient. Adverse working conditions and high physical stress because of manually carrying heavy stocks (e. g., cold store) can be reduced or avoided by using automated handling solutions [4, 7]. However, the advent of new technologies may alleviate these challenges; especially, computer vision using PMD technology or structured light sensors and embedded systems are particularly promising.

In this chapter, two different new approaches are proposed to detect the loading situation on euro pallets for automated de-palletizing of packages. The new approaches combine high cost efficiency, high robustness of the detection process, high time efficiency inter alia by reducing the search domain and high flexibility in terms of supporting objects with variable dimensions and different material properties. In the first instance, packages are automatically stored on a loading device (e. g., euro pallet) by means of a handling unit (e. g., consisting of an articulated robot and a gripper). In a further process step shown in Fig. 1, single unit loads will be detached (de-palletized) to the final destination or for order picking. In all cases, the outline of the loading unit has to be checked prior to de-palletizing to prevent collisions. To perform this task, the actual position of at least every object in the loading unit's top layer has to be detected.

Therefore, a new challenging method of contour checking, providing a reliable and fast gripping position to a handling unit, is highly desirable—especially if it is economically competitive to known solutions. Two iterative algorithms adapted to the special problem of de-palletizing euro pallet loads will be presented in this chapter. Both algorithms match the object geometry directly with the object's representing point cloud and limit the possible transformations to three degrees of freedom.

This chapter is structured in five sections. Following this introduction containing a general motivation for this work, state-of-the-art is presented for automated de-palletizing and sensors for the detection of loading conditions in Sect. 2. Next,

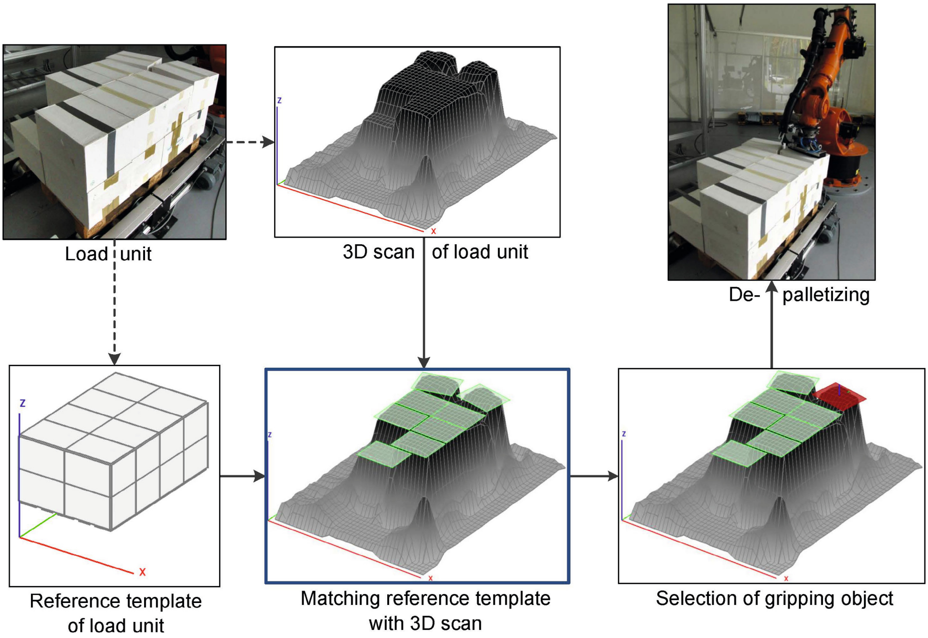


Fig. 1. Basic concept for fast de-palettizing [21]

the developed concept of detection with a 2.5D approach using a PMD sensor will be described in Sect. 2.3. An alternative approach, using a registered and colored 3D model of the pallet which was created by a structured light sensor is shown in Sect. 3. The core aspect of this paper, the evaluation of the novel automated 3D detection of the euro pallet load, is presented in Sect. 4. Finally, Sect. 5 summarizes the main statements of the chapter and gives an outlook on future work.

2 Background

In this section, a brief introduction to the state-of-the-art concerning de-palletizing is given and applicable sensors for low-cost load detection are presented. A comprehensive survey of the technical background is given in [21] and [8].

2.1 De-palletizing

Unstacking of palletized goods (de-palletizing) is usually done manually. High demands for flexibility (e. g., different size and shape of goods) are just one reason against fully automation. High-performance solutions are often too cost-intensive to make them economically justifiable. Generally, required information for (automatic) de-palletizing is acquired by sensors like laser scanners or expensive computer-vision systems.

Accurate and stable handling of goods is no longer technical challenging but the detection of an exact gripping position is a tedious task. Especially, after removing typically used load securing measures (e. g., wrapped film etc.), transportation of the pallet could cause translational and/or rotatory displacements of the load. Hence, known nominal packing positions are no longer valid and deviations have to be detected individually [23].

The developed solution presented in section 2.3 detects and verifies the complete (visible) euro pallet load by using PMD based contour checking and visualizes it afterwards. This allows an early intervention of the user in case of unacceptable deviations according to the reference positions of the euro pallet load and in case of the absence of particular parcels respectively.

2.2 Detection

Complex 3D-vision systems, costly laser scanners or throughput decreasing local sensor solutions integrated in grippers are substituted by a low-cost PMD sensor and a low-cost structured light sensor for the findings of this book chapter.

A PMD sensor measures the distance to an object by emitting modulated infrared light followed by determining the phase shift between the emitted and reflected light. The output of the sensor is a two-dimensional depth map where typical resolutions reach from 16 × 1 to 352 × 288 depth measurement points. The PMD-sensor ifm O3D201, that is used for the findings of the 2.5D approach in this chapter, has a resolution of 64 × 50 depth pixels and a field of view of 30° (horizontal) and 40° (vertical). In addition to the distance measurements, the sensor also delivers a grey scale image with the same resolution.

By combining PMD technology and a predetermined model of loading situations, stored during assembling the pallet, this approach can compensate for the drawbacks of each respective system. An essential part of the approach are computer-graphics methods specific to the given problem to both detect the deviation between the nominal and the actual loading position and, if necessary, to initiate an automated correction of the packaging scheme.

Within a second approach for de-palletiting (presented in Sect. 3), the ASUS Xtion Pro Live structured light sensor is used to acquire a registered 3D model of the pallet. The sensor emits an infrared speckle pattern, the reflected light pattern is captured by an infrared sensor and then compared with a reference pattern to estimate a per-pixel depth of the environment. Using this measurement principle, the Xtion sensor produces a two-dimensional depth map with a resolution of 640 × 480 px, covering a field of view of 58° (horizontal) and 45° (vertical). The depth map is combined with the calibrated image of a RGB camera which is also part of the Xtion sensor. Using multiple measurements of the sensor, a colored 3D model of the pallet load is created which is later on used for de-palletizing.

2.3 2.5D Contour Check

The concept of the automated de-palletizing system using a 2.5D approach with a PMD sensor is illustrated in Fig. 2. It consists of the following components:

Sensor Actor System, *Detection Algorithm*, *Graphical User Interface (GUI)* and the *De-palletizing Process*. In order to allow for a real automated robot application, the components are integrated in an usable software system (cf. Fig. 3). It combines the stored loading scheme and the real data from the PMD sensor to an augmented reality view that easily allows an operator to intervene in the de-palletizing process if necessary. At the beginning of the process, the software establishes connections with the RFID reader (that readouts the loading scheme), robot and PMD sensor. Afterwards, the operator can start the automated process that moves particular boxes or all boxes from the pallet to a materials handling device (e. g., a conveyor belt). Due to possible displacements of the euro pallet load, a matching of the expected load model and the current load model given by the PMD sensor is examined based on the developed contour checking algorithm which is roughly outlined in the following. A more detailed description of this algorithm and the camera calibration techniques can be found in [10].

The input data for the Detection Algorithm (cf. Fig. 2) consists of: (1) the *transformation matrix* between the camera- and robot-world coordinate system determined by a hand-eye calibration (refer to [10] for details), (2) the current *PMD data* and (3) the expected *load model* delivered from the RFID transponder fixed on the euro pallet. Subject of the preprocessing step is to compensate systematic depth measurement errors of the PMD sensor. Afterwards, the PMD data is transformed into the robot-world coordinate system in order to provide the operator an intuitive reference to the real environment. The iterative load detection algorithm starts with an initial point for the model which is calculated as the arithmetic average of all significant points of the PMD data that represent the euro pallet load. During the following *Load model coarse fitting* step, the position of the whole load model based on the previous initialization and the set of the current significant points is refined. This procedure follows the basic ideas of the Iterative Closest Points (ICP) concept [1]. In order to correct possible displacements of particular parcels afterwards, the *Load model fine fitting* step matches each parcel of the model on the top layer individually. The fitting process for each parcel is basically the same as in the coarse fitting step. Furthermore, collision tests for each possible transformation are performed since no intersections of the parcels are possible.

The result of the fine fitting step is a matched model with individually refined parcels on the load top layer in direct relationship to the robot's world coordinate system, the *Detected Load Model*. In order to ensure robustness and to prevent passing of maybe uncorrectable load information to the following process chain, the results have to be verified. The number of significant points which do not support the model indicate a detection failure, e. g., in case of additional objects on the load. In case of a failed verification process (*Detection Failure*), the automated de-palletizing process is stopped and manually operator intervention is required for continuing.

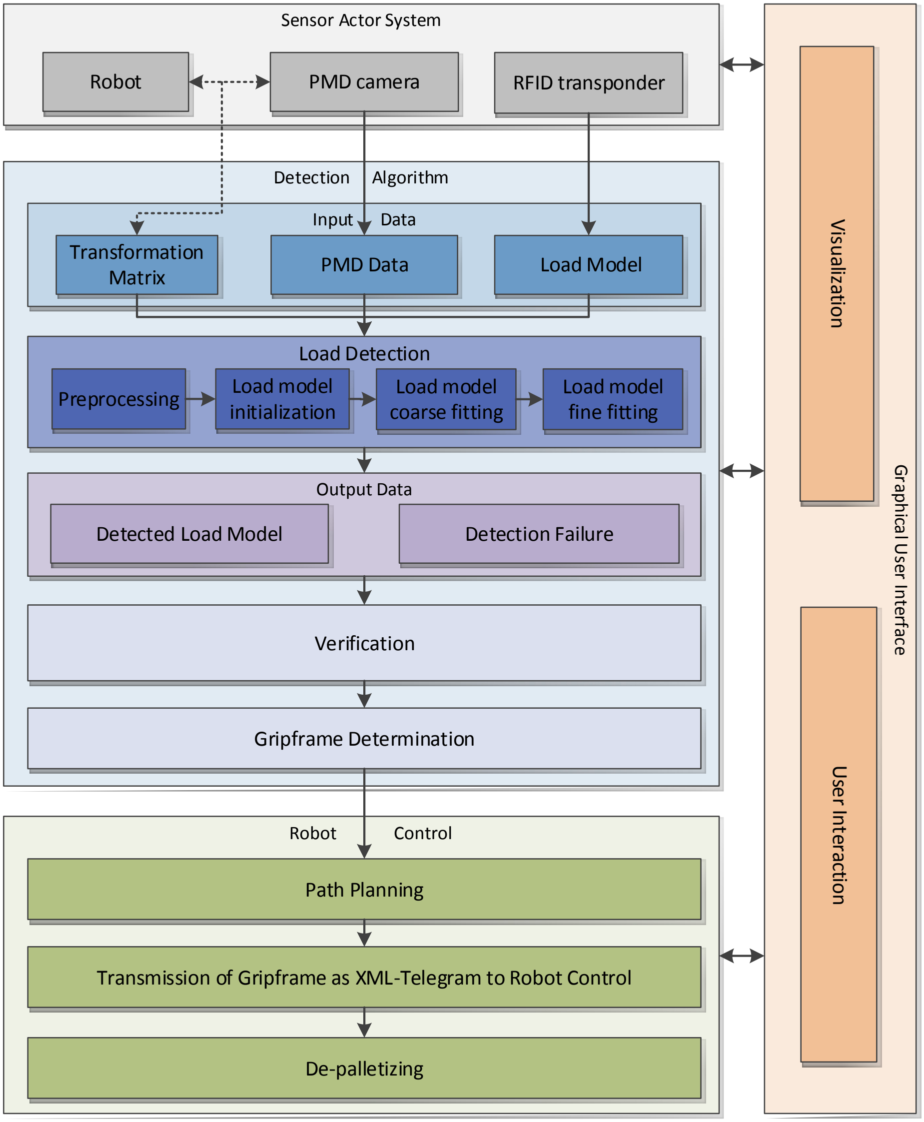
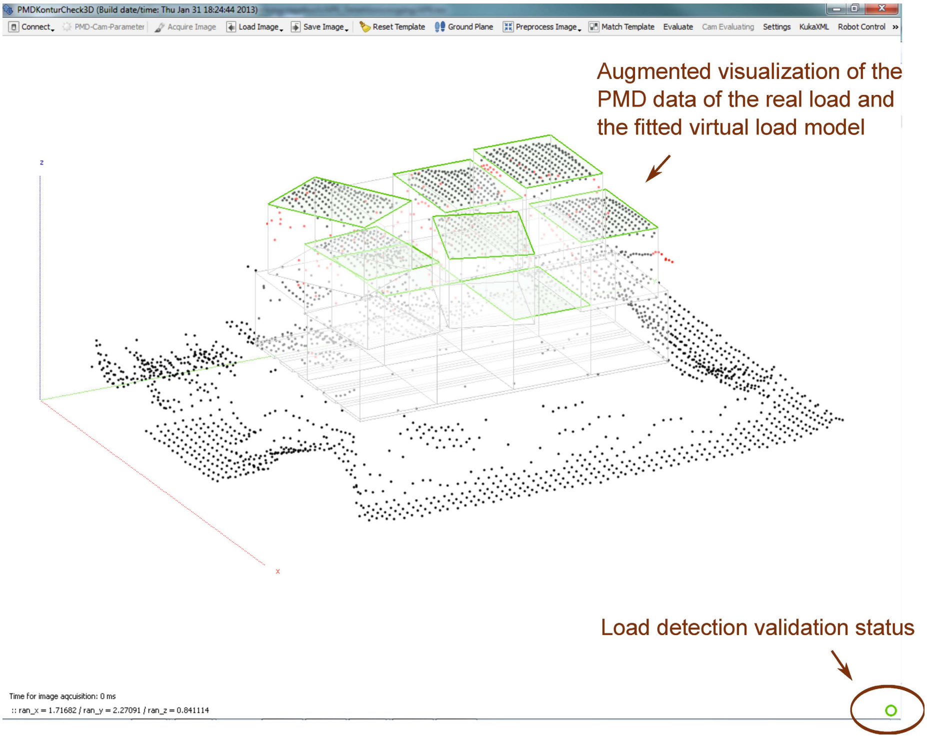


Fig. 2. Concept of the automated de-palletizing system

After a final *Verification* step, e. g., to detect possible foreign objects on the top most layer, either the operator manually picks a parcel to de-palletize (manual mode) or the next parcel to de-palletize is determined automatically (automatic mode) by the system and in both cases a gripframe for the robot is calculated for the selected parcel (*Gripframe Determination*). Finally, the gripframe (and optionally the discrete path points) is transmitted to the robot's control device in terms of XML telegrams via TCP based communication forcing



Augmented visualization of the PMD data of the real load and the fitted virtual load model

Load detection validation status

```
Time for image aqcuisition: 0 ms
 :: ran_x = 1.71682 / ran_y = 2.27091 / ran_z = 0.841114
```

Fig. 3. Graphical User Interface of the automated de-palletizing software

the robot to de-palletize the selected parcel (shown as the *Robot Control* box in Fig. 2).

In automatic mode, the process is continued in a loop until the euro pallet load is completely de-palletized. During the whole de-palletizing process the operator retains control and can intervene in the process at any time.

Evaluation of Accurancy

For evaluation purposes, detection and gripping operations were processed in a real picking cell environment consisting of an industrial robot, a vacuum gripper, and parcels with the dimension 300 × 240 × 400 mm in order to determine the accuracy and effectiveness of the calibration and detection algorithms. During the test procedure, parcels in different layers of the pallet load are selected, detected and the gripping position was reached by a checking mandrel mounted on the robot hand. The grip point of a particular parcel is defined by the barycenter of its top layer. A representative result (9 out of 18 packages) is shown in Tab. 1. The evaluation results confirm the potential of our detection algorithm for automatic unloading systems in practice. The detection results with the current low cost camera meets the accuracy requirements from the logistics' point of view as the critical z-value is less than 10 mm (tolerance of the gripping system). The

Table 1. Achieved accuracy with an ifm O3D201 PMD sensor

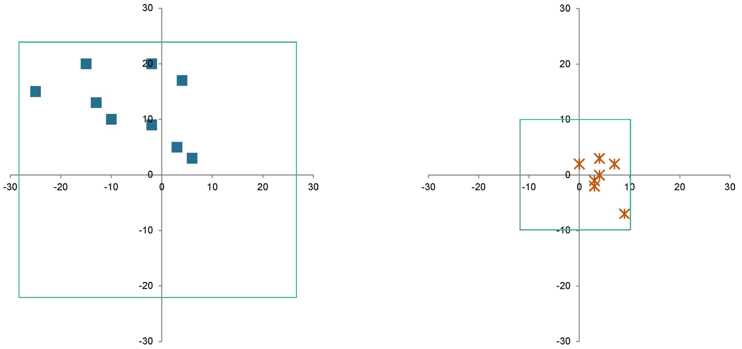
Package number	1	2	3	4	5	6	7	8	9
x**-deviation (mm)**	-15	3	-25	4	6	-2	-13	-2	-10
y**-deviation (mm)**	20	5	15	17	3	20	13	9	10
z**-deviation (mm)**	10	2	0	4	4	0	4	10	3

Table 2. Technical properties of the MESA and ifm sensor

Camera	**MESA SwissRanger 4000**	**ifm O3D201**
Depth resolution (px)	176 × 144	64 × 50
Field of view (degree)	43.6 × 34.6	40 × 30
Max. frame rate (fps)	50	20

planar deviation in x- and y-direction allows more tolerance. Since the dimensions of the parcels are known, a deviation in x- and y-direction less than 26 mm causes no damage of the load due to unexpected collisions and is acceptable.

In addition to the used ifm O3D201 sensor, a camera with a significant higher resolution (MESA SwissRanger 4000) was tested to show the potential of the detection algorithm. Table 2 visualizes the different technical specifications of the compared cameras. With 25 632 px (178 × 144 px), the resolution of the MESA camera is about 8 times higher than the ifm camera (64 × 50 px). Both sensors offer approximately the same field of view. The maximum frame rate of the MESA sensor is 2.5 times higher than the frame rate of the ifm sensor which is only relevant for dynamic applications. Finally, contemporary cost for the MESA sensor exceed the ifm camera about 4 times.



(a) IFM camera: deviation in x- and y-direction (mm)

(b) MESA camera: deviation in x- and y-direction (mm)

Fig. 4. Comparison of gripping accuracy for the **(a)** ifm and **(b)** MESA camera

Figure 4 pictures the detection results of the two cameras in comparison. As expected, the MESA sensor accomplishes a higher accuracy (less than 1 cm) compared with the ifm sensor (less than 2.6 cm). Given the higher resolution, this is a predictable result but the improvement of the accuracy is not linear in the ratio of resolution (8/1) between the two sensors. Due to missing optimizations of the calibration algorithm concerning the MESA sensor, a further improvement compared to the ifm sensor is expected. Hence robust detection and handling of significantly smaller objects is possible. Because of falling prices for high resolution PMD camera, this technology offers a wide range of applications—especially in the field of logistics.

3 3D Dynamic Densor Positioning

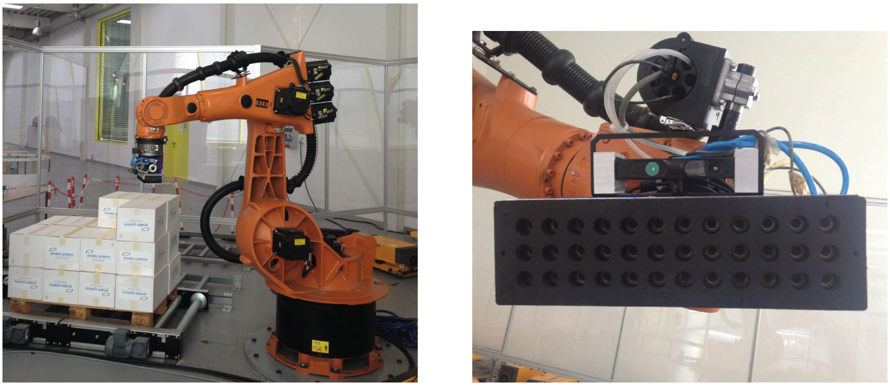
In this chapter, a novel approach for automatic load detection is presented based on a detailed 3D reconstruction of the environment. With regard to the evaluation purposes, a demonstrator was designed as visualized in Fig. 5(a). The demonstrator consist of an industrial robot *KUKA KR 125/3* with a mounted vacuum gripper in order to handle the load units and a depth camera *Asus Xtion Pro Live* mounted at the robot's gripper as illustrated in Fig. 5(b). The key idea of the concept is to generate captures of the loading scene from different views with the depth camera and reconstruct a detailed 3D map of the scene. Afterwards single quad like loading units are detected with regard to an automatic handling by the robot. The pipeline of the novel load detection pipeline is illustrated in Fig. 6. An essential step is the system calibration. As the robot base coordinate system serves as the world coordinate system of the reconstructed 3D map of the environment, a relationship of the camera coordinate system and the robot coordinate system must be established in advance, known as hand-eye-calibration (s. Sect. 3.2). In the "Acquire Stage" single captures of the loading scene with the depth camera are taken from different views by positioning the robot to determined poses. After filtering, the single captures are transformed to the world coordinate system resulting in a detailed 3D map. Finally the input data for the loading detection algorithm is reduced in two steps. First, the data outside the RoI is discarded based on a known region of interest (ROI). Second, the data undergoes a downsampling process. The particular steps of the pipeline are described in detail in the following.

3.1 Preprocessing

The subject of the preprocessing step is on the one hand the compensation of sensor noise and on the other hand the preparation of the input data with regard to an acceleration of the subsequent processing steps.

Outlier Removal

Structured light sensors produce failures at object bounds and on highly reflective materials which are in the following denoted as outliers. In particular, the



(a) Demonstrator setup

(b) Depth sensor attached to the robot's gripper

Fig. 5. The left image shows the used industrial robot KUKA KR 125/2 in front of a euro pallet load which has to be automatically de-palletized. The right image illustrates the vacuum gripper and the structured light depth sensor (Asus Xtion Pro Live) attached to the robot's gripper

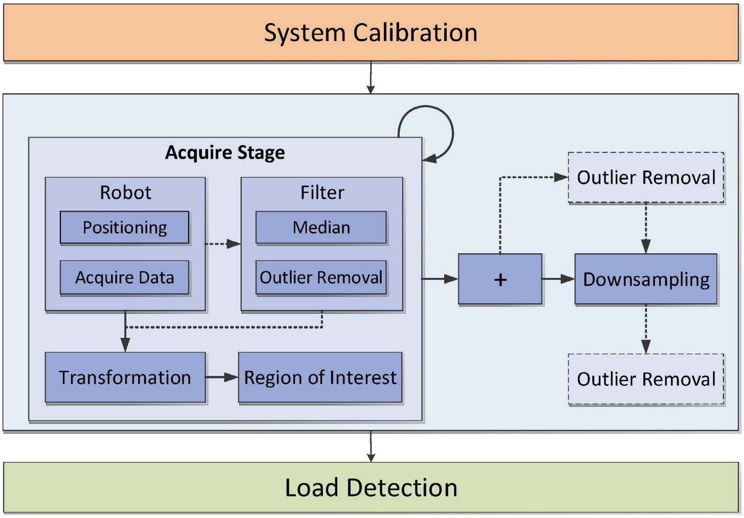


Fig. 6. Concept of load detection with dynamic sensor positioning

contrast between foreground and background at object bounds leads to strong outliers. In contrast to valid point values, a characteristic property of outliers is reduced density in the local neighborhood of these points. To take advantage of this characteristic, a method based on statistical assumptions [12] is used. Let f be a function which assigns a point $\mathbf{p}$ the next k neighbors of a point cloud $\mathcal{P}$:

$$f_{\mathcal{P},k}(\mathbf{p}) = \{\mathbf{t} \in \mathcal{T} \subseteq \mathcal{P} \mid \forall \mathbf{x} \in \mathcal{P} \setminus \mathcal{T} \colon \|\mathbf{p} - \mathbf{t}\| \leq \|\mathbf{p} - \mathbf{x}\| \wedge |\mathcal{T}| = k\} . \quad (1)$$

Each cluster $\mathcal{C}_{\mathbf{p},k} = f_{\mathcal{P},k}(\mathbf{p})$ possesses a local mean value

$$\mu_{\mathbf{p},k} = \frac{1}{k} \sum_{q \in \mathcal{C}_{\mathbf{p},k}} \|\mathbf{p} - \mathbf{q}\|. \tag{2}$$

The global mean value

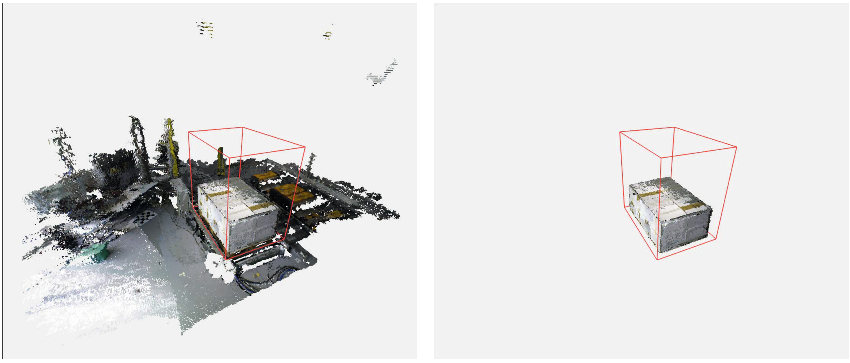
$$\mu_{\mathcal{P},k} = \frac{1}{|\mathcal{P}|} \sum_{p \in \mathcal{P}} \mu_{\mathbf{p},k} \tag{3}$$

with respect to the entire point cloud $\mathcal{P}$ is the mean value of all local mean values $\mu_{\mathbf{p},k}$. Given the assumption that the local mean values underlie a normal distribution with mean value μ and standard deviation σ, a point is declared as outlier if the local mean value deviates about $\mu \pm \alpha \cdot \sigma$. The parameter α is used for adaptation with respect to the parameter k. Experiments verified the choice of $\alpha = 1$ leading to robust results with a residual noise component of approximately 1 %.

Region of Interest (ROI)
The point cloud representing the environment is composed of particular point clouds acquired from different camera poses. According to the current pose, the content of significant information representing the load situation and background varies. With regard to acceleration of the following processing steps, the goal is to eliminate points which are not relevant for the load detection process. The ROI encapsulating the relevant points is modeled through a global definition of a convex hull using boundary planes as visualized in Fig. 7. While a single boundary plane divides space in two subspaces, the combination of several boundary planes enables the definition of a finite subspace. The definition of a plane in Hessian normal form supports efficient testing to determine on which side of the plane the point lies. By defining the ROI through a cube represented by six boundary planes, the significant information can be efficiently extracted.

3.2 3D Reconstruction

The depth sensor is attached to the robot gripper, as visualized in Fig. 5(b) which enables a controlled acquisition of the loading situation with defined sensor poses. Subject of the 3D reconstruction is to compose the particular point clouds (2.5D) acquired from individual sensor poses into a global model. The efficient fusion of two point clouds require the knowledge of the relative transformation between the local sensor coordinate systems at the time point of capturing related to the global world coordinate system. While the robot coordinate system servers as the global world coordinate system, the relative transformation between the sensor origin and the coordinate system of the robot's tool point has to be determined. This is known as hand-eye calibration [18]. The fusion of point clouds leads to a high density point cloud with non-homogeneous overlapping local regions. With regard to processing time, of the algorithms in the load detection step (s. Sect. 3.2), the information redundancy is reduced by means of downsampling the fused point clouds.



(a) Before ROI extraction (b) After ROI extraction

Fig. 7. The left image illustrates a 3D scene composed of 16 single point clouds captured from different views. The right image shows the result after excluding points outside the ROI which is highlighted by the red quad

Hand-Eye Calibration

In robotics, hand-eye calibration is a necessary task whenever a camera is mounted at a robot arm. The knowledge of the relative transformation between the camera and the robot's tool point coordinate system establishes the two significant tasks [6] which are essential with regard to the 3D reconstruction:

Conversion of sensor measurements to the global coordinate system. This procedure enables an efficient fusion of camera data acquired from different views.

Accurate positioning of the robot tool relative to the sensor. In order to get a detailed model of the environment, it is necessary to capture sensor data from specific views. Therefore the sensor must be positioned in a specific pose.

The state of the art comprises two approaches for the hand-eye calibration. The hand-eye calibration problem statement according to Tsai [18], as illustrated in Fig. 8, can be formulated as

$$\mathbf{AX} = \mathbf{XB}. \tag{4}$$

Based on two different sensor poses (tool point poses respectively) the matrix

$$\mathbf{B} = {}^{O}_{C_2}\mathbf{T}^{-1} \cdot {}^{O}_{C_1}\mathbf{T} = {}^{C_2}_{C_1}\mathbf{T} \tag{5}$$

represents the transformation between two different coordinate systems ${}^{C_1}K$, ${}^{C_2}K$ resulting from the current camera pose. The parameters are known as extrinsic camera parameters. The matrix

$$\mathbf{A} = {}^{T_2}_{B}\mathbf{T} \cdot {}^{T_1}_{B}\mathbf{T}^{-1} = {}^{T_2}_{T_1}\mathbf{T} \tag{6}$$

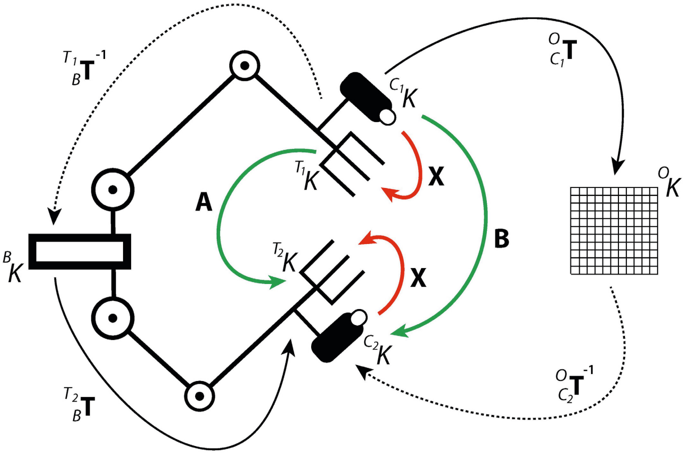


Fig. 8. The hand-eye calibration according to Tsai [18] considers two poses of the robot tool point with respect to the camera

is known and contains the transformation of the two robot tool points. The matrix $\mathbf{X} = {}^{T}_{C}\mathbf{T}$ is unknown.

A variant of the hand-eye calibration [20, 24] is defined by the equation system

$$\mathbf{AX} = \mathbf{YB}, \tag{7}$$

whereupon in addition to the calculation of the hand-eye calibration, an estimation of the robot pose relating to the world coordinate system takes place. The transformation $\mathbf{X} = {}^{C}_{T}\mathbf{T}$ is the unknown hand-eye transformation, $\mathbf{A} = {}^{O}_{C}\mathbf{T}$ the camera pose relating to the world coordinate system ${}^{O}K$ and $\mathbf{B} = {}^{B}_{T}\mathbf{T}$ the pose of the robot tool relating to the robot's base coordinate system ${}^{B}K$. Both $\mathbf{A}$ and $\mathbf{B}$ are known. Further unknown is the transformation

$$\mathbf{Y} = {}^{O}_{B}\mathbf{T} \tag{8}$$

which correlates the world coordinate system with the robot base coordinate system. The variant of the hand-eye calibration considers the kinematic chain of a single view point resulting in two unknown magnitudes. In contrast, the hand-eye calibration according to Tsai one unknown magnitude is eliminated by considering two separate view points. For a detailed description of the hand-eye calibration and the solution of the problem statement the reader is referred to [6].

The literature confirms that the variant of the hand-eye calibration [20, 24] produces more accurate results [15] in comparison to the hand-eye calibration according to Tsai. The Camera Calibration Toolbox for Matlab [2] and DLR CalLab [16] tools were used to perform the hand-eye calibration. Both tools support the described variant of the hand-eye calibration.

Downsampling

The information redundancy arises from fusing the point clouds during the 3D reconstruction process; it has to be reduced in order to ensure efficiency within the subsequent calculation steps. Let $\mathcal{P}$ be a point cloud and $\mathbf{p}_{\min} \in \mathcal{P}$ the point with the smallest coordinate values of all points in $\mathcal{P}$. Analogously, let $\mathbf{p}_{\max}$ be the point with the largest values. Then $\mathbf{p}_{\min}$ and $\mathbf{p}_{\max}$ span a cube H which can be split in disjunctive quads Q_i,

$$H = \dot{\bigcup}_{i \in I} Q_i, \tag{9}$$

resulting in a regular grid in three dimensional space. The reduction of the point cloud is performed by substituting the points in each quad by their centroid

$$\operatorname{cent}(Q_i) = \frac{1}{|Q_i|} \cdot \sum_{m=1}^{|Q_i|} \mathbf{p}_m, \quad \mathbf{p}_m \in Q_i. \tag{10}$$

The calculated centroids form the resulting (reduced) point cloud [13].

3.3 Load Detection

The typical package unit in logistics has a quad-like shape. As a loading situation consists of a composition of multitude of package units, an atomic loading unit in terms of a quad-like form constitutes an efficient transport solution. The detection of the atomic loading unit is described in the following.

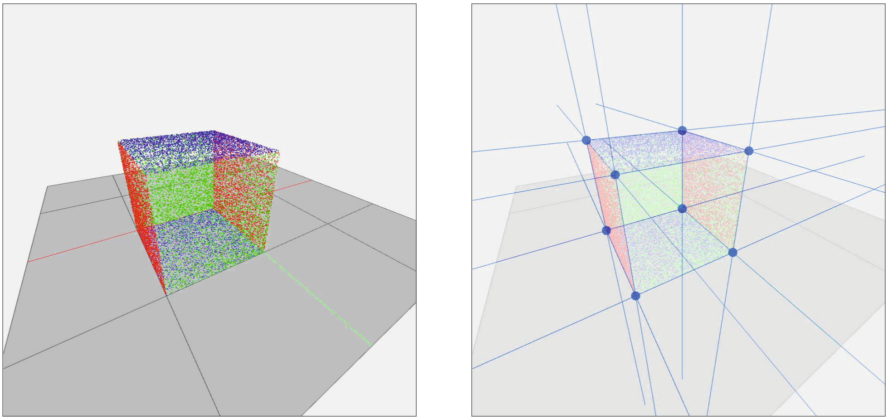
A quad configuration Q is represented as $Q = (\alpha, \beta, \gamma, s_x, s_y, s_z, t_x, t_y, t_z)$. The pose of the quad is described by the Eulerian angles α, β, γ and the translation vector $\mathbf{t} = (t_x, t_y, t_z)^T$. The scaling parameters s_x, s_y, s_z represent the quad dimensions. The quad detection algorithm is based on the extraction of the isolated side faces in terms of planes. Afterwards the quad corner points are extracted by intersection calculation of the planes, followed by an extrapolation of the corner points. Figure 9 exemplary illustrates the process of quad extraction.

Let $\mathcal{P}$ be a point cloud representing a quad. In the first step, the Ranndom Sample Consensus-algorithm (RANSAC) [5] is applied in order to split the point cloud into subsets $\mathcal{P}_i$

$$\mathcal{P} = \dot{\bigcup}_{i \in I} \mathcal{P}_i \cup \mathcal{Q}, \tag{11}$$

each of them representing an individual plane. An characteristic property of the RANSAC algorithm is the rejection of outliers. Three planes are necessary for a quad detection at the minimum. The RANSAC algorithm assigns each subset $\mathcal{P}_i$ of $\mathcal{P}$ a plane model $E_i : \mathbf{n}_i \cdot \mathbf{x} = d$, represented in hessian normal form. A modification of the RANSAC algorithm, the M-estimator Sample Consensus (MSAC) [17], enhances the accuracy of the fitted models without additional computational effort.

Based on the extracted plane models for the side faces, the corner points of the quad are determined. A corner point can be calculated by selecting three pairwise



(a) Synthetic point cloud (b) Corner detection

Fig. 9. The images illustrate the load detection using the example of a synthetically constructed point cloud quad (a). After detecting the side faces and extracting edge points (blue spheres), the result of the load detection is shown in the right image

orthogonal plane models and computing their intersection. The determination of all corner points assumes the extraction of all plane models for the six side faces of the quad. In practice it is often an impossible task to capture a loading unit completely due to occlusions. Hence, in these cases an extrapolation of missing corner points is performed.

4 Evaluation

This section deals with the experimental analysis of the hand-eye calibration (Sect. 4.1) and load detection (Sect. 4.2) algorithms with respect to accuracy and run time.

The analysis was carried out using the KUKA KR 125/2 robot which is equipped with a vacuum gripper to gasp the loading units (s. Fig. 5(a)). On the gripper, the Asus Xtion Pro Live is mounted in order to be able to capture the load configuration in different views (s. Fig. 5(b)). The robot and the camera are connected to an off-the-shelf x86 compatible desktop computer for controlling the robot's movement and running the proposed algorithms. It is equipped with an AMD Phenom II X4 955 BE quad core processor, 8 GB main memory, an AMD Radeon HD 7750 graphics card and Microsoft Windows 7 (64 bit).

4.1 Hand-Eye Calibration

The hand-eye calibration (s. Sect. 3.2) was computed and qualitatively compared using the Camera Calibration Toolbox for Matlab [2] (employing the methodology described in [22]) and the DLR Camera Calibration Toolbox (CalLab and

CalDe) [16] (using the methods explained in [15]). The methodologies have already been presented in Sect. 3.2.

The floor plane is henceforth used as the basis of comparison which also serves as the mounting point for the robot. For simplicity, it is assumed that the floor plane coincides with the xy-plane of the robot's base coordinate system. Additionally, the base coordinate system is supposed to be located in the angular point of the robot's bottom attachment. Given the merged point cloud $\mathcal{P}$, the orthogonal projection $\mathbf{p}_i$ of $\hat{\mathbf{p}}_i \in \mathcal{P}$ onto the xy-plane (which is assumed to be the floor plane) implies a distance $\|\mathbf{p}_i - \hat{\mathbf{p}}_i\|$. Using the distances for all points of $\mathcal{P}$, the mean squared error (MSE) is defined as

$$\varepsilon_{\text{MSE}} = \frac{1}{|\mathcal{P}|} \cdot \sum_{i=1}^{|\mathcal{P}|} \|\mathbf{p}_i - \hat{\mathbf{p}}_i\| \,. \tag{12}$$

To give stronger weightings to larger distances, the root mean squared error (RMSE)

$$\varepsilon_{\text{RMSE}} = \frac{1}{|\mathcal{P}|} \cdot \sum_{i=1}^{|\mathcal{P}|} (\mathbf{p}_i - \hat{\mathbf{p}}_i)^2 \tag{13}$$

is even more suitable. If the floor plane actually coincides with the xy-plane, ε_{MSE} and $\varepsilon_{\text{RMSE}}$ become zero. Unfortunately, the aforementioned assumptions are only partly applicable in practice. First, the floor plane does not exactly coincide with the xy-plane since the robot is mounted on a circular disk (having a height of approx. 45 mm) which, in turn, is mounted on the floor plane. To roughly compensate this issue, the xy-plane is translated by 45 mm along the z-axis. Additionally, sensor noise and asperities of the floor plane and the robot's base mounting further void the assumption of coincidence. However, since these issues affect all computations in the same way, eqn. 12 and 13 can still be used to assess and compare the performance of different hand-eye calibration techniques.

Table 3 depicts the errors ε_{MSE} and $\varepsilon_{\text{RMSE}}$ (cf. last two columns) for the hand-eye calibration using the Matlab and DLR tool. The first column "TEST" states the number of the test series for reference. "METHOD" lists the 2 calibration methods along with the uncalibrated result for the sake of comparison. They have been used to compute the hand-eye calibration matrix $\mathbf{X}$ (s. eqn. 4) given the number of "RECORDINGS" (3rd column). "MAX" indicates the distance of the point having the largest distance to the xy-plane. Each row employs a test set of 6 captured views of the floor plane which are shown in Fig. 10. The merged point clouds have 1 637 067 points in total.

As it can be seen from Tab. 3, the results of the Camera Calibration Toolbox for Matlab are considerably inferior to the results of the DLR tool even though more input images for computing the hand-eye calibration have been provided to Matlab (s. 3rd column). Also note that, within the test series 2 and 3, 6 and 8 input images have been rejected during the computation of the hand-eye calibration respectively. This is caused by the inability of appropriately detecting and localizing the features of the calibration pattern. The fourth test series is

Table 3. Calibration errors using a (fixed) test set of 6 views of the floor plane depicted in Fig. 10; "TEST" states the reference number of the test series. "METHOD" lists the 2 calibration methods along with the uncalibrated result for comparison. "RECORDINGS" contains the number of input images from different views that have been used for calibration. "MAX" indicates the distance of the point having the largest distance to the xy-plane. The last two columns state the calibration error in terms of eqn. 12 and 13

Test	Method	Recordings	Max. (mm)	ε_{MSE} (mm)	ε_{RMSE} (mm)
0	Uncalibrated	none	2333.00	1011.94	1032.45
1	Matlab [2]	8 (8)	210.43	32.93	40.90
2		18 (26)	78.71	34.99	37.29
3		26 (32)	69.22	20.53	21.93
4	DLR [16]	3 (3)	1488.70	575.45	753.62
5		5 (5)	52.57	13.52	15.31
6		10 (10)	51.28	11.33	13.19

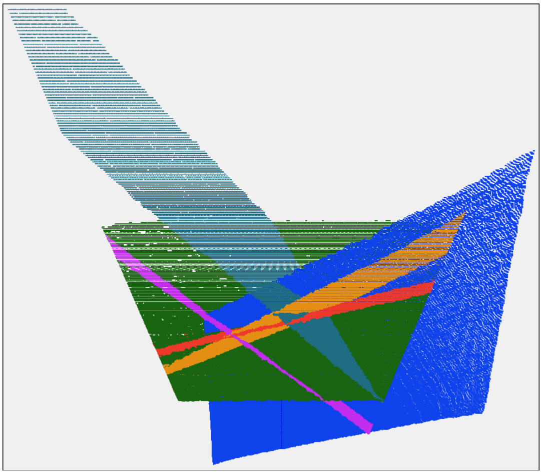
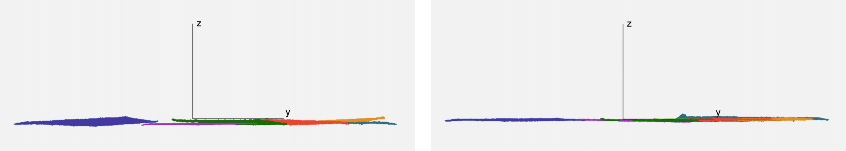
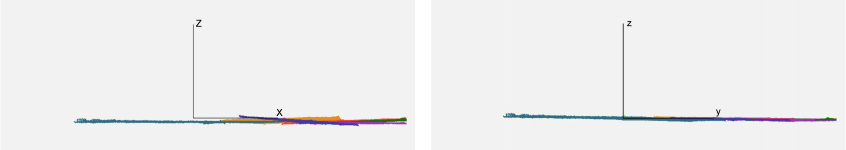


Fig. 10. Isometric side view of the 6 point clouds representing the floor plane captured from different views using the ASUS Xtion Pro mounted on the robot's gripper tool. Each point cloud is colored individually

an exception of this observation and can be justified by the insufficient number of (just 3) input images. From the fifth and sixth test series, one may conclude that 5 recordings are sufficient since the computed errors are approximately the same. In general, the DLR tool outperforms Matlab which may also be visually confirmed by Fig. 11. It shows the isometric representations of the lateral and frontal views of the calibration results by means of test series 2 (Matlab) and 5 (DLR). In summary, test series 6 achieved the best results with an root mean squared error of about 1.3 cm.



(a) Matlab: lateral view (b) DLR: lateral view



(c) Matlab: frontal view (d) DLR: frontal view

Fig. 11. Exemplary isometric representation (lateral and frontal view) of the calibration results (i. e., merged point clouds) using Matlab (test 2) and DLR (test 5). The DLR tool clearly outperforms Matlab which may be confirmed visually, cf. Fig. 11(b) and (d)

4.2 Load Detection

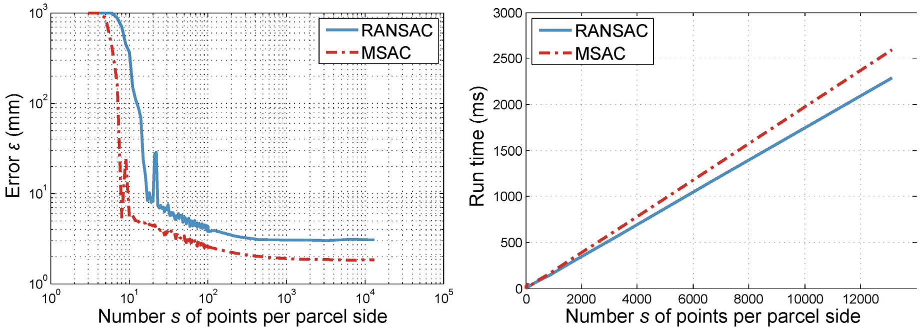
Recall the load detection algorithm of Sect. 3.3 which is essentially composed of (1) plane (model) fitting using RANSAC [5] or MSAC [17], and (2) the computation and verification of intersecting points given the fitted planes. Within this section, its run time and accuracy is analyzed. Additionally, trade-off considerations are presented and the impact of different parameter values is discussed.

The foundation of the analysis is a unit cube of size 1 × 1 × 1 m whose faces have been sampled randomly to generate point clouds for evaluation purposes. The unit cube simulates a parcel on a euro pallet. The use of synthetically generated data allows for many variations during the test, simplifies repeatability and ensures reproducibility. In addition, the detected results (corner points) can simply be checked by comparing the corner points of the (known) input cube.

Let therefore $\mathcal{C}_d$ be the point set of detected corner points and $\mathcal{C}_t$ the set of truth corner points respectively. Then the error ε of the detected cube is given by

$$\varepsilon = \frac{1}{|\mathcal{C}_d|} \cdot \sum_{\forall \mathbf{c}_d \in \mathcal{C}_d} \left(n(\mathbf{c}_d) - \mathbf{c}_d\right)^2 \tag{14}$$

where $n(\mathbf{c}_d) \in \mathcal{C}_t$ denotes the nearest neighbor of $\mathbf{c}_d$ in $\mathcal{C}_t$. Thus, eqn. 14 computes the quadratic mean of distances from detected corner points $\mathbf{c}_d$ to their nearest neighbor $n(\mathbf{c}_d)$. This way, ε specifies the similarity (in millimeter) of the detected cube compared to the generated input cube. If the detection fails (e. g., due to an insufficient number of detected planes), ε is set to the constant penalty value 1000 mm.



(a) Accuracy vs. #points per side

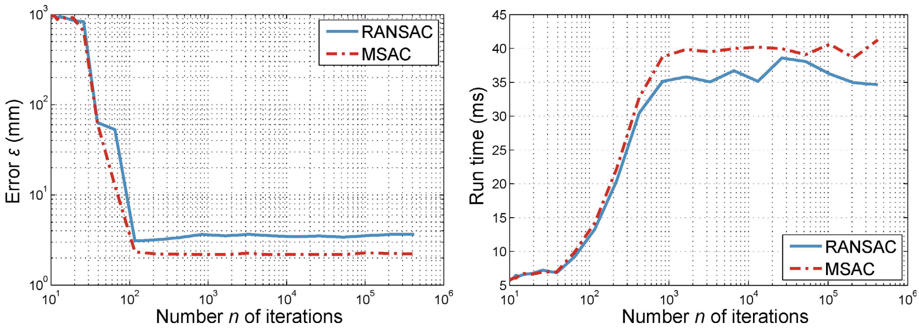
(b) Run time vs. #points per side

Fig. 12. Trade-off considerations: point cloud size vs. (a) accuracy and (b) run time. The number s of points per parcel face varied from 3 to 13 000

For the sake of brevity, a set of default parameters for the detection algorithm is used throughout this section if not explicitly stated otherwise. More precisely, the number n of RANSAC (or MSAC) iterations is set to 1500. The distance threshold Θ defaults to 30 mm. Since each face of the input cube is sampled randomly, the default number s of sample points per (parcel) side defaults to 200. Finally, each experiment was repeated 100 times and the results (i. e., the error ε and the run time) were averaged to account for variations during the model fittings (RANSAC or MSAC) and different system loads.

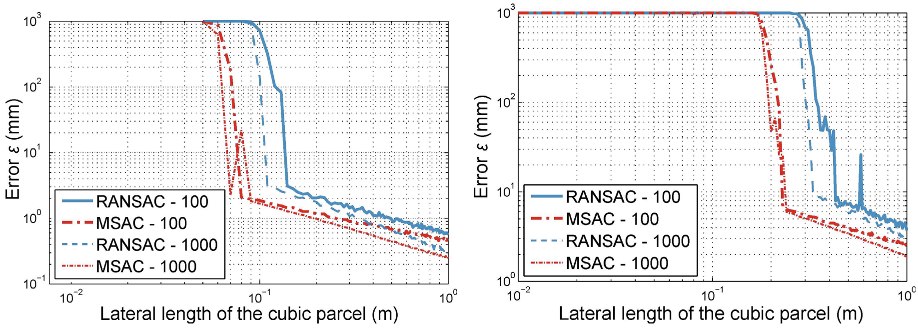
Trade-off: Point Cloud Size vs. Accuracy and Run Time In a first step, the influence of varying point cloud sizes is analyzed. Results are depicted in Fig. 12. Recall that $\varepsilon = 1000$ mm denotes the fixed penalty value, i. e., no detection was possible in such cases. The number s of points per parcel varied from 3 to 13 000 and the residual parameters have been set to their defaults. On the one hand, Fig. 12(a) reveals that the error ε (s. eqn. 14) steadily decreases when s approaches 1000. On the other hand, Fig. 12(b) shows that the run time scales linearly with s. For $s = 1000$, the run time stays beneath 250 ms and the error attains its minimum. Comparing RANSAC (blue) and MSAC (red) with each other shows that MSAC approaches a slightly higher accuracy but also demands more time. The latter becomes even more evident in case of large numbers of points per parcel side (e. g., $s = 12\,000$). It can be concluded from the figures that both RANSAC and MSAC provide stable results for $s \geq 100$ with an error less than 5 mm.

Trade-off: Number of Iterations vs. Accuracy and Run Time Figure 13 presents trade-off considerations with respect to the number of iterations for RANSAC (and MSAC) vs. the resulting accuracy and processing time. Consequently, the number n of iterations varied from 1 to 420 000 while the remaining parameters were set to their defaults. Figure 13(a) illustrates how the accuracy roughly remains constant after $n = 100$ iterations. Nevertheless, Fig. 13(b) shows that



(a) Accuracy vs. iteration count (b) Run time vs. iteration count

Fig. 13. Trade-off considerations: number of iterations vs. (a) accuracy and (b) run time. The number of points per parcel face is $s = 200$ and the number n of iterations in the RANSAC (blue) and MSAC (red) varied from 1 to 420 000



(a) $\Theta = 10\,\text{mm}$ (b) $\Theta = 30\,\text{mm}$

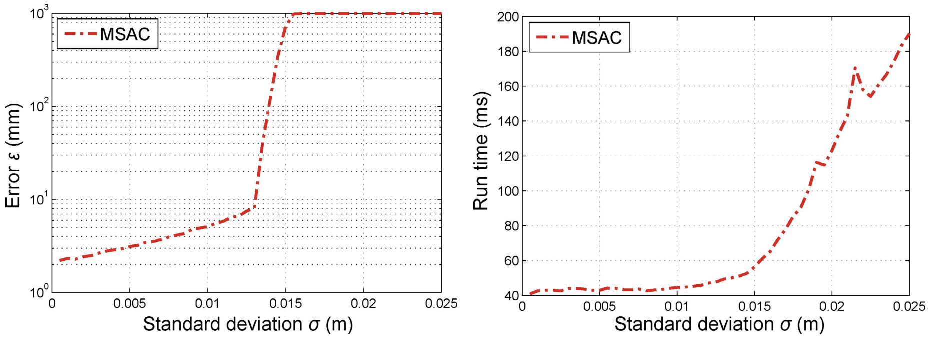
Fig. 14. Trade-off considerations: parcel dimensions vs. accuracy. The number s of points per parcel was set to 100 (solid) and 1000 (dashed). In (a), the threshold for RANSAC / MSAC was set to $\Theta = 10\,\text{mm}$ and the results in (b) were produced using $\Theta = 30\,\text{mm}$. The cubic parcel dimensions varied from 0.01 to 1 m

the run time still increases until $n = 1000$ (for both RANSAC and MSAC). Afterwards, the processing time settles down to approx. 40 ms. This observation may suggest to set $n \approx 100$ since no more accuracy enhancements are expected.

Trade-off: Parcel Dimensions vs. Accuracy Figure 14 depicts the accuracy in terms of varying parcel dimensions. More specifically, the size of the generated cube varied from 0.01 to 1 m and n was set to 1500. In Fig. 14(a) Θ was set to 10 mm and in Fig. 14(b) $\Theta = 30\,\text{mm}$ was used. Additionally, s was set to 100 (solid) and 1000 (dashed). MSAC fails if the parcel length approximately approaches 23 cm while RANSAC already fails when the parcel length reaches 35 cm (s. Fig. 14(b)). Figure 14(a) roughly shows the same results except that both RANSAC and MSAC require a relatively larger parcel length in order to

Table 4. Robustness in terms of randomly transformed point clouds representing the parcel which has to be detected; in total, 1000 transformations were used. The number of points per parcel face was set to $s = 100$. The results demonstrate the invariance of the parcel position towards the detection rate

Algorithm	**Transformed**	**Mean (mm)**	**Std. dev. (mm)**
RANSAC [5]	✗	4.09	1.38
	✓	4.14	1.44
MSAC [17]	✗	2.58	0.46
	✓	2.61	0.47



(a) Accuracy vs. Gaussian noise (b) Run time vs. Gaussian noise

Fig. 15. Robustness in terms of simulated Gaussian noise by varying the standard deviation σ

work properly. This may be justified by the reduced accuracy which is associated with the higher value of Θ (allowing more points to be used for setting up the consensus set). The impact of s is neglectable here.

Trade-off: Randomly Transformed Point Clouds vs. Accuracy Table 4 depicts the results of analyzing how accuracy is influenced when transforming the merged point cloud $\mathcal{P}$ which represents the final (fused) view of the generated unit cube. This way, the cube's point cloud may exhibit arbitrary poses (i. e., positions and orientations). The first column states the two algorithms for the model fitting and the second column indicates whether a random transformation has been applied. Note that the tests were repeated 1000 times and the results were averaged. Comparing the last two columns with each other clearly proves the invariance towards pose changes.

Trade-off: Simulated Gaussian Sensor Noise vs. Accuracy and Run Time To account for varying sensor noise, Fig. 15 visualizes the results of applying Gaussian noise during the generation of the unit cube. More specifically, each point was modified alongside its normal by employing a normal distribution with

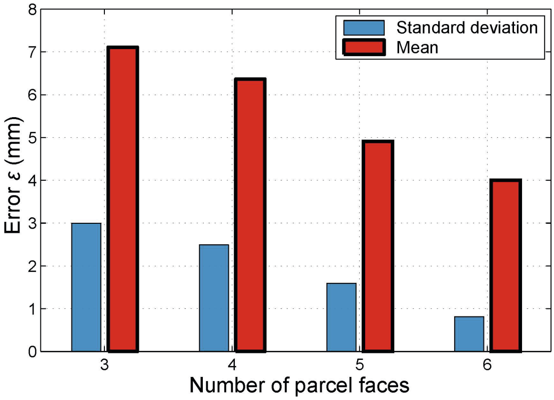


Fig. 16. Accuracy in terms of varying numbers of detected parcel faces; the number of points per parcel face was set to $s = 800$ and Gaussian noise was added to each point alongside the normal vector with a standard deviation of 10 mm

varying standard deviation σ. With increasing standard deviation up to approx. $\sigma = 0.013$ m, the error increases continuously but does not exceed 10 mm. When the standard deviation reaches approx. half of the threshold value Θ (i. e., $\Theta = 15$ mm), a correct detection is not possible anymore (visualized in (a) by the constant penalty value $\varepsilon = 1000$ mm). Also notice how the run time increases when σ increases as well. This is justified by the fact that the consensus set converges more slowly. Hence, more iterations are needed which explains the increase of run time.

Trade-off: Number of Detected Parcel Faces vs. Accuracy Finally, the extrapolation performance in case of missing corner points is assessed (s. Sect. 3.3). Figure 16 summarizes the errors ε (eqn. 14) for different numbers of detected parcel faces. The detection error is significantly affected by the number of given parcel faces. However, even in case of only 3 given parcel faces (the minimum), the mean detection error stayed beneath 7 mm. As expected, the best results are achieved in case of 6 given parcel faces since no extrapolation is needed.

5 Conclusion and Outlook

Two different novel approaches for the detection of parcel loading positions on a pallet were presented. Based on the predetermined model of the loading situation and the raw data of low cost 3D sensors, the innovative registration algorithms produce accurate results regarding the position detection. The first approach (s. Sect. 2.3) is broadly evaluated [21, 10, 9] and constitutes a competitive alternative in comparison to existing complex and costly 3D-vision systems. The presented solution reduces the costs of contour checking in the automated depalletizing process from an economic point of view concurrently having a moderate run time (less than 1 s on the employed test system). Within the scope of application, an accuracy of 10 mm (MESA) and 26 mm (ifm) regarding the gripping position is adequate for a robust handling of 300 × 400 mm parcels. Main advantage is the high frame rate (25 fps) of the static mounted sensor as well as the parallel processing of the detection pipeline (s. Fig. 2) while the robot is handling the loading objects.

In the second approach (s. Sect. 3), the Asus Xtion is mounted at the robot's gripper, acquiring data from dynamic poses. Comparing two different camera calibration methods, the evaluation shows better results for the DLR tool with an root mean squared error of about 1.3 cm. Within the evaluation a deviation of 7 mm was ascertained for the accuracy of gripping points while using data from a single picture of the sensor (with 3 parcel faces). This inaccuracy could even be reduced to 5 mm by taking an additional picture to get 5 parcel faces into focus. Concerning run times the evaluation shows that an error minimizing number of points per side could be processed in less than 250 ms. A full 3D mapping of the detected object, including undercuts and lateral gabs, is the main advantage of this approach.

Referring to industrial applications in logistic systems overall cycle times (performance) and systems cost are crucial. Additional movement of the handling device (robot / sensor) to acquire the necessary data for a full 3D scan does not seem reasonable for a rather simple task like contour check of box-shaped objects. But using the developed approximation method even one picture is sufficient to provide an accuracy higher than the first presented approach. Even though time for data acquisition (depends on the actual position of the robot) and time for processing the detection pipeline have to be added, achievable handling speed seems to promise a potential competitive solution. Always taking into account that only synthetic generated data are the basis of the performed tests.

Besides the evaluation of the second approach in a real environment, a combined (parallel) positioning of the gripper and dynamic data acquisition should be investigated in the future to proof the practical potential. Additionally, use cases with more complex detection tasks will be researched to evaluate other application fields for the presented method.

Acknowledgement. The presented results have been obtained during the solving of IGF research project 16916N supported by the German Federal Ministry of Economics and Technology.

References

[1] Besl, P., McKay, N.D.: A Method for Registration of 3-D Shapes. IEEE Transactions on Pattern Analysis and Machine Intelligence 14(2), 239–256 (1992)

[2] Bouguet, J.: Camera calibration toolbox for matlab (2010), `http://www.vision.caltech.edu/bouguetj/calib_doc/` (last visited April 15, 2014)

[3] De Meyer, A.: Manufacturing Operations in Europe: Where Do We Go Next? European Management Journal 16(3), 262–271 (1998)

[4] Echelmeyer, W., Kirchheim, A., Wellbrock, E.: Robotics-logistics: Challenges for automation of logistic processes. Automation and Logistics 1-3, 2099–2103 (2008)

[5] Fischler, M., Bolles, R.: Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. Communications of the ACM 24(6), 381–395 (1981)

[6] Horaud, R., Dornaika, F.: Hand-eye calibration. International Journal of Robotics Research 14(3), 195–210 (1995)

[7] Kaiser, R., Baier, G.: Neue Lösungen zum Einsatz von Robotik bei der Handhabung konplexer Güter. In: Roboter in der Intralogistik, pp. 91–105. Fraunhofer IPA Technologieforum F (2009)

[8] Prasse, C., Skibinski, S., Weichert, F., Stenzel, J., Müller, H., ten Hompel, M.: Concept of automated load detection for de-palletizing using depth images and rfid data. In: Computing and Engineering (ICCSCE) Control System, pp. 249–254 (2011)

[9] Prasse, C., Skibinski, S., Weichert, F., Stenzel, J., Müller, H., ten Hompel, M.: Concept of Automated Load Detection for De-Palletizing Using Depth Images and RFID Data. In: 2011 IEEE International Conference on Control Systems, Computing and Engineering (ICCSCE 2011) (2011)

[10] Prasse, C., Stenzel, J.R.B., Weichert, F., Müller, H., ten Hompel, M.: Low cost contour check of loading units using pmd sensors. In: International Conference on Sensing Technology, ICST 2013 (2013)

[11] Ruske, K.D., Kauschke, P., Reuter, J., von der Gracht, H., Gnatzy, T., Darkow, I.L.: Transportation & Logistics 2030 - Volume 3: Emerging Markets New hubs, new spokes, new industry leaders? Tech. rep., SMI (Supply Chain Management Institute) and Price Waterhouse Cooper (2010)

[12] Rusu, R.B., Beetz, M., Marton, Z.C., Blodow, N., Dolha, M.: Towards 3d point cloud based object maps for household environments. Robotics and Autonomous Systems Journal (Special Issue on Semantic Knowledge) (2008)

[13] Rusu, R.B., Cousins, S.: 3d is here: Point cloud library (pcl). In: IEEE International Conference on Robotics and Automation (ICRA), Shanghai, China (2011), `http://pointclouds.org/`

[14] Schuh, G., Wemhöner, N., Friedrich, C.: Scenario-based Lifecycle Analysis of Manufacturing Systems. CIRP - Journal of Manufacturing Systems 35(2) (2006)

[15] Strobl, K., Hirzinger, G.: Optimal hand-eye calibration. In: IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 4647–4653 (2006)

[16] Strobl, K.H., Hirzinger, G.: More accurate pinhole camera calibration with imperfect planar target. In: 2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops), pp. 1068–1075 (2011)

[17] Torr, P., Zisserman, A.: MLESAC: A New Robust Estimator with Application to Estimating Image Geometry. Computer Vision and Image Understanding 78(1), 138–156 (2000)

[18] Tsai, R.Y., Lenz, R.K.: A new technique for fully autonomous and efficient 3d robotics hand/eye calibration. IEEE Transactions on Robotics and Automation 5(3), 345–358 (1989)

[19] VDMA: VDMA Forum Intralogistik erwartet für 2011 Umsatzplus von neun Prozent, `http://www.vdma.org` (last visited April 15, 2014)

[20] Wang, C.C.: Extrinsic calibration of a vision sensor mounted on a robot. IEEE Transactions on Robotics and Automation 8(2), 161–175 (1992)

[21] Weichert, F., Skibinski, S., Stenzel, J., Prasse, C., Kamagaew, A., Rudak, B., ten Hompel, M.: Automated detection of euro pallet loads by interpreting pmd camera depth images. Logistics Research 6(2-3), 99–118 (2013)

[22] Wengert, C.: Fully automatic camera and hand to eye calibration, `http://www.vision.ee.ethz.ch/software/calibration_toolbox/calibration_toolbox.php` (last visited April 15, 2014)

[23] Zhang, B., Skaar, S.: Robotic de-palletizing using uncalibrated vision and 3D laser-assisted image analysis. In: Intelligent Robots and Systems, October 10-15, pp. 3820–3825 (2009)

[24] Zhuang, H., Roth, Z.S., Sudhakar, R.: Simultaneous robot/world and tool/ange calibration by solving homogeneous transformation equations of the form ax=yb. IEEE Transactions on Robotics and Automation 10(4), 549–554 (1994)

Circularly Moving Sensor for Use of Modulation Effect - CAROUSEL

Masako Kishida and Yusuke Hioka

University of Canterbury,
Private Bag 4800,
Christchurch 8140, New Zealand
`masako.kishida@canterbury.ac.nz`

Abstract. Estimation of the Direction Of Arrival (DOA) of a signal is essential in various applications such as phased array antennas for communications, microphone arrays for audio and room acoustics, and Sonars for underwater acoustics. In general, sensor arrays used in those applications consist of several to hundreds of spatially distributed sensors, which could be expensive. This chapter is devoted for a recently introduced DOA estimation architecture, CAROUSEL (CirculARly mOving sensor for USE of moduLation effect). The unique feature of CAROUSEL is its requirement of only a single sensor. DOA estimation is achieved by moving the sensor in a circular path to cause a phase modulation in the signal so that the observed signal contains the DOA information. Algorithms for CAROUSEL have been developed and simulated for the application to estimate the DOA of sinusoidal and harmonic signals. (This chapter is based on the authors' conference papers [1,2].)

Keywords: DOA estimation, sensor array, single sensor, circular motion, Doppler effect, phase modulation.

1 Introduction

Direction Of Arrival (DOA) estimation is a problem to find the direction of the source of arriving signals such as electromagnetic, acoustic, mechanical, hydroacoustic, and seismic signal. The DOA estimation problems appear in various applications; phased array antennas for communications [3], microphone arrays for audio and room acoustics [4], and Sonars for underwater acoustics [5], to name a few.

Conventional approaches to DOA estimation compare the phase of signals observed at different locations at the same time (Fig. 1). This idea utilizes the fact that the phase of a signal depends on the distance between the sensor and the signal source. For this reason, DOA estimation requires a sensor array that consists of a number of spatially distributed sensors to collect signals, which are then used in algorithms such as beamforming [6], generalized cross-correlation [7], MUSIC [8] and ESPRIT [9] to find DOA.

A. Mason et al. (eds.), *Sensing Technology: Current Status and Future Trends IV*,
Smart Sensors, Measurement and Instrumentation 12, DOI: 10.1007/978-3-319-12898-6_11

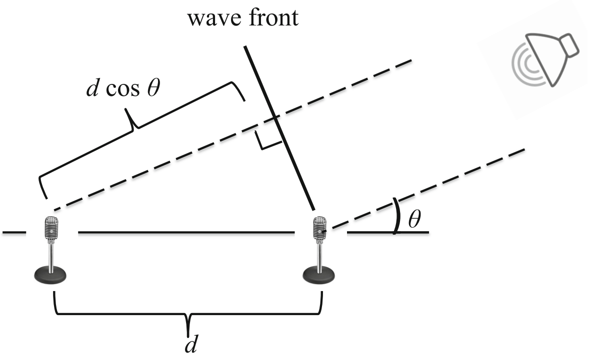


Fig. 1. DOA estimation by sensor array

However, one of the critical drawbacks of using sensor arrays is its cost; using a large number of sensors requires a large number of associated elements such as power supplies, cabling, and AD/DA converters (given that the technique is implemented by digital signal processing). Computational cost for the algorithms to process the large volume of data is also not negligible. This has constrained the application of sensor array technologies for commercial products, especially for personal use, e.g. microphone arrays for hands-free communication. Various commercial products utilizing sensor array technologies have recently appeared in the market thanks to the lowering prices of transducers and digital signal processors, yet, it is still challenging to achieve satisfactory performance using a small number of sensors.

This motivated recent study of using sensor arrays in motion, which aims at mitigating the detrimental effect of spatial aliasing [10]. One could imagine that if a sensor continuously moves in space and observes signals in different locations, then it would eventually play a role as a sensor array. Schasse et. al [11], [12] proposed *circular sampling* which samples data of a circular array in a round robin fashion to induce a nonlinear modulation in the observed signal.

This chapter is devoted for the recently introduced DOA estimation architecture, *CAROUSEL* (CirculARly mOving sensor for USE of moduLation effect). The idea is to utilize the Doppler effect, the change in frequency of a signal when the observer and source are relatively in motion (Fig. 2). To this end, CAROUSEL has a single sensor moving along a circumference of a circle. Similar to Schasse's approach, sensing by CAROUSEL induces phase modulation in the observed signal that contains information of the DOA in its message, which is then demodulated to estimate the DOA.

The rest of this chapter is organized as follows. After introducing CAROUSEL architecture along with signal models in Section 2, algorithms to estimate DOA by using CAROUSEL are proposed in Section 3. In Section 4, discrete-time

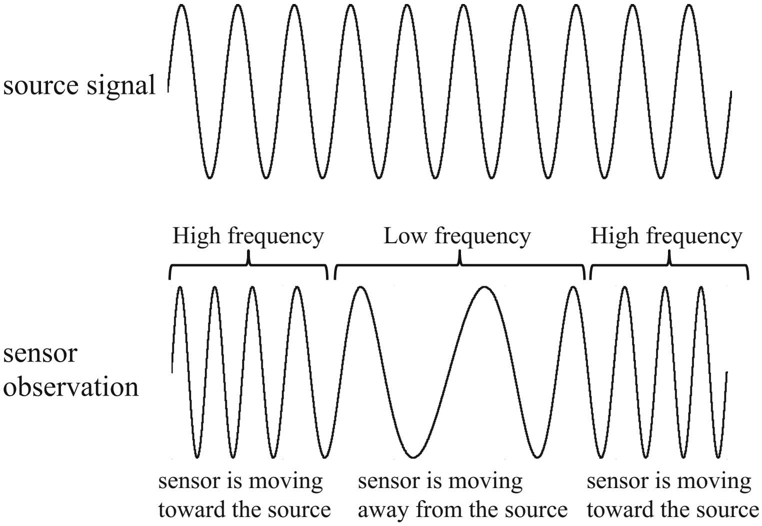


Fig. 2. By the Doppler effect, the observed signal frequency changes if the sensor is moving. With the CAROUSEL architecture, one could estimate the DOA of the signal by detecting the amount of the phase shift in signals.

implementation is discussed. Section 5 presents simulation results that show the effectiveness of the proposed method and finally Section 6 concludes the chapter.

Notations

Notations are standard. Bold font indicates a vector. $\mathbf{p}^T$ denotes the transpose of a vector $\mathbf{p}$, $\dot{\mathbf{p}}$ denotes the time derivative of $\mathbf{p}$, $\mathrm{Re}\{s\}$ is the real part of s. mod $(a, b) := a-(b\lfloor a/b \rfloor)$, with $\lfloor x \rfloor$ denoting the largest integer not greater than x. c is the speed of the wave propagation.

2 Modeling: CAROUSEL

This section introduces CAROUSEL architecture and how to use CAROUSEL to estimate DOA.

2.1 Sensor Setup

Schematic of CAROUSEL setup is shown in Fig. 3. A sensor moves along the circumference of a circle of radius r centered at the origin of the xy-plane. Let the x-axis direction be the direction from the origin to the sensor at $t = 0$ so that the sensor's azimuth angle at $t = 0$ is $\theta(0) = 0$. The sensor location at time $t \geq 0$ can be written in terms of the azimuth angle as

$$\mathbf{p}(t) = [r\cos\theta(t),\ r\sin\theta(t),\ 0]^T. \tag{1}$$

The look direction vector from the origin to a source located in the direction of the azimuth angle θ_s and the polar angle ϕ_s is

$$\mathbf{u} = [\cos\theta_s \sin\phi_s,\ \sin\theta_s \sin\phi_s,\ \cos\phi_s]^T. \tag{2}$$

For simplicity, it is assumed that the source location is fixed. However if the moving speed of the source is slow enough so that the observed frequency at the origin is negligible, approaches in this chapter can be used to track a moving source without modification.

2.2 Sinusoidal Signal Modeling

Suppose that there is a signal source located sufficiently far from the sensor that emits a sinusoidal signal, which has a constant amplitude A and a constant frequency f_0 Hz so that the signal's angular frequency is $\omega_0 = 2\pi f_0$ rad/s. Then, the equation of the signal observed at the origin is described by the plane wave model [6],

$$s_{\text{origin}}(t) = \text{Re}\{A\exp\left(j\psi_{\text{origin}}(t)\right)\}, \tag{3}$$

where

$$\psi_{\text{origin}}(t) = \omega_0 t + \psi_0,$$

and ψ_0 is the phase angle of the signal at the origin at time $t = 0$. The time difference of arrival of the signal between the origin and the position of the sensor can be modeled from geometry, and the signal observed by the moving sensor at time t with respect to the phase of the signal at the origin $s(t)$ is

$$s(t) = \text{Re}\{A\exp\left(j\psi(t)\right)\}, \tag{4}$$

where $\psi(t)$ is the instantaneous phase angle of the signal received by the moving sensor

$$\begin{aligned}\psi(t) &= \omega_0\left[t + \frac{\mathbf{p}^T(t)\mathbf{u}}{c}\right] + \psi_0\\ &= \omega_0\left[t + \frac{r}{c}\sin(\phi_s)\cos(\theta(t) - \theta_s)\right] + \psi_0.\end{aligned}$$

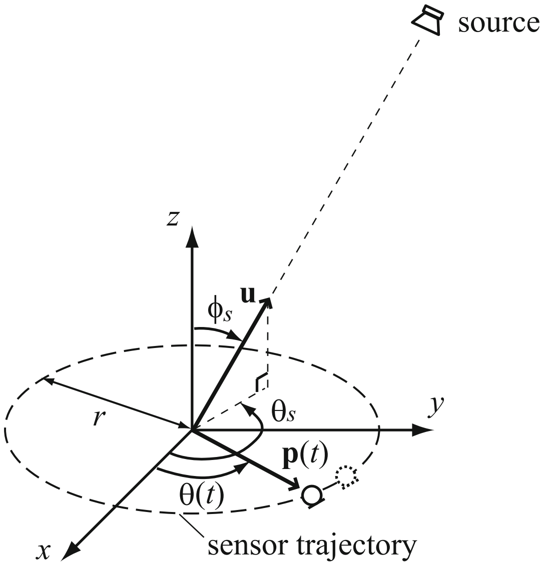


Fig. 3. DOA estimation problem using CAROUSEL

This implies that the observed signal is in the form of a phase modulated signal

$$s(t) = \mathrm{Re}\{A \exp\left[j\left(\omega_0 t + \psi_0 + m(t)\right)\right]\},$$

where its carrier is

$$A \exp\left[j\left(\omega_0 t + \psi_0\right)\right],$$

and its message is

$$m(t) = \frac{r\omega_0}{c} \sin(\phi_s) \cos(\theta(t) - \theta_s). \tag{5}$$

This phase modulation is caused by the circular motion of the sensor. If the sensor is stationary, $\theta(t)$ is constant yielding a constant $m(t)$, and there is no modulation.

2.3 Harmonic Signal Modeling

Now consider the case where the source emits a stationary sinusoid with its $L-1$ harmonic components. Then, the signal observed at the origin can be modeled by the superposition of the pitch and its $L-1$ harmonic components as

$$s_{\mathrm{origin}}(t) = \sum_{l=1}^{L} A_l \cos\left(\omega_0 l t + \psi_l\right), \tag{6}$$

where A_l and ψ_l are the amplitude and the phase angle of the signal at the origin at time $t = 0$ for the l-th order harmonic component, respectively, and ω_0 is the pitch frequency of the signal. Due to the stationarity of the signal, all the A_l and ω_0 are constant.

Similar to Section 2.2, the signal observed by the moving sensor at time t with respect to the phase of the signal at the origin is

$$\begin{aligned} s(t) &= \sum_{l=1}^{L} A_l \cos\left(\omega_0 l \left[t + \frac{\mathbf{p}^T(t)\mathbf{u}}{c}\right] + \psi_l\right) \qquad (7) \\ &= \sum_{l=1}^{L} A_l \cos\left(\omega_0 l \left[t + \frac{r}{c}\sin(\phi_s)\cos(\theta(t) - \theta_s)\right] + \psi_l\right). \end{aligned}$$

For the l-th order harmonic component, the carrier and messages are, respectively,

$$\begin{aligned} &A_l \exp\left[j\left(\omega_0 l t + \psi_l\right)\right], \\ &m_l(t) = \frac{r\omega_0 l}{c}\sin(\phi_s)\cos(\theta(t) - \theta_s). \end{aligned}$$

3 Algorithms for Direction of Arrival Estimation

From the observed signal, various existing demodulation techniques, such as a phase lock loop demodulator, can find the contained message (see e.g., [13,14,15]). Once the message is found, estimation of the source direction is straightforward if the parameters r, c, ω_0 and $\theta(t)$ are known *a-priori*. This section summarizes the CAROUSEL DOA estimation problem.

3.1 CAROUSEL DOA Estimation Problem

Find

- the azimuth angle θ_s rad of the source direction, and
- the polar angle ϕ_s rad of the source direction,

from the phase modulated signal $s(t)$ observed by CAROUSEL as shown in Fig. 3, under the following assumptions

- the radius of the circular motion r m, the angular location $\theta(t)$ rad and the angular velocity $\dot{\theta}(t)$ rad/s of the sensor are known *a-priori*,
- the source is located sufficiently far from the sensor (i.e., plane wave model can be used) and is fixed in time, and
- the angular speed of the sensor $|\dot{\theta}(t)|$ is much slower than the signal frequency ω_0.

In the following sections, "ˆ" is used to represent the estimated value of the argument.

3.2 Sinusoidal Signal

Suppose that the source signal is a sinusoid with a known constant frequency $f_0 = \omega_0/2\pi$ Hz and a known constant amplitude A.

Signal Analysis. Theoretically, the source direction can be found from the message as follows. Recall the message $m(t)$ contained in the observed message is

$$m(t) = \frac{r\omega_0}{c} \sin(\phi_s) \cos(\theta(t) - \theta_s). \tag{8}$$

The time derivative of $m(t)$ is

$$\dot{m}(t) = \frac{dm(t)}{dt} = -\frac{r\omega_0}{c} \dot{\theta}(t) \sin(\phi_s) \sin(\theta(t) - \theta_s). \tag{9}$$

From (8) and (9),

$$\phi_s = \sin^{-1}\left(\frac{c}{r\omega_0} \sqrt{m^2(t) + \left(\dot{m}(t)/\dot{\theta}(t) \right)^2} \right), \tag{10}$$

$$\theta_s = \text{mod } (\theta(t) - \bar{\theta}(t), 2\pi), \tag{11}$$

where

$$\bar{\theta}(t) = \begin{cases} \cos^{-1}\left(\dfrac{m(t)}{\sqrt{m^2(t) + \left(\dot{m}(t)/\dot{\theta}(t) \right)^2}} \right), & \dot{m}/\dot{\theta} \le 0 \\ 2\pi - \cos^{-1}\left(\dfrac{m(t)}{\sqrt{m^2(t) + \left(\dot{m}(t)/\dot{\theta}(t) \right)^2}} \right), & \dot{m}/\dot{\theta} > 0 \end{cases}. \tag{12}$$

Here, $\dot{m}(t) = 0$ when $\dot{\theta}(t) = 0$ so that $\dot{m}(t)/\dot{\theta}(t)$ is well-defined. Note that although the time variable appears in the expressions of right-hand-sides of both equations (10) and (11), those equations represent time-independent values because the time-dependent factors cancel out. Also notice that whether $\theta(t) - \theta_s$ is located in the upper half of the plane (i.e., the first or the second quadrant) or in the lower half of the plane (i.e., the third or the fourth quadrant) can be determined by the sign of $\sin(\theta(t) - \theta_s)$, which is related to $\dot{m}/\dot{\theta}$ as seen from (10).

Estimation Procedure

Step 1: Estimation of ψ_0. Due to the assumption that $|\dot{\theta}(t)| \ll \omega_0$, the message is assumed to be constant for a short period of one cycle of the source signal (carrier), $1/f_0 = 2\pi/\omega_0$,

$$\widehat{m(t) + \psi_0} = \tan^{-1}\left(\frac{-\int_{t-2\pi/\omega_0}^{t} s(\tau)\sin(\omega_0 \tau) d\tau}{\int_{t-2\pi/\omega_0}^{t} s(\tau)\cos(\omega_0 \tau) d\tau} \right). \tag{13}$$

Equation (13) can be derived by using the orthogonality of sinusoidal functions (see Appendix for details).

Due to the periodicity of the message originating from the circular motion of the sensor, the phase shift angle ψ_0 can be estimated by taking an integral of one cycle of the message

$$\hat{\psi}_0 = \frac{\int_{\theta_0}^{\theta_0+2\pi} (\widehat{m(t)+\psi_0}) d\theta}{\int_{\theta_0}^{\theta_0+2\pi} d\theta} = \frac{\int_{t}^{t+R} (\widehat{m(\tau)+\psi_0}) \dot{\theta}(\tau) d\tau}{\int_{t}^{t+R} \dot{\theta}(\tau) d\tau}, \tag{14}$$

where R is the time for the sensor to complete one cycle of its circular motion, and θ_0 can be any angle of the sensor location $\theta_0 = [0, 2\pi)$.

Step 2: Estimation of $m(t)$. With ω_0 and $\hat{\psi}_0$ in hand, the message is estimated by

$$\hat{m}(t) = \tan^{-1}\left(\frac{-\int_{t-2\pi/\omega_0}^{t} s(\tau)\sin(\omega_0\tau + \hat{\psi}_0) d\tau}{\int_{t-2\pi/\omega_0}^{t} s(\tau)\cos(\omega_0\tau + \hat{\psi}_0) d\tau} \right). \tag{15}$$

Step 3: Estimation of ϕ_s and θ_s. The final DOA estimates, $\hat{\phi}_s$ and $\hat{\theta}_s$, are obtained by substituting $\hat{m}(t)$ and $\dot{\hat{m}}(t)$ to $m(t)$ and $\dot{m}(t)$, respectively, in (10) and (11).

Remark: Estimation of A and ω_0. Although the source frequency ω_0 is assumed to be known *a-priori*, one could estimate it as well as the source amplitude A by the following procedure. Let the sensor be at rest when the sampling starts, i.e., $\dot{\theta}(0) = 0$. During an initial time period, keep the angular velocity of the sensor small, so that enough data can be observed for $\theta(t) \approx \theta(0)$. Thus for the initial period of observation, the phase of the observed signal can be approximated by a constant

$$\psi(t) \approx \omega_0 t + \underbrace{\frac{r\omega_0}{c}\sin(\phi_s)\cos(\theta(0) - \theta_s) + \psi_0}_{\text{constant}}, \tag{16}$$

hence

$$s(t) \approx A\cos\left(\omega_0 t + \text{constant}\right), \tag{17}$$

and A and ω_0 can be identified quickly, for example, by using any fitting algorithm such as least square method with a sinusoid (e.g., Matlab `fit` function).

3.3 Harmonic Signal

Now suppose that the source signal is a sum of sinusoids with a known constant pitch frequency $f_0 = \omega_0/2\pi$ Hz and its harmonics with different known constant amplitudes A_l for $l = 1, \cdots, L$.

Signal Analysis. Similar to the sinusoidal signal case, the followings hold for each l. The message and its derivative are, respectively,

$$m_l(t) = \frac{r\omega_0 l}{c} \sin(\phi_s) \cos(\theta(t) - \theta_s), \tag{18}$$

$$\dot{m}_l(t) = -\frac{r\omega_0 l}{c} \dot{\theta}(t) \sin(\phi_s) \sin(\theta(t) - \theta_s). \tag{19}$$

From (18) and (19),

$$\phi_s = \sin^{-1}\left(\frac{c}{r\omega_0 l}\sqrt{m_l^2(t) + \left(\dot{m}_l(t)/\dot{\theta}(t)\right)^2}\right), \tag{20}$$

$$\theta_s = \text{mod }(\theta(t) - \bar{\theta}(t), 2\pi), \tag{21}$$

where

$$\bar{\theta}(t) = \begin{cases} \cos^{-1}\left(\dfrac{m_l(t)}{\sqrt{m_l^2(t) + \left(\dot{m}_l(t)/\dot{\theta}(t)\right)^2}}\right), & \dot{m}_l/\dot{\theta} \leq 0 \\ 2\pi - \cos^{-1}\left(\dfrac{m_l(t)}{\sqrt{m_l^2(t) + \left(\dot{m}_l(t)/\dot{\theta}(t)\right)^2}}\right), & \dot{m}_l/\dot{\theta} > 0. \end{cases} \tag{22}$$

Again, $\dot{m}_l(t) = 0$ when $\dot{\theta}(t) = 0$ so that $\dot{m}_l(t)/\dot{\theta}(t)$ is well-defined.

Estimation Procedure. The DOA estimation can be obtained by using any harmonic l.

Step 1: Estimation of ψ_l. Under the assumption that $|\dot{\theta}(t)| \ll \omega_0$, the message is assumed to be constant during one fundamental cycle of the source signal. With this approximation along with orthogonality of trigonometric functions,

$$\widehat{m_l(t) + \psi_l} = \tan^{-1}\left(\frac{-\int_{t-2\pi/\omega_0 l}^{t} s(\tau)\sin(\omega_0 l\tau)d\tau}{\int_{t-2\pi/\omega_0 l}^{t} s(\tau)\cos(\omega_0 l\tau)d\tau}\right). \tag{23}$$

Due to the periodicity of the message originating from the circular motion of the sensor, the phase shift angle ψ_l can be estimated by taking an integral of one cycle of the message. Further taking the normalized integral for one cycle of the sensor motion will produce the estimate of the phase of the l-th order harmonic component,

$$\hat{\psi}_l = \frac{\int_{\theta_0}^{\theta_0+2\pi} \widehat{m_l(t)+\psi_l} d\theta}{\int_{\theta_0}^{\theta_0+2\pi} d\theta} = \frac{\int_t^{t+R} \left(\widehat{m_l(\tau)+\psi_l}\right) \dot{\theta}(\tau) d\tau}{\int_t^{t+R} \dot{\theta}(\tau) d\tau}, \tag{24}$$

where R is the time for the sensor to complete one cycle of its circular motion, and θ_0 can be any angle of the sensor location $\theta_0 = [0, 2\pi)$.

Step 2: Estimation of $m_l(t)$. With ω_0 and $\hat{\psi}_l$ in hand, the message is estimated by

$$\hat{m}_l(t) = \tan^{-1}\left(\frac{-\int_{t-2\pi/\omega_0 l}^{t} s(\tau)\sin(\omega_0 l\tau + \hat{\psi}_l) d\tau}{\int_{t-2\pi/\omega_0 l}^{t} s(\tau)\cos(\omega_0 l\tau + \hat{\psi}_l) d\tau}\right). \tag{25}$$

Step 3: Estimation of ϕ_s and θ_s. The final DOA estimates, $\hat{\phi}_s$ and $\hat{\theta}_s$, are calculated by substituting $\hat{m}_l(t)$ and $\dot{\hat{m}}_l(t)$ to $m_l(t)$ and $\dot{m}_l(t)$, respectively, in (20) and (21).

4 Implementation in Discrete-Time

So far, this chapter discussed the estimation procedures in continuous-time. However, in practice, the algorithms need to be applied to discrete-time signals. Care must be taken when computing integrals and derivatives with discrete-time signals. A single sinusoidal signal is used for illustration for simplicity purposes, however, similar results apply to the harmonic signals as well.

4.1 Integral and Derivative in Discrete-Time

The signal with a square bracket denotes a discrete-time signal and n is the sample index. In implementing the processes in the *Step 2* and *Step 3*, both the integrals in (13) – (15), and the derivative in (12) are approximated by summation and difference, respectively, as follows;

$$\widehat{m[n]+\psi_0} = \tan^{-1}\left(\frac{-\sum_{n-\lfloor\frac{2\pi}{\omega_0 T}\rfloor}^{n} s[n]\sin(\omega_0 nT)}{\sum_{n-\lfloor\frac{2\pi}{\omega_0 T}\rfloor}^{n} s[n]\cos(\omega_0 nT)}\right), \tag{26}$$

$$\hat{\psi}_0 = \frac{\sum_{n-\frac{R}{T}}^{n} (\widehat{m[n]+\psi_0})\dot{\theta}(nT)}{\sum_{n-\frac{R}{T}}^{n} \dot{\theta}(nT)}, \tag{27}$$

$$\hat{m}[n] = \tan^{-1}\left(\frac{-\sum_{n-\lfloor\frac{2\pi}{\omega_0 T}\rfloor}^{n} s[n]\sin(\omega_0 nT + \psi_0)}{\sum_{n-\lfloor\frac{2\pi}{\omega_0 T}\rfloor}^{n} s[n]\cos(\omega_0 nT + \psi_0)}\right), \tag{28}$$

$$\dot{\hat{m}}[n] = \frac{\hat{m}[n] - \hat{m}[n-1]}{T}, \tag{29}$$

where $s[n] := s(nT)$ is the sampled signal at time nT for the sampling period $T = 1/F$ for the sampling frequency F.

As T approaches to 0, the discrete-time estimation approaches to the continuous-time estimation. However, non-zero T due to limited sampling frequency always causes errors in estimation. This is because the integrals and derivatives are approximated by summations and difference equations as seen in the right-hand-sides of (26) – (29). Fig. 4 shows $\widehat{m[n]+\psi_0}$ obtained by (26). Although theoretically $m(t)$ contains only the instantaneous frequency of $\theta(t)$, the result includes some small perturbation which is apparently caused by the discretization of the signal.

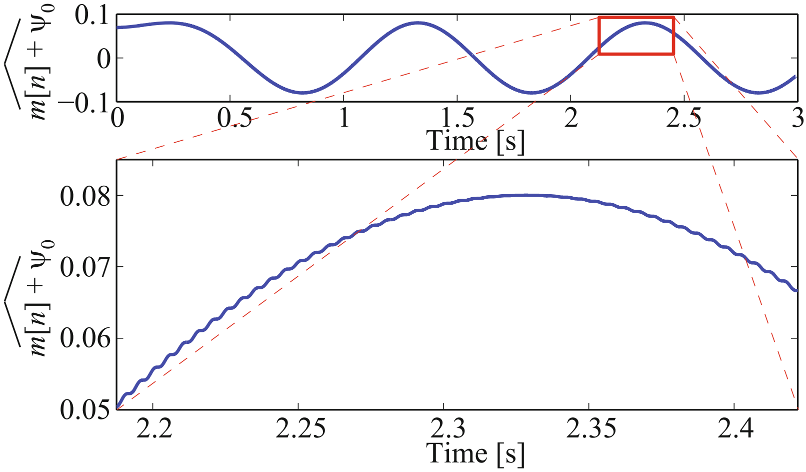


Fig. 4. Message signal retrieved from discrete-time signal

4.2 Sinusoidal Curve Fitting

To mitigate the detrimental effect of the error caused by signal discretization, sinusoidal curve fitting can be applied to the raw discrete-time message $m[n]$ to obtain the refined message $\bar{m}[n]$.

$$\bar{m}[n] = \bar{\lambda}\cos(\dot{\theta}(\infty)nT - \bar{\sigma}), \tag{30}$$

where

$$\{\bar{\lambda}, \bar{\sigma}\} = \arg\min_{\lambda,\sigma} \sum_n \left(\lambda \cos\left(\dot{\theta}(\infty)nT - \sigma\right) - m[n]\right)^2,$$

$\dot{\theta}(\infty)$ is the asymptotic value of the sensor speed, and the range of the summation is chosen appropriately. The fitting function, $\lambda \cos\left(\dot{\theta}(\infty)nT - \sigma\right)$, is chosen from the observation that the frequency of the message is equivalent to the sensor speed which is assumed to be a constant, i.e., in the ideal sampling, hence the message should be in the form of $\lambda \cos\left(\dot{\theta}(\infty)nT - \sigma\right)$. Once the parameters $\bar{\lambda}$ and $\bar{\sigma}$, which are the functions of DOA, are obtained, the continuous message can be constructed as

$$\bar{m}(t) = \bar{\lambda}\cos(\dot{\theta}(\infty)t - \bar{\sigma}), \tag{31}$$

and can be used to find DOA as in *Step 3* of Section 3.2

5 Simulations

This section evaluates the performance of CAROUSEL implemented to a discrete-time signal by simulation. Additive white Gaussian noise (AWGN) which simulates the internal noise of the sensor was superimposed on the signal at various signal-to-noise ratio (SNR). Parameters used in simulations are summarized in Table 1. Plots are obtained by averaging the results of 100 trials with different noise signals for each parameter set.

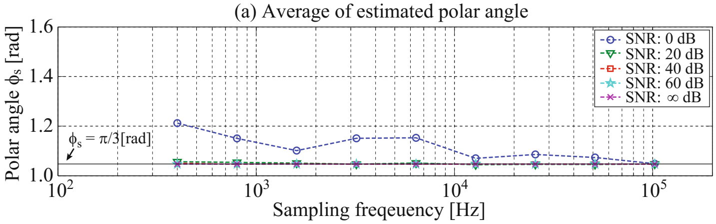
Table 1. Parameter values used in the simulation

Radius of sensor motion: r	0.05	m
Wave speed: c	340	m/s
Source amplitude: A	2000	
Duration of signal	3	sec
Azimuth angle: θ_s	$\frac{\pi}{6}$	rad
Polar angle: ϕ_s	$\frac{\pi}{3}$	rad
Angular speed of sensor: $\dot{\theta}$	2π	rad/s

5.1 Sinusoidal Signal

First, in order to see the effect of sampling frequency on the accuracy of DOA estimation, the sampling frequency was varied by the octave band of $f_0 = 100$ Hz for a sinusoidal signal. Fig. 5 and Fig. 6 show the estimated angles and the mean squared value of the errors from the true DOA, respectively. Note that SNR $= \infty$ means the observation was noise-free (i.e., no AWGN was added to the signal).

Fig. 5 shows that the CAROUSEL is capable of estimating both the azimuth and the polar angles accurately provided the sampling frequency is relatively higher than the frequency of the signal. Fig. 6 shows that increasing the sampling frequency improves the accuracy of the DOA estimation, this is because the discretization error is reduced.


(a) Average of estimated polar angle

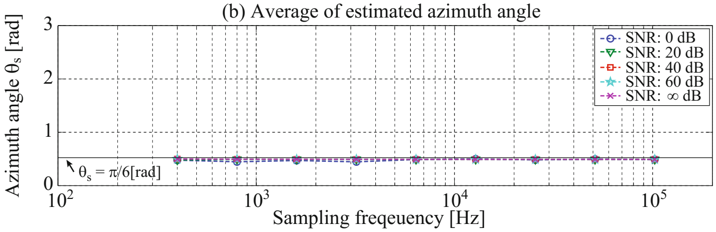
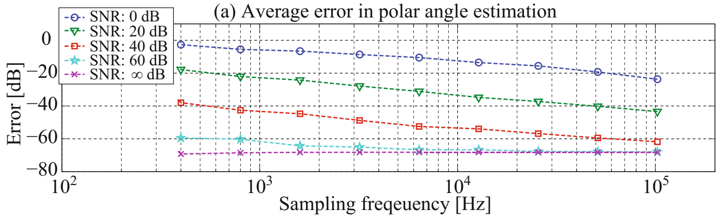

(b) Average of estimated azimuth angle

Fig. 5. Average of estimated DOA: (a) polar angle, (b) azimuth angle


(a) Average error in polar angle estimation

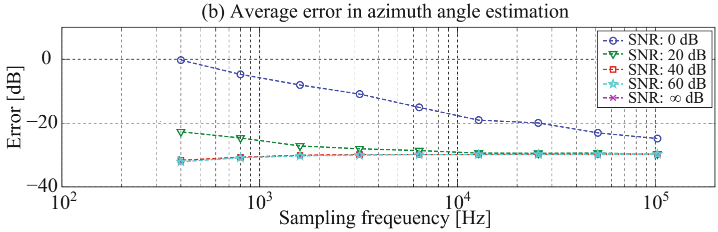

(b) Average error in azimuth angle estimation

Fig. 6. Average error in estimated DOA: (a) polar angle, (b) azimuth angle

5.2 Harmonic Signal

Next, the accuracy of the estimated DOA for harmonic signals with different pitch (fundamental) frequencies were investigated with a fixed sampling frequency $F = 25600$ Hz. Fig. 7 shows the absolute error of the estimated DOA for signals where the amplitude of the harmonics had different decay along with the order of the harmonics, i.e., $A_l = \alpha^{l-1}$ where α was varied from 0 to 1. Note the signal was a pure sinusoid when α was set to zero whereas all the harmonic components had the same amplitude as that of the pitch component with $\alpha = 1$. The maximum order of the harmonics was set to 4 in order to avoid spectral aliasing at the highest pitch frequency, i.e., 32000 Hz considering that the sampling frequency of 25600 Hz. The estimation error decreased for both the

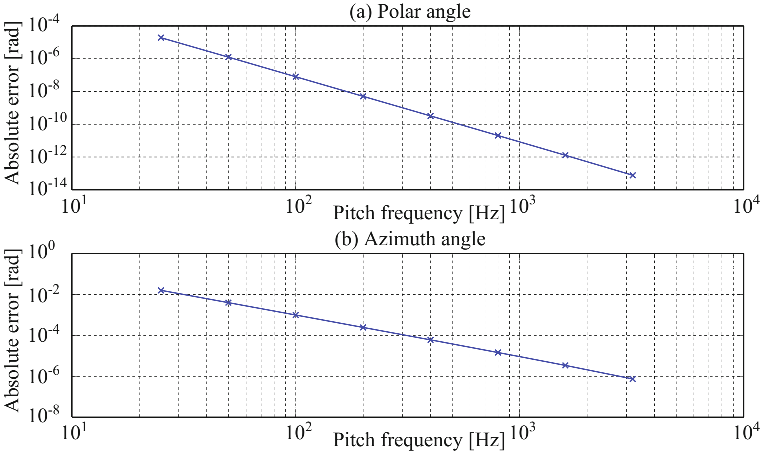


Fig. 7. Average error of estimated angles for different pitch frequencies

azimuth and the elevation angles as the pitch frequency increased. This trend can be explained by looking at the amount of frequency shift caused by the Doppler effect. By comparing equations (6) and (7), it can be seen that the amount of frequency shift is proportional to the pitch frequency, thus the proportion of the errors in the phase shift becomes smaller when the pitch frequency is higher. Thus, the more accurate DOA estimation can be achieved by focusing on the higher frequency component in the signal.

In order to verify the capability of CAROUSEL for practical signals, DOA of five different voiced phonemes in speech signals, i.e., /a/, /e/, /i/, /o/, and /u/, generated by the formant synthesizer [16] were estimated. The formant frequencies and bandwidths used for synthesizing the phonemes are listed in the Table 2, where up to the third formant was considered for generating each phoneme. Fig. 8 and Fig. 9 show the absolute error of the estimated DOA for each phoneme. Slightly better estimation accuracy can be seen in the phonemes with higher pitch frequency, which agrees with Fig. 7. The estimation accuracy varied depending on the phonemes when the input SNR was low, e.g. estimated DOA of /a/ was more affected by the noise than that of /u/. Because the DOA is estimated by the extraction of $m_l(t)$ from the observation by (25), its accuracy will be degraded if the l-th order harmonic in the observed signal is contaminated by the noise. Since the first formant frequency of /a/ is much higher than its pitch frequency, the lower level of the pitch component caused the low DOA estimation accuracy. Fig. 10 summarizes the estimation error of the same signals in Fig. 9, i.e., phonemes with 250 Hz pitch frequency, but the DOA information was extracted from the third harmonic component of the observed signal ($l = 3$). The result is self-explanatory that the phonemes whose formant frequencies were closer to the frequency of the extracted harmonic, 750 Hz, provided more accurate estimation even though the input SNR was low. This concludes that the harmonic order l for extracting $m_l(t)$ should be chosen so that the harmonic with high SNR is utilized for the DOA estimation.

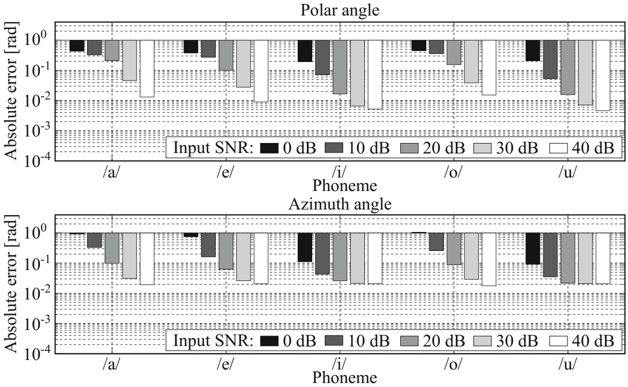


Fig. 8. Average error of estimated angles for five different phonemes of pitch frequency 150 Hz. The DOA was extracted from the pitch component ($l = 1$).

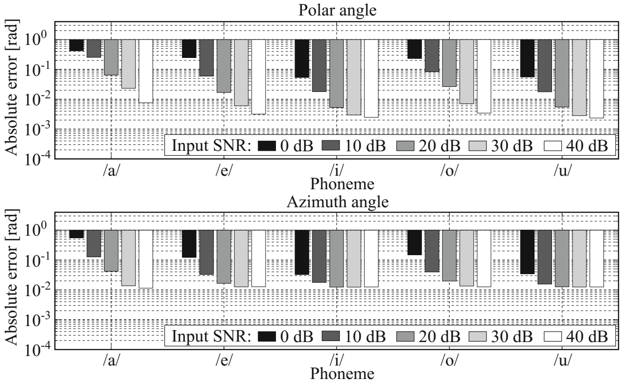


Fig. 9. Average error of estimated angles for five different phonemes of pitch frequency 250 Hz. The DOA was extracted from the pitch component ($l = 1$).

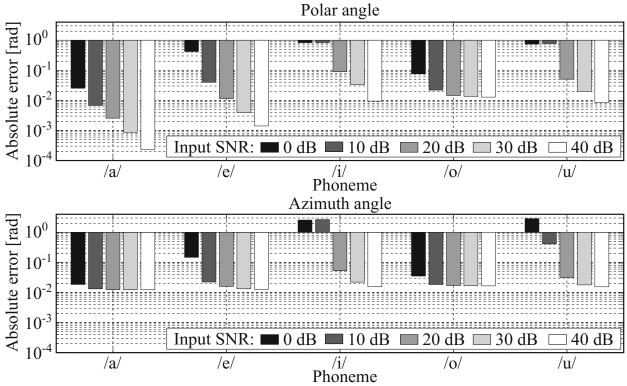


Fig. 10. Average error of estimated DOAs for five different phonemes of pitch frequency 250 Hz. The DOA was extracted from the third harmonics ($l = 3$).

Table 2. Parameters used for synthesizing vocal phonemes

Phoneme	Formant frequencies	Formant bandwidth
/a/	$700, 1220, 2600$	$50, 60, 100$
/e/	$450, 1900, 2400$	$50, 70, 90$
/i/	$250, 2100, 3100$	$40, 50, 170$
/o/	$450, 900, 2600$	$40, 50, 60$
/u/	$250, 1400, 2200$	$50, 50, 50$

6 Conclusions

This chapter presented recently proposed sensing architecture, CAROUSEL, which consists of a single sensor moving along the circumference of a circle while observing signals. The sensor motion induces a modulation effect to the observed signal whose message contains the DOA of the signal that can be estimated by demodulation. Algorithms for DOA estimation were presented and implementation methods were discussed for sinusoidal and harmonic signals. Simulation results show that CAROUSEL can estimate the DOA with accuracy provided the sampling frequency is reasonably high compared to the signal frequency. Experiments with the proposed method implemented to practical sensors is a future subject.

References

1. Kishida, M., Hioka, Y.: Circularly moving sensor for use of modulation effect. In: Proceedings of 7th International Conference on Sensing Technology, pp. 242–246 (2013)
2. Hioka, Y., Kishida, M.: Direction of arrival estimation of harmonic signal using single moving sensor. In: Proceedings of 8th IEEE Sensor Array and Multichannel Signal Processing Workshop (to appear, 2014)
3. Hansen, R.: Phased Array Antennas, 2nd edn. Wiley (2007)
4. Brandstein, M., Ward, D.: Microphone Arrays. Springer (2001)
5. Sherman, C., Butler, J.: Transducers and Arrays for Underwater Sound. Springer (2007)
6. Johnson, D.H., Dudgeon, D.E.: Array signal processing: concepts and techniques. Prentice Hall (1993)
7. Knapp, C., Carter, G.C.: The generalized correlation method for estimation of time delay. IEEE Transactions on Acoustics, Speech and Signal Processing 24(4), 320–327 (1976)
8. Schmidt, R.: Multiple emitter location and signal parameter estimation. IEEE Transactions on Antennas and Propagation 34(3), 276–280 (1986)
9. Roy, R., Kailath, T.: Esprit-estimation of signal parameters via rotational invariance techniques. IEEE Transactions on Acoustics, Speech and Signal Processing 37(7), 984–995 (1989)

10. Cigada, A., Lurati, M., Ripamonti, F., Vanali, M.: Moving microphone arrays to reduce spatial aliasing in the beamforming technique: theoretical background and numerical investigation. Journal of the Acoustical Society of America 124(6), 3648–3658 (2008)
11. Schasse, A., Martin, R.: Localization of acoustic sources based on the teager-kaiser energy operator. In: Proceedings of 18th European Signal Processing Conference, pp. 2191–2194 (2010)
12. Schasse, A., Tendyck, C., Martin, R.: Source localization based on the doppler effect. In: International Workshop on Acoustic Signal Enhancement, pp. 1–4 (2012)
13. Sakrison, D.J.: Communication theory, transmission of waveforms and digital information. John Wiley & Sons (1968)
14. Libbey, R.: Signal and Image Processing Sourcebook. Van Nostrand Reinhold (1994)
15. Chitode, J.S.: Communication Theory. Technical Publications (2008)
16. Furui, S.: Digital Speech Processing: Synthesis, and Recognition, 2nd edn. Marcel Dekker (2000)

Appendix

In (13), the phase of the observation $s(t)$ is estimated by utilizing the periodicity of both the carrier and the message. Since the carrier is defined by

$$s(t) = A\cos(\omega_0 t + m(t) + \psi_0)$$

and ω_0 is already known, the term $\omega_0 t$ of the phase of $s(t)$ can be separated by

$$\begin{aligned} s(t) &= A\cos(\omega_0 t + m(t) + \psi_0) \\ &= A\left[\cos(\omega_0 t)\cos(m(t) + \psi_0) - \sin(\omega_0 t)\sin(m(t) + \psi_0)\right]. \end{aligned}$$

Due to the assumption that the frequency of the carrier wave is much higher than the rotation speed of the sensor, i.e., $\omega_0 \gg |\dot{\theta}|$, the term $m(t) + \psi_0$ can be treated as a constant. From the orthogonality of sine and cosine functions, the following approximations hold.

$$\begin{aligned} &\int_{t-\frac{2\pi}{\omega_0}}^{t} s(\tau)\sin(\omega_0\tau)d\tau \\ &= \int_{t-\frac{2\pi}{\omega_0}}^{t} A\sin(\omega_0\tau)\cos(\omega_0\tau)\cos(m(\tau)+\psi_0)d\tau - \int_{t-\frac{2\pi}{\omega_0}}^{t} A\sin^2(\omega_0\tau)\sin(m(\tau)+\psi_0)d\tau \\ &\approx A\cos(m(t)+\psi_0)\int_{t-\frac{2\pi}{\omega_0}}^{t}\sin(\omega_0\tau)\cos(\omega_0\tau)d\tau - A\sin(m(t)+\psi_0)\int_{t-\frac{2\pi}{\omega_0}}^{t}\sin^2(\omega_0\tau)d\tau \\ &= -A\pi\sin(m(t) + \psi_0). \end{aligned} \tag{32}$$

Similarly,

$$\int_{t-\frac{2\pi}{\omega_0}}^{t} s(\tau)\cos(\omega_0\tau)d\tau = A\pi\cos(m(t)+\psi_0). \tag{33}$$

From (32) and (33), it follows that

$$m(t)+\psi_0 = \tan^{-1}\left(\tan(m(t)+\psi_0)\right) = \tan^{-1}\left(\frac{-\int_{t-\frac{2\pi}{\omega_0}}^{t} s(\tau)\sin(\omega_0\tau)d\tau}{\int_{t-\frac{2\pi}{\omega_0}}^{t} s(\tau)\cos(\omega_0\tau)d\tau}\right).$$

Magnetotransport at High Frequency of Soft Magnetic Amorphous Ribbons

L. González-Legarreta[1], A. Talaat[2], M. Ipatov[2], V. Zhukova[2],
A. Zhukov[2,3], J. González[2], and B. Hernando[1,*]

[1] Dpt. Physics, University of Oviedo, Av. Calvo Sotelo s/n, 33007 Oviedo, Spain
[2] Dpt. Materials Physics, University of the Basque Country,
Paseo Manuel de Lardizábal 3, 20018 San Sebastián, Spain
[3] IKERBASQUE Foundation, Bilbao, Spain
`grande@uniovi.es`

Abstract. Magnetic properties of amorphous alloy ribbons have been extensively studied for nearly a quarter of a century since their first production. Much interest in these materials has been stimulated by their remarkable magnetomechanical and magnetotransport properties. It is amazing to notice with respect to the so-called giant magnetoimpedance effect (GMI) effect that the rapid advantages in the understanding of the underlying physical mechanisms of GMI have allowed the development of practical devices and applications using this effect. Therefore, the large influence of magnetic fields and also mechanical stresses in determining the magnetic permeability and electrical impedance of such soft ferromagnetic alloys make these materials very suitable for sensing magnetic field, current and stress. In this chapter we will try to report the recent studies on GMI response of near-zero magnetostriction Co-based amorphous ribbons in as-cast (with different wide dimension) state, and exhibiting a macroscopic uniaxial magnetic anisotropy induced by different stress-annealing treatment (300 MPa applied tensile stress at different temperature i.e., 340, 360, and 400 ºC, during 1 h) in the frequency range from 10 MHz up to 1000 MHz. Comparison among GMI effect of as-cast state and stress-annealed ribbons will be discussed. It is remarkable that a more spectacular and defined GMI effect is observed in the stress-annealed ribbons owing to the presence of a macroscopic uniaxial transverse magnetic anisotropy developed with the stress-annealing treatment that enhances the transverse component of the magnetic susceptibility.

Keywords: Amorphous Materials, Soft Magnetic Behavior, Magnetoimpedance Effect.

1 Introduction

From the discovery of the giant magnetoimpedance (GMI) effect in 1994 by Panina and Mohri [1]and Beach and Berkowitz [2] in non-magnetostrictive soft magnetic

---

[*] Corresponding author.

A. Mason et al. (eds.), *Sensing Technology: Current Status and Future Trends IV*,
Smart Sensors, Measurement and Instrumentation 12, DOI: 10.1007/978-3-319-12898-6_12

amorphous wires (diameter around 120 μm), there was an intensively research activity by reason of promising technological applications based on the GMI phenomenon [3, 4]. Such scientific research has dealt with several aspects concerning the intrinsic magnetotransport properties (i.e., frequency range, intensity of the effect, magnetic field to observe possible maximum, noise, etc.) as well as those related with microstructural (mainly amorphous or nanocrystalline) or geometrical character (as it was mentioned, initially in wire, but GMI has been also reported in glass-coated microwire [5, 6], ribbon [7], micro-patterned ribbon [8], multilayers [9]) and, therefore, GMI is actually opening a new branch of research combining the micromagnetics of soft magnets with the classical electrodynamics.

Recently, the development of high-performance magnetic sensors has benefited from the discovery of this new magnetic phenomenon, giant magnetoimpedance in metal-based amorphous alloys. It has been demonstrated that magnetic sensors based upon the GMI effect offer several advantages over conventional magnetic sensors. The key factor is the ultra-high sensitivity of GMI sensors. When compared with a GMR sensor that has a sensitivity of ~1%/Oe, the field sensitivity of a typical GMI sensor can reach a value as high as 500%/Oe [10]. But the development of GMI sensors is still at an early stage, for that, it is necessary to understand the GMI phenomenon and the properties of GMI materials, in order to improve the magnetic sensor with the GMI effect for technological applications.

1.1 Fundamental Aspects of GMI

When a soft ferromagnetic conductor is subjected to a small alternating current (ac), a large change in the ac complex impedance of the conductor can be achieved under an applied magnetic field. This is known as the giant magnetoimpedance effect [10]. The relative change of the impedance (Z) with applied field (H), which is defined as the GMI effect, is expressed by:

$$\Delta Z/Z\,(\%) = 100\% \times \frac{Z\,(H) - Z\,(H_{\max})}{Z\,(H_{\max})} \tag{1}$$

where $H_{\max}$ is usually the external magnetic field sufficient to saturate the sample magnetization. In practice, the value of $H_{\max}$ is usually available for a given experimental equipment.

The complex impedance of a linear electronic element $Z = R + \mathrm{j}\omega L$ (R and L are the resistance and inductance, respectively) is given by the ratio of the voltage amplitude V_{ac}, measured on the element, to the current amplitude I_{ac} of a sinusoidal current $I_{ac} \sin \omega t$, passing through it. For a conductor with length L and uniform cross-section with area q (see Fig. 1) the impedance can be expressed by the formula [11]:

$$Z = \frac{V_{ac}}{I_{ac}} = \frac{L\,E_z(S)}{q\,\langle j_z \rangle_q} = R_{dc}\,\frac{j_z(S)}{\langle j_z \rangle_q} \tag{2}$$

where E_z and j_z are the longitudinal components of electric field and current density, respectively, and R_{dc} is the dc electrical resistance. S is the value at the surface, and $\langle j_z \rangle_q$ is the j_z average value over the cross section q.

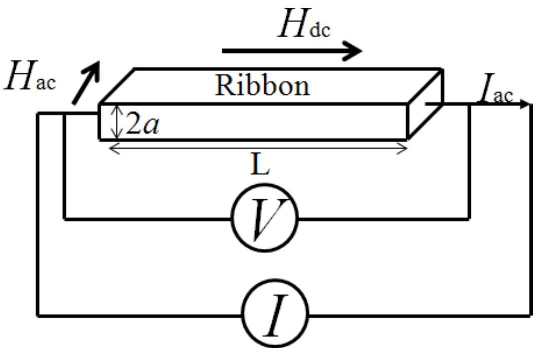


Fig. 1. The definition of impedance

The current density $j_z(r)$ in eq. (2) of the conductor can generally be obtained within the framework of classical electrodynamics of continuous media, by solving simultaneously the reduced Maxwell equation:

$$\nabla^2\mathbf{H} - \frac{\mu_0}{\rho}\dot{\mathbf{H}} = \frac{\mu}{\rho}\dot{\mathbf{M}} - \operatorname{grad}\operatorname{div}\mathbf{M} \tag{3}$$

and the Landau-Lifshitz equation for the motion of the magnetization vector

$$\mathbf{M} = \gamma\mathbf{M} \times \mathbf{H}_{\mathrm{eff}} - \frac{\alpha}{\mathrm{M_s}}\mathbf{M} \times \dot{\mathbf{M}} - \frac{1}{\tau}(\mathbf{M} - \mathbf{M}_0) \tag{4}$$

where γ is the gyromagnetic ratio, M_S the saturation magnetization, M_0 the static magnetization, H_{eff} the effective magnetic field and α the damping parameter [11, 12].

Analytically, it is difficult to obtain the exact solution of the problem by solving simultaneously the Maxwell eq. (3) and the Landau–Lifshitz equation of motion eq. (4). But at low frequency, the problem can be treated with a quasistatic approximation. In this case, using the simple material relation $\mathrm{B} = \mu\mathrm{H}$ in the Maxwell equations, the current density in an infinite film conductor can be easily calculated [11, 12]. And substituting for $j_z(r)$ into eq. (2) the impedance of an infinite film conductor is:

$$Z = \mathrm{R}_{\mathrm{dc}} \cdot \mathrm{j}ka \coth(\mathrm{j}ka) \tag{5}$$

where $2a$ is the thickness of the ribbon, R_{dc} is the electrical resistance for a dc and $k = (1+\mathrm{j})/\delta_{\mathrm{m}}$ with imaginary unit j. δ_{m} is the penetration depth in a magnetic medium with transverse permeability (μ_{T}) and for the ribbon [13]:

$$\delta_m = \frac{c}{\sqrt{4\pi^2 f\,\sigma\,\mu_T}} \tag{6}$$

where c is the speed of light, σ the electrical conductivity, and $f = \omega/2\pi$ is the frequency of the ac (I_{ac}) flowing along the sample.

At high frequencies, the behavior of the magnetization must be described by the Landau-Lifshitz-Gilbert equation or similar dynamical equations and ferromagnetic resonance (FMR) effects can appear, displaying large permeability and impedance peaks [14]. The complete description of GMI including dynamical effects has been evaluated by several authors, with different degrees of approximations [11, 13].

As seen previously, the giant magnetoimpedance depends on the frequency (*f*) of the driving ac. So the GMI can be classified into three frequency regimes:

1. *Low-frequency regime*: when the frequency is too low, the skin effect is very weak. This is due to the skin effect that is higher than the thick of the sample. Then, the changes in voltage at the sample's ends are mainly due the so-called magnetoinductive effect (imaginary part of Z) [15]. The limit in this regime depends on the sample's thickness, being the typical values up to a few kHz for thicknesses around some tens of microns (in ribbons and wires) and some tens of MHz for thicknesses below 1 μm in thin films.
2. *Intermediate frequency regime*: (between ~100 kHz up to a few MHz), where GMI originates mainly from the variation of the skin depth due to strong changes of the effective magnetic permeability caused by the applied dc magnetic field. It is noted here that depending on the sample geometry, the GMI profile can reach its peak in the intermediate frequency range, as a consequence of the contribution of the permeability from both domain wall motion and magnetization rotation to GMI. Reduction in GMI at higher frequencies is related to the domain walls becoming strongly damped by eddy currents and only magnetization rotation contributes to GMI.
3. *High-frequency regime*: (several MHz up to GHz), where the origin of GMI is believed to be related to the gyromagnetic effect and ferromagnetic relaxation. The maxima in GMI profiles are shifted towards higher fields, where the samples are already saturated magnetically [10]. Strong changes in skin depth are caused by the same mechanism as in the ferromagnetic resonance.

2 Experimental Techniques

Amorphous ribbons of next nominal composition $Co_{66.5}Fe_{3.5}Si_{12.0}B_{18.0}$ (in the as-cast state, saturation magnetostriction at room temperature is of the order of $\lambda_S \approx -0.10 \times 10^{-6}$) and $Co_{66.3}Fe_{3.7}Si_{12.0}B_{18.0}$ (in the as-cast state, saturation magnetostriction at room temperature is of the order of $\lambda_S \approx -0.18 \times 10^{-6}$) were fabricated by the melt-spinning technique using a Fe wheel. First alloy ribbons were cast in a 10^{-3} mbar vacuum and with different (0.35, 0.60, 0.70, and 0.90 mm) width, and 20 μm thick. $Co_{66.3}Fe_{3.7}Si_{12.0}B_{18.0}$ alloy ribbon was quenched in the same conditions but with 0.70 mm in width and 20 μm thick. Different pieces of this ribbon were submitted to the following stress-annealing treatment:

(a) Sample 1: tensile stress annealing at 400 ºC with 300 MPa during 1 h, without pre-annealing.

(b) Sample 2: pre-annealing at 340 ºC during 1 h, followed by a tensile stress annealing at 340 ºC with 300 MPa during 1 h.

(c) Sample 3: pre-annealing at 360 ºC during 1 h, followed by a tensile stress annealing at 360 ºC with 300 MPa during 1 h.

In all treatments the tensile stress was applied along the ribbon axis. As the most important consequence, these thermal treatments develop a macroscopic uniaxial magnetic anisotropy that will play a drastic role regarding the magnetoimpedance (MI) effect of these treated ribbons. After annealing, the saturation magnetostriction coefficient obtained by the small angle magnetization rotation (SAMR) [16] technique changed to a positive value, for all samples, which is close to zero, a particular characteristic of materials exhibiting a good MI response.

Hysteresis loops were measured at 50 Hz and 112 Hz in 10 cm long samples using the technique described in [4]. The impedance of either as-cast and/or with different stress induced anisotropy 1 cm long samples was measured with a network analyzer at moderate and high frequencies (10-1000 MHz), as has been previously described in detail [6, 17]. A schematic diagram of the experimental set-up [18] is shown in Fig. 2. The ribbon was placed in a specially designed microstrip sample holder, which is placed inside a long solenoid that creates a homogeneous magnetic field H_E up to 15 kA/m along the ribbon axis. The system allows the application of a circular/transversal bias field H_B created by a dc bias current I_B applied to the sample through a bias-tee element. The longitudinal impedance Z_{zz} is determined with an Agilent N5230A vector network analyzer (VNA) from the reflection coefficient S_{11} measurements, and the off-diagonal impedance can also be measured from the transmission coefficient S_{21} as a voltage induced in the pick-up coil wounded over the ribbon. Such off-diagonal impedance vs. applied field measurements allow the determination of both the magnitude and the sign of the external magnetic field bellow the anisotropy field value of the sample. However, by measuring the longitudinal impedance of a sample with helical anisotropy and applying the dc bias current I_B, the sensor magnetic field range can be increased up to fields higher than the anisotropy sample one [18]. The analyzer power output in all measurements here presented was 10 dBm that corresponds to a 1.4 mA high frequency driving current passing through the ribbon.

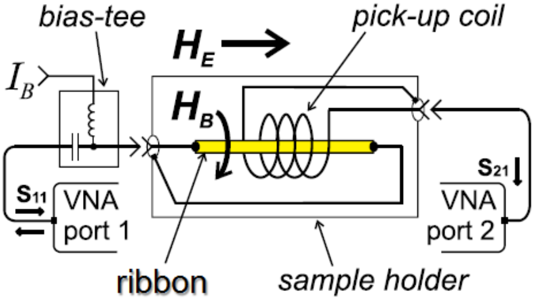


Fig. 2. Experimental set-up for measuring the ribbon impedance and the MI effect [18]

3 Experimental Results and Discussion

3.1 Hysteresis Loops Measurements

3.1.1 As-Quenched $Co_{66.5}Fe_{3.5}Si_{12.0}B_{18.0}$ Ribbons with Different Width

Firstly, it has been studied four $Co_{66.5}Fe_{3.5}Si_{12.0}B_{18.0}$ amorphous alloy ribbons with different widths (0.35, 0.60, 0.70 and 0.90 mm) to see how this parameter affects the hysteresis loops.

The hysteresis loops of the ribbons were measured along the ribbon main axis at a low frequency value of 112 Hz. Fig. 3 (a) shows field dependence of magnetization *M(H)* at room temperature for the four $Co_{66.5}Fe_{3.5}Si_{12.0}B_{15.0}$ studied samples. From this figure, the coercivity for all ribbons was obtained (see Fig. 3 (b–e) and Tab. 1). It can be observed that all the samples exhibit a very low coercivity. Furthermore, the lowest coercivity field detected corresponds to the widest ribbon; consequently, this sample presents a softer magnetic behavior. On the one hand, this value corresponds to a half of the coercivity found in other ribbons, as it can be observed in Tab. 1. On the other hand, it has been found that the saturation magnetization, M_S, ranges from 470 kA/m to 510 kA/m for all the samples (see Tab. 1). These differences in the M_S values can be ascribed to the fact that the studied ribbons were fabricated in four different batches.

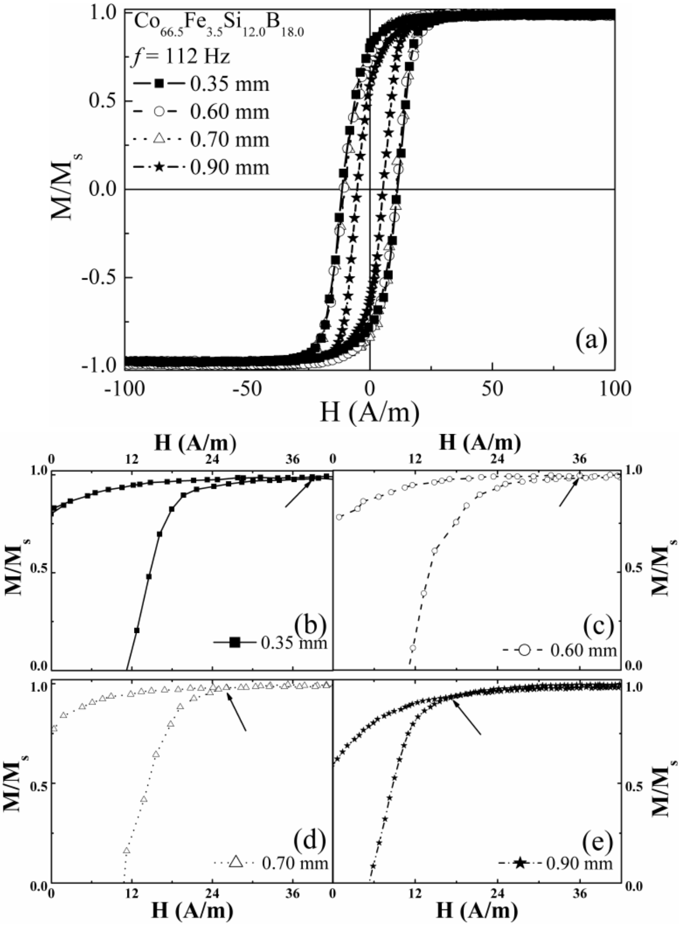


Fig. 3. Normalized hysteresis loops at 112 Hz measured parallel to the ribbon axis (a), and positive part of the hysteresis loops (b), (c), (d) and (e), of the four $Co_{66.5}Fe_{3.5}Si_{12.0}B_{18.0}$ ribbons with different widths. Arrows indicate the anisotropy field for each sample.

Table 1. Coercivity field, saturation magnetization and effective anisotropy field for the four ribbons of $Co_{66.5}Fe_{3.5}Si_{12.0}B_{18.0}$ with different widths

Widths of the ribbons	H_c (A/m)	M_S (kA/m)	H_K (A/m)
0.35 mm	11.3	510	38.2
0.60 mm	11.1	470	36.0
0.70 mm	10.9	494	26.0
0.90 mm	5.5	486	17.1

The other parameter estimated from the hysteresis loops has been the effective anisotropy field. The values have been estimated from the knee area just before to the approach of the magnetic saturation as is indicated by the arrows in Fig. 3 (b–e), with an accuracy of the 0.5% for all ribbons. The data of the effective anisotropy fields for all samples are shown in Tab. 1. These values are averaged over the whole sample but the dominant contribution corresponds to the bulk ribbon. Moreover, it can be observed that the smallest effective anisotropy field belongs to the widest ribbon, which indicates us that the 0.90 mm width sample has the highest permeability.

3.1.2 Stress-Annealed $Co_{66.3}Fe_{3.7}Si_{12.0}B_{18.0}$ Alloy Ribbons

In addition, it has been also investigated three $Co_{66.3}Fe_{3.7}Si_{12.0}B_{18.0}$ amorphous alloy ribbons of the same batch after being submitted to three different stress annealings. Hysteresis loops for all ribbons were measured along the ribbon main axis at a low frequency value of 50 Hz. Fig. 4 shows *M(H)* at room temperature for the three studied $Co_{66.3}Fe_{3.7}Si_{12.0}B_{15.0}$ samples. It can be observed that all hysteresis loops are very similar in shape. The magnetization curves show an almost ideal behavior, with a constant slope nearly up to saturation and quasi absence of hysteresis effect. Then, all samples exhibit a homogeneous induced anisotropy perpendicular to the ribbon axis. In addition, a saturation magnetization, $\mu_0 M_S$, of 424 kA/m was obtained for all samples.

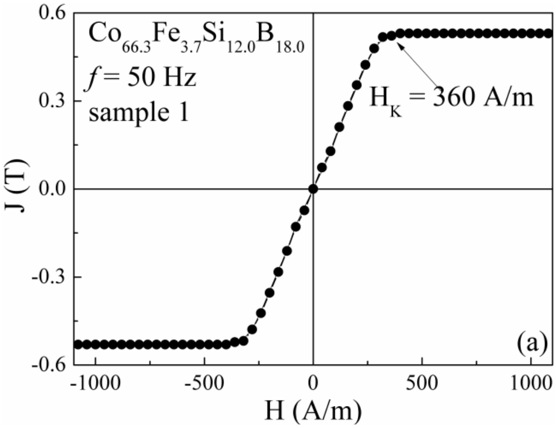


Fig. 4. Hysteresis loops at 50 Hz measured parallel to the ribbon axis for three $Co_{66.3}Fe_{3.7}Si_{12.0}B_{18.0}$ amorphous ribbons after annealing. Sample 1, stress annealing at 400ºC with 300 MPa during 1h without pre-annealing (a), sample 2, 1h stress relaxation at 340ºC followed by 1h stress annealing at 340ºC with 300 MPa (b), and sample 3, 1h stress relaxation at 360ºC followed by 1h stress annealing at 360ºC with 300 MPa (c).

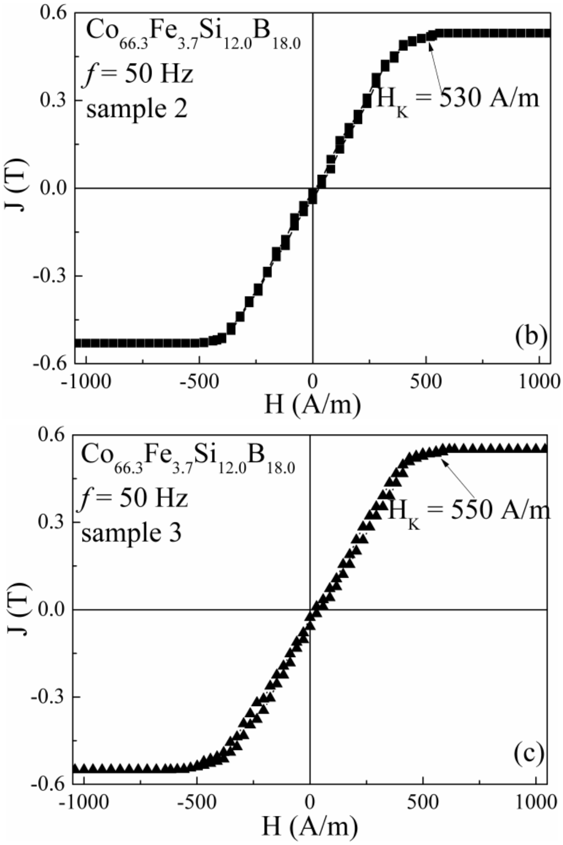


Fig. 4. (*Continued*)

Additionally, the analysis of the effective anisotropy field shows that the sample 1 has the smallest value, being this 360 A/m and, for the samples 2 and 3, the effective anisotropy fields are 530 A/m and 550 A/m, respectively. These values correspond to the bulk ribbons. The anisotropy field values were estimated with an accuracy of the 0.5 % in all cases.

3.2 Magnetoimpedance Response

3.2.1 Ribbon Width Influence on the MI Response for the $Co_{66.5}Fe_{3.5}Si_{12.0}B_{18.0}$ Alloy

First of all, it has been researched the geometric dependence of the magnetoimpedance effect in $Co_{66.5}Fe_{3.5}Si_{12.0}B_{18.0}$ amorphous ribbons. In Fig. 5 (a)-(d) are shown the 3D variation of the longitudinal electrical impedance, $Z_{zz}(H, f)$, being H the applied magnetic field and f the frequency, between 10–1000 MHz, of an ac electrical current flowing through the ribbon with different widths, 0.35, 0.60, 0.70 and 0.90 mm, respectively. It is observed that as the frequency is increased, it emerges more clearly a two-peak behavior, symmetric with respect to H. This behavior is clearly appreciated in all ribbons, with the exception of the narrower one, which will be explained later. Besides, the MI response is more relevant when the width of the ribbon increases.

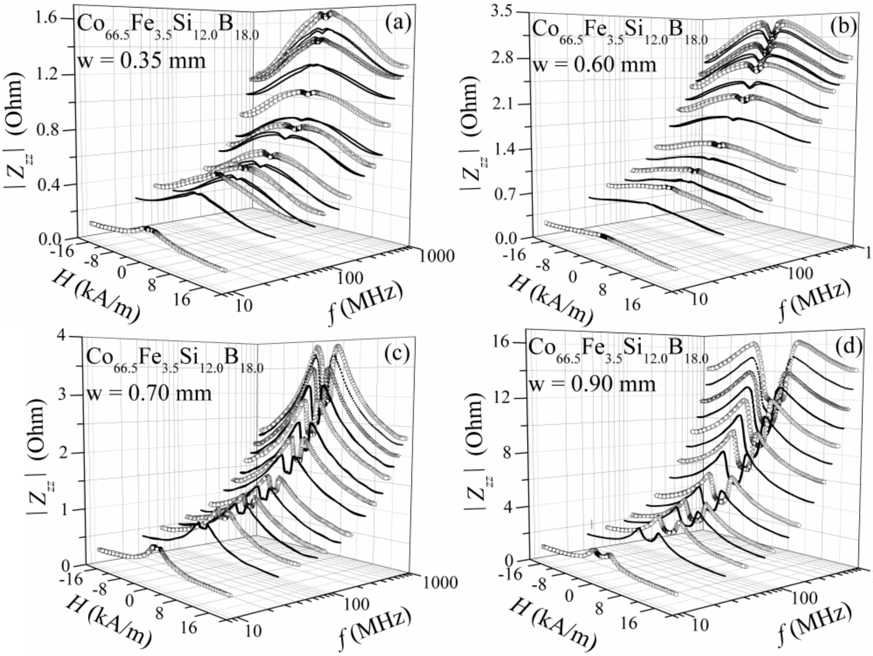


Fig. 5. Module of the impedance response of samples with different values of the width, 0.35 mm (a), 0.60 mm (b), 0.70 mm (c), and 0.90 mm (d), as a function of the frequency and of the magnetic field applied in the ribbon direction

The value of H corresponding to the maximum of Z_{zz} is linked to the average value of the anisotropy field, H_K, at a certain frequency and to the surface anisotropy distribution in the samples [1, 19]. Moreover, in the Fig. 5 (a)-(d), it is observed that the value of H at the peak in Z_{zz} increases with the frequency, remaining the two-peak behavior. This shift on the MI peaks with the frequencies is higher for 0.90 mm width ribbon. The shift is related to the change in skin depth with the frequency-dependent magnetic permeability. As increasing the driving frequency of the ac electric current throughout the ribbon, a decrease of the surface layer thickness is produced and the permeability changes. The application of a longitudinal magnetic field large enough to saturate the sample modifies the transverse permeability by reorientation of the static magnetization and the intrinsic field change of permeability, simultaneously. As a result, the effective skin depth increases, providing an impedance decreasing.

Below, it has checked the dependence of the anisotropy field, H_K, with the frequency (from 500 to 1000 MHz) for all ribbons. In order to do that, we will represent the anisotropy field as a function of the frequency; the data are shown in Fig. 6. Comparing these values of anisotropy field obtained from Fig. 5 with the data obtained from the analysis of the hysteresis loops (Tab. 1), we may observe that they are very different. This is due to the skin effect since with the increasing of the ac electrical current frequency, the surface layer thickness decreases and, as a consequence, the magnetic permeability varies drastically. Therefore, it can be assumed that in this frequency range, the decreasing of the surface layer thickness is owing to the effective skin depth, which steeply affects the MI response.

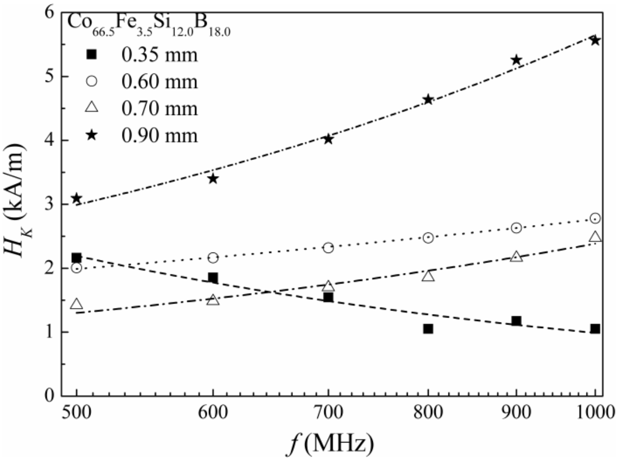


Fig. 6. Frequency dependence of the anisotropy field for all ribbons with different width values. Symbols denote experimental data and lines denote theoretical fittings.

The data of the Fig. 6 have been fitted using the following expression:

$$H_K = a + b f^c \tag{7}$$

where f is the frequency expressed in MHz, H_K, the anisotropy field expressed in A/m and a, b and c are the parameters of fitting. Note that a corresponds to H_K value for $f = 0$, which can be direct correlated with the H_K estimated from the hysteresis loops (Fig. 3). The equations obtained from the fitting, as well as the anisotropy fields at 112 Hz for comparison with the values obtained from the hysteresis loops (see Fig. 3), are shown in Tab. 2:

Table 2. Fitting equations for the anisotropy field as a function of the frequency in the range 500–1000 MHz for the four ribbons with different width, and the anisotropy field at 112 Hz obtained from fitting and from hysteresis loop measurements

Ribbon width (mm)	Fitting equation	H_K (112 Hz) (fitting) (A/m)	H_K (112 Hz) (hysteresis loop) (A/m)
0.35	$H_K = (39.2 \pm 0.6) + (3.3 \pm 0.3)\times 10^6 f^{(-1.18 \pm 0.16)}$	…	38.2
0.60	$H_K = (35.6 \pm 0.3) + (98 \pm 8)\cdot f^{(0.482 \pm 0.013)}$	36.8	36.0
0.70	$H_K = (25.9 \pm 0.4) + (5 \pm 3)\cdot f^{(0.90 \pm 0.10)}$	25.9	26.0
0.90	$H_K = (16.8 \pm 0.2) + (10 \pm 3)\cdot f^{(0.92 \pm 0.05)}$	16.8	17.1

It is observed (Tab. 2) that the values of H_K deduced from the hysteresis loops, which are understood as H_K for the bulk ribbon, are in good agreement with those obtained from eq. (7) for all ribbons apart from the narrower one that presents a clear different behavior of H_K as a function of f.

The narrower ribbon presents a different behavior of H_K with f due to the influence of the shape anisotropy. When the ribbon width decreases, the edge roughness worsens the

MI effect at the low frequency range, but enhances it in the high frequency range (1–3.5 GHz). Certainly, the cross section of this narrower ribbon is still quite large when compared to that of glass-coated microwires, where a linear dependence of H_K with the square of the frequency was reported in the range of 0.5–15 GHz [20], indicating that the reduction of cross section changes the aforementioned tendency related to the appearance of the ferromagnetic resonance phenomenon. In fact, analyzing the frequency dependence of H_K for this ribbon, in the range of 1.0–3.0 GHz (see Fig. 7), a different behavior can be observed when compared to the low frequency range of Fig. 6. In fact, if these data are fitted using the eq. (7), the value of H_K at f = 112 Hz is similar to that deduced from the hysteresis. The results are shown in the following Fig. 7:

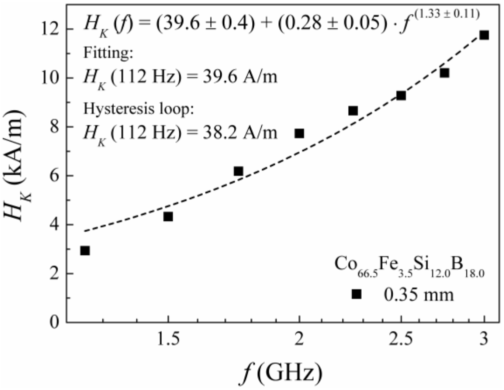


Fig. 7. Frequency dependence of the anisotropy field for the ribbon with 0.35 mm width. Symbols denote experimental data and the line denotes the theoretical fitting.

3.2.2 Induced Anisotropies by Stress Annealing in $Co_{66.3}Fe_{3.7}Si_{12.0}B_{18.0}$ Ribbon Samples. Influence on the MI Properties

The effect of induced anisotropy by stress annealing has been also explored in the magnetoimpedance response for the three different $Co_{66.3}Fe_{3.7}Si_{12.0}B_{18.0}$ amorphous alloy ribbons. 3D variation of the longitudinal electrical impedance, $Z_{zz}(H, f)$, versus the applied magnetic field and the frequency (10–1000 MHz) of the ac electrical current flowing through the samples 1, 2 and 3 are shown in the Fig. 8. In Fig. 8 (a), it is observed that for the sample 1, which does not have any pre-annealing, the impedance evolves in symmetric two-peaks with respect to H as the frequency increases. Meanwhile, the impedance for samples 2 and 3 (Fig. 8 (b) and (c), respectively), which do have a pre-annealing followed by stress annealing at different temperatures, displays two clear and different behavior. On the one hand, in the sample 2 it is noted that the MI response shows a less defined two-peaks pattern at low frequencies, below 150 MHz, and above this frequency the two-peaks pattern emerges clearly. On the other hand, the sample 3 shows an evident asymmetric behavior, which vanishes as the frequency increases. These different features observed in ribbons can be explained by the influence of the anisotropy induced by annealing under applied tensile stress on the MI effect. The reason behind such behaviors will be described in the following lines.

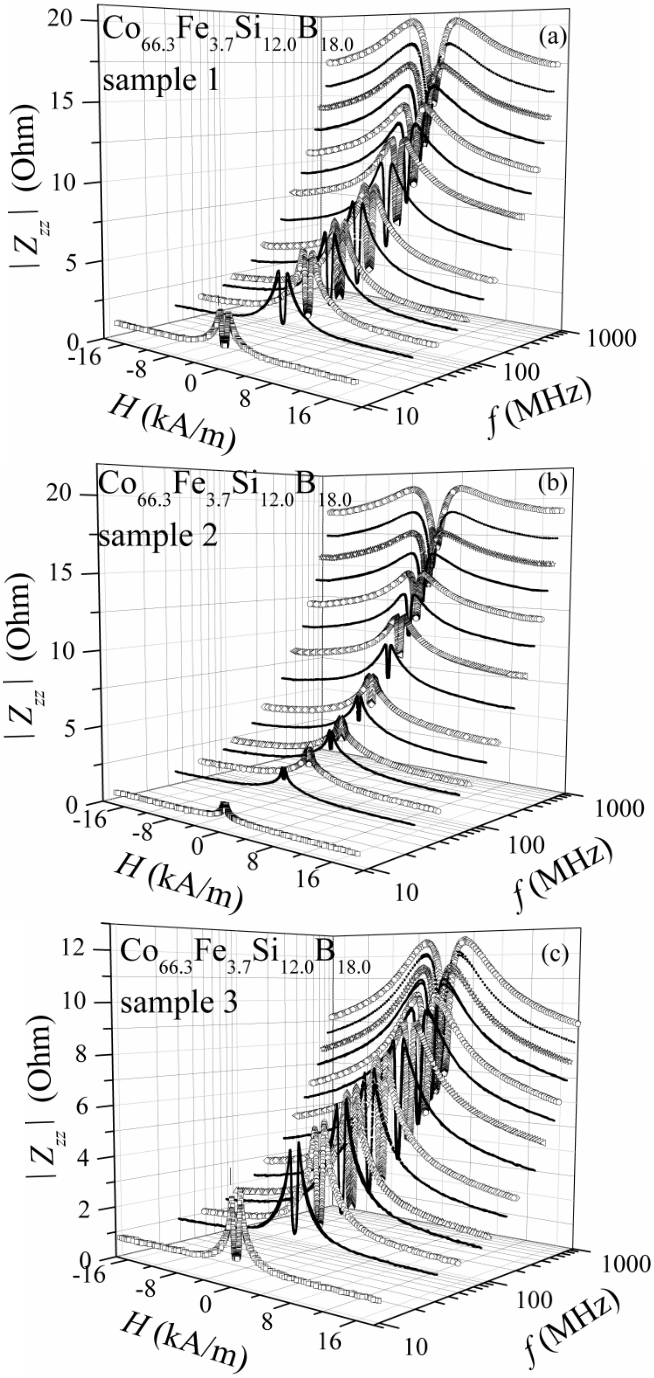


Fig. 8. Module of the impedance of the samples 1, 2 and 3 as a function of the magnetic field applied in the ribbon direction, for different frequency values in the range 10–1000 MHz.

It is well known that the magnetic anisotropy induced by stress annealing in Co-rich metallic glasses exhibits a remarkable behavior [21]. As a matter of fact, this kind of anisotropy can be originated even when annealing is performed above the Curie temperature, what occurs for the ribbons studied here, which have a T_c around 250 ºC. Moreover, this kind of anisotropy implies in a presence of a structural anisotropy due to the existence of internal stresses. It is important to notice that metallic glasses exhibit essentially an elastic behavior at low stresses and temperatures, while at higher temperatures and/or higher levels of stress, an inelastic and plastic behavior is observed. The inelastic response is reversible whereas the plastic response is irreversible. The amplitude of both inelastic and plastic deformation depends on the time and temperature of the performed annealing. Therefore, amorphous ribbons annealed under applied stress undergo a viscoelastic creep. Consequently, the stress annealing induces an uniaxial anisotropy, K_u, which is composed by at least two contributions:

$$K_u = K_{an} + K_{pl}$$

where K_{an} is recoverable and originates from inelastic deformations making the ribbon axis the hard magnetization direction, while K_{pl} is non-recoverable and arises from plastic deformations transforming the ribbon axis an easy magnetization direction. Both contributions present different signs that depend basically on the composition of the amorphous ribbon [21]. Also, an anisotropy originated by residual internal stresses could be present. It is outstanding that the pre-annealing leads to a relaxation process of internal stresses and, consequently, the anisotropy differences between the pre-annealed and as-quenched samples can be considered as a plastic component.

Accordingly, the stress annealing in these ribbons develops an effective uniaxial magnetic anisotropy with the easy axis perpendicular to the length of the ribbons. This induced anisotropy is more intense than that of the as-quenched state and could be expected to be more homogeneously distributed along the cross section and, consequently, the peaks of $Z_{zz}(H, f)$ in the stress annealed ribbons are quite different compared with those of the as-quenched ribbons, see Fig. 5.

Therefore, by analyzing the samples 1, 2 and 3 of this section (Fig. 8), it can be observed that the two-peaks behavior is well-defined and the MI response has its maximum value in the sample 1 treated at 400 ºC without a previous pre-annealing, which could indicate that the optimum conditions for the process of inducing anisotropy corresponds to this annealing temperature in agreement with previous works on stress-induced anisotropy in this kind of amorphous alloys [21, 22]. By using this higher temperature and omitting the preceding stress relief performed in the samples 2 and 3, the ribbon's easy axis anisotropy K_{pl} is promoted in a way that this one may compensate partially or totally the ribbon's hard axis K_{an}. In fact, the sample 1 possesses the lower anisotropy field, see Fig. 4 (a), in comparison with the two other samples (Fig. 4 (b) and (c)) and the lowest magnetostriction value too, which is one of the pre-requisite to observe large GMI response in soft magnetic materials.

In the case of sample 2 (Fig. 8 (b)), submitted to a stress-annealing with a previous pre-annealing, there is a near absence of the two-peaks behavior in the $Z_{zz}(H, f)$ curves at low frequencies. This anomaly in the frequency dependence of the impedance can be related to some inhomogeneous distribution of the induced anisotropy across the ribbon section and along the ribbon length [23]. Due to the pre-annealing process applied to the ribbon, a more stable amorphous phase during the stress annealing is expected, giving rise to a smaller plastic strain and a smaller contribution to K_u from K_{pl}. This explains the higher value of the anisotropy field measured at low frequency (50 Hz) in comparison with sample 1 (see Fig. 4). However, the MI effect is a sensitive probe to detect inhomogeneous surface anisotropies as is reflected in $Z_{zz}(H, f)$ curves for sample 2.

Finally, in the case of the sample 3 (Fig. 8 (c)), also submitted to a stress annealing after a pre-annealing, $Z_{zz}(H, f)$ curves present an asymmetric character. It is worthy to mention that the asymmetric GMI has been previously observed in Co-based amorphous ribbons annealed in air and in the presence of a weak magnetic field [24, 25, 26], which was attributed to the surface crystallization that takes place at this annealing temperature (at 380 ºC). Forsooth, this type of annealing produces an asymmetric hysteresis loops in amorphous ribbons due to the exchange interaction of amorphous bulk with the magnetically harder crystalline surface layers [23]. In the pre-annealing process, it is expected that an unidirectional magnetic anisotropy should be induced in the crystalline layers with an asymmetric character in the two ribbon surfaces. Because the crystalline phase is the hard magnetic one, it remains magnetically ordered within a relatively wide range of magnetic applied fields. The features of the amorphous and the crystalline phase in field-annealed Co-based amorphous ribbons analyzed separately were revealed due to the exchange coupling [27]. The coupling between the crystalline and amorphous phase produces an effective exchange bias field resulting in an asymmetry in the dc magnetization that is responsible for the asymmetric GMI response in pre-annealed stress-annealed ribbon. Hence, the GMI effect becomes asymmetric with regard to the mentioned pre-annealing treatment, and this asymmetry can be assigned to the inelastic stress-induced anisotropy component to be rather different in the two ribbon surfaces, owing to the crystallization effect starts to be significant. Therefore, this GMI behavior could result from magnetization of the amorphous ribbon, suggesting that there is an antiferromagnetic coupling of these two phases in the ribbon surface layer between the magnetization of the crystalline one with the magnetization of the inner bulk amorphous phase [24].

Concerning the frequency dependence of the field corresponding to the Z_{zz} maximum, this value of H is linked to the average value of the anisotropy, H_K, at each frequency and to the effective anisotropy distribution in the samples. In the Fig. 8, it is observed that the MI peaks (Z_{zz} maximum) are shifted to higher magnetic fields as the frequency is raised. This shift is related to the variation of the skin depth with the frequency-dependent magnetic permeability. In fact, the presence of stress-induced anisotropy with the easy magnetization direction perpendicular to the ribbon axis leads to significantly high value of the transverse magnetic permeability for these samples. Increasing the driving frequency of the ac electric current through the ribbon, the decreasing of the surface layer thickness is produced partially by the high value of the transverse magnetic permeability, according to the classical electrodynamics theory of the skin effect.

Subsequently, the analyzed results for the frequency dependence of the magnetic anisotropy field, H_K, obtained from $Z_{zz}(H)$ curves in the range from 500 MHz up to 1000 MHz, for the three samples displayed in Fig. 8, are shown in Fig. 9.

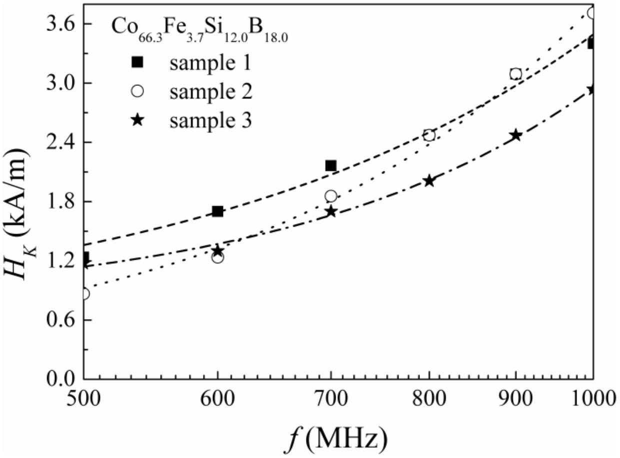


Fig. 9. Frequency dependence (range: 500–1000 MHz) of the anisotropy field for three $Co_{66.3}Fe_{3.7}Si_{12.0}B_{18.0}$ amorphous ribbons with different stress-annealing induced magnetic anisotropy. Symbols denote experimental data and lines denote theoretical fittings.

The data of the Fig. 9 have been fitted in accordance with eq. (7). Expressions obtained from this fitting as well as the anisotropy fields at 50 Hz are shown in the following Tab. 3:

Table 3. Fitting equations for the anisotropy field as a function of the frequency, in the range 500–1000 MHz, of stress-annealed ribbons

samples	Fitting equation	H_K (50 Hz) (fitting) A/m	H_K (50 Hz) (hysteresis loops) A/m
1	$H_K = 512 + (6.72)\times10^{-4} f^{2.1}$	512	360
2	$H_K = 74 + (2.01)\times10^{-4} f^{2.3}$	74	530
3	$H_K = 732 + (3.93)\times10^{-4} f^{2.4}$	732	550

In Tab. 3 the values of H_K deduced from the hysteresis loop are depicted and they are about the same order of magnitude but less than those deduced from eq. (7) for samples 1 and 3, while in the sample 2 the H_K value derived from the hysteresis loops is around seven times higher than the value obtained from the fitting. Such discrepancies are ascribed to the effect of the induced anisotropy in each ribbon, as it was previously discussed above.

4 Conclusions

Magnetoimpedance response of Co-rich amorphous $Co_{66.5}Fe_{3.5}Si_{12.0}B_{18.0}$ and $Co_{66.3}Fe_{3.7}Si_{12.0}B_{18.0}$ alloys ribbons results to be very sensitive to both the width parameter and stress-annealing treatment, as a consequence of the average process of the random distribution of magnetic moments through the width dimension in the first case, and to the induction of a uniform uniaxial anisotropy transverse to the ribbon axis in the second case, affecting drastically the soft magnetic character of the sample. Two-peak behavior of the GMI effect is observed in all as-cast studied ribbons, appearing the sharper peaks for a width equal to 0.70 mm although the effect is enhanced as increasing the value of the ribbon width. This behavior is explained in terms of the skin depth penetration and the dispersion of easy axes through the ribbon thickness.

This GMI response is sharper and narrower in the stress annealed ribbons than that exhibited by as-cast ribbons, which has to be considered as a quality factor for potential sensing applications.

Acknowledgements. This work has been supported by EU ERA-NET program under project “SoMaMicSens” (MANUNET-2010- Basque-3). Authors from UPV/EHU acknowledge financial support from the Department of Industry of the Basque Government (Programme SAIOTEK2011, Projects: S-PE11UN013, S-PE11UN087). FPI grant was acknowledged by L.G.-L. Technical and human support provided by SGIker is gratefully acknowledged.

References

[1] Panina, L.V., Mohri, K.: Magneto-impedance effect in amorphous wires. Appl. Phys. Lett. 65, 1189–1191 (1994)

[2] Beach, R.S., Berkowitz, A.E.: Giant magnetic field dependent impedance of amorphous FeCoSiB wire. Appl. Phys. Lett. 64, 3652 (1994)

[3] Kraus, L., Vázquez, M., Knobel, M.: Giant Magnetoimpedance. K. H. J. Buschow, Nort-Holland (2003)

[4] Zhukova, V., Chizhik, A., Zhukov, A., Torccunov, A., Larin, V., González, J.: Optimization of Giant Magnetoimpedance in Co-Rich Amorphous Microwires. IEEE Trans. Magn. 38, 3090 (2002)

[5] Zhukov, A., González, J., Blanco, J.M., Vázquez, M., Larin, V.: Microwires coated by glass: A new family of soft and hard magnetic materials. J. Mater. Res. 15, 2107 (2000)

[6] Ipatov, M., Zhukova, V., Zhukov, A., Gozález, J.: Magnetoimpedance sensitive to dc bias current in amorphous microwires. Appl. Phys. Lett. 97, 252507 (2010)

[7] Hernando, B., Sánchez, M.L., Prida, V.M., Tejedor, M., Vázquez, M.: Magnetoimpedance effect in amorphous and nanocrystalline ribbons. J. Appl. Phys. 90, 4783 (2001)

[8] Chen, L., Zhou, Y., Lei, C., Zhou, Z.-M., Ding, W.: Effect of meander structure and line width on GMI effect in micro-patterned Co-based ribbon. J. Phys. D: Appl. Phys. 42, 145005 (2009)

[9] García, C., Florez, J.M., Vargas, P., Ross, C.A.: Asymmetrical giant magnetoimpedance in exchange-biased NiFe. Appl. Phys. Lett. 96, 232501 (2010)

[10] Phan, M.-H., Peng, H.-X.: Giant magnetoimpedance materials: Fundamentals and applications. Progress in Materials Science 53, 323–420 (2008)

[11] Kraus, L.: Theory of giant magneto-impedance in the planar conductor with uniaxial magnetic anisotropy. J. Mag. Mag. Mater. 195, 764–778 (1999)
[12] Landau, L.D., Lifshitz, E.M.: Electrodynamics of Continuous Media. Pergamon, Oxford (1975)
[13] Panina, L.V., Mohri, K., Uchiyama, T., Noda, M.: Giant magneto-impedance in Co-rich amorphous wires and films. IEEE Trans. Magn. 31, 1249–1260 (1995)
[14] Barandiarán, J.M., Gacía-Arribas, A., de Cos, D.: Transition from quasistatic to ferromagnetic resonance regime in giant magnetoimpedance. J. Appl. Phys. 99, 103904 (2006)
[15] Mohri, K., Kohzawa, T., Kawashima, K., Yoshida, H., Panina, L.V.: Magneto-inductive effect (MI effect) in amorphous wires. IEEE Trans. Magn. 28, 3150–3152 (1992)
[16] Narita, K., Yamasaki, J., Fukunaga, H.: Measurement of Saturation Magnetostriction of a Thin Amorphous Ribbon by Means of Small-Angle Magnetization Rotation. IEEE Trans. Magn. 16, 435 (1980)
[17] Ipatov, M., Zhukova, V., Zhukov, A., Gonzalez, J., Zvezdin, A.: Low-field hysteresis in the magnetoimpedance of amorphous microwires. Phys. Rev. B 81, 134421 (2010)
[18] Ipatov, M., Zhukova, V., Zhukov, A., González, J.: Expanding the longitudinal magnetoimpedance sensor range by direct bias current. J. Appl. Phys. 113, 203902 (2013)
[19] Panina, L.V., Mohri, K.: High-frequency giant magneto-impedance in Co-rich amorphous wires and films. J. Mag. Soc. Japan 19, 265–268 (1995)
[20] García-Miquel, H., Esbrí, M.J., Andrés, J.M., García, J.M., García-Beneytez, J.M., Vázquez, M.: Power absorption and ferromagnetic resoonance in Co-rich metallic glasses. IEEE Trans. Magn. 37, 561 (2001)
[21] Nielsen, O.V.: Effects of longitudinal and torsional stress annealing on the magnetic anisotropy in amophous ribbon materials. IEEE Trans. Magn. 21, 2008 (1985)
[22] González, J., Kulakowski, K.: Stress and field induced magnetic anisotropies in Co-rich amorphous alloys. J. Magn. Magn. Mater. 82, 94 (1989)
[23] Záveta, K., Nielsen, O.V., Jurek, K.: A domain study of magnetization processes in a stress-annealed metallic glass ribbon for fluxgate sensors. J. Magn. Magn. Mater. 117, 61 (1992)
[24] Rheem, Y.W., Kim, C.G., Kim, C.O., Yoon, S.S.: Hysteretic characteristics of giant magnetoimpedance due to the exchange coupling in annealed amorphous materials. J. Appl. Phys. 91, 7433 (2002)
[25] Kim, C.G., Kim, C.O., Yoon, S.S.: The role of exchange coupling on the giant magnetoimpedance of annealed amorphous materials. J. Magn. Magn. Mater. 249, 293 (2002)
[26] Shin, K.H., Graham, J.C.D., Zhou, P.Y.: Asymmetric hysteresis loops in Co-based ferromagnetic alloys. IEEE Trans. Magn. 28, 2772 (1992)
[27] Kim, C.G., Rheem, Y.W., Kim, C.O., Shalyguina, E.E., Ganshina, E.A.: Magnetostatic prperties of heterogeneous Co-based amorphous/crystalline phases. J. Magn. Magn. Mater. 262, 412 (2003)

Hybrid Artificial Intelligent Control for Feedback Control of Outer Rotor Spherical Actuator

J. Chu, N. Niguchi, and K. Hirata*

Department of Adaptive Machine Systems, Graduate School of Engineering,
Osaka University, Osaka, Japan
k-hirata@ams.eng.osaka-u.ac.jp

Abstract. Artificial intelligent control methods using fuzzy control, neural networks, and genetic algorithm have recently been recognized as important algorithms that can improve the performances of electronic devices. These methods use their algorithms for solving problems of devices of nonlinear condition. An adaptive control algorithm adjusts the nonlinear system with various parameters for demand performance. In this chapter, we propose a feedback control method using the adaptive neuro-fuzzy inference system (ANFIS), which is a hybrid algorithm that combines fuzzy control, neural network control and adaptive control, for an outer rotor spherical actuator. The proposed algorithm has the combined advantage of expert knowledge of the fuzzy inference system (FIS), learning capabilities of the neural networks (NN) for control of a nonlinear system and adaptabilities of adaptive control for adjusting various parameters. An example application is actuator drives which require adaptability and robustness. This control method using ANFIS is expected to produce more accurate rotor positions compared to other control methods. In order to verify this, experiments using a prototype of an outer rotor spherical actuator fitted with rotary encoders are conducted using a dSPACE controller with MATLAB/Simulink, and we compare the accuracy of the ANFIS control method with the traditional PID control method for feedback control of the outer rotor spherical actuator.

Keywords: hybrid control, artificial intelligent control, feedback control, adaptive neuro-fuzzy inference system, rotary encoder, spherical actuator.

1 Introduction

Conventionally, multi-DOF (Degree Of Freedom) devices use a number of single-DOF actuators. Therefore traditional multi-DOF devices have a complicated and consequentially heavy structure [1]. In addition, these devices have a slow response and low positioning precision due to their inevitable problems such as a large backlash, non-linear friction and kinematic singularities in the operation range. The number of actuators and the weight can be reduced and the structure can be simplified

---

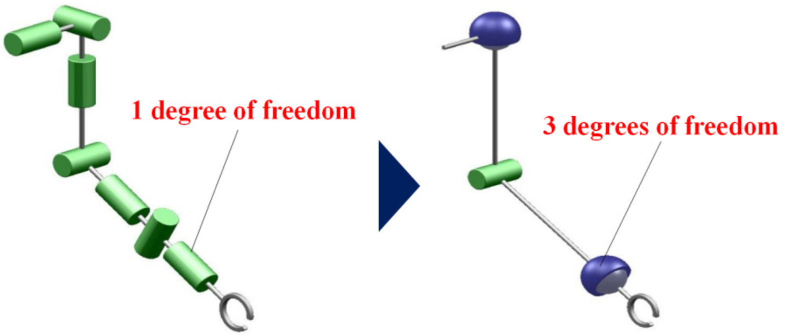
* Corresponding author.

A. Mason et al. (eds.), *Sensing Technology: Current Status and Future Trends IV*,
Smart Sensors, Measurement and Instrumentation 12, DOI: 10.1007/978-3-319-12898-6_13

by constructing a single multi-DOF actuator. The difference in between using single and spherical multi-DOF actuators in a multi-DOF mechanical structure such as a robot arm is shown in Fig. 1, where the spherical actuator is particularly suitable for applications such as robot wrists, elbows, shoulders and eyes [2]. Other applications include satellite attitude control [3] and in automobiles such as active suspension adaptive head lamps and side-view mirrors. Due to the numerous possibilities, various spherical actuators have been proposed [4-13]. For example, new 3-DOF inner rotor spherical actuators have been proposed [14, 15]. Among the many types of spherical actuators, the outer rotor spherical actuator can produce higher torques than the inner rotor spherical actuator of the same size [16]. Control of existing spherical actuators is very complicated, because spherical actuators are nonlinear devices. Therefore, complicated feedback control systems of the 3-DOF spherical actuator using an image sensor [17], consecutive single-axis encoders with a serial or parallel mechanism [18, 19] and field-feedback control of the spherical actuator using special embedded sensors [20] are used for more accurate feedback control.

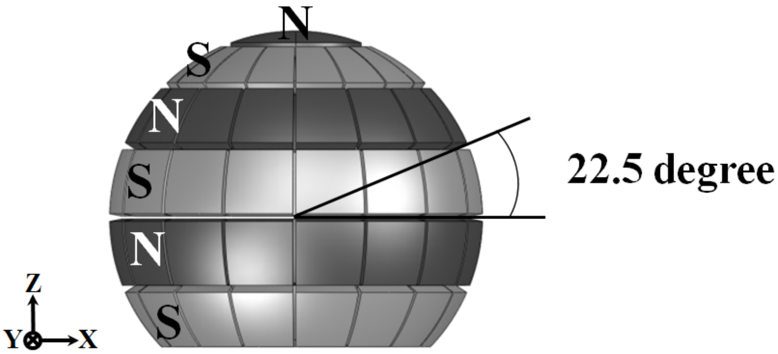
Many feedback control algorithms for actuator control have been proposed such as PID (Proportional, Integral, and Differential) control [21], fuzzy logic control [22], neural network control [23] and hybrid intelligent control. A PID controller is a control loop feedback mechanism controller widely used in industrial control systems. The PID controller calculates an error value as the difference between a measured process variable and a desired value using fixed gains which has to be obtained through many experiments, which is a disadvantage. Fuzzy logic and neural networks are natural complementary tools in building intelligent control systems. While neural networks are low level computational structures that draw good performances when dealing with raw data, fuzzy logic deals with inference on a higher level using linguistic information acquired from the domain experts. Hybrid intelligent control combines at least two intelligent technologies such as the ANFIS controller for torque ripple control of translational meshing motors [24] or the parallel fuzzy controller for high-performance brushless motors [25]. The integrated neuro-fuzzy systems can combine the parallel computation and learning abilities of the neural networks with human-like knowledge representation and explanation abilities of the fuzzy systems. As a result, the neural networks become more transparent, while the fuzzy systems become capable of learning. The Adaptive Neuro-Fuzzy Inference System (ANFIS), which was developed in the early 1990s by Jang [26], combines the concepts of the fuzzy logic and neural networks to form a hybrid intelligent system that enhances the ability to automatically learn and adapt. The adaptive network is a network of nodes and directional links. These networks learn the relationship between inputs and outputs using learning rules such as back propagation [28]. Therefore, a control method using ANFIS is expected to produce more accurate results compared to other control methods.

In this chapter, we propose a feedback control method using ANFIS for controlling an outer rotor spherical actuator and conduct experiments using ANFIS with a dSPACE controller and MATLAB/Simulink.

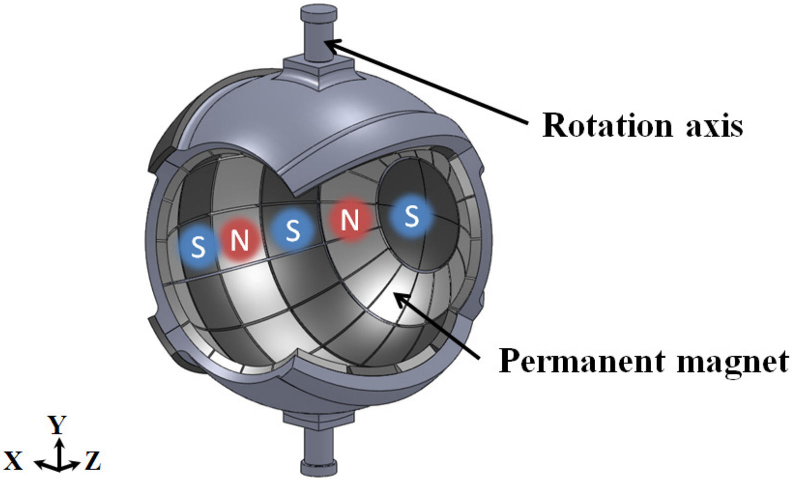


(a) Used 1-DOF actuator (b) Used multi-DOF actuator

Fig. 1. Mechanical structure of multi-DOF device



(a) Magnets arrangement of rotor

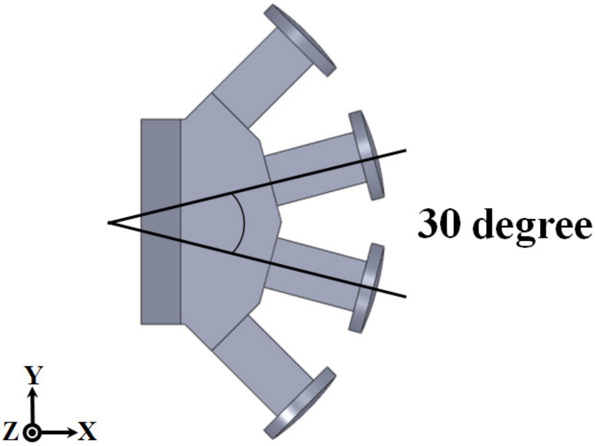


(b) Structure of rotor

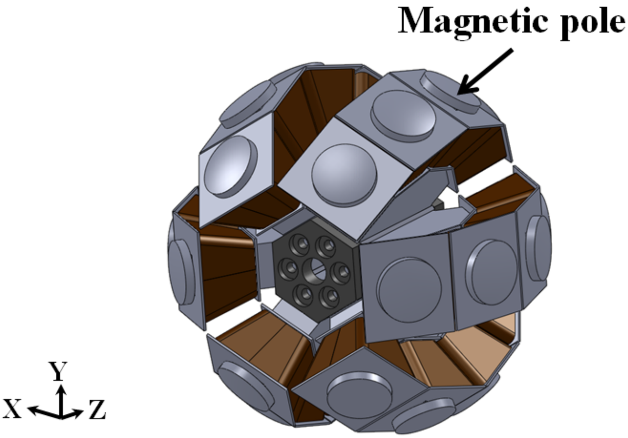
Fig. 2. Rotor of the outer rotor spherical actuator

2 Basic Structure of Outer Rotor Spherical Actuator

The outer rotor spherical actuator is composed of a stator and rotor as shown in Fig. 2 and 3. In the rotor, six rows of identically polarized N and S pole shell-shaped permanent magnets are arranged around the Z-axis. On the other hand, the stator has 24 teeth with 310 turn concentrated windings on each pole. They are also arrayed around the Z-axis at even intervals as shown in Fig. 4. This structure makes it similar to controlling a 16-pole-12-slot synchronous motor around the X- and Y-axes.



(a) Magnetic pole of stator



(b) Structure of magnetic poles

Fig. 3. Stator of the outer rotor spherical actuator

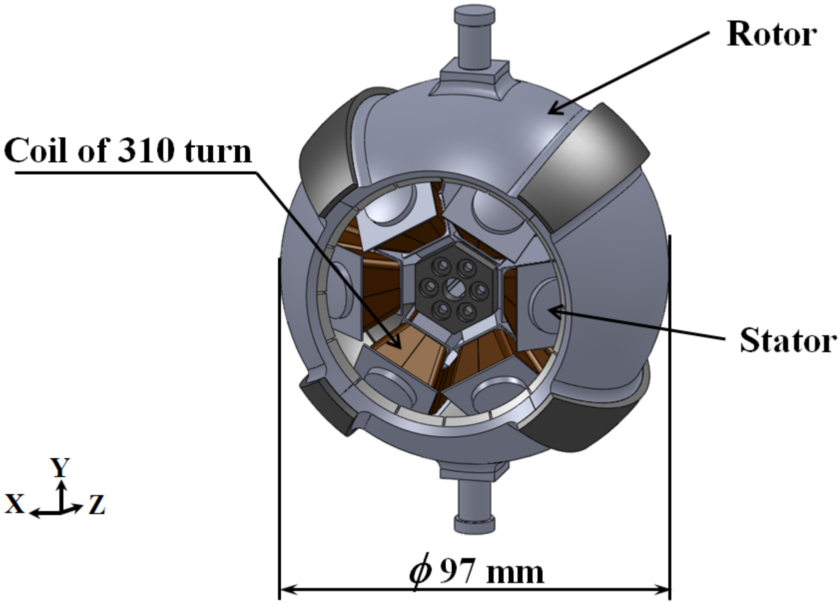


Fig. 4. Outer rotor spherical actuator in this study

3 Control Method

ANFIS is a class of adaptive networks that are functionally equivalent to the fuzzy inference systems and consists of 5 layers. The architecture of ANFIS is shown in Fig. 5, which is composed of an input layer (layer 1), fuzzification (layer 2), application of the sugeno-type fuzzy rule (layer 3), normalization (layer 4), defuzzification (layer 5) and the output.

We assume the fuzzy inference system under consideration has input x and y and one output z. For a first order Sugeno fuzzy model, a typical rule set with two fuzzy IF-THEN rules can be expressed as

$$If\ \ x\ is\ A_1\ and\ \ y\ is\ B_1\quad THEN\ \ f_1 = p_1 x + q_1 y + r_1 \tag{1}$$

$$If\ \ x\ is\ A_2\ and\ \ y\ is\ B_2\quad THEN\ \ f_2 = p_2 x + q_2 y + r_2 \tag{2}$$

Fig. 6 illustrates the reasoning mechanism for this Sugeno model [27].

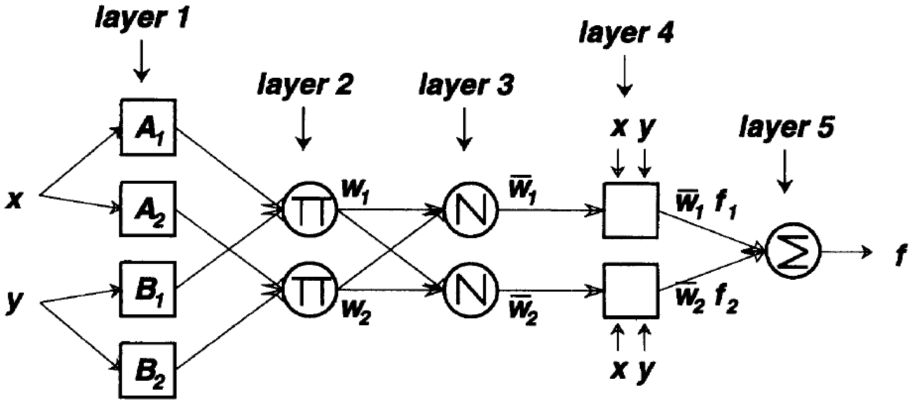


Fig. 5. ANFIS architecture

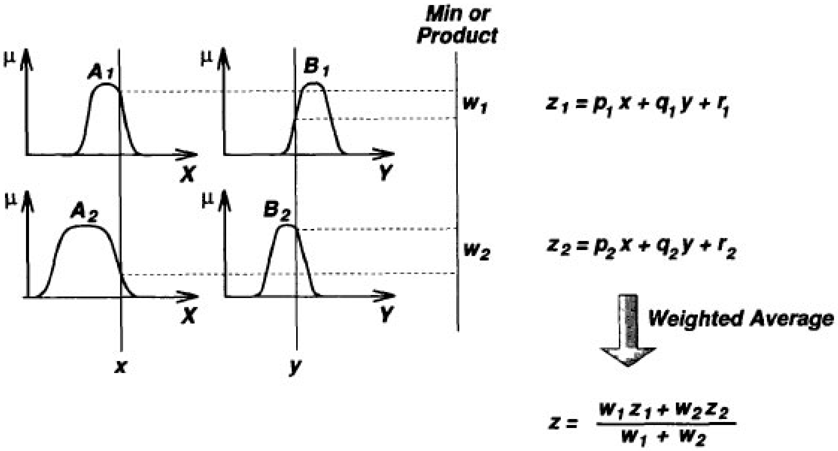


Fig. 6. Two-input first-order Sugeno fuzzy model with two rules

Layer 1: Each node *i* in this layer is an adaptive node with a node output defined by (3).

$$O_{1,i} = \mu_{A_i}(x) \qquad for \quad i = 1,2 \tag{3}$$

$$O_{1,i} = \mu_{B_{i-2}}(y) \qquad for \quad i = 3,4$$

where x (or y) is an input to the node and A_i (or B_{i-2}) is a fuzzy set associated with this node. In other words, outputs of this layer are membership values of the premise part. The membership functions for A_i and B_i can be any appropriate parameterized membership functions introduced in Section 2. A_i can be characterized by a generalized bell function (4).

$$\mu_A(x) = \frac{1}{1 + \left| \frac{x - c_i}{a_i} \right|^{2b_i}} \tag{4}$$

where $\{a_i, b_i, c_i\}$ is a parameter set. Parameters in this layer are referred to as premise parameters.

Layer 2: Each node in this layer is a fixed node labeled $\prod$, which multiplies the incoming signals and outputs the product by (5).

$$O_{2,i} = w_i = \mu_{A_i}(x) \times \mu_{B_i}(y), \quad i = 1, 2 \tag{5}$$

Each node output represents the firing strength of the rule.

Layer 3: Each node in this layer is a fixed node labeled N. The *i*-th node calculates a ratio of the i-th rule's firing strength to the sum of all rules' firing strengths. The output of each node in this layer is given by (6).

$$O_{3,i} = \overline{w}_i = \frac{w_i}{w_1 + w_2}, \quad i = 1, 2. \tag{6}$$

The outputs of this layer will be called the normalized firing strengths.

Layer 4: Each node *i* in this layer is an adaptive node with a node function by (7)

$$O_{4,i} = \overline{w_i} f_i = \overline{w_i}(p_i x + q_i y + r_i) \tag{7}$$

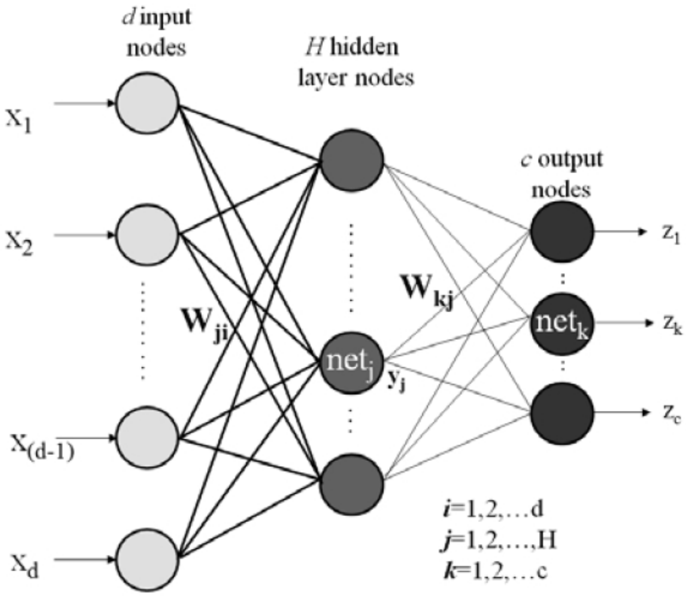
where $\overline{w_i}$ is an output of layer 3 and $\{p_i, q_i, r_i\}$ is the parameter set. Parameters in this layer will be referred to as consequent parameters.

Layer 5: the single node in this layer is a fixed node labeled $\sum$, which computes the overall output as a summation of all incoming signals. The output of this single node is given by (8).

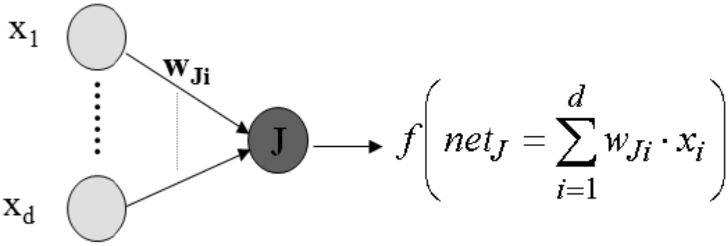
$$O_{5,i} = \sum_i \overline{w_i} f_i = \frac{\sum_i w_i f_i}{\sum_i w_i} \tag{8}$$

The basic learning rule is a back propagation gradient descendent, which calculates error signals recursively from the output layer backward to the input nodes. This learning rule is exactly the same as the back propagation learning rule used in the common feed-forward neural networks without feedback or loop connection.

Fig. 7 shows the typical multilayer neural network, whose layers are classified into the input, output, and hidden layers. The back propagation algorithm propagates errors inversely from the output layer to the hidden layer.



(a) Generalized multilayered neural networks



(b) Basic input/output expression of multilayered neural networks

Fig. 7. Basic concept of multilayered neural networks

Using Fig. 7 (b), the input and output signals can be defined at each layer, as follows:

$$y_j = f(net_j) = f\left(\sum_{i=1}^{d} w_{ji}\, x_i\right) \tag{9}$$

$$z_k = f(net_k) = f\left(\sum_{j=1}^{H} w_{kj}\, y_j\right)$$

The weighted value *w* can be expressed as follows, based on gradient descent error minimization by (10).

$$w(k+1)=w(k)+\Delta w(k) \tag{10}$$

$$\Delta w=-\eta\frac{\partial J(w)}{\partial w}$$

where η is the learning rate and J can be expressed as follows:

$$J(w)=\frac{1}{2}\sum_{k=1}^{c}(t_k-z_k)=\frac{1}{2}\|t-z\|^2 \tag{11}$$

By arranging the output and hidden layers using the chain rule, the equation can be expressed as follows [28]:

$$\Delta w_{ji}=-\eta\frac{\partial J}{\partial w_{ji}}=\eta\left[\sum_{k=1}^{c}\delta_k w_{kj}\right]f'(net_i)x_i=\eta\cdot\delta_j\cdot x_i \tag{12}$$

The connection weight of the back propagation algorithm w_{ji} was generally initialized at a value of −0.5 ~ 0.5 [28].

4 Experimentation

4.1 Proposed Controller Scheme and Prototype

A schematic diagram of the experimental system using the ANFIS controller is shown in Fig. 8. In this system, feedback control is conducted by ANFIS which adds a modification signal into the desired trajectory before it is input into the current generator.

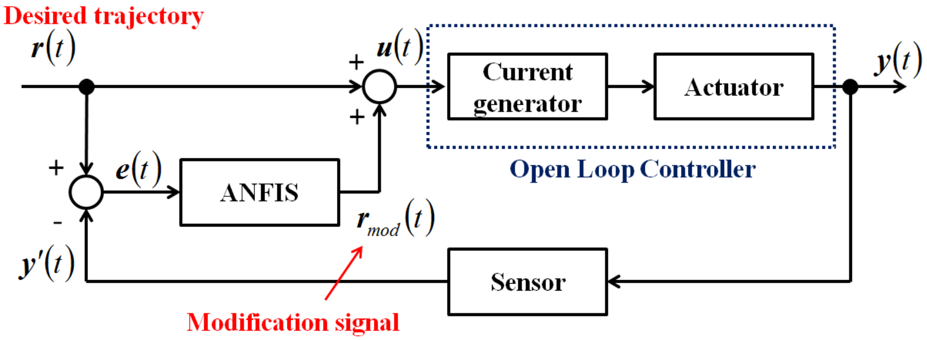


Fig. 8. Block diagram of the proposed controller scheme

Current patterns of the current generator are decided by using the block diagram shown in Fig. 9 [16]. When a desired signal is given as a trajectory, this system produces current patterns for all coils to track the given trajectory. The block diagram is explained below.

When a desired trajectory is given, the system calculates a current pattern for all coils as follows:

$$I_i = -\sin(8\alpha_i) \tag{13}$$

where the subscript i expresses the i-th pole, I_i and A_i are the current and amplitude, respectively, and α_i is the mechanical angle between the i-th EM pole and unknown parameters A_i and α_i are derived from the following equations:

$$A_i = \frac{1}{\left\| \boldsymbol{k}'' \times \boldsymbol{p}_i'' \right\|} \left(k'' \times \boldsymbol{p}_i'' \right)_{x''} \tag{14}$$

$$\alpha_i = \tan^{-1}\left(\frac{p''_{zi}}{\sqrt{p''^2_{xi} + p''^2_{yi}}} \right) \tag{15}$$

where $\boldsymbol{k}''$ and $\boldsymbol{P}_i''=[p_{ix}''\ p_{iy}''\ p_{iz}'']^{\mathrm{T}}$ are unit vectors of the axis Z'' axis and position vector observed in the rotor coordinate, respectively. To calculate those two parameters, the position vector $\mathbf{p}_i''$ should be known and it is calculated by using the following equations:

$$\boldsymbol{p}_i = {}^{WORLD}T_{AXIS}\ {}^{AXIS}T_{ROTOR}\, \boldsymbol{p}_i'' \tag{16}$$

where $\boldsymbol{P}_{\mathrm{i}}$ is a position vector of the i-th vector observed in the world coordinate, ${}^{WORLD}T_{AXIS}$ and ${}^{AXIS}T_{ROTOR}$ are the transformation matrixes from the world coordinate ($X\ Y\ Z$) to the axis coordinate ($X'\ Y'\ Z'$) and from the axis coordinate to the rotor coordinate($X''\ Y''\ Z''$), respectively.

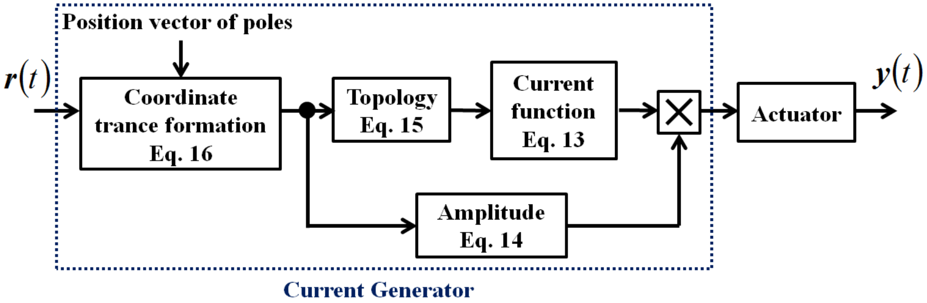


Fig. 9. Block diagram of the open loop controller

A schematic of the control system using a prototype of the actuator is shown Fig. 10. The prototype of the actuator used in this research is shown Fig. 11. The rotor is supported by a two-DOF gimbal structure. Rotary encoders are installed in both axes of the gimbal structure, and the dynamic characteristics are measured.

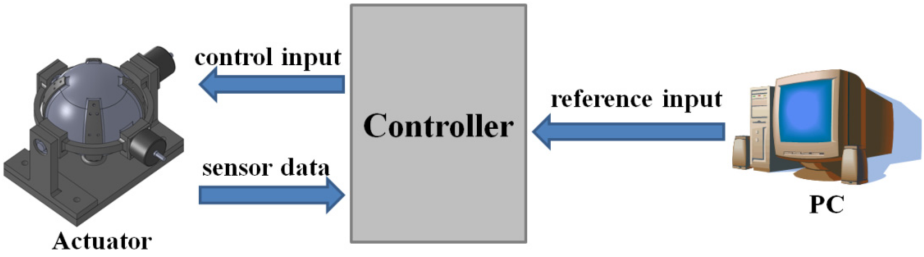


Fig. 10. Schematic of the control system

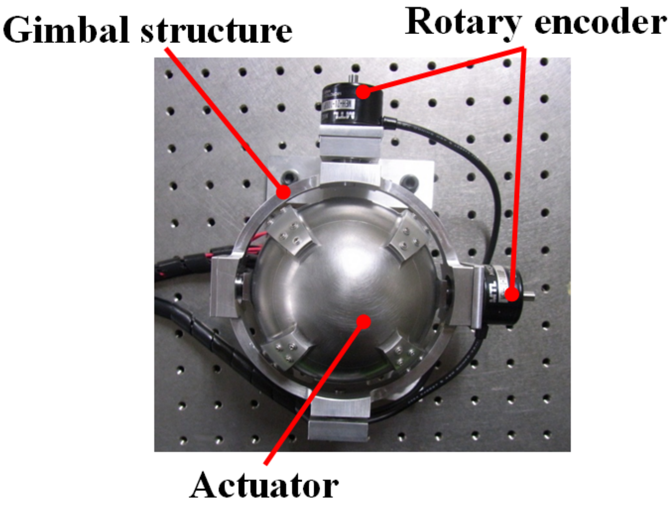


Fig. 11. Prototype of the actuator

4.2 Experimental Result

The angle of rotation on the X and Y axes is ±30 degree, as shown in Fig. 12. Basically, the rotor is designed to have a ±35 degree working range. However, to prevent bumping the coil cover and the end of the yoke, the proposed spherical actuator should be tested in the safe region.

All actual rotation angles of actuator are shown with solid lines obtained from the two rotary encoders of the gimbal structure.

The result of the experimentation under open loop control is shown in Fig. 12 and Fig. 13 where both rotation angles obtained from the rotary encoder are far from the desired values. These results indicate that a control method without sensor feedback systems makes it impossible to obtain the desired values. When the actuator changes the direction, the shape of solid line is rough because the open loop control system does not use the feedback values.

The result of the experimentation under PID control is shown in Fig. 14 and 15 where the position feedback system performs rather well, although small errors can be observed. The actual rotation angle of PID control has more smooth line like the shape of desired angle. But, the actual angles are not accurate values.

The result of the experimentation under ANFIS control is shown in Fig. 16 and 17 where the ANFIS control method is much more accurate and the error value between the desired and actual angles are smaller than that of open loop control and PID control.

The average errors when open loop, PID and ANFIS control are used are shown in Table 1. The average errors of open loop, PID and ANFIS control are about 5 degree, about 2 degree and about 1 degree.

As mentioned above, the ANFIS control is superior in terms of accuracy.

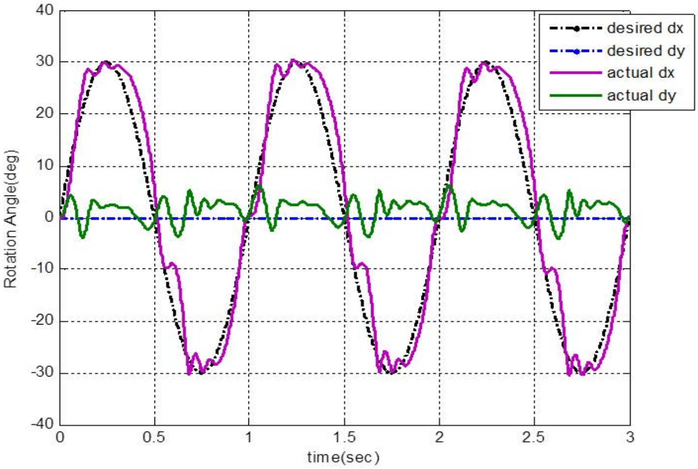


Fig. 12. Result of the open loop control, around X-axis

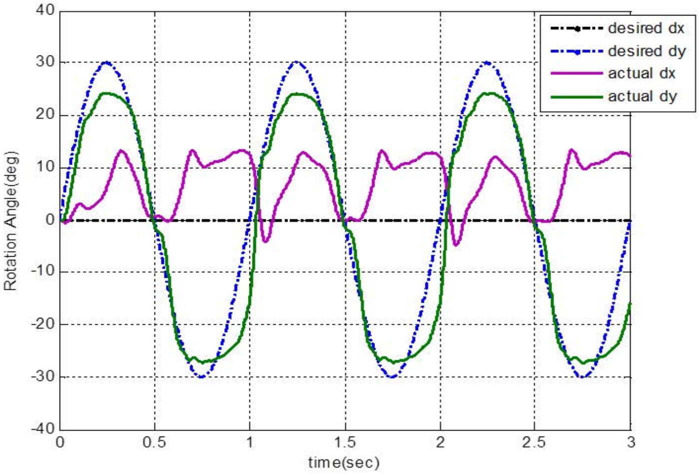


Fig. 13. Result of the open loop control, around Y-axis

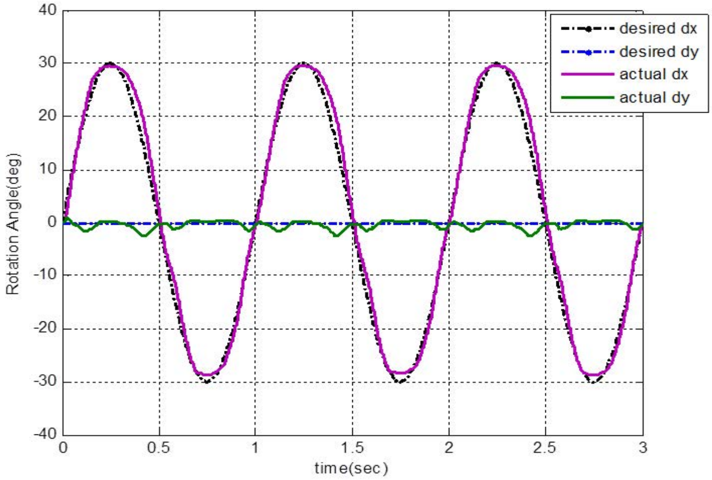


Fig. 14. Result of the PID control, around X-axis

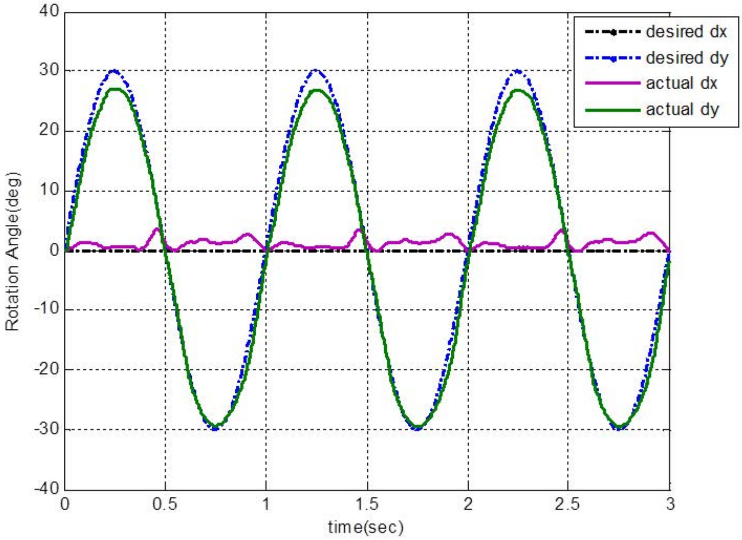


Fig. 15. Result of the PID control, around Y-axis

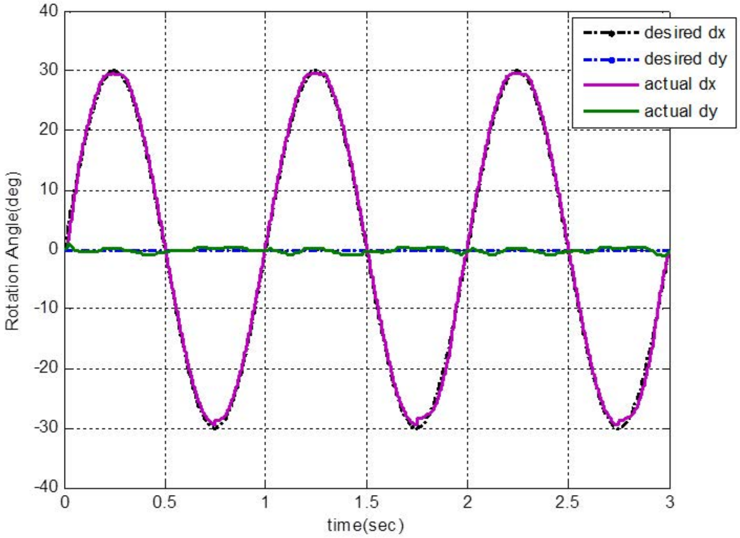


Fig. 16. Result of the ANFIS control, around X-axis

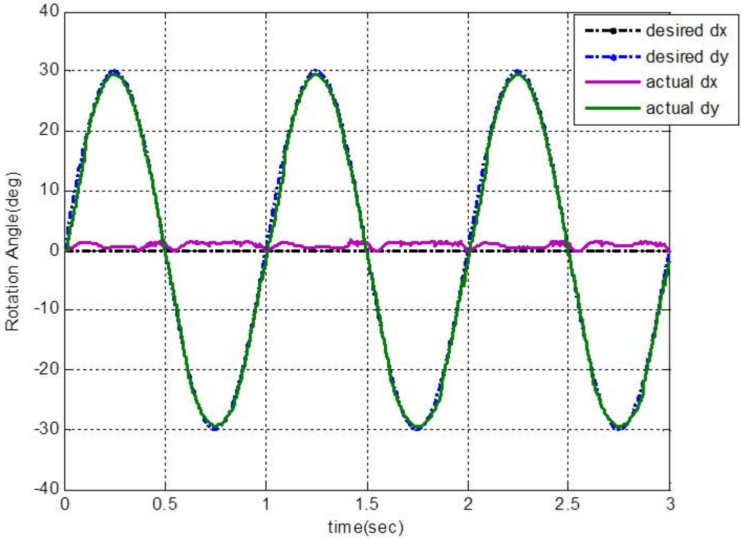


Fig. 17. Result of the ANFIS control, around Y-axis

Table 1. Average error of each control method

Control Method	Average error around X axis (degree)	Average error around Y axis (degree)
Open loop control	4.377	5.515
PID control	2.000	2.032
ANFIS control	0.740	1.036

5 Conclusion

This chapter proposed a feedback control method for an outer rotor spherical actuator using an adaptive neuro-fuzzy inference system. The proposed neuro-fuzzy based controller incorporates fuzzy logic laws with a five-layer artificial neural network. Applying ANFIS controller was very easy to design of controller parameters, because those parameters obtained automatically to throw learning process. The best response of feedback controller of spherical actuator base ANFIS was obtained which was the bell function with 5 numbers of membership function, hybrid learning and linear premise parameter. The results of the experiments proved the effectiveness of the feedback control method using the adaptive neuro-fuzzy inference system. We confirmed that the rotor accurately followed the desired value. In the near future, we will extend ANFIS in order to minimize errors of the controller by combining genetic algorithm as well.

References

[1] Mckerrow, P.J.: Introduction to Robotics. Addison Wesley Publishing Company (1991)

[2] Rashid, M.K., Khalil, Z.A.: Configuration Design and Intelligent Stepping of a Spherical Motor in Robotic Joint. Journal of Intelligent and Robotic Systems 40, 165–181 (2004)

[3] Rossini, L., Ch'etelat, O., Onillon, E., Perriard, Y.: Force and Torque Analytical Models of a Reaction Sphere Actuator Based on Spherical Harmonic Rotation and Decomposition. IEEE/ASME Transactions PP, 1–13 (2012)

[4] Weck, E.H.M., Reinartz, T., Henneberger, G., De Doncker, R.W.: Design of a spherical motor with three degree of freedom. Annals CIRP 49, 289–294 (2000)

[5] Xia, C., Li, H., Shi, T.: 3-D Magnetic Field and Torque Analysis of a Novel Halbach Array Permanent-Magnet Spherical Motor. IEEE Transactions on Magnetics 44(8), 2016–2020 (2008)

[6] Chen, W., Zhang, L., Yan, L., Liu, J.: Design and control of a three degree-freedom permanent magnet spherical actuator. Sens. Actuators A: Phys., 75–86 (2012)

[7] Yan, L., Chen, I.-M., Lim, C.K., Yang, G., Lin, W., Lee, K.-M.: Design and Analysis of a Permanent Magnet Spherical Actuator. IEEE/ASME Transactions on Mechatronics 13(2), 239–248 (2008)

[8] Gofuku, A., Sasaki, R., Yano, T., Wada, Y., Shibata, M.: Development of a spherical stepping motor rotating around six axes. Int. J. Appl. Electromagn. Mechan. 39, 905–911 (2012)

[9] Wang, J., Mitchell, K., Jewell, G.W., Howe, D.: Multi-Degree-of-Freedom Spherical Permanent Magnet Motor. In: Proc. Int. Conf. Robot, pp. 1798–1805, Autom. (2001)

[10] Yan, L., Chen, I.-M., Yang, G., Lee, K.-M.: Analytical and experimental investigation on the magnetic field and torque of a permanent magnet spherical actuator. IEEE/ASME Transactions on Mechatronics 11(4), 409–419 (2006)

[11] Shigeki, T., Guoqiang, Z., Osamu, M.: Development of new generation spherical ultrasonic motor. In: Proceedings of the IEEE International Conference on Robotics and Automation 1996, vol. 2873, pp. 2871–2876 (1996)

[12] Purwanto, E., Toyama, S.: Control Method of a spherical ultrasonic motor. In: IEEE/ASME International Conference on Advanced Intelligent Mechatronics, vol. 1322, pp. 1321–1326 (2003)

[13] Zhao, C., Li, Z., Huang, W.: Optimal design of the stator of a three-DOF ultrasonic motor. Sensors and Actuators A: Physical 121(2), 494–499 (2005)

[14] Ikejiri, S., Hirata, K., Maeda, S.: Proposal of electromagnetic spherical actuator with 3-DOF. Int. J. COMPEL 29(4), 994–1003 (2010)

[15] Maeda, S., Hirata, K.: Characteristics Analysis of a New Spherical Actuator Employing 3-D FEM. In: Proc. IEEE COMPUMAG PB4.1, pp. 29–30 (2011)

[16] Tsukano, M., Sakaidani, Y., Hirata, K., Niguchi, N., Maeda, S., Zaini, A.: Analysis of 2-Degree of Freedom Outer Rotor Spherical Actuator Employing 3-D Finite Element Method. IEEE Transactions on Magnetics 49(5), 2233–2236 (2013)

[17] Maeda, S., Hirata, K., Tong, M.: Feedback control of electromagnetic actuator with three degrees of freedom using optical image sensors. Electr. Eng. Jpn. 181(2) (2012)

[18] Lee, K.-M., Roth, R.B., Zhou, Z.: Dynamic Modeling and Control of a Ball-Joint-Like Variable-Reluctance Spherical Motor. Journal of Dynamic Systems Measurement and Control 118, 29–40 (1996)

[19] Lee, K.-M., Sosseh, R.A., Wei, Z.: Effects of the torque model on the control of a VR spherical motor. Control Engineering Practice 12, 1437–1449 (2004)

[20] Bai, K., Lee, K.-M.: Magnetic Field Model for Direct Field-Feedback Control of a Permanent Spherical Motor. In: IEEE/ASME International Conference on Advanced Intelligent Mechatronics, pp. 768–773 (2013)

[21] Ang, K.H., Li, C.G.Y.: PID control system analysis, design, and technology. IEEE Transactions on Control Systems Technology 13(4), 559–576 (2005)

[22] Lee, C.C.: Fuzzy logic in control systems: fuzzy logic controller.I. IEEE Transactions on Systems, Man and Cybernetics 20(2), 404–418 (1990)

[23] Narendra, K.S., Parthasarathy, K.: Identification and control of dynamical systems using neural networks. IEEE Transactions on Neural Networks 1(1), 4–27 (1990)

[24] Wen, X., Liao, Q., Wei, S., Xu, Q.: Research of Translational Meshing Motor Control Based on Adaptive Neuro-Fuzzy Inference System. In: International Symposium on Intelligent Information Technology Application (2009)

[25] Rubaai, A., Castro-Sitiriche, M.J., Ofoli, A.R.: Design and Implementation of Parallel Fuzzy PID Controller for High-Performance Brushless Motor Drives: An Integrated Environment for Rapid Control Prototyping. IEEE Transactions on 44(4) (2008)

[26] Jang, J.-S.R., Sun, C.-T.: Neuro-fuzzy modeling and control. Proceedings of the IEEE 83(3), 378–406 (1995)

[27] Takagi, T., Sugeno, M.: Derivation of fuzzy control rules from human operator's control actions. In: Proceedings IFAC Symposium Fuzzy Information, Knowledge Representation and Decision Analysis, pp. 55–60 (July 1983)

[28] Kim, D.S.: Neural Networks: Theory and application. Jinhan M&B (2010)

Author Index