RECENT ADVANCES on ELECTROSCIENCE and COMPUTERS

Proceedings of the International Conference on Systems, Control, Signal Processing and Informatics (SCSI 2015)

Proceedings of the International Conference on Electronics and Communication Systems (ECS 2015)

Barcelona, Spain
April 7-9, 2015

RECENT ADVANCES on ELECTROSCIENCE and COMPUTERS

Proceedings of the International Conference on Systems, Control, Signal Processing and Informatics (SCSI 2015)

Proceedings of the International Conference on Electronics and Communication Systems (ECS 2015)

Barcelona, Spain
April 7-9, 2015

All papers of the present volume were peer reviewed by no less than two independent reviewers. Acceptance was granted when both reviewers' recommendations were positive.

Series: Recent Advances in Electrical Engineering Series | 46

ISSN: 1790-5117
ISBN: 978-1-61804-290-3

RECENT ADVANCES on ELECTROSCIENCE and COMPUTERS

Proceedings of the International Conference on Systems, Control, Signal Processing and Informatics (SCSI 2015)

Proceedings of the International Conference on Electronics and Communication Systems (ECS 2015)

Barcelona, Spain
April 7-9, 2015

Organizing Committee

Editors:
Professor Nikos E. Mastorakis, Technical University of Sofia, Bulgaria
Professor Imre Rudas, Obuda University, Budapest, Hungary
Professor Marina V. Shitikova, Voronezh State University of Architecture and Civil Engineering, Russia
Professor Yuriy S. Shmaliy, Universidad de Guanajuato, Salamanca, Mexico

Program Committee:
Prof. Sonia Tarragona (Univerdidad de Le?n, Spain)
Prof. Lotfi Zadeh (IEEE Fellow,University of Berkeley, USA)
Prof. Leon Chua (IEEE Fellow,University of Berkeley, USA)
Prof. Michio Sugeno (RIKEN Brain Science Institute (RIKEN BSI), Japan)
Prof. Dimitri Bertsekas (IEEE Fellow, MIT, USA)
Prof. Demetri Terzopoulos (IEEE Fellow, ACM Fellow, UCLA, USA)
Prof. Georgios B. Giannakis (IEEE Fellow, University of Minnesota, USA)
Prof. George Vachtsevanos (Georgia Institute of Technology, USA)
Prof. Abraham Bers (IEEE Fellow, MIT, USA)
Prof. Brian Barsky (IEEE Fellow, University of Berkeley, USA)
Prof. Aggelos Katsaggelos (IEEE Fellow, Northwestern University, USA)
Prof. Josef Sifakis (Turing Award 2007, CNRS/Verimag, France)
Prof. Hisashi Kobayashi (Princeton University, USA)
Prof. Kinshuk (Fellow IEEE, Massey Univ. New Zeland),
Prof. Leonid Kazovsky (Stanford University, USA)
Prof. Narsingh Deo (IEEE Fellow, ACM Fellow, University of Central Florida, USA)
Prof. Kamisetty Rao (Fellow IEEE, Univ. of Texas at Arlington,USA)
Prof. Anastassios Venetsanopoulos (Fellow IEEE, University of Toronto, Canada)
Prof. Steven Collicott (Purdue University, West Lafayette, IN, USA)
Prof. Nikolaos Paragios (Ecole Centrale Paris, France)
Prof. Nikolaos G. Bourbakis (IEEE Fellow, Wright State University, USA)
Prof. Stamatios Kartalopoulos (IEEE Fellow, University of Oklahoma, USA)
Prof. Irwin Sandberg (IEEE Fellow, University of Texas at Austin, USA),
Prof. Michael Sebek (IEEE Fellow, Czech Technical University in Prague, Czech Republic)
Prof. Hashem Akbari (University of California, Berkeley, USA)
Prof. Yuriy S. Shmaliy, (IEEE Fellow, The University of Guanajuato, Mexico)
Prof. Lei Xu (IEEE Fellow, Chinese University of Hong Kong, Hong Kong)
Prof. Paul E. Dimotakis (California Institute of Technology Pasadena, USA)
Prof. Martin Pelikan (UMSL, USA)
Prof. Patrick Wang (MIT, USA)
Prof. Wasfy B Mikhael (IEEE Fellow, University of Central Florida Orlando,USA)
Prof. Sunil Das (IEEE Fellow, University of Ottawa, Canada)
Prof. Panos Pardalos (University of Florida, USA)
Prof. Nikolaos D. Katopodes (University of Michigan, USA)
Prof. Bimal K. Bose (Life Fellow of IEEE, University of Tennessee, Knoxville, USA)
Prof. Janusz Kacprzyk (IEEE Fellow, Polish Academy of Sciences, Poland)
Prof. Sidney Burrus (IEEE Fellow, Rice University, USA)
Prof. Biswa N. Datta (IEEE Fellow, Northern Illinois University, USA)
Prof. Mihai Putinar (University of California at Santa Barbara, USA)
Prof. Wlodzislaw Duch (Nicolaus Copernicus University, Poland)
Prof. Tadeusz Kaczorek (IEEE Fellow, Warsaw University of Tehcnology, Poland)
Prof. Michael N. Katehakis (Rutgers, The State University of New Jersey, USA)
Prof. Pan Agathoklis (Univ. of Victoria, Canada)
Dr. Subhas C. Misra (Harvard University, USA)
Prof. Martin van den Toorn (Delft University of Technology, The Netherlands)

Prof. Malcolm J. Crocker (Distinguished University Prof., Auburn University,USA)
Prof. Urszula Ledzewicz, Southern Illinois University , USA.
Prof. Dimitri Kazakos, Dean, (Texas Southern University, USA)
Prof. Ronald Yager (Iona College, USA)
Prof. Athanassios Manikas (Imperial College, London, UK)
Prof. Keith L. Clark (Imperial College, London, UK)
Prof. Argyris Varonides (Univ. of Scranton, USA)
Prof. S. Furfari (Direction Generale Energie et Transports, Brussels, EU)
Prof. Constantin Udriste, University Politehnica of Bucharest , ROMANIA
Dr. Michelle Luke (Univ. Berkeley, USA)
Prof. Patrice Brault (Univ. Paris-sud, France)
Prof. Jim Cunningham (Imperial College London, UK)
Prof. Philippe Ben-Abdallah (Ecole Polytechnique de l'Universite de Nantes, France)
Prof. Photios Anninos (Medical School of Thrace, Greece)
Prof. Ichiro Hagiwara, (Tokyo Institute of Technology, Japan)
Prof. Andris Buikis (Latvian Academy of Science. Latvia)
Prof. Akshai Aggarwal (University of Windsor, Canada)
Prof. George Vachtsevanos (Georgia Institute of Technology, USA)
Prof. Ulrich Albrecht (Auburn University, USA)
Prof. Imre J. Rudas (Obuda University, Hungary)
Prof. Alexey L Sadovski (IEEE Fellow, Texas A&M University, USA)
Prof. Amedeo Andreotti (University of Naples, Italy)
Prof. Ryszard S. Choras (University of Technology and Life Sciences Bydgoszcz, Poland)
Prof. Remi Leandre (Universite de Bourgogne, Dijon, France)
Prof. Moustapha Diaby (University of Connecticut, USA)
Prof. Brian McCartin (New York University, USA)
Prof. Elias C. Aifantis (Aristotle Univ. of Thessaloniki, Greece)
Prof. Anastasios Lyrintzis (Purdue University, USA)
Prof. Charles Long (Prof. Emeritus University of Wisconsin, USA)
Prof. Marvin Goldstein (NASA Glenn Research Center, USA)
Prof. Costin Cepisca (University POLITEHNICA of Bucharest, Romania)
Prof. Kleanthis Psarris (University of Texas at San Antonio, USA)
Prof. Ron Goldman (Rice University, USA)
Prof. Ioannis A. Kakadiaris (University of Houston, USA)
Prof. Richard Tapia (Rice University, USA)
Prof. Milivoje M. Kostic (Northern Illinois University, USA)
Prof. Helmut Jaberg (University of Technology Graz, Austria)
Prof. Ardeshir Anjomani (The University of Texas at Arlington, USA)
Prof. Heinz Ulbrich (Technical University Munich, Germany)
Prof. Reinhard Leithner (Technical University Braunschweig, Germany)
Prof. Elbrous M. Jafarov (Istanbul Technical University, Turkey)
Prof. M. Ehsani (Texas A&M University, USA)
Prof. Sesh Commuri (University of Oklahoma, USA)
Prof. Nicolas Galanis (Universite de Sherbrooke, Canada)
Prof. S. H. Sohrab (Northwestern University, USA)
Prof. Rui J. P. de Figueiredo (University of California, USA)
Prof. Valeri Mladenov (Technical University of Sofia, Bulgaria)
Prof. Hiroshi Sakaki (Meisei University, Tokyo, Japan)
Prof. Zoran S. Bojkovic (Technical University of Belgrade, Serbia)
Prof. K. D. Klaes, (Head of the EPS Support Science Team in the MET Division at EUMETSAT, France)
Prof. Emira Maljevic (Technical University of Belgrade, Serbia)
Prof. Kazuhiko Tsuda (University of Tsukuba, Tokyo, Japan)
Prof. Milan Stork (University of West Bohemia , Czech Republic)
Prof. C. G. Helmis (University of Athens, Greece)
Prof. Lajos Barna (Budapest University of Technology and Economics, Hungary)

Prof. Nobuoki Mano (Meisei University, Tokyo, Japan)
Prof. Nobuo Nakajima (The University of Electro-Communications, Tokyo, Japan)
Prof. Victor-Emil Neagoe (Polytechnic University of Bucharest, Romania)
Prof. P. Vanderstraeten (Brussels Institute for Environmental Management, Belgium)
Prof. Annaliese Bischoff (University of Massachusetts, Amherst, USA)
Prof. Virgil Tiponut (Politehnica University of Timisoara, Romania)
Prof. Andrei Kolyshkin (Riga Technical University, Latvia)
Prof. Fumiaki Imado (Shinshu University, Japan)
Prof. Sotirios G. Ziavras (New Jersey Institute of Technology, USA)
Prof. Constantin Volosencu (Politehnica University of Timisoara, Romania)
Prof. Marc A. Rosen (University of Ontario Institute of Technology, Canada)
Prof. Thomas M. Gatton (National University, San Diego, USA)
Prof. Leonardo Pagnotta (University of Calabria, Italy)
Prof. Yan Wu (Georgia Southern University, USA)
Prof. Daniel N. Riahi (University of Texas-Pan American, USA)
Prof. Alexander Grebennikov (Autonomous University of Puebla, Mexico)
Prof. Bennie F. L. Ward (Baylor University, TX, USA)
Prof. Guennadi A. Kouzaev (Norwegian University of Science and Technology, Norway)
Prof. Eugene Kindler (University of Ostrava, Czech Republic)
Prof. Geoff Skinner (The University of Newcastle, Australia)
Prof. Hamido Fujita (Iwate Prefectural University(IPU), Japan)
Prof. Francesco Muzi (University of L'Aquila, Italy)
Prof. Claudio Rossi (University of Siena, Italy)
Prof. Sergey B. Leonov (Joint Institute for High Temperature Russian Academy of Science, Russia)
Prof. Arpad A. Fay (University of Miskolc, Hungary)
Prof. Lili He (San Jose State University, USA)
Prof. M. Nasseh Tabrizi (East Carolina University, USA)
Prof. Alaa Eldin Fahmy (University Of Calgary, Canada)
Prof. Gh. Pascovici (University of Koeln, Germany)
Prof. Pier Paolo Delsanto (Politecnico of Torino, Italy)
Prof. Radu Munteanu (Rector of the Technical University of Cluj-Napoca, Romania)
Prof. Ioan Dumitrache (Politehnica University of Bucharest, Romania)
Prof. Corneliu Lazar (Technical University Gh.Asachi Iasi, Romania)
Prof. Nicola Pitrone (Universita degli Studi Catania, Italia)
Prof. Miquel Salgot (University of Barcelona, Spain)
Prof. Amaury A. Caballero (Florida International University, USA)
Prof. Petar Popivanov (Bulgarian Academy of Sciences, Bulgaria)
Prof. Alexander Gegov (University of Portsmouth, UK)
Prof. Lin Feng (Nanyang Technological University, Singapore)
Prof. Colin Fyfe (University of the West of Scotland, UK)
Prof. Zhaohui Luo (Univ of London, UK)
Prof. Wolfgang Wenzel (Institute for Nanotechnology, Germany)
Prof. Weilian Su (Naval Postgraduate School, USA)
Prof. Phillip G. Bradford (The University of Alabama, USA)
Prof. Ray Hefferlin (Southern Adventist University, TN, USA)
Prof. Gabriella Bognar (University of Miskolc, Hungary)
Prof. Hamid Abachi (Monash University, Australia)
Prof. Karlheinz Spindler (Fachhochschule Wiesbaden, Germany)
Prof. Josef Boercsoek (Universitat Kassel, Germany)
Prof. Eyad H. Abed (University of Maryland, Maryland, USA)
Prof. F. Castanie (TeSA, Toulouse, France)
Prof. Robert K. L. Gay (Nanyang Technological University, Singapore)
Prof. Andrzej Ordys (Kingston University, UK)
Prof. Harris Catrakis (Univ of California Irvine, USA)
Prof. T Bott (The University of Birmingham, UK)

Prof. T.-W. Lee (Arizona State University, AZ, USA)
Prof. Le Yi Wang (Wayne State University, Detroit, USA)
Prof. Oleksander Markovskyy (National Technical University of Ukraine, Ukraine)
Prof. Suresh P. Sethi (University of Texas at Dallas, USA)
Prof. Hartmut Hillmer(University of Kassel, Germany)
Prof. Bram Van Putten (Wageningen University, The Netherlands)
Prof. Alexander Iomin (Technion - Israel Institute of Technology, Israel)
Prof. Roberto San Jose (Technical University of Madrid, Spain)
Prof. Minvydas Ragulskis (Kaunas University of Technology, Lithuania)
Prof. Arun Kulkarni (The University of Texas at Tyler, USA)
Prof. Joydeep Mitra (New Mexico State University, USA)
Prof. Vincenzo Niola (University of Naples Federico II, Italy)
Prof. Ion Chryssoverghi (National Technical University of Athens, Greece)
Prof. Dr. Aydin Akan (Istanbul University, Turkey)
Prof. Sarka Necasova (Academy of Sciences, Prague, Czech Republic)
Prof. C. D. Memos (National Technical University of Athens, Greece)
Prof. S. Y. Chen, (Zhejiang University of Technology, China and University of Hamburg, Germany)
Prof. Duc Nguyen (Old Dominion University, Norfolk, USA)
Prof. Tuan Pham (James Cook University, Townsville, Australia)
Prof. Jiri Klima (Technical Faculty of CZU in Prague, Czech Republic)
Prof. Rossella Cancelliere (University of Torino, Italy)
Prof. Dr-Eng. Christian Bouquegneau (Faculty Polytechnique de Mons, Belgium)
Prof. Wladyslaw Mielczarski (Technical University of Lodz, Poland)
Prof. Ibrahim Hassan (Concordia University, Montreal, Quebec, Canada)
Prof. Stavros J.Baloyannis (Medical School, Aristotle University of Thessaloniki, Greece)
Prof. James F. Frenzel (University of Idaho, USA)
Prof. Vilem Srovnal,(Technical University of Ostrava, Czech Republic)
Prof. J. M. Giron-Sierra (Universidad Complutense de Madrid, Spain)
Prof. Walter Dosch (University of Luebeck, Germany)
Prof. Rudolf Freund (Vienna University of Technology, Austria)
Prof. Erich Schmidt (Vienna University of Technology, Austria)
Prof. Alessandro Genco (University of Palermo, Italy)
Prof. Martin Lopez Morales (Technical University of Monterey, Mexico)
Prof. Ralph W. Oberste-Vorth (Marshall University, USA)
Prof. Vladimir Damgov (Bulgarian Academy of Sciences, Bulgaria)
Prof. P.Borne (Ecole Central de Lille, France)

Additional Reviewers

Francesco Zirilli	Sapienza Universita di Roma, Italy
Sorinel Oprisan	College of Charleston, CA, USA
Xiang Bai	Huazhong University of Science and Technology, China
Philippe Dondon	Institut polytechnique de Bordeaux, France
Yamagishi Hiromitsu	Ehime University, Japan
Frederic Kuznik	National Institute of Applied Sciences, Lyon, France
George Barreto	Pontificia Universidad Javeriana, Colombia
Takuya Yamano	Kanagawa University, Japan
Imre Rudas	Obuda University, Budapest, Hungary
Tetsuya Shimamura	Saitama University, Japan
M. Javed Khan	Tuskegee University, AL, USA
Eleazar Jimenez Serrano	Kyushu University, Japan
Valeri Mladenov	Technical University of Sofia, Bulgaria
Jon Burley	Michigan State University, MI, USA
Andrey Dmitriev	Russian Academy of Sciences, Russia
Moran Wang	Tsinghua University, China
Jose Flores	The University of South Dakota, SD, USA
Hessam Ghasemnejad	Kingston University London, UK
Santoso Wibowo	CQ University, Australia
Kazuhiko Natori	Toho University, Japan
Konstantin Volkov	Kingston University London, UK
Kei Eguchi	Fukuoka Institute of Technology, Japan
Abelha Antonio	Universidade do Minho, Portugal
Tetsuya Yoshida	Hokkaido University, Japan
Matthias Buyle	Artesis Hogeschool Antwerpen, Belgium
Deolinda Rasteiro	Coimbra Institute of Engineering, Portugal
Masaji Tanaka	Okayama University of Science, Japan
Bazil Taha Ahmed	Universidad Autonoma de Madrid, Spain
Zhong-Jie Han	Tianjin University, China
James Vance	The University of Virginia's College at Wise, VA, USA
Angel F. Tenorio	Universidad Pablo de Olavide, Spain
Genqi Xu	Tianjin University, China
João Bastos	Instituto Superior de Engenharia do Porto, Portugal
Miguel Carriegos	Universidad de Leon, Spain
Shinji Osada	Gifu University School of Medicine, Japan
Ole Christian Boe	Norwegian Military Academy, Norway
Lesley Farmer	California State University Long Beach, CA, USA
Dmitrijs Serdjuks	Riga Technical University, Latvia
Alejandro Fuentes-Penna	Universidad Autónoma del Estado de Hidalgo, Mexico
Francesco Rotondo	Polytechnic of Bari University, Italy
Stavros Ponis	National Technical University of Athens, Greece
José Carlos Metrôlho	Instituto Politecnico de Castelo Branco, Portugal
Minhui Yan	Shanghai Maritime University, China

Table of Contents

ISBN: 978-1-61804-290-3

ISBN: 978-1-61804-290-3

ISBN: 978-1-61804-290-3

Plenary Lecture 1

Extended Unbiased FIR Filtering for Indoor Robot Self-Localization

Professor Yuriy S. Shmaliy
Department of Electronics Engineering
DICIS, Universidad de Guanajuato,
Salamanca, 36885, Mexico
E-mail: shmaliy@ugto.mx

Abstract: Mobil robot self-localization in diverse environments is a key problem for many industrial applications. We consider a novel estimation technique called extended unbiased finite impulse response (EFIR) filtering which has several advantages against the traditional extended Kalman filter (EKF): better robustness against uncertainties, lower sensitivity to noise, and smaller round-off errors. A fast iterative EFIR localization algorithm utilizing recursions is discussed as a rival to the EKF. Unlike the EKF, the EFIR filter completely ignores the noise statistics. Instead, it requires an optimal horizon of N_{opt} points in order for the localization performance to be acceptably suboptimal. It is shown that N_{opt} can be specialized via measurements with much smaller efforts and cost than for the noise statistics required by the EKF. Overall, EFIR filtering is more successful in accuracy than the EKF under the uncertain conditions. Extensive investigations of the approach are conducted in applications to indoor mobile robot self-localization via triangulation and in radio frequency identification (RFID) tag grid environments. Better performance of the EFIR filter is demonstrated when the noise statistics are not known exactly. As a special inference, it is shows that the EKF diverges not only due to large nonlinearities and large noise as was previously known from the Kalman filter theory, but also due to errors in the imprecisely defined noise statistics. In contrast, the EFIR filter does not demonstrate divergence in this case.

Brief Biography of the Speaker:

Dr. Yuriy S. Shmaliy has been a full professor in Electrical Engineering of the Universidad de Guanajuato, Mexico, since 1999. He received the B.S., M.S., and Ph.D. degrees in 1974, 1976 and 1982, respectively, from the Kharkiv Aviation Institute, Ukraine. In 1992 he received the Dr.Sc. (technical) degree from the Soviet Union Government. In March 1985, he joined the Kharkiv Military University. He serves as full professor beginning in 1986 and has a Certificate of Professor from the Ukrainian Government in 1993. In 1993, he founded and, by 2001, had been a director of the Scientific Center “Sichron” (Kharkiv, Ukraine) working in the field of precise time and frequency. His books *Continuous-Time Signals* (2006) and *Continuous-Time Systems* (2007) were published by Springer, New York. His book *GPS-based Optimal FIR Filtering of Clock Models* (2009) was published by Nova Science Publ., New York. He also edited a book *Probability: Interpretation, Theory and Applications* (Nova Science Publ., New York, 2012) and contributed to several books with invited chapters. Dr. Shmaliy has authored more than 300 Journal and Conference papers and 80 patents. He is IEEE Fellow; was rewarded a title, Honorary Radio Engineer of the USSR, in 1991; and was listed in Outstanding People of the 20th Century, Cambridge, England in 1999. He currently serves on the Editorial Boards of several International Journals and is a member of the Program Committees of various Int. Symposia. His current interests include statistical signal processing, optimal estimation, and stochastic system theory.

ISBN: 978-1-61804-290-3

ISBN: 978-1-61804-290-3

Parameter uncertainty modeling in nonlinear dynamical system for guaranteed interval parameter estimation

Qiaochu Li[1] and Carine Jauberthie [2] and Lilianne Denis-Vidal[1] and Zohra Cherfi[1]

***Abstract*—This paper deals with state and parameter estimation in a bounded error context. Different from the standard stochastic approach, the measurement errors are considered bounded but otherwise unknown. Using interval analysis, parameter estimation problem can be formulated as set computation, and a branch and bound based algorithm has been proposed in literature. But this technique has a natural inefficiency problem during bisection process, so we propose a new method which takes the bisection operation with one parameter at a time by introducing the parameter variation error function to remodel the measurement error. This parameter error model is based on the sensitivities. It is especially useful to models when their parameters are many and the estimation time is crucial. After a brief introduction of interval analysis, our original algorithm is proposed. At last, an illustrative example has been studied. Properties of this approach are discussed and illustrated based on this example. This method could be potentially used as on-line diagnosis considered its little time cost.**

***Index Terms*—Continuous-time systems, Parameter estimation, Nonlinear systems, Bounded noise, Interval analysis.**

I. Introduction

With mathematical tools, the physical phenomena can be modeled for purpose of process control, optimization, model calibration, etc. Due to the measurement error, unknown properties or the mistakes, the accuracy of the models used in practice suffers from uncertainty. A key problem in developing models is parameter estimation which influence a large domain of application. In dynamical system, an exact estimation of parameter is needed so that a best fit model can be achieved and used for automatic control or diagnosis for example.

Parameter estimation is conducted by minimizing some aim function that may be purely heuristic or deduced from information about noise corrupting the data and possibly about the prior distribution for the parameter vector.

Bounded error parameter estimation does not use this optimization like method but aims instead at characterizing the set of all values of the parameter vector that are consistent with bounds on the errors, selecting the one which is acceptable between the behavior of the model and that of the system to be modeled. The first work have been done by F.C. Schweppe [23] on state estimation for linear models. And then, this type of parameter or state bounding is often referred to as guaranteed estimation. In this proposition, one could envelop the solution set by ellipsoids, while other types of containers than ellipsoids could and have been used, such as boxes, paralleloptopes, zonotopes [15] or other limited complexity polytopes.

Parameter or state bounding when the model is nonlinear is much more complicated, and nonlinear ellipsoidal calculus remains a larges open subject, a recent progress refer to [4]. So the linearization method has been proposed. Instead of using this approach, one may use the tools provided by interval analysis to compute an approximate but guaranteed characterization of sets directly in the nonlinear case. And the parameter estimation problem could be reformulated as finding the feasible or unfeasible sets which are or not consistent with the measurements by using evaluation functions. The feasible or the unfeasible sets will be judged by set inversion via interval analysis (SIVIA) which has been described in detail [8]. But the efficiency problem remains a big headache in such algorithm especially when the number of parameters is big. In this paper, a one at a time (OAT) parameter estimation procedure is proposed based on changing the uncertainty of parameter to the output, so that at every iteration, only one parameter will be reduced at a time. And the improved parameter bounds will be used directly in the next iteration of parameter estimation process until no further improvement is made.

This paper is organized as follows: Section II will give a global picture on the treating problem. Section III explains the basic algorithm of interval analysis. In section IV, our one at a time (OAT) parameter estimation is explained and a preliminary algorithm is proposed. As the intrinsic problem of interval analysis, some application constraints have been discussed about. In section V, a case study is proposed, the OAT method has been compared with results obtained by classic approach. At last, the section VI will give a brief explanation about OAT parameter estimation. Some notices of use has been delivered.

II. NOTATION AND PROBLEM FORMULATION

This paper follows the standard notation of interval analysis [9]. Suppose we want to estimate the unknown state x for a nonlinear dynamic system of the following form:

$$\begin{cases} \dot{\boldsymbol{x}}(t) = \boldsymbol{f}(\boldsymbol{x}(t), \boldsymbol{p}), \\ \boldsymbol{y}(t) = \boldsymbol{h}(\boldsymbol{x}(t)), \quad \boldsymbol{x}(0) \ \in \boldsymbol{X}_0. \end{cases} \tag{1}$$

[1] Qiaochu Li, Lilianne Denis-Vidal and Zohra Cherfi is with the Sorbonne University, Université de Technologie de Compiègne, 60 203 Compiègne cedex qiaochu.li@utc.fr, lilianne.denis-vidal@math.univ-lille1.fr, zohra.cherfi@utc.fr

[2] Carine Jauberthie is with LAAS, CNRS, 7 avenue du Colonel Roche, F-31400 Toulouse, France Université de Toulouse, UPS, LAAS, F-31400 Toulouse, France cjaubert@laas.fr

ISBN: 978-1-61804-290-3

where $\boldsymbol{x}(t) \in \mathbb{IR}^n$ and $\boldsymbol{y}(t) \in \mathbb{IR}^m$ denote respectively the state variables and the measured outputs. The initial conditions $\boldsymbol{x}(0)$ is supposed to belong to an initial bounded "box" $\boldsymbol{X}_0 = [\underline{x_0}, \overline{x_0}]$. The parameter vector p is constant and is assumed to belong to a bounded "box" $\boldsymbol{P}_0 = [\underline{p_0}, \overline{p_0}]$. Time t is assumed to belong to $[0, t_{max}]$. The functions f and h are nonlinear functions. f is real and analytic on M for every $p \in [p_0]$, where M is an open set of $\mathbb{R}^n$ such that $x(t) \in M$ for every $p \in \boldsymbol{p}_0$ and $t \in [0, t_{max}]$). Moreover the function f is assumed to be sufficiently differentiable in the domain M.

In interval context, the purpose of parameter estimation is to find $\boldsymbol{p}$ such that $\boldsymbol{y}_m(\boldsymbol{p})$ fits best in an inclusion test, the subscribe m indicates the model output. To be specified, the parameters are considered consistent if the error $v(t_i)$ is assumed to satisfy:

$$\boldsymbol{y}(t_i) - \boldsymbol{y}_m(t_i, \boldsymbol{p}) \in \boldsymbol{v}(t_i) = [\underline{v}(t_i), \overline{v}(t_i)], \quad i = 1, ..., N. \quad (2)$$

where $\boldsymbol{y}(t_i)$ represents the measured output with certain parameters and $\boldsymbol{y}_m(t_i)$ represents the model output, this inclusion is to be taken component-wisely in every dimension. We assume that $\underline{v}(t_i)$ and $\overline{v}(t_i)$ are known as lower and upper bounds for the acceptable output errors. Such bounds may, for instance, correspond to a bounded measurement noise. The integer N is the total number of sample times.

Solving dynamical systems with given initial states can be considered as solving an ordinary differential equation (ODE) of initial value problem. In interval arithmetic, many tools are available which could be used to generate guaranteed bounds for the solution of (1) at the sampling times $\{t_1, t_2, ..., t_N\}$. With the output of model system $\boldsymbol{y}_m$, the parameter estimation could be done by projecting the intervals variables to the solution space, if pessimism exists, we could partition these variables into smaller intervals by executing a finite time bisection process in every parameter dimension, or one could use the set inversion algorithm with a stop criterion in bisection step to generate rigorous bounds for it. In the following section, the interval tools and a set inversion bounding principal will be introduced.

III. INTERVAL ANALYSIS FOR STATE ESTIMATION

Interval analysis provides tools for computing with sets which are described using outer-approximations formed by union of non-overlapping boxes. The following results are mainly taken from [7].

A. Basic tools

A real interval $\boldsymbol{u} = [\underline{u}, \overline{u}]$ is a closed and connected subset of $\mathbb{IR}$ where $\underline{u}$ represents the lower bound of $\boldsymbol{u}$ and $\overline{u}$ represents the upper bound. The width of an interval $\boldsymbol{u}$ is defined by $w(u) = \overline{u} - \underline{u}$, and its midpoint by $m(u) = (\overline{u} + \underline{u})/2$.

The set of all real intervals of R is denoted $\mathbb{IR}$.
Two intervals $\boldsymbol{u}$ and $\boldsymbol{v}$ are equal if and only if $\underline{u} = \underline{v}$ and $\overline{u} = \overline{v}$. Real arithmetic operations are extended to intervals [16].

Arithmetic operations on two intervals $\boldsymbol{u}$ and $\boldsymbol{v}$ can be defined by:

$$\circ \in \{+, -, *, /\}, \quad \boldsymbol{u} \circ \boldsymbol{v} = \{x \circ y \mid x \in [u],\ y \in [v]\}.$$

An interval vector (or box) $\boldsymbol{x}$ is a vector with interval components and may equivently be seen as a cartesian product of scalar intervals:

$$\boldsymbol{x} = \boldsymbol{x}_1 \times \boldsymbol{x}_2 ... \times \boldsymbol{x}_n.$$

The set of $n-$dimensional real interval vectors is denoted by $\mathbb{IR}^n$.

An interval matrix is a matrix with interval components. The set of $n \times m$ real interval matrices is denoted by $\mathbb{IR}^{n \times m}$. The width $w(.)$ of an interval vector (or of an interval matrix) is the maximum of the widths of its interval components. The midpoint $m(.)$ of an interval vector (resp. an interval matrix) is a vector (resp. a matrix) composed of the midpoint of its interval components.

Classical operations for interval vectors (resp. interval matrices) are direct extensions of the same operations for punctual vectors (resp. punctual matrices) [16].

Let $f : \mathbb{IR}^n \to \mathbb{IR}^m$, the range of the function f over an interval vector $\boldsymbol{u}$ is given by:

$$f(\boldsymbol{u}) = \{f(x) | x \in \boldsymbol{u}\}.$$

The interval function $\boldsymbol{f}$ from $\mathbb{IR}^n$ to $\mathbb{IR}^m$ is an inclusion function for f if:

$$\forall \boldsymbol{u} \in \mathbb{IR}^n, \quad f(\boldsymbol{u}) \subseteq \boldsymbol{f}(\boldsymbol{u}).$$

An inclusion function of f can be obtained by replacing each occurrence of a real variable by its corresponding interval and by replacing each standard function by its interval evaluation. Such a function is called the natural inclusion function. In practice the inclusion function is not unique, it depends on the syntax of f. A commonly used inclusion function is the mean value form, which can envelope all the possible solution of x by computing its Jacobian matrix and its middles value. If $\boldsymbol{f}$ is differentiable over $\boldsymbol{x}$, we have:

$$\forall x \in \boldsymbol{x}, \boldsymbol{f}(\boldsymbol{x}) \in \boldsymbol{f}(m) + \boldsymbol{J}(\boldsymbol{x})(\boldsymbol{x} - m) \quad (3)$$

where $\boldsymbol{J}$ is the Jacobian matrix of $\boldsymbol{f}$ and m the middle value of $\boldsymbol{x}$.

B. State Estimation by using Taylor expansions

This section concerns the integration of (1). Thus, the objective of this section is to estimate the state vector x at the sampling times $\{t_1, t_2, ..., t_N\}$ corresponding to the measurement times of the outputs. We note $\boldsymbol{x}_j$ the box $\boldsymbol{x}(t_j)$ where t_j represents the sampling time, $j = 1, ..., N$ and x_j represents the solution of (1) at t_j. State estimation for dynamical nonlinear systems can be solved efficiently by considering methods based on Taylor expansions [3], [16], [20] or [22]. These methods consist in two parts: the first one verifies the existence and uniqueness of the solution by using the fixed point theorem and the Picard-Lindelöf operator. At a time t_{j+1}, an a priori box $\tilde{\boldsymbol{x}}_j$ containing all solutions corresponding to all possible trajectories between t_j and t_{j+1} is computed. In the second part, the solution at t_{j+1} is computed by using a Taylor expansion, where the remainder term is $\tilde{\boldsymbol{x}}_j$.

ISBN: 978-1-61804-290-3

However, in practice, the set $\tilde{\boldsymbol{x}}_j$ often fails to contain the true solution [19]. Thus, the classical technique used consists in inflating this set until it verifies the following inclusion [13]:

$$\boldsymbol{x}_j + h_j f(\tilde{\boldsymbol{x}}_j) \subseteq \tilde{\boldsymbol{x}}_j, \tag{4}$$

where h_j denotes the integration step and $\boldsymbol{x}_j$ the first solution.

This method is performed in the *Enclosure algorithm* and developed in [17].

To generate these guaranteed bounds, there are many available free solvers ready to use, like VSPODE [12] or VNODE-LP [18], etc. The package of VNODE-LP implements particularly algorithm corresponding to high order enclosure and Hermite-Obreschkoff method to conquer the rapping effect, and VSPODE use another approach to fight for the pessimism.

C. Set inversion via interval analysis

Consider the problem of determining a solution set for the unknown quantities u defined by:

$$\mathrm{S} = \{u \in \mathrm{U} \mid \Phi(u) \in \boldsymbol{y}\} = \Phi^{-1}(\boldsymbol{y}) \cap \mathrm{U}, \tag{5}$$

where $[y]$ is known a priori, U is an a priori search set for u and Φ a nonlinear function not necessarily invertible in the classical sense. (5) involves computing the reciprocal image of Φ and is known as a set inversion problem which can be solved using the algorithm Set Inversion Via Interval Analysis (denoted SIVIA). The algorithm SIVIA proposed in [8] is a recursive algorithm which explores all the search space without losing any solution. This algorithm makes it possible to derive a guaranteed enclosure of the solution set S as follows:

$$\underline{\mathrm{S}} \subseteq \mathrm{S} \subseteq \overline{\mathrm{S}}.$$

The inner enclosure $\underline{\mathrm{S}}$ is composed of the boxes that have been proved feasible. To prove that a box $\boldsymbol{u}$ is feasible it is sufficient to prove that $\Phi(\boldsymbol{u}) \subseteq \boldsymbol{y}$. Reversely, if it can be proved that $\Phi(\boldsymbol{u}) \cap \boldsymbol{y} = \emptyset$, then the box $\boldsymbol{u}$ is unfeasible. Otherwise, no conclusion can be reached and the box $\boldsymbol{u}$ is said undetermined. The latter is then bisected and tested again until its size reaches a user-specified precision threshold $\varepsilon > 0$. Such a termination criterion ensures that SIVIA terminates after a finite number of iterations.

IV. Guaranteed parameter estimation

In the literature, the set inversion computation is wildly used by many practitioners. With interval tools in hands, one could handle problems such as parameter estimation, parameter validation, and system calibration things at one time with the help of so called guaranteed properties from interval analysis by using SIVIA. But the big disadvantage of this technique is its efficiency, which is referring to huge time consumption and memory needs in nowadays computer. In this section, we propose a new point of view in interval parameter estimation, considering the parameter variation as an output error added to the measurements during estimation.

To estimate the parameters with a mathematical model in interval analysis, the set-membership computation requires all elements in the parameter vector $\boldsymbol{p} \in \mathbb{P}$ must satisfy:

$$\boldsymbol{y}_m(t, \boldsymbol{p}) \in \boldsymbol{y}(t) = [\underline{y}(t), \overline{y}(t)], \tag{6}$$

where $\boldsymbol{y}$ is the measurement value with $\boldsymbol{p}$ to be estimated. Here, $\boldsymbol{y}$ and $\boldsymbol{y}_m$ is an interval vector, the inclusion is valid componentwisely between measurement and model output.

Based on these, with an explicit expression of $\boldsymbol{y}_m$, short notation of $\boldsymbol{y}_m(t, \boldsymbol{p})$. One could use an interval constraint propagation technique to bound the $\boldsymbol{p}$ [6]. However, if the function $\boldsymbol{y}_m$ has no explicit expression, a mean value form for $\boldsymbol{y}_m(\boldsymbol{p})$ could be taken to compute its bounds [10]. Suppose $\mathbb{S}$ is the set of admissible parameters and $\boldsymbol{y}_m$ is differentiable over p. For kth component $\boldsymbol{y}_k^m$ of $\boldsymbol{y}_m(\boldsymbol{p})$, for all $\boldsymbol{p} \in \mathbb{S} \subset \boldsymbol{P}$ and $m \in \boldsymbol{p}$, m is the middle point of $\boldsymbol{p}$, we have :

$$\boldsymbol{y}_k^m \in \boldsymbol{y}_k^m(m) + \sum_{j=1}^{n_p} (\boldsymbol{p}_j - m_j)[\frac{\partial \boldsymbol{y}_k^m}{\partial \boldsymbol{p}_j}](\boldsymbol{p}) \in \boldsymbol{y}_k. \tag{7}$$

From the point of view by interval analysis with mean value form, the rayon of uncertainty is from the parameters and their joint sensitivities in Equation 7. Note that this uncertainty is symmetric of 0. Suppose one of the parameter is known, which is a constant, no matter how its sensitivity it is, the bring in uncertainty from this parameter is 0, as the radius of this parameter is 0. So, one possible way to model the uncertainty from the parameter, is using the element of sum in our context.

Let take:

$$w(t, \boldsymbol{p}) = \sum_{j=1}^{n_p} (\boldsymbol{p}_j - m_j)[\frac{\partial \boldsymbol{y}_k^m}{\partial \boldsymbol{p}_j}](t, \boldsymbol{p}), \tag{8}$$

as the uncertainty brought by parameters to $\boldsymbol{y}_k$. So, for estimating the kth parameter, the $\boldsymbol{p}^v$ must satisfy:

$$\boldsymbol{y}_m(t, \boldsymbol{p}^v) \in \boldsymbol{y}(t) + w(t, \boldsymbol{p}_k^v), \tag{9}$$

where $\boldsymbol{p}_k^v := [\boldsymbol{p}_1, \ldots, m_k, \ldots, \boldsymbol{p}_{n_p}]$ and $\boldsymbol{p}^v := [\boldsymbol{p}_1, \ldots, \boldsymbol{p}_{n_p}]$. This means the model output with parameter $\boldsymbol{p}_k$ can be bounded by the new measurement adding the parameter variation errors on the measurement errors.

As illustrated in Figure 1, the bounding of the output of system can be dissembled by its interval uncertainty from sensitivities, its overestimation is $O(w(\boldsymbol{p})^2)$ according to mean value theory.

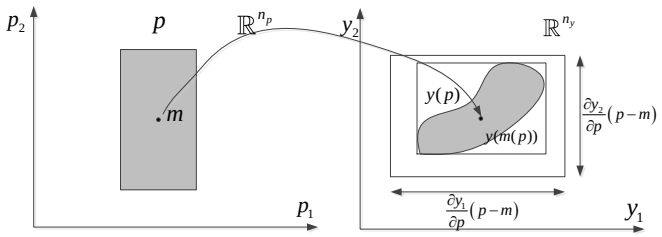


Fig. 1: Mean value inclusion function's interval uncertainty source

Based on this approach, we propose the following algorithm to solve the problem between the efficiency and time consuming conflict during estimation of parameter. Added in a parameter estimation algorithm, this kind of bounding the output change according to parameters' variation could allow us estimating one parameter at a time. The full algorithm is summarized as following:

Algorithm 1 Parameter estimation ($\boldsymbol{y}$, $\mathcal{P}_{feasible}$, ε): single run

Require: $\boldsymbol{x}(0)$, $\boldsymbol{p}(0)$;
Ensure: $\mathcal{P}_{admis}$, $\mathcal{P}_{uncertain}$, $\mathcal{P}_{rejected}$;
initialization: $P_{list} := \boldsymbol{p}(0)$, $x_e(0) := (\boldsymbol{x}(0), \boldsymbol{p}(0))$;
for k := 1 : n_p **do**
  calculate the sensitivities with $\boldsymbol{p}_k^v := [\boldsymbol{p}_1, \ldots, m(\boldsymbol{p}_k), \ldots, \boldsymbol{p}_{n_p}]$, $\boldsymbol{p} \in \mathcal{P}_{list}$
  generation uncertainty added measurements: $\boldsymbol{y} := \boldsymbol{y} + w(\boldsymbol{p}_{i \neq k})$
  while $P_{list}(k) :\neq \emptyset$ **do**
    $\boldsymbol{p}^v := [\boldsymbol{p}_1, \ldots, Pop(\mathcal{P}_{list}(k)), \ldots, \boldsymbol{p}_{n_p}]$;
    $i := 1$;
    while $i <= N$ **do**
      $x_e(i) := ODE(x_e(i-1))$;
      $j := i$;
      $i := i + 1$;
    end while
    if $\boldsymbol{h}(\boldsymbol{x}_e(1:j)) \subseteq \boldsymbol{y}(1:j)$ **then**
      $\mathcal{P}_{admis}(k) := \mathcal{P}_{admis}(k) \cup \boldsymbol{p}_k$;
    else if $\boldsymbol{h}(\boldsymbol{x}_e(1:j)) \cap \boldsymbol{y}(1:j) := \emptyset$ **then**
      $\mathcal{P}_{rejected}(k) := \mathcal{P}_{rejected}(k) \cup \boldsymbol{p}_k$;
    else if $w(\boldsymbol{p}_k) < \varepsilon$ **then**
      $\mathcal{P}_{uncertain}(k) := \mathcal{P}_{uncertain}(k) \cup \boldsymbol{p}_k$;
    else
      $bisectBox(\boldsymbol{p_k}) \rightarrow \{\boldsymbol{p}_{k_1}, \boldsymbol{p}_{k_2} \mid \boldsymbol{p}_{k_1} \cup \boldsymbol{p}_{k_2} = \boldsymbol{p}_k\}$;
      $\mathcal{P}_{list}(k) := \mathcal{P}_{list}(k) \cup \boldsymbol{p}_{k_1}$, $\mathcal{P}_{list}(k) := \mathcal{P}_{list}(k) \cup \boldsymbol{p}_{k_2}$;
    end if
  end while
end for

The notation ODE indicates the usage of a validated integrator from interval analysis, which could be VNODE-LP or VSPODE for example. $\mathcal{P}_{list}(k)$ indicates the kth component in the stack list, same notation for $\mathcal{P}_{admis}$, $\mathcal{P}_{rejected}$ and $\mathcal{P}_{uncertain}$. The feasible sets are redefined as the union of admissible sets and uncertain sets in this algorithm.

This single run algorithm can be iterated many times until there is no further improvement in the bounded parameters. At each execution, one should use the estimated results from previews computation to the next refinement. The key step is the calculation of sensitivity for each parameter, this can be done by using the extended state space model, which combines the sensitivity function with its state form. To be clear, suppose the observation of state is full $\boldsymbol{y} = \boldsymbol{x}$.

We note $\frac{\partial \boldsymbol{x}(t,\boldsymbol{p})}{\partial \boldsymbol{p}} = s_{jk}(t, \boldsymbol{p})$, with $j = 1, \ldots, dim\boldsymbol{x}$ and $k = 1, \ldots, dim\boldsymbol{p}$. To calculate the sensitivity, we take the derivative of 1 with respect to p_k:

$$s'_{jk} = \frac{\partial f_j(\boldsymbol{x}, \boldsymbol{p})}{\partial x_j} s_{jk} + \frac{\partial f_j(\boldsymbol{x}, \boldsymbol{p})}{\partial p_k}. \tag{10}$$

Since the $\boldsymbol{x}_0$ is supposed to be known, the initial values of sensitivities are :

$$s_{jk}(t_0) = \frac{\partial \boldsymbol{x}(t_0)}{\partial p_k} = 0.$$

Combined with Equation 1, both states and sensitivities can be obtained with this extended form by an ODE.

Note that by using the mean value form for inclusion function as the error additive relation, one has to presume the solution sets of parameters are convex.

The following section is dedicated to the application example taken from aerospace domain. The proposed method will be considered as method 1 and the former estimation method [11] is called method 2. Both of these methods will receive the guaranteed results by uncertain sets in a finite time.

V. APPLICATION

The case study that we consider is the longitudinal motion of a glider, one could easily choose other dynamical systems to study. The projection of the general equations of motion onto the aerodynamic reference frame of the aircraft and the linearization of aerodynamic coefficients give the following dynamic system:

$$\begin{cases} \dot{V} = & -g\sin(\theta-\alpha) - \frac{1}{2m}\rho S V^2 (C_x^0 \\ & + C_{x\alpha}(\alpha-\alpha_0) + C_{x\delta_m}(\delta_m - \delta_{m_0})), \\ \dot{\alpha} = & \frac{2}{2mV + \rho S l V C_{z\dot{\alpha}}} \Big\{ mVq + mg\frac{\cos(\theta-\alpha)}{V} \\ & - \frac{1}{2}\rho S V^2 (C_z^0 + C_{z\alpha}(\alpha-\alpha_0) \\ & + C_{zq}\frac{ql}{V} + C_{z\delta_m}(\delta_m - \delta_{m_0})) \Big\}, \\ \dot{q} = & \frac{1}{2B}\rho S l V^2 \Big\{ C_m^0 + C_{m\alpha}(\alpha-\alpha_0) + C_{mq}\frac{ql}{V} \\ & + C_{m\dot{\alpha}} \frac{2l}{2mV^2 + \rho S l V^2 C_{z\dot{\alpha}}} \Big[mVq \\ & + mg\frac{\cos(\theta-\alpha)}{V} \\ & - \frac{1}{2}\rho S V^2 (C_z^0 + C_{z\alpha}(\alpha-\alpha_0) + C_{zq}\frac{ql}{V} \\ & + C_{z\delta_m}(\delta_m - \delta_{m_0})) \Big] + C_{m\delta_m}(\delta_m - \delta_{m_0}) \Big\}, \\ \dot{\theta} = & q. \end{cases} \tag{11}$$

In these equations, the state vector x is given by $(V, \alpha, q, \theta)^\top$, the observation y is full (i.e., $y = x$), the input u is δ_m (δ_{m_0} represents the initial condition). The variable V denotes the speed of the aircraft, α the angle of attack, α_0 the trim value of α, θ the pitch angle, q the pitch rate, δ_m the elevator deflection angle, ρ the air density, g the acceleration due to gravity, l a reference length and S the area of a reference surface. B represents a moment of inertia. The parameters are $p = (C_{z\dot{\alpha}}, C_{zq}, C_{m\dot{\alpha}}, C_{mq})$, which are assumed to be uncertain. The other coefficients correspond to the dynamic stability derivatives are supposed to be known.

The state estimation of the aircraft system is performed by using VSPODE with the parameter estimation algorithm above, which we call it here method 1 and the algorithm in [11] which we will call it method 2. The major difference between these two methods is the first one use the bisection operation one at a time, concentrated at each parameter in one loop, whereas the second one bisect the largest component in an interval vector. In this one at a time algorithm, the sensitivity can be obtained by using the extended state-space model, which combined the sensitivity functions with state functions [10].

The initial conditions $\boldsymbol{x}_0$ are supposed to belong to:

$$\boldsymbol{X}_0 = \begin{bmatrix} 28.48 & 28.52 \\ 6.363 & 6.393 \\ 0.1092 & 0.2392 \\ 2.4064 & 2.4074 \end{bmatrix}. \quad (12)$$

The parameters to be estimated are $p = (C_{z\dot{\alpha}}, C_{zq}, C_{m\dot{\alpha}}, C_{mq})$, which are not well known but are supposed to be included in:

$$\boldsymbol{P}_0 = \begin{bmatrix} 1.62 & 1.98 \\ 4.5 & 5.5 \\ -5.5 & -4.5 \\ -24.2 & -19.8 \end{bmatrix}. \quad (13)$$

The output error (2) is supposed to be given by:

$$\boldsymbol{v} = \begin{bmatrix} -0.0447 & 0.0447 \\ -0.0044 & 0.0044 \\ -0.0044 & 0.0044 \\ -0.0044 & 0.0044 \end{bmatrix}. \quad (14)$$

The measurements have been generated by using the parameters equal to $(1.8, 5, -5, -22)$ and initial states $\boldsymbol{x}(0)$ and also the input. The test duration is fixed at one second. The stop criterion for SIVIA is $\epsilon = [0.001, 0.005, 0.005, 0.01]$ for each parameter corresponded.

The input of model is chosen as following:

$$u(t) = \delta_{m0} - a_5 H(t - t_{0_5}) + 2a_5 H(t - t_{2_5}) - 2a_5 H(t - t_{3_5}). \quad (15)$$

with $a_5 = 1.6$ degrees, $i = 0, \ldots, 5$ and $t_{0_5} = 0\,s$, $t_{1_5} = 0.2\,s$, $t_{2_5} = 0.4\,s$, $t_{3_5} = 0.6\,s$, $t_{4_5} = 0.8\,s$. The function H represents the Heaviside function.

After computation, the feasible sets represented by Figure 2 and Figure 3 have been obtained.

The Figure is presented by a polygon with four edges. Each side corresponds to its initial interval. The red line is the admissible range for parameters. Notice that the four sides do not have the same scale, so their original bounds are settled to fit each sides of the square. Only the feasible values are indicated in the figures.

The time consumption are summarized in table I, the product of width for each parameter has been also delivered.

TABLE I: Comparison between two methods

Method	Time (s)	Width product
1	589	2.645^{-4}
2	152458	1.516^{-6}

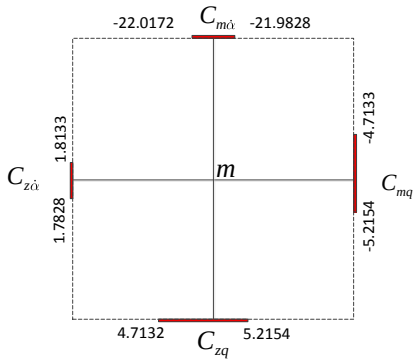


Fig. 2: Admissible range of each parameter: method 1

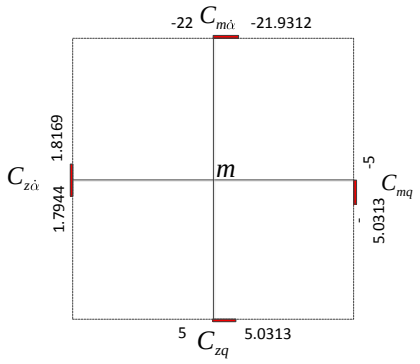


Fig. 3: Admissible range of each parameter: method 2

Clearly, the method 2 obtained the finest results by consuming much more time than our method. The method 1 has obtained a slightly larger results by saving lots of time. This is understandable as the OAT parameter estimation introduced a third information on sensitivities, a large amount of unfeasible sets could be excluded easily. With updating the obtained feasible sets, one could compute another time for more precise parameter bounds. The computation of sensitivity is a key procedure, so a good estimation is urgent. One could use VSPODE to generate these or by partitioning the original sets, VNODE-LP could also be considerable.

The previous simulation has not be made in a real-time context. But the computational requirement is now compatible with the real-time constraints for the case study. Same algorithms were successful employed for example in robotics [25] or more recently in [14].

VI. CONCLUSION

In this contribution, a new procedure for parameter estimation in a bounded-error context is pointed out. This original method allows us to estimate parameter one at a time with considering other parameter's uncertainty as an output error. As in interval analysis, the pessimism always exists, the quality of the so called parameter uncertainty error on the output is essential to use this technique. During the calculation process,

ISBN: 978-1-61804-290-3

with the progression in estimated parameters, smaller interval range could be obtained gradually. The time efficiency is improved obviously compared to our former approach without lose much solution sets in term of precision. By the nature of mean value inclusion function, one has to face the overestimation of derivative of f, so when the solution sets are non convex, this methods could have difficult in bounding them. VSPODE is more powerful in bounding the interval sets, for the comparison between VNODE-LP and VSPODE, please refer to [12].

This parameter estimation method has potential for being used for fault detection and diagnosis problems in continuous-time systems or hybrid systems. Fault detection mechanisms using bounded uncertainty models present the advantage to guarantee absence of false alarms [2]. They have been investigated in the last few years and some comparative analysis works exist [2], [24]. The drawback of these methods is the missing alarms problem, which is due to overestimated results. However, recent methods such as the one used in this paper or [1], [21] should provide significant improvement in this direction.

Another direction of future work consists in extending state estimation in a bounded-error context to state and parameter estimation for complete aircraft model. This opens new perspectives for fault diagnosis, for instance using faults models [5].

REFERENCES

[1] O. Adrot and S. Ploix. Fault detection based on set-membership inversion. In Zhang Zhang, editor, *Supervision and Safety of Technical Processes*, volume 6 of *1*, pages 575 – 580, Beijing, P.R. China, 2006. IFAC.

[2] J. Armengol, L. Travé-Massuyès, Vehi J., and J.Ll. de la Rosa. A survey on interval model simulators and their properties related to fault detection. In *Annual Reviews in Control*, volume 24, Oxford, UK, 2000. Elsevier Science.

[3] M. Berz and K. Makino. Verified integration of odes and flows using differential algebraic methods on high-order taylor models. *Reliable Computing*, 4(4):361–369, 1998.

[4] B Houska, ME Villanueva, and B Chachuat. A validated integration algorithm for nonlinear odes using taylor models and ellipsoidal calculus. *Proceedings of the IEEE Conference on Decision and Control*, pages 484–489, 2013.

[5] C. Jauberthie, N. Verdière, and L. Travé-Massuyès. Fault detection and identification relying on set-membership identifiability. *Annual Reviews in Control*, 37:129–136, 2013.

[6] L. Jaulin. Interval constraint propagation with application to bounded-error estimation. *Automatica*, 36(10):1547–1552, 2000.

[7] L. Jaulin. *Applied interval analysis: with examples in parameter and state estimation, robust control and robotics*, volume 1. Springer, 2001.

[8] L. Jaulin and E. Walter. Set inversion via interval analysis for nonlinear bounded-error estimation. *Automatica*, 29:1053 – 1064, 1993.

[9] R.B. Kearfott, M.T. Nakao, A. Neumaier, S.M. Rump, S.P. Shary, and P. Van Hentenryck. Standardized notation in interval analysis. 2005.

[10] M. Kieffer and E. Walter. Guaranteed estimation of the parameters of nonlinear continuous-time models: Contributions of interval analysis. *International Journal of Adaptive Control and Signal Processing*, 25(3):191–207, 2011.

[11] Q. Li, C. Jauberthie, L. Denis-vidal, and Z. Cherfi. Guaranteed state and parameter estimation for nonlinear dynamical aerospace models. In *Informatics in Control, Automation and Robotics (ICINCO), 2014 11th International Conference on*, volume 01, pages 519–527, Sept 2014.

[12] Y. Lin and M.A. Stadtherr. Guaranteed state and parameter estimation for nonlinear continuous-time systems with bounded-error measurements. *Industrial and Engineering Chemistry Research*, pages 7198–7207, 2007.

[13] R.J. Lohner. Enclosing the solutions of ordinary initial and boundary value problems. pages 9 – 28. Wiley Teubner series in computer science, 1987.

[14] J.P. Merlet. Interval analysis and reliability in robotics. *International Journal of Reliability and Safety*, 3(1):104–130, 2009.

[15] S.H. Mo and J.P. Norton. Fast and robust algorithm to compute exact polytope parameter bounds. *Mathematics and Computers in Simulation*, 32(5-6):481 – 493, 1990.

[16] R.E. Moore. *Interval analysis*. Prentice Hall, Englewood cliffs, NJ, 1966.

[17] N.S. Nedialkov. Some recent advances in validated methods for ivps for odes. *Applied numerical Mathematics*, 42:269 – 284, 2002.

[18] N.S. Nedialkov. Implementing a rigorous ode solver through literate programming. Technical report, Dept. of Computing and Software, 2010.

[19] N.S. Nedialkov, K.R. Jackson, and G.F. Corliss. Validated solutions of initial value problems for ordinary differential equations. *Applied Mathematical Computing*, 105:21 – 68, 1999.

[20] N.S. Nedialkov, R. Kenneth, and A. Jackson. An effective high-order interval method for validating existence and uniqueness of the solution of an ivp for an ode. *Computing*, 17, 2001.

[21] H. Niemann. Active fault diagnosis in closed-loop uncertain systems. In Zhang Zhang, editor, *Supervision and Safety of Technical Processes*, volume 6 of *1*, pages 587 – 592, Beijing, P.R. China, 2006. IFAC.

[22] R. Rihm. Interval methods for initial value problems in odes. In *IMACS-GAMM International Workshop on Validated Computations*, Amsterdam, 1994. Elsevier.

[23] F.C. Schweppe. Recursive state estimation: unknown but bounded errors and system inputs. *Automatic Control, IEEE Transactions on*, 13(1):22–28, 1968.

[24] S. Tornil, T. Escobet, and L. Travé-Massuyès. Robust fault detection using interval models. Cambridge, UK, 2003. 12th European Control Conference ECC-03.

[25] E. Walter, L. Jaulin, and M. Kieffer. Interval analysis for guaranteed and robust non-linear estimation in robotics. *Nonlinear Analysis: Theory, Methods & Applications*, 47(1):191–202, 2001.

INVITED-DIMITROVA

Qiaochu LI obtained B.Eng from the Xi'an Jiaotong University. He is currently pursuing a Ph.D degree in the Department of Computer Engineering and Automatic Control, Université de Technologie de Compiègne (UTC). His research interests include interval analysis, parameter estimation and diagnostic for complex nonlinear systems.

Carine Jauberthie is Associate Professor at University Paul Sabatier of Toulouse (France) since 2005. She is a researcher at the CNRS Laboratoire dAnalyse et dArchitecture des Syst'emes (LAAS) (http://www.laas.fr) in the Diagnosis and Supervisory Control (DISCO) research team (http://www.laas.fr/DISCO/). She obtained a Ph.D. degree in 2002 in Applied Mathematics, specialized in system control, from University of Technology of Compi'egne, France, at ONERA Center of Lille in collaboration with the Laboratory of Applied Mathematics. Her research interests concern fault detection and diagnosis based on interval analysis, the use of constraint satisfaction approaches, and the analysis of the related properties, including identifiability of linear and nonlinear systems.

Lilianne Denis-vidal was teacher in high school from 1975 to 1991. She received the "Doctorat de troisième cycle" in 1981 in pure Mathematics from the University of Lyon, France, the Ph.D. degree in Control of systems in 1993 from the University of Compiègne, (UTC, France) and the Research Habilitations in Applied Mathematics in 2004 from the University of Compiègne. In 1991 she became Assistant professor in Mathematics at the Department of Mathematics of the Teacher Education Institute, University of Sciences and Technologies of Lille, Villeneuve D'Ascq, France. Her main research is in the field of mathematical modeling of pharmacokinetics, population dynamics, aircraft systems, automotive air pollution, parameter identifiability, system identification, differential and computer algebra.

Zohra Cherfi received the Ph.D. degree in mathematics from the University Pierre et Marie Curie, Paris, France, in 1988. She is currently a Professor with the Department of Mechanics, Compiègne University of Technology, Compiègne, France. Her current research interests concern reliability engineering, optimization and modelling for control of product quality.

Design & Study of a Low Power High Speed 8 Transistor Based Full Adder Using Multiplexer & XOR Gates

Biswarup Mukherjee, Aniruddha Ghoshal

***Abstract*— In this paper, we propose a new technique for implementing a low power high speed full adder using 8 transistors. Full adder circuits are used comprehensively in Application Specific Integrated Circuits (ASICs). Thus it is desirable to have high speed operation for the sub components. The explored method of implementation achieves a high speed low power design for the full adder. Simulated results indicate the superior performance of the proposed technique over conventional 28 transistor CMOS full adder. Detailed comparison of simulated results for the conventional and present method of implementation is presented.**

Keywords: High speed Low power full adder, 2-T MUX, 3-T XOR, 8-T FA, Pass transistor logic, CMOS (Complementary Metal Oxide Semiconductor)

I. INTRODUCTION

With the tremendous progress of recent electronic gadgets like laptops, mobile phones etc. and the evolution of the nanotechnology, the low-power & high speed microelectronic devices has come to the forefront. Today, there are an increasing number of portable applications requiring high speed, small-area low-power high throughput circuitry. Therefore, circuits with high speed, low-power consumption become the major consideration for design of system-components. Now a day logic circuits are designed using pass transistor logic techniques. It has recently been proposed with the objective of improving speed and power consumption [4, 6, 7]. Two of them, simultaneously developed by Hitachi CPL [4] and DPL [6], are the most notable. The Double Pass-Transistor Logic, developed by Hitachi 1993 demonstrated a 1.5ns 32-bit ALU in 0.25µm CMOS technology [6] and 4.4ns 54X54 bit multiplier [7].

Biswarup Mukherjee is an Assistant professor, Department of Electronics & Communication Engineering, Neotia Institute of Technology, Management And Science, Jhinga, D.H. Road, West Bengal, India, biswarup80@gmail.com
Aniruddha Ghosal is an associate professor in Institute of Radio Physics and Electronics, University of Calcutta, Kolkata, India, aghosal2008@gmail.com

The main objective of our work is to implement the low power high speed full adder & to draw a detailed comparative study with conventional full adder. The purpose of implementing the low power full adder is to project that using less number of transistors in comparison to the conventional full adder, the propagation delay time & power consumption gets reduced. It also helps in reducing the layout area thereby decreasing the entire size of a device where this adder is used. Power consumption is increasingly becoming the blockage in the design of ICs in advanced process technologies. We evaluate them from an industrial product development perspective. We also give a brief outlook to proposals on other levels in the design flow and to future work.

II. THEORY

The sum and carry out signals of the full adder are defined as the following two combinational Boolean functions of the three input variables A, B, and C.

$$Sum = A \oplus B \oplus C \qquad \text{-------eqn.1}$$

$$Carry = AB + BC + CA \qquad \text{-------eqn.2}$$

Accordingly the functions can be represented by CMOS logic as follows in fig. 1,

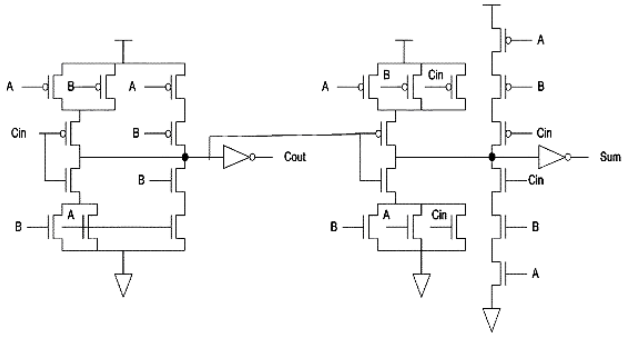


Fig. 1. Conventional 28-T CMOS 1 bit full adder

This work presents full adder using pass transistor logic MUX & XOR gate, which containing lesser transistors and achieving better performance. In comparison with conventional CMOS, pass transistor logic analogy shows performance improvement of up to 43% while the improvement in size ranges around 20%. In comparison with pass transistor logic analogy, there is 15% to 50% improvement in speed. As per as the power consumption is

ISBN: 978-1-61804-290-3

concerned, comparison between pass transistor logic analogy and conventional CMOS shows a 30% to 50% savings in favor of pass transistor logic analogy [4]. This full adder using pass transistor logic has advantages over CMOS and is characterized by excellent speed and low power. In electronics, pass transistor logic (PTL) describes several logic families used in the design of integrated circuits. It reduces the count of transistors used to make different logic circuits, by eliminating redundant transistors. Transistors are used as switches to pass logic levels between nodes of a circuit, instead of as switches connected directly to supply voltages [1]. This reduces the number of active devices, but has the disadvantage that output levels can be no higher than the input level. Each transistor in series has a lower voltage at its output than at its input [2]. For proper operation, design rules restrict the arrangement of circuits, so that sneak paths, charge sharing, and slow switching can be avoided. Simulation of circuits may be required to ensure adequate performance [3].

Double-pass transistor logic eliminates some of the inverter stages required for complementary pass transistor logic by using both N and P channel transistors, with dual logic paths for every function [1]. While it has high speed due to low input capacitance, it has only limited capacity to drive a load. Pass transistor logic has become important for the design of low-power high-performance digital circuits due to the smaller node capacitances and reduced transistors count it offers. However, the acceptance and application of this logic depends on the availability of supporting automation tools, e.g. timing simulators, which can accurately analyze the performance of large circuits at a speed, significantly faster than that of SPICE based tools [2].

The objective of this work is to develop and the synthesis methods for full adder based on pass transistor logic which will minimize the number of transistors used and yet preserve the speed of the logic. We have used 250nm technology to test our examples and to make comparisons.

III. ARCHITECTURE

The basic architecture of proposed 1 bit 8-Transistored full adder (8-T FA) is composed of 2 XOR gate to implement the sum out & another two input MUX to implement the carry out. The basic structure of the 2:1 MUX using pass gate transistor logic is shown in figure 2 [5]. In this configuration we have connected PMOS and NMOS along with a SEL line, as in MUX. As we know that PMOS works on ACTIVE LOW and NMOS works on ACTIVE HIGH. So, when the SELECT input is low (0) then the PMOS get activated, and show the input IN0 in the output and due to low input (0) the NMOS stands idle, as it is activated in high input. Same for the case, when the input is high (1) then the NMOS get activated, and show the input IN1 in the output. Thus this circuitry behaves as a 2-input MUX using SEL line, and shows the favorable output as 2:1MUX.

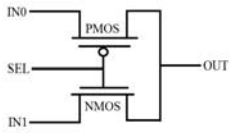


Fig.2. Basic view of 2T MUX

The second component required to design a full adder with 8 transistors is XOR gate. Conventional XOR gate can be fabricated using TG logic needs more than 3 transistors. But here is new design of a XOR gate with 3 transistors as shown in figure 3 below,

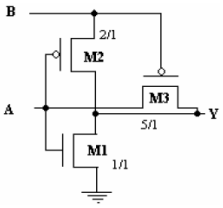


Fig.3. Basic view of 3T XOR gate

The design is based on a customized version of a CMOS inverter and a PMOS pass transistor. When the input B is at logic 1, the inverter on the left operates like a normal CMOS inverter. So the output Y is the complement of input A. When the input B is at logic 0, the CMOS inverter output is at high impedance. But, the pass transistor M3 is on and the output Y gets the same logic value as input A. The operation of the whole circuit is thus like a 2 input XOR gate [8].

Using 2 input MUX & 2 input XOR gate we can design a full adder as shown in figure 4 below,

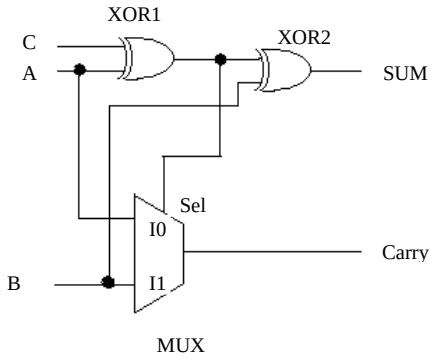


Fig.4. Full Adder based on MUX & XOR gates

V. LOGIC ANALYSIS

The digital circuit shown in the fig. 3 can be analyzed logically with the help of simple Boolean algebra. The inputs to XOR 1 gate are A & C. So the output is A $\oplus$ C. This output is again XORed with B so the output of XOR2 is,

Sum =A $\oplus$ B $\oplus$ C ……eqn. 3(same as eqn. 1)

ISBN: 978-1-61804-290-3

Regarding carry generation the inputs to the MUX are A & B respectively & the select line is A ⊕ C. So the output becomes,

$$\begin{aligned}\text{Carry} &= \overline{(A \oplus C)}.A + (A \oplus C).B \\ &= (\overline{A}C+A\overline{C})A+(\overline{A}C+A\overline{C})B \\ &= AC+AB\overline{C}+\overline{A}BC \\ &= AC(B+\overline{B})+AB\overline{C}+\overline{A}BC \\ &= ABC+A\overline{B}C+AB\overline{C}+\overline{A}BC \\ &= ABC+ABC+ABC+AB\overline{C}+A\overline{B}C+\overline{A}BC \\ &= ABC+AB\overline{C}+ABC+\overline{A}BC+ABC+A\overline{B}C \\ &= AB(C+\overline{C})+BC(A+\overline{A})+AC(B+\overline{B}) \\ &= AB+BC+CA \end{aligned}$$

------eqn.4
(same as eqn. 2)

Accordingly the truth table becomes,

TABLE I TRUTH TABLE OF PROPOSED 8-T FA

INPUTS			INOUT	OUTPUTS	
A	B	C	XOR1	XOR2 (SUM)	MUX (CARRY)
0	0	0	0	0	0
0	0	1	1	1	0
0	1	0	0	1	0
0	1	1	1	0	1
1	0	0	1	1	0
1	0	1	0	0	1
1	1	0	1	0	1
1	1	1	0	1	1

There are three major sources of power dissipation in a digital CMOS circuit: logic transition, short-circuit current and leakage current [6], [7]. The short-circuit current is the direct current passing through the supply and the ground, when both the NMOS and the PMOS transistors are simultaneously active [2], [6]. As the proposed 8-T adder does not have direct connections to or port (voltage connections to the back gate terminals are not considered), the probability of a direct path formation from positive voltage supply to the ground during switching can be substantially reduced; that is, the power consumption due to short circuit current is considered negligibly small. Furthermore, in the new 8-T adder, all of its internal gate nodes are directly excited by the fresh input signals (and), leading to a much faster transition (low rise and fall times) in its output signals. As a result, the power consumption of the following buffer stage can benefit from faster/cleaner Sum and Cout outputs.

V. SIMULATION

The act of simulating something generally entails on behalf of certain key characteristics or behaviors of a selected physical or abstract system. Simulation can be used to get the functional and timing analysis of the circuit models. Here conventional & our proposed full adder circuits are analyzed in standard simulator using 250 nm technology. We implement the conventional full adder using 28T (transistor) & simulate out its transient response, power consumption & delay analysis. Then we again design & simulate out the low power full adder using the concept of pass-gate transistor & XOR logic & thereby implementing the design with 8-transistors. First the circuits are simulated in schematic editor by providing different input combinations shown in table 1. The input specifications are tabulated in the table 2. The schematic diagrams of conventional 28-T FA and proposed 8-T FA are shown in figure 5 & 6 respectively.

TABLE II INPUT SPECIFICATIONS FOR 250NM TECHNOLOGY SIMULATION

Source Type	Bit
Zero Value	0V
One Value	5V dc
Period of Waveform	20 ns
Rise time	1 ns
Fall time	1 ns
Bit Parameter (input A)	00001111
Bit Parameter (input B)	00110011
Bit Parameter (input Cin)	01010101
Stop Time	80 ns

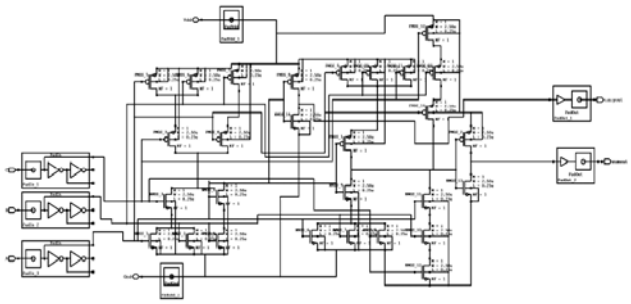


Fig 5: Schematic diagram of conventional 28-T full adder

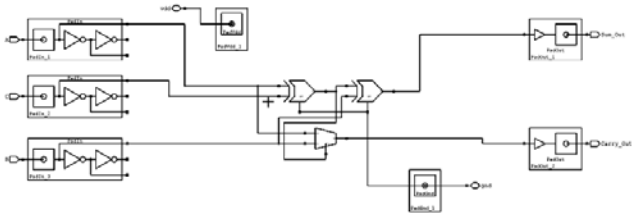


Fig. 6. Schematic diagram of proposed 8-T full adder

VI. SIMULATION RESULTS & ANALYSIS

The schematics shown in fig. 5 & fig.6 are simulated with the help of standard simulator tool to get the timing,

ISBN: 978-1-61804-290-3

power analysis. The transient responses of these schematics are shown below in fig.7 & fig. 8.

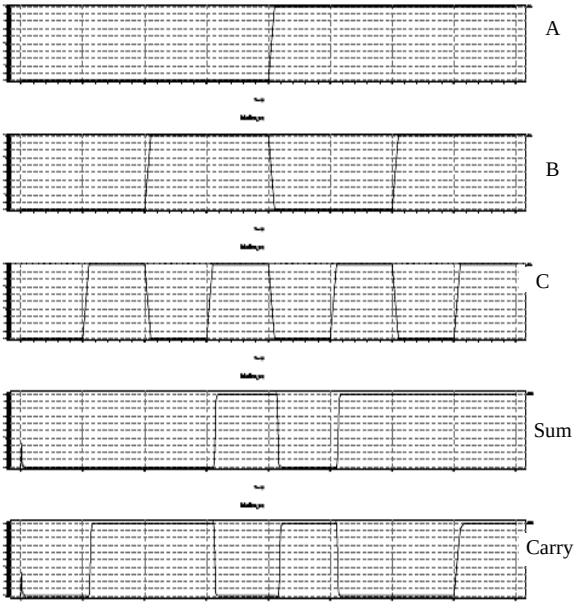


Fig. 7. Transient response of conventional 28-T FA

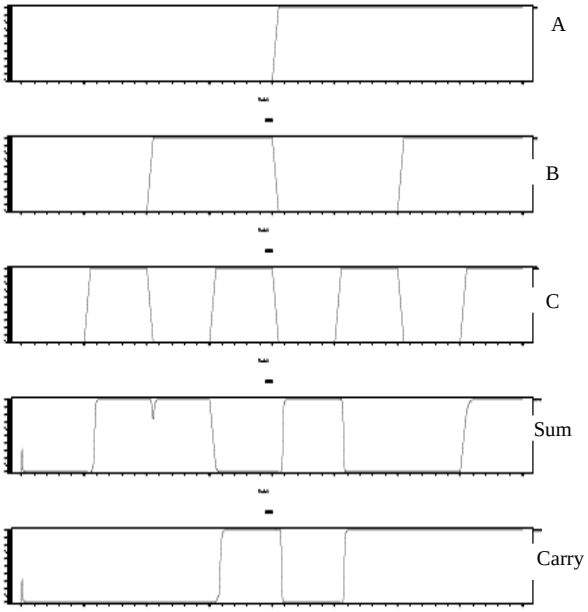


Fig. 8. Transient response of proposed 8-T FA

From the fig. 8 it can be easily understood that the proposed 8-T full adder is having the same transient response as like the conventional 28-T FA. The power & timing analysis is tabulated in the table below,

TABLE 3. COMPARISN BETWEEN CONVENTIONAL 28T FA & PROPOSED 8-T FA

28-T FA		8-T FA	
Power Analysis	Transient Delay Analysis	Power Analysis	Delay Analysis
Max power 354.6744 mw Min power .01124084 mw	Rise time delay: 0.85ns Fall Time Delay: 1.34 ns Transient Delay: 1.09 ns	Max power 298.2324 mw Min power .2377583uw	Rise time delay: 26.9 ps Fall Time Delay: 1.18 ns Transient Delay: 0.60 ns

The output waveform of proposed 8-T full adder is consisting of glitches in the sum and the carry part. These glitches are present due to the use of pass transistor logic, as in pass transistor the NMOS is a strong (0) element where as PMOS is a strong (1) element. So, there is a disturbance in the output.

This can be shorted out by using PAD IN and PAD OUT in the circuitry. This will give the buffered output and free of glitches in the output waveform.

Regarding area consideration it can be easily understood that this proposed 8-T FA core will be of much lower size compared to conventional 28-T FA.

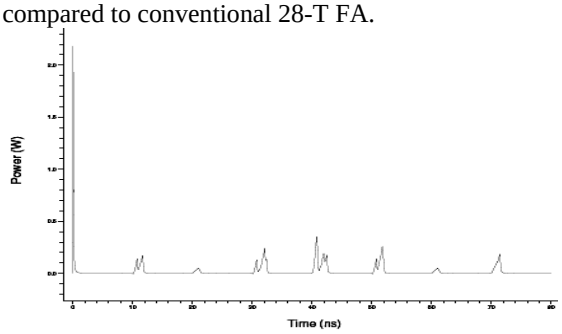


Fig. 9: Power consumption of conventional 28-T FA

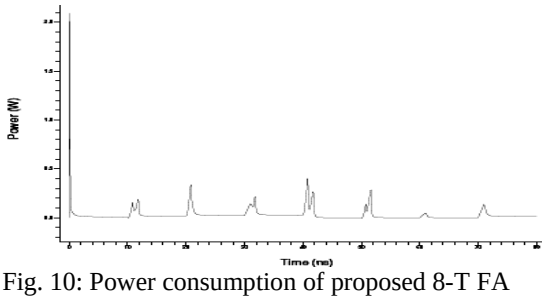


Fig. 10: Power consumption of proposed 8-T FA

ISBN: 978-1-61804-290-3

CONCLUSION

From the above results it can concluded that our proposed full adder has got better performance in speed, power and area consideration in comparison with conventional full adder. It turns out that in contrast to older process technologies, this approach is more suitable for industrial usage in advanced process technologies.

REFERENCES

[1] Jaume Segura, Charles F. Hawkins *"CMOS electronics: how it works, how it fails"*, Wiley-IEEE, 2004, page 132

[2] Clive Maxfield *Bebop to the Boolean boogie: an unconventional guide to electronics* Newnes, 2008, pp. 423-426

[3] Albert Raj/Latha *VLSI Design* PHI Learning Pvt. Ltd. pp. 150-153

[4] Yano, K, et al, "A 3.8 ns CMOS 16X16b multiplier using complementary pass transistor logic", *IEEE J. Solid State Circuits, Vol 25, p388-395, April 1990*

[5] Yingtao Jiang, Abdulkarim Al-Sheraidah, Yuke Wang, Edwin Sha, and Jin-Gyun Chung, "A Novel Multiplexer-Based Low-Power Full Adder" *IEEE Transaction on circuits and systems-II: Express Brief, Vol. 51, No. 7,p-345, July- 2004*

[6] Makoto Suzuki, et al, "A 1.5 ns 32 b CMOS ALU in double pass transistor logic", *ISSCC Dig. Tech. Papers, p90-91, February 1993*

[7] N. Ohkubo, et al, "A 4.4 ns CMOS 54X54 b multiplier using pass transistor multiplexer", *Proceedings of the IEEE 1994 Custom Integrated Circuit Conference, May 1-4 1994, p599-602, San Diego, California.*

[8] Shubhajit Roy Chowdhury, Aritra Banerjee, Aniruddha Roy, Hiranmay Saha, "A high Speed 8 Transistor Full Adder Design using Novel 3 Transistor XOR Gates", *International Journal of Electronics, Circuits and Systems 2, p-218, 2008*

A Real-Time Production Scheduling Framework based on Autonomous Agents

Kwan Hee Han, Yongsun Choi and Sung Moon Bae

***Abstract*—**The function of production scheduling is to provide the release and execution of orders according the conditions of production planning to meet customer requirements. Production scheduling is a difficult problem, particularly when it takes place in an open, dynamic environment. Conventional static scheduling system cannot deals with this problem effectively. Agent technology is particularly appealing to model and solve production scheduling problems. Proposed in this paper is real-time production rescheduling framework based on autonomous agents. As production line supervisor manages the continuously changing shop floor situations which disturb initial schedule, agent behaves like human in generating a realistic and easy-to-understand real-time production schedule.

***Keywords*—**Autonomous Agent, Decision-making, Production Scheduling, Production System

I. Introduction

PRODUCTION scheduling is the process of selecting and assigning manufacturing resources for specific time periods to the set of manufacturing processes in the plan [1]. Recently, emerging requirement of production scheduling is real-time scheduling capability to cope with continuous disturbances in today's JIT (Just-In-Time) production environment.

Production scheduling is a difficult problem, particularly when it takes place in an open, dynamic environment. In a manufacturing system, rarely do things go as expected. The set of things to do is generally dynamic. The system may be asked to do additional tasks that were not anticipated, and is sometimes allowed to omit certain tasks. The resources available to perform tasks are subject to change. Certain resources can become unavailable, and additional resources introduced. The beginning time and the processing time of a task are also subject to variation. A task can take more time than anticipated or less time than anticipated, and tasks can arrive early or late [2].

During two or more decades, centralized computer-based information system such as MRP (Material Requirement Planning) and ERP (Enterprise Resource Planning) which has features of top-down and sequential nature are mostly applied as a tool for production planning and scheduling (PP&S) in manufacturing industry. It is claimed that these batch-mode PP&S systems revealed critical shortcomings because they assume that the capacity is infinite [3]. Besides this limitation, they could not deal with production disturbances properly. In real shop floor, schedule changes occur frequently to reduce the negative impact of disturbances such as machine failure, delay of material supply, and employee absence. However, schedule change is a difficult task in the conventional scheduling system to adapt to dynamic real situations.

In summary, conventional centralized batch-mode scheduling system have limitations as bellows: 1) Static scheduling systems do not reflect dynamic changes of shop floor in real-time. 2) Conventional scheduling system does not achieve global optimization through coordinating various conflicting performance criteria. 3) It has not capability of sensitivity analysis in case of parameter changes of production environment.

Therefore, in order to increase the predictability of shop floor progress and to ensure the reliability of production schedule, a new approach to production scheduling is needed to overcome the limitations of existing batch-mode static scheduling. The new requirements are as follows: First, a new approach can monitor and reflect the real situation of shop floor in real-time. Second, it can model the various parameters and performance objectives of shop floor in a scheduling system with accuracy.

Because of its highly combinatorial aspects, its dynamic nature and its practical interest for manufacturing systems, the scheduling problem has been widely studied in the literature by various methods: heuristics, constraint propagation techniques, constraint satisfaction problem formalism, simulated annealing, Taboo search, genetic algorithms, neural networks, etc.

Agent technology has recently been used in attempts to resolve this problem as a promising way to provide optimization [2, 4]. Agents help to capture individual interests, local decision making using incomplete information, autonomy, responsiveness, robustness and modular, distributed, reconfigurable organizational structures [5].

The most important common properties of computational agents are as follows [6]: 1) Agents act on behalf of their designer or the user they represent in order to meet a particular purpose. 2) Agents are autonomous in the sense that they control both their internal state and behavior in the environment. 3)

Kwan Hee Han is with the Department of Industrial & Systems Engineering, Gyeongsang National University, Korea (phone: +82-55-772-1702; fax: +82-55-772-1699; e-mail: hankh@ gnu.ac.kr).

Yongsun Choi (Corresponding Author) is with Department of System Management & Engineering, Inje University, Korea (phone: +82-55-320-3117; fax: +82-55-322-3632; e-mail: yschoi@ inje.ac.kr).

Sung Moon Bae is with the Department of Industrial & Systems Engineering, Gyeongsang National University, Korea (phone: +82-55-772-1705; fax: +82-55-772-1699; e-mail: bsm@ gnu.ac.kr).

Agents exhibit some kind of intelligence, from applying fixed rules to reasoning, planning and learning capabilities. 4) Agents interact with their environment, and in a community, with other agents. 5) Agents are ideally adaptive, i.e., capable of tailoring their behavior to the changes of the environment without the intervention of their designer.

In real shop floor, field workers make decisions with knowledge and experiences about shop floor to cope with unexpected events. If they cannot solve these problems by themselves, they cooperate with workers of pre-/ post-process or consult a supervisor on that problem. In other words, to build or adjust a production schedule requires sophisticated interaction between participants. Software agent, which is a rule-based software object, is suitable for modeling field worker's behavior with complex interactions. Therefore, it is a powerful tool to solve the dynamic problem of production scheduling.

The aim of this paper is to propose an agent-based scheduling framework. To do this, design framework is shown, and based on this proposed framework, a prototype system is implemented to show the usefulness of proposed approach.

The rest of the paper is organized as follows. Section 2 describes a proposed agent-based production scheduling framework. Section 3 describes an implemented prototype system for automotive parts industry as a case study. Finally, the conclusion and suggestions for further research are found in section 4.

II. Agent-based Real-Time Production Scheduling Framework

Scheduling is a decision-making process that is used on a regular basis in many manufacturing and service industries. It deals with the allocation of resources to tasks over given time periods and its goal is to optimize one or more objectives. The objectives can also take many different forms. One objective may be the minimization of the completion time of the last task and another may be the minimization of the number of tasks completed after their respective due dates. Scheduling, as a decision-making process, plays an important role in most manufacturing systems [7].

In real world, as depicted in Figure 1, production is started by issuing production order to shop floor. Shop floor consists of multiple workers and machines. Eventually, these resources make finished products. During a production, when unexpected events occur, production schedule must be adjusted by a supervisor using accumulated domain knowledge and experiences.

This real world situation can be mapped to a production model using autonomous agents as follows, which is also depicted in Figure 2: Production order is modeled by task agent; Product is modeled by product agent; human/machine is modeled by workstation agent; Supervisor is divided and modeled by 3 manager agent type as follows: Task manager agent interacts with task agent instances. Product manager agent interacts with product agents. Workstation manager agent interacts with workstation agent instances. Three types of manager agent also interact with each other.

Harmonious interactions through the cooperation among agents enable dynamic real-time scheduling, which is difficult to attain in the conventional centralized rule-based approach [8, 9].

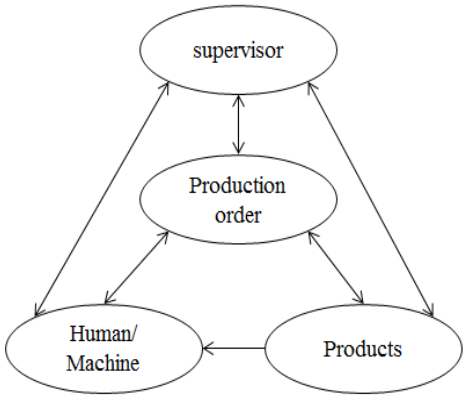


Figure 1. Elements of real production system

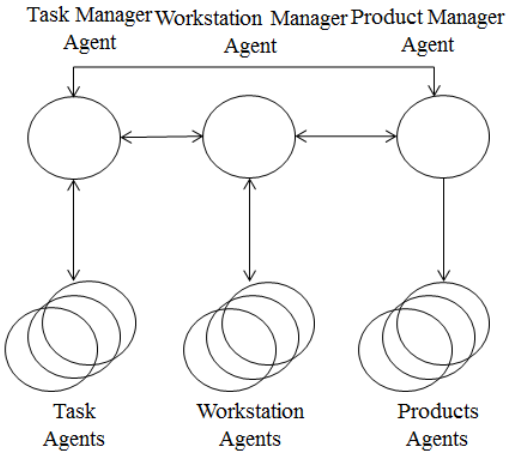


Figure 2. Elements of agent-based production system

For dynamic adaptation to continuously changing production environment, functional division of agent is necessary as follows: 1) decision making part, 2) communication part, 3) planning and coordination part, 4) monitoring part. Functional structure of autonomous agent is shown in Figure 3. To attain the goal of production, decision making part is further classified into two sub-types: self-model maintaining domain knowledge deals with local decisions, whereas acquaintance model deals with global decision, which manages the information about other agent. These two elements are essential parts of agent for real-time scheduling [10].

High level communication part, which is implemented in CORBA (Common Object Request Broker Architecture) object**,** determines communication mode between agents. There are two messaging modes as follows: 1) Broadcasting is a messaging where a piece of information is sent from one point to all other points. In this case, there is just one sender, but the information is sent to all connected receivers. 2) Multicasting is a messaging where a piece of information is sent from one or more points to

ISBN: 978-1-61804-290-3

a set of other points. In this case, there is may be one or more senders, and the information is distributed to a set of receivers (there may be no receivers or any other number of receivers).

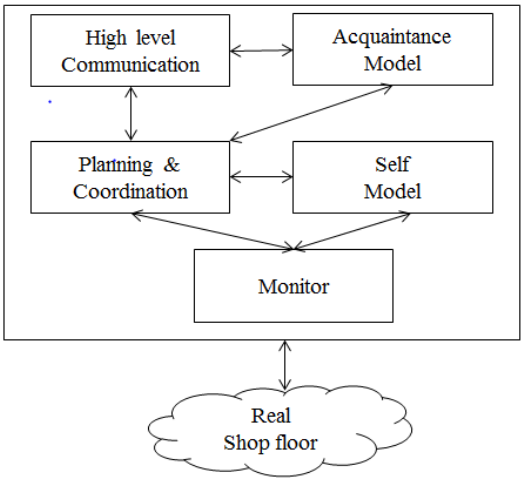


Figure 3. Functional structure of autonomous agent

Planning and coordination part defines knowledge according to the selected decision-making strategy for cooperation. In this paper, bidding mechanism was adopted as a decision making strategy. Monitor part handles input from human or machine.

To sum up, each agent is comprised of 5 parts, and each part consists of attributes and its operations.

During the pursuit of their goal, autonomous agents act independently with intelligence in ordinary times. If necessary, they make decisions through the cooperation, competition and negotiation with other agents like humans. Therefore, they belong to an agent community, which has characteristics as follows: 1) Join and withdrawal to community should be allowed to each agent when each agent is created or deleted. 2) Within a community, each agent ought to notify its capability represented as self-model to other agents.

In this paper, agent community is composed of several sub-group communities and their corresponding supervisor agent rather than one whole community as depicted in Figure 4. This type of agent community structure has advantages as follows: 1) communication overhead is reduced for the cooperation among many agents. 2) Addition and deletion of agents to an agent community is flexible without having impact on whole agent community.

Each autonomous agent has domain knowledge within self-model individually, which is different form conventional central rule base. Therefore, cooperation among agents is indispensable to decision making because necessary knowledge for good decision-making is distributed.

There are two approaches to agent decision making: First is a sequential pairwise comparison of all alternatives. It is suitable for small scale agent community, because it causes considerable computational and communicational overhead in the large scale community. Second is a bidding method, which is effective in the large scale agent community [11]. In this method, there exist one bidding controller and many bidders. Bidding controller requests a bid, and each bidder proposes a bid. Bidding controller selects the bidder with best bid value.

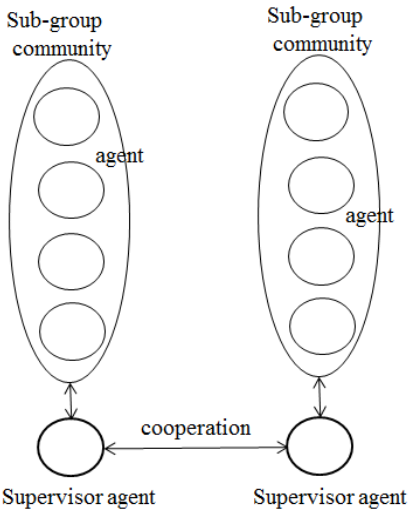


Figure 4. Structure of agent community

Communication among agents is usually accomplished by message transmission and receipt. To provide inter-operability between heterogeneous agents, a commonly understood agent communication language (ACL) is used: examples include KQML (Knowledge Query and Manipulation Language), Arcol, and FIPA's ACL [12]. The first ACL was the KQML that included many performatives, assertives and directives which agents use for telling facts, asking queries, subscribing to services and/or finding other agents. In this paper, CORBA (Common Object Request Broker Architecture) is adopted as a message calling among agents. In real production shop floor, message passing is occurred frequently within one communication node rather than communications between nodes, and has characteristics of small classes and many instances. In this situation, CORBA is an effective method to facilitate the communication of agents which is resulted in reducing messaging overhead.

III. Application of Agent-based Real-Time Production Scheduling System

Building a detailed schedule that can efficiently utilize resource capacity requires careful consideration of many interacting constraints. Conventional scheduling is often so time-consuming that only one possible schedule can be produced and no other scenarios can be tried. Worse still, when order changes come along or a machine goes down, the whole thing may have to be re-worked. In this paper, to improve the limitation of conventional scheduling method, and to generate a realistic production schedule, agent-based framework for real-time scheduling system is proposed.

As a case study for implementing a proposed framework, a prototype system was applied to the press line of automotive parts company. A press line which produces stamped panels is comprised of blanking sub-line and stamping sub-line in series. Finished product is moved to post-process which is called a panel assembly line. In a press line, production schedule and

production order is made for the stamping sub-line, and the schedule of blanking sub-line is made by backward scheduling.

Generally, the goal of production scheduling is to generate an optimal schedule meeting post-process's requirements. The need of real-time scheduling is for adapting to disturbances due to the dynamic fluctuations of production environment.

A scheduling system developed in this paper generates a re-schedule if the following conditions are met: 1) inventory level is changed (production report of stamped panel (+), delivery of stamped panel to post-process (-), adjustment of inventory (±). 2) machine failure and repair are occurred, 3) emergency order is issued.)

The above conditions cause the adjustment of current schedule. So, in real shop floor, line supervisor solves this situation by rescheduling with cooperation with other workers. Agents imitate these human behaviors realistically.

Major reason of rescheduling is due to the low stock level of stamped panel. So, in case of above 1), safety stock level is calculated as follows: SSL=PCPH * PLT where SSL=safety stock level, PCPH=panel consumption in the post-process per hour, PLT=production lead time of press line. If (SSL > current stock level), new production order is generated, i.e., reasonable production reschedule and priority of new production order is determined with consideration to existing production orders. Production schedule of new order has time stamp such as earliest start time, latest start time, earliest finish time and latest finish time. After the schedule of stamping sub-line is determined, the schedule of blanking sub-line is calculated automatically by backward scheduling.

As an example, in case of low stock level due to the delivery of stamped panel to post-process, detailed schedule generation procedure for new order in a press line is as follows:

1) The event of 'stamped panel (part number: #101) consumption' is notified to workstation manager agent.

2) If (safety stock level of #101> current stock level of #101), workstation manager agent notifies this event to product manager agent and task manager agent.

3) Product manager agent sends this event to product agent for #101 stamped panel within product agent community.

4) Product agent for #101 sends its stock level to task manager agent

5) Task manager agent creates the task agent for # 101 stamped panel, and registers it to agent community managed by task manager agent.

6) Task agent for #101 requests for bid to workstation agents to determine the production line. The selection criterion of bid is an EST (Earliest Start Time) of each production line to produce this stamping panel.

7) Workstation agents that can process this task agent participate in bidding, and propose a bid value.

8) Task agent for #101 selects the workstation agent which proposes best value of EST.

9) After workstation agent is determined, workstation manager agent determines the priority of task #101 by considering the relationship with existing production orders. For example, if there is an existing production order (called A) sharing same stamping die with new order (called B), the priority of B is adjusted to let it next to A.

IV. Conclusions and Further Research

Manufacturing industries are under great pressure caused by the rising costs of energy, materials, labor, capital, and intensifying worldwide competition. In other words, external environment of enterprise are rapidly changing brought about majorly by global competition, cost and profitability pressures, and emerging new technology.

In particular, frequent change of customer requirements is a tough challenge to manufacturing company. Conventional static batch-mode scheduling system cannot deals with this problem effectively. Agent technology is particularly appealing to model and solve production planning and control problems in manufacturing. Proposed in this paper is a real-time production rescheduling framework based on autonomous agents, and prototype system was implemented in the press line of automotive parts company. As production line supervisor manages the continuously changing shop floor situations which disturb initial schedule, agent behaves like human in generating real-time production schedule effectively. The main advantage of proposed framework is to generate a realistic and easy-to-understand schedule by imitating real-world decision-making process.

Further research is the integration of real-time production scheduling (RPS) system and simulation system: Current prototype system cannot ensure whether generated schedule is optimal or not. In the integrated structure, RPS system generates schedule alternatives, and simulations system evaluates each alternative in terms of pre-defined performance indicators, and the best one is selected as a new schedule.

Acknowledgment

This Work was supported by Academy-oriented Research Funds of Development Fund Foundation, Gyeongsang National University, 2014.

References

[1] W. Shen, L. Wang and Q. Hao, Agent-based distributed manufacturing process planning and scheduling: A state-of-the-art survey, *IEEE Transactions on Systems, Man, and Cybernetics–Part C: Applications and Reviewers*, Vol. 36, No. 4, pp. 563–577, 2006.

[2] W. Shen and H. N. Douglas, Agent-Based Systems for Intelligent Manufacturing: A State-of-the-Art Survey, *Knowledge and Information Systems*, Vol.1, No. 2, pp. 129-156, 1999.

[3] K. Udeda, Intelligent Manufacturing Systems-From Knowledge-base to Emergence-type, *Journal of the Japan Society of Precision Engineering*, Vol.5, No.11, pp. 5-10, 1993.

[4] F. T. S. Chan, and J. Zhang, A multi-agent-based agile shop floor control system, *International Journal of Advanced Manufacturing Technology*, Vol. 19, No. 10, pp. 764–774, 2002.

[5] J. Váncza and A. MaÂrkus, An agent model for incentive-based production scheduling, *Computers in Industry*, Vol. 43, No. 2, pp. 173-187, 2000.

[6] L. Monostori, J. Váncza and S.R.T. Kumara, Agent-Based Systems for Manufacturing, Annals of the CIRP, Vol. 55, No. 2, pp. 697-720, 2006.

[7] M. L. Pinedo, *Scheduling-Third Edition*, Springer, New York, USA, 2008.
[8] M. Wooldridge and N. R. Jennings, Agent theories, architectures, and languages: A survey, *Lecture Notes in Computer Science*, Vo. 890, pp. 1-39, 1995.
[9] Y. Zhang, G. Q. Huang, S. Sun and T. Yang, Multi-agent based real-time production scheduling method for radio frequency identification enabled ubiquitous shop floor environment, *Computers & Industrial Engineering*, Vol. 76, pp. 89–97, 2014.
[10] T. Wittig, *ARCHON: An Architecture for Multi-agent Systems*, pp. 21-22, Ellis Horwood Upper Saddle River, NJ, USA, 1992.
[11] M.K. Lim, Z. Zhang and, W.T. Goh, An iterative agent bidding mechanism for responsive manufacturing, *Engineering Applications of Artificial Intelligence*, Vol. 22, No. 7, pp. 1068–1079, 2009.
[12] J. Pitt and A. Mamdani, A Protocol-Based Semantics for an Agent Communication Language, *Proceedings of 16th International Joint Conference on artificial intelligence*, pp. 486-491, 1999.

Positivity and linearization of a class of nonlinear fractional continuous-time systems by state-feedbacks

Tadeusz Kaczorek

***Abstract*—**The positivity and linearization of a class of nonlinear fractional continuous-time system by nonlinear state-feedbacks are addressed. Necessary and sufficient conditions for the positivity of the class of fractional nonlinear systems are established. A method for linearization of the nonlinear systems by nonlinear state-feedbacks is presented. It is shown that by suitable choice of state-feedbacks it is possible to obtain asymptotically stable and controllable fractional linear system and if the closed-loop system is positive then it is unstable.

***Keywords*—**fractional, positive, nonlinear, system, linearization, state-feedback.

I. Introduction

In positive systems inputs, state variables and outputs take only non-negative values. Examples of positive systems are industrial processes involving chemical reactors, heat exchangers and distillation columns, storage systems, compartmental systems, water and atmospheric pollution models. A variety of models having positive linear behavior can be found in engineering, management science, economics, social sciences, biology and medicine, etc. Positive linear systems are defined on cones and not on linear spaces. Therefore, the theory of positive systems is more complicated and less advanced. An overview of state of the art in positive systems theory is given in the monographs [6, 15]. Positive linear systems consisting of n subsystems with different fractional orders have been addressed in [16, 18]. Minimum energy control of positive discrete-time and continuous-time linear system has been considered in [11-14]. The theory of geometrical approach to analysis of nonlinear systems based on the Lie algebra has been given in [2, 10, 19]. The problem of linearization of nonlinear systems by nonlinear state-feedbacks has been investigated in [1, 3-5, 7, 8, 10, 19-23].

In this paper the positivity and linearization of a class of nonlinear fractional continuous-time systems by nonlinear state-feedbacks will be addressed. The paper is organized as follows. In section 2 necessary and sufficient conditions for the positivity of a class of fractional nonlinear systems are established. The linearization of the nonlinear system by nonlinear state-feedbacks is addressed in section 3. An example illustrating the considerations is given in section 4. Concluding remarks are given in section 5.

The following notation will be used: $\Re$ - the set of real numbers, $\Re^{n\times m}$ - the set of $n\times m$ real matrices and $\Re^n=\Re^{n\times 1}$, $\Re_+^{n\times m}$ - the set of $n\times m$ matrices with nonnegative entries and $\Re_+^n=\Re_+^{n\times 1}$, M_n - the set of $n\times n$ Metzler matrices (with nonnegative off-diagonal entries), I_n - the $n\times n$ identity matrix.

This work was supported by National Science Centre in Poland under work No. 2014/13/B/ST7/03467.

T. Kaczorek is with the Bialystok University of Technology, Wiejska 45A, Bialystok, POLAND; e-mail: Kaczorek@isep.pw.edu.pl.

II. Positivity of nonlinear systems

Consider the nonlinear system

$$ {}_0D_t^{\alpha}x(t)=Ax+f(x)+Bu,\ 0<\alpha<1 \tag{1a}$$

where

$$ {}_0D_t^{\alpha}x(t)=\frac{d^{\alpha}x(t)}{dt^{\alpha}}=\frac{1}{\Gamma(n-\alpha)}\int_0^{\infty}\frac{f^{(n)}(\tau)}{(t-\tau)^{\alpha+1-n}}d\tau,\ f^{(n)}(\tau)=\frac{d^nf(\tau)}{d\tau^n}, \tag{1b}$$

$n-1<\alpha<n$, $n\in W=\{1,2,...\}$ is the Caputo definition of $\alpha\in\Re$ order derivative of $x(t)$ and

$$ \Gamma(\alpha)=\int_0^{\infty}e^{-t}t^{\alpha-1}dt \tag{1c}$$

is the Euler gamma function and

$$ x=\begin{bmatrix}x_1\\ \vdots\\ x_n\end{bmatrix},\ A=\begin{bmatrix}0 & 1 & 0 & \dots & 0\\ 0 & 0 & 1 & \dots & 0\\ \vdots & \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & 0 & \dots & 1\\ 0 & 0 & 0 & \dots & 0\end{bmatrix},\ f(x)=\begin{bmatrix}f_1(x_1)\\ f_2(x_1,x_2)\\ \vdots\\ f_n(x_1,\dots,x_n)\end{bmatrix},\ B=\begin{bmatrix}0\\ \vdots\\ 0\\ 1\end{bmatrix} \tag{2}$$

$x=x(t)\in\Re^n$, $u=u(t)\in\Re$ are the state and input vectors ,respectively.

ISBN: 978-1-61804-290-3

It is assumed that the functions $f_k(x_1,...,x_k)$, $k = 1,2,\ldots,n$ are continuously differentiable for all their arguments.

Definition 1. The nonlinear system (1) is called (internally) positive if $x(t) \in \Re_+^n$ for all $x(0) \in \Re_+^n$, $t \geq 0$ and every $u(t) \in \Re_+$, $t \geq 0$.

Theorem 1. The nonlinear system (1) is positive if and only if

$$f_k(\bar{x}) \in \Re_+ \text{ for } \bar{x} = [x_1,...,x_{j-1},0,x_{j+1},...,x_k(t)]^T \in \Re_+,$$
$$j = 1,2,\ldots,k \text{ and } u(t) \in \Re_+,\ t \geq 0. \quad (3)$$

Proof. For given $f(x)$ the solution of (1) has the form

$$x(t) = \Phi_0(t)x(0) + \int_0^t \Phi(t-\tau)[f(x(\tau)) + Bu(\tau)]d\tau \quad (4a)$$

where

$$\Phi_0(t) = E_\alpha(At^\alpha) = \sum_{k=0}^{\infty} \frac{A^k t^{k\alpha}}{\Gamma(k\alpha+1)}, \quad (4b)$$

$$\Phi(t) = \sum_{k=0}^{\infty} \frac{A^k t^{(k+1)\alpha-1}}{\Gamma[(k+1)\alpha]}. \quad (4c)$$

The linear system obtained from (1) for $f_k(x_1,...,x_k) = 0$, $k = 1,2,\ldots,n$ is positive since the matrix A is a Meltzer matrix, $B \in \Re_+^n$. Using the well-known Picard method the k-approximation of the solution of (1) can be found from the formula

$$x_{k+1}(t) = \Phi_0(t)x(0) + \int_0^t \Phi(t-\tau)[f(x_k(\tau)) + Bu(\tau)]d\tau,\ k = 1,2,\ldots (5)$$

The Lipschitz conditions for (1) are satisfied since by assumption the functions $f_k(x_1,...,x_k)$, $k = 1,2,\ldots,n$ are continuously differentiable. Using the method given in [9] it is easy to show that the equation (1) has nonnegative solution $x(t) \in \Re_+^n$, $t \geq 0$ if and only if the conditions (3) are satisfied. □

The proof can be also accomplished using the method presented in [17].

III. Linearization by State-Feedback

For the nonlinear system (1) we introduce the following new state variables (the components of the new state vector $z = [z_1 \quad ... \quad z_n]^T$)

$$\begin{aligned}
z_1 &= x_1,\\
z_2 &= x_2 + f_1(x_1),\\
z_3 &= x_3 + f_2(x_1,x_2) + \frac{\partial f_1}{\partial x_1}[x_2 + f_1(x_1)] = x_3 + \bar{f}_2(x_1,x_2),\\
z_4 &= x_4 + f_3(x_1,x_2,x_3) + \frac{\partial \bar{f}_2}{\partial x_1}[x_2 + f_1(x_1)] + \frac{\partial \bar{f}_2}{\partial x_2}[x_3 + f_2(x_1,x_2)] \quad (6)\\
&= x_4 + \bar{f}_3(x_1,x_2,x_3),\\
&\vdots\\
z_n &= x_n + \bar{f}_{n-1}(x_1,...,x_{n-1}).
\end{aligned}$$

The relations (6) can be written shortly as $z = \phi(x)$. From (6) we have

$$\begin{aligned}
x_1 &= z_1,\\
x_2 &= z_2 - f_1(z_1),\\
x_3 &= z_3 - \bar{f}_2(z_1,z_2), \quad (7)\\
&\vdots\\
x_n &= z_n - \bar{f}_{n-1}(z_1,...,z_{n-1})
\end{aligned}$$

and shortly $x = \phi^{-1}(z)$. The nonlinear system (1) in the new state variables (6) has the form

$$\begin{aligned}
\frac{d^\alpha z_1}{dt^\alpha} &= \frac{d^\alpha x_1}{dt^\alpha} = x_2 + f_1(x_1) = z_2,\\
\frac{d^\alpha z_2}{dt^\alpha} &= \frac{d^\alpha x_2}{dt^\alpha} + \frac{\partial f_1}{\partial x_1}\frac{d^\alpha x_1}{dt^\alpha} = x_3 + f_2(x_1,x_2) + \frac{\partial f_1}{\partial x_1}[x_2 + f_1(x_1)] = z_3,\\
&\vdots\\
\frac{d^\alpha z_{n-1}}{dt^\alpha} &= x_n + f_{n-1}(x_1,...,x_{n-1}) + \frac{\partial \bar{f}_{n-2}}{\partial x_1}[x_2 + f_1(x_1)]\\
&\quad + ... + \frac{\partial \bar{f}_{n-2}}{\partial x_{n-2}}[x_{n-1} + f_{n-2}(x_1,...,x_{n-2})] = z_n,\\
\frac{d^\alpha z_n}{dt^\alpha} &= f_n(x_1,...,x_n) + u + \frac{\partial \bar{f}_{n-1}}{\partial x_1}[x_2 + f_1(x_1)]\\
&\quad + ... + \frac{\partial \bar{f}_{n-1}}{\partial x_{n-1}}[x_n + f_{n-1}(x_1,...,x_{n-1})]\Big|_{x=\phi^{-1}(z)}\\
&= -a_0 z_1 - a_1 z_2 - ... - a_{n-1} z_n + v
\end{aligned} \quad (8)$$

where

$$\begin{aligned}
v &= u + g(x),\\
g(x) &= \sum_{i=0}^{n-1} a_i z_{i+1}\Big|_{z=\phi(x)} + f_n(x_1,...,x_n) + \frac{\partial \bar{f}_{n-1}}{\partial x_1}[x_2 + f_1(x_1)] \quad (9)\\
&\quad + ... + \frac{\partial \bar{f}_{n-1}}{\partial x_{n-1}}[x_n + f_{n-1}(x_1,...,x_{n-1})].
\end{aligned}$$

ISBN: 978-1-61804-290-3

The equations (8) can be written in the form

$$\frac{d^{\alpha}z}{dt^{\alpha}} = Az + Bv, \quad z(0) = \phi[x(0)] \in \Re^n \tag{10}$$

where

$$\overline{A} = \begin{bmatrix} 0 & 1 & 0 & ... & 0 \\ 0 & 0 & 1 & ... & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & ... & 1 \\ -a_0 & -a_1 & -a_2 & ... & -a_{n-1} \end{bmatrix}, \quad B = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}. \tag{11}$$

Note that applying to the nonlinear system (8) the nonlinear state-feedback

$$u = v - g(x) \tag{12}$$

we obtain linear closed-loop system described by the equation (10). The coefficients a_k, $k = 0,1,...,n-1$ can be chosen so that the linear system (10) is asymptotically stable.

Note that for all values of the coefficients a_k, $k = 0,1,...,n-2$ the pair (11) is controllable since [8, 9]

$$\text{rank}[B \quad \overline{A}B \quad ... \quad \overline{A}^{n-1}B] = n. \tag{13}$$

The fractional linear system (10) with (11) is positive if and only if $a_k = 0$, $k = 0,1,...,n-2$. In this case the fractional linear system is unstable.

Therefore the following theorems have been proved.

Theorem 2. The fractional nonlinear system (1) can be linearized by the nonlinear state-feedback (12) and for suitable choice of the coefficients a_k, $k = 0,1,...,n-1$ the linear closed-loop system (10) is asymptotically stable and controllable.

Theorem 3. The fractional nonlinear system (1) can be linearized by the nonlinear state-feedback (12), so that the closed-loop system (10) for $a_k = 0$, $k = 0,1,...,n-1$ is positive but unstable.

IV. Example

Consider the fractional nonlinear system described by the equation (1) for $n = 3$ with

$$A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}, \quad B = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, \quad f(x) = \begin{bmatrix} x_1^2 \\ x_1 x_2 \\ x_2 x_3 \end{bmatrix}, \quad x(0) \in \Re_+^3. \tag{14}$$

The system (14) is positive since the conditions (3) are satisfied and $u = u(t) \in \Re_+$, $t \geq 0$. In this case the new state variable z_k, $k = 1,2,3$ are defined as follows

$$z = \begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 + x_1^2 \\ x_3 + 3x_1x_2 + 2x_1^3 \end{bmatrix} = \phi(x) \tag{15}$$

and

$$x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} z_1 \\ z_2 - z_1^2 \\ z_3 - 3z_1z_2 + z_1^3 \end{bmatrix} = \phi(x)^{-1}. \tag{16}$$

The nonlinear system (14) in the new state variables is described by the equation

$$\frac{d^{\alpha}z}{dt^{\alpha}} = \frac{d^{\alpha}}{dt^{\alpha}} \begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix}$$

$$= \begin{bmatrix} z_2 \\ z_3 \\ -z_1^5 - 2z_1^4 + 4z_1^3z_2 - 3z_1^2z_2 - z_1^2z_3 + 3z_2^2 - 3z_1z_2^2 + 3z_1z_3 + z_2z_3 + u \end{bmatrix}. \tag{17}$$

To linearize the nonlinear system (17) we apply the nonlinear state-feedback (12) of the form

$$\begin{aligned} u = v - g(z) = v - a_0z_1 - a_1z_2 - a_2z_3 + z_1^5 + 2z_1^4 - 4z_1^3z_2 \\ + 3z_1^2z_2 + z_1^2z_3 - 3z_2^2 + 3z_1z_2^2 - 3z_1z_3 - z_2z_3 \end{aligned} \tag{18}$$

and we obtain the linear system (10) with

$$A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -a_0 & -a_1 & -a_2 \end{bmatrix}, \quad B = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}. \tag{19}$$

The linear system is controllable for all values of the coefficients a_k, $k = 0,1,2$ and it is asymptotically stable if and only if $a_k > 0$, $k = 0,1,2$ and $a_1a_2 > a_0$.

The linear system (10) with (19) is positive if and only if $a_k = 0$, $k = 0,1$ since in this case the matrix

$$A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} \in M_3. \tag{20}$$

and the linear system is unstable.

Remark 1. The presented method can be easily extended to the fractional nonlinear systems described by the equation (1) with controllable pair $(\overline{A}, \overline{B})$ and the monomial transformation matrix P such that $A = P\overline{A}P^{-1}$, $B = P\overline{B}$ and A and B have the form (11).

V. Concluding Remarks

The positivity and linearization of a class of nonlinear system by nonlinear state-feedbacks have been addressed.

Necessary and sufficient conditions for the positivity of the class of nonlinear systems (Theorem 1) have been established. It has been shown that the nonlinear systems can be linearized by nonlinear feedbacks so that the linear close-loop system is asymptotically stable and controllable (Theorem 2) and positive but unstable (Theorem 3). The considerations are illustrated by an example.

REFERENCES

[1] J.L.M. Aguilar, R.A. Garcia, C.E. D'Attellis, "Exact linearization of nonlinear systems: trajectory tracking with bounded control and state constrains," *Proc. of the 38th Midwest Symposium on Circuits and systems*, Rio de Janeiro, 1995, pp. 620-622.

[2] R.W. Brockett, "Nonlinear systems and differential geometry," *Proc. of IEEE*, vol. 64, no. 1, 1976, pp. 61-71.

[3] B. Charlet, J. Levine, R. Marino, "Sufficient conditions for dynamic state feedback linearization," *SIAM J. Contr. Optimization*, vol. 29, no. 1, 1991, pp. 38-57.

[4] C. Daizhan, T. Tzyh-Jong, A. Isidori, "Global external linearization of nonlinear systems via feedback," *IEEE Trans. on Autom. Contr.*, 1985, pp. 808-811.

[5] B. Fang, A.G. Kelkar, "Exact linearization of nonlinear systems by time scale transformation," *Proc. of the American Control Conf.*, Denver-Colorado 2003, pp. 3555-3560.

[6] L. Farina and S. Rinaldi, *Positive Linear Systems; Theory and Applications*, J. Wiley, New York, 2000.

[7] B. Jakubczyk, "Introduction to geometric nonlinear control; controllability and Lie bracket," *Summer Schools on Mathematical Control Theory*, Triest, 2001.

[8] B. Jakubczyk, W. Respondek, "On linearization of control systems," *Bull. Pol. Acad. Sci. Tech.*, vol. 28, 1980, pp. 517-521.

[9] D. Idczak and R. Kamocki, "On the exsitance and uniqueness and formula for the solution of R-L fractional Couchy problem in Rn," *Fractional Calculus and Applications*, vol. 14, no. 4, 2011, pp. 538-553.

[10] A. Isidori, *Nonlinear Control Systems*, Springer Verlag, Berlin, 1989.

[11] T. Kaczorek, "Minimum energy control of descriptor positive discrete-time systems," *COMPEL*, vol. 33, no. 3, 2014, pp. 976-988.

[12] T. Kaczorek, "En extension of Klamka's method of minimum energy control to fractional positive discrete-time linear systems with bounded inputs," *Bull. Pol. Acad. Sci. Tech.*, vol. 62, no. 2, 2014, pp. 227-232.

[13] T. Kaczorek, "Necessary and sufficient conditions for minimum energy control of positive discrete-time linear systems with bounded inputs," *Bull. Pol. Acad. Sci. Tech.*, vol. 62, no. 1, 2014, pp. 85-90.

[14] T. Kaczorek, "Minimum energy control of fractional positive continuous-time linear systems with bounded inputs," *Int. J. Appl. Math. Comput. Sci.*, vol. 24, no. 2, 2014, pp. 335-340.

[15] T. Kaczorek, *Positive 1D and 2D Systems*, Springer Verlag, London, 2002.

[16] T. Kaczorek, "Positive linear systems consisting of n subsystems with different fractional orders," *IEEE Trans. Circuit and Systems*, vol. 58, no. 6, 2011, pp. 1203-1210.

[17] T. Kaczorek, "Positivity and linearization of a class of nonlinear continuous-time systems by state-feedback," *AMCS*, 2015 (in Print).

[18] T. Kaczorek, *Selected Problems of Fractional System Theory*, Springer Verlag, Berlin, 2012.

[19] W. Malesza and W. Respondek, "State-linearization of positive nonlinear systems; applications to Lotka-Volterra controlled dynamics," *Taming Heterogeneity and Complexity of Embedded Control*, John Wiley, Newport Beach, CA USA, 2007, pp. 451-473.

[20] W. Malesza, "Geometry and equivalence of linear and nonlinear control systems invariant on corner regions," *Phd Thesis*, Warsaw University of Technology, Warsaw 2008.

[21] R. Marino, P. Tomei, *Nonlinear Control Design – geometric, adaptive, robust*, Prentice Hall, London, 1995.

[22] K. Melham, M. Saad, S.C. Abou, "Linearization by redundancy and stabilization of nonlinear dynamical systems: a state transformation approach," *IEEE Int. Symposium on Industrial Electronics*, 2009, pp. 61-68.

[23] J.H. Taylor, A.J. Antoniotti, "Linearization algorithms for computer-aided control engineering," *Control Systems Magazine*, 1993, pp. 58-64.

[24] G. Wei-Bing, W. Dang-Nan, "On the method of global linearization and motion control of nonlinear mechanical systems," *Proc. of the Int. Conf. on Industrial Electronics, Control, Instrumentation and Automation*, San Diego, 1992, pp. 1476-1481

Radar Equation Applied to SAW Tag Sensing

Guatavo Cerda-Villafana and Yuriy. S. Shmaliy

***Abstract*—**Surface acoustic wave sensors and ID tags keep expanding their range of applications. But their limited reading range, in passive sensing, hinders their full-scale implementation. The radar equation is used in SAW tag sensing mostly to define the maximum reading distance between the reader and tag. This paper discusses effect of the radar equation parameters on the SAW system range and refers to practical values reported in the literature. Thereby, we predict possible ranges for SAW tag systems taking into account all the elements integrated to.

***Keywords*—**Passive device, radar equation, sensors, surface acoustic wave.

I. Introduction

THE proliferation of passive surface acoustic wave (SAW) devices [1], [2] has made an equation which relates the reading distance to the system parameters an important and necessary tool. There have been developed several forms of the radar equation to SAW systems [3]-[6] referring to different factors. This paper reviews some SAW tag system applications using the radar equation, analyses effect of the system parameters on the reading range, and provides an insight on how such systems work.

II. Principles of Operation

Passive SAW tag systems employ transducer mechanisms, one for transforming radio waves to electric signals and the other to transform electric signals to acoustic waves. Such systems are two-way, to mean that the acoustic waves are further transformed to electric signals and to radio waves.

The operational principle of SAW tags is based on the piezoelectric effect of some materials, such as quartz crystal, generating acoustic waves by applying an electric field and vice-versa, and generating an electric field when receiving acoustic waves, see Fig. 1. Piezoelectric materials used as substrate for SAW devices are quartz, lithium niobate (ideal for temperature sensing applications), lithium tantalite, lanthanum gallium silicate, langasite, etc. These substrates can be used to measure pressure, torque, vibration, acceleration, displacement, etc., by applying reflective delay lines [3]. For sensing other physical variables, the substrate is commonly coated with materials affected by the physical variable of interest. For chemical sensors chemosensitive materials that absorbs selectively the targeted molecule are applied, even Carbon nanotubes have been reported for detecting volatile organic compounds (VOC) [7].

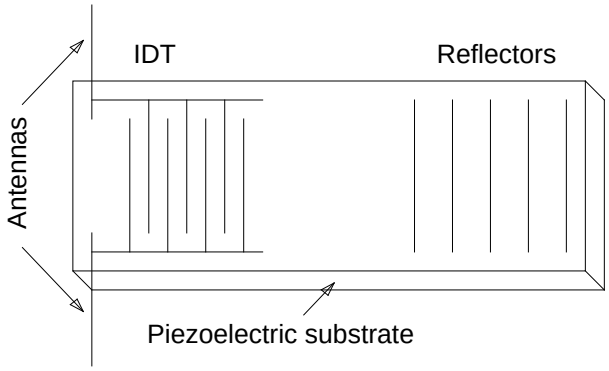


Fig. 1. A passive SAW tag device.

III. The Radar Equation

For the following analysis, we consider an open space and clear path without backscattering. The radar equation can be derived from the nondirectional power density S_u,

$$S_u = \frac{P_T}{4\pi R_1^2}, \tag{1}$$

where P_T is the transmitted power (W) from the interrogator, S_u is the nondirectional power density (W/m2), and R_1 is the range from the antenna of the interrogator to the tag (m). If we include the antenna gain to (1) as

$$S_g = S_u \cdot G_{rt}, \tag{2}$$

where S_g is the directional power density, S_u is the nondirectional power density, and G_{rt} is the transmitting antenna gain from the interrogator. We now take the captured power P_D by the antenna at the tag device as

$$P_D = S_g \cdot G_{dr} \cdot A_{dr} \tag{3}$$

in which P_D is the captured power (W) by the device, A_{dr} is the effective area of the receiving antenna (m^2), and G_{dr} is the receiver antenna gain.

By substituting S_g in (3), the equation becomes

G. Cerda-Villafana is with the Electronic Engineering Department, University of Guanajuato, Gto. 64885, Mexico (phone: 52-464-6479940, ext. 2403; e-mail: gcerdav@ugto.mx).

Y. S. Shmaliy, is with the Electronic Engineering Department, University of Guanajuato, Gto. 64885, Mexico (e-mail: shmaliy@ugto.mx).

$$P_D = \frac{P_T}{4\pi R_1^2} \cdot G_{rt} \cdot G_{dr} \cdot A_{dr} \quad . \tag{4}$$

Likewise, the power density yielded at the receiver can be found to be

$$S_e = \frac{P_G}{4\pi R_2^2} \quad , \tag{5}$$

where S_e is the power density at receiving place (W/m^2), P_G is the transmitted power by the sensor device (W), R_2 is the range tag-interrogator (m).

The power transmitted by the sensor device is related to the captured power as

$$P_G = \frac{P_D}{D} G_{dt} \quad , \tag{6}$$

where D is the loss at the sensor element and G_{dt} is the transmitting antenna gain.

Hence, S_e can be represented as

$$S_e = \frac{P_D G_{dt}}{4\pi R_2^2 D} \quad . \tag{7}$$

At the interrogator antenna, the received power P_R can be found to be

$$P_R = S_e \cdot A_r \quad , \tag{8}$$

where P_R is the power (W) received by the interrogator and A_r is effective area (m2) of the receiver antenna. The power received by the tag can thus be calculated as

$$P_R = \frac{P_D G_{dt}}{4\pi R_2^2 D} \cdot A_r \quad . \tag{9}$$

If we now take the effective antenna aperture for a dipole antenna as given by [7]

$$A_r = \frac{G_{rr} \cdot \lambda^2}{4\pi} \quad , \tag{10}$$

We then can transform (9) to

$$P_R = \frac{P_D G_{dt} \lambda G_{rr}}{(4\pi)^2 R_2^2 D} \quad . \tag{11}$$

Now, by substituting P_D, we have

$$P_R = \frac{P_T G_{rt} G_{dt} A_{dr} \lambda^2 G_{rr}}{(4\pi)^3 R_2^2 R_1^2 D} \quad . \tag{12}$$

Taking into account that $A_{dr} = G_{dr} \lambda^2 / 4\pi$ and $R_1 = R_2$, we next get

$$P_R = \frac{P_T G_{rt} G_{dr} G_{dt} \lambda^4 G_{rr}}{(4\pi)^4 R^4 D} \quad . \tag{13}$$

Equation (13) can now be solved for range R,

$$R = \frac{\lambda}{4\pi} \sqrt[4]{\frac{P_T G_{rt} G_{dr} G_{dt} G_{rr}}{P_R D}} \quad . \tag{14}$$

If we further accept that $G_{rt} = G_{rr}$ and $G_{dr} = G_{dt}$, we come up with the following expression

$$R = \frac{\lambda}{4\pi} \sqrt[4]{\frac{P_T G_r^2 G_d^2}{P_R D}} \quad . \tag{15}$$

The power received by the interrogator P_R which represents the power of the desired signal S_{in} can be expressed in terms of the system noise temperature and the signal-to-noise ratio [8]

$$\gamma_{in} = \frac{S_{in}}{N_{sys}} = \frac{S_{in}}{k T_{sys} B} \quad . \tag{16}$$

Otherwise, we can write

$$P_R = S_{in} = \gamma_{in} k T B \quad , \tag{17}$$

where B is the bandwidth, k is the Boltzmann's constant, and T_{sys} is the reference temperature in Kelvin degrees. Finally, measuring γ at the end of the amplifier circuits in the receiver, given in terms of γ_{in} and the receiver noise figure F,

$$\gamma = \frac{\gamma_{in}}{F} \tag{18}$$

that gives us

$$\gamma_{in} = \gamma F \quad . \tag{19}$$

Substituting γ_{in}, the radar equation for SAW tag systems finally becomes

$$R = \frac{\lambda}{4\pi} \sqrt[4]{\frac{P_T G_r^2 G_d^2}{\gamma F k T_{sys} B D}} \quad . \tag{20}$$

ISBN: 978-1-61804-290-3

IV. APPLICATIONS

The radar equation similar to (20) has been applied to SAW tag/sensor systems and used by many researchers. Below, we discuss several such applications. Note that not each application has enough information about the system performance. Hence, we use possible values of the missed parameters in order to predict the reader range.

A. Electromagnetic Wavelength (λ)

The most used value for λ is 0.122m which corresponds to an operation frequency of 2.45 GHz, which is one of the two unlicensed frequency bands allotted to low-power devices. These bands are suitable for SAW devices (433.07-434.77 MHz and 2.4-2.483 GHz). In [3], they use λ equal to 0.691. This value corresponds to the frequency of 433.9 MHz. Other frequencies such as 70MHz [9] and 69.7MHz [10] can also be exploited.

B. Interrogation Unit Power (P_T)

The allowed effective isotropically radiated power (EIRP), or, alternatively, equivalent isotropic radiation power (EIRP) for the unlicensed frequencies is 25 mW. This maximum power is mentioned in two works [3] and [13]. Other authors refer to 10 mW.

C. Interrogator Unit Gain (G_r)

This parameter represents the ratio between the directional power density S_u and the nondirectional power density S_g. The interrogator unit gain has been defined in [8] as

$$G_r = \frac{\text{maximum power intensity}}{\text{average power intensity over } 4\pi \text{ steradians}}$$

and it ranges from 1.64 [9], [10] to 12 [3], [11] and [13].

D. SAW Tag Unit Gain (G_d)

The gain for the SAW tag unit is equivalent to the gain for the interrogator unit. The possible values of this gain range from 1.64 [9], [10] to 8 in [5].

E. Signal to Noise Ratio (γ)

The minimal detected signal can be specified via the signal-to-noise ratio (SNR). The ample variation of the SNR values reported by the authors range from 3 dB [11] to 50 dB [10]. It is in part due to the use of the coherent integration of the signals from multiple request cycles, pulse compression or other techniques which allow the safety operation of systems with such low values.

F. Bandwidth (B)

As a part of the thermal noise power (k T_0 B), where k is the Boltzmann's constant, and T_0 is the reference temperature in Kelvin degrees, the bandwidth B should be as narrow as possible. The rule of thumb says that its minimum value is given by

$$B = \frac{1}{\tau},$$

where τ is the pulse duration of the highest sine wave frequency component carried by the signal. The values of B reported by the authors in the literature range from 0.6 MHz for systems with the longest λ [9], [10] to 83.5 MHz for systems with the shortest λ.

G. Receiver Noise Figure (F)

The noise attributable to the receiver part of the interrogator unit is composed by different sources. It includes the flicker noise, thermal noise, and shot noise. The combined noise was reported to range from 3 dB [9], [10] to 5 dB in others sources.

H. Insertion Attenuation of the SAW sensor (D)

This parameter comprises all the attenuation a signal sustains in the SAW tag and depends mainly on the frequency and substrate material. The reported values went from 6.85 dB in [9] to 50 in [3], [13], and [14].

Table 1. Parameters reported by various authors. (1) This parameter was missed in the considered article.

Parameter	[3] - *	[9] - x	[10]-□	[11]-◊	[12]-Δ	[13]-o
λ	0.691 0.125	4.286	4.3	0.122	0.122	0.122
P_T	0.025	0.001	0.001	0.01	0.01	0.025
G_r	12	1.64	1.64	12	8.5	12
G_d	6	1.64	1.64	6	8	6
γ	10	6	50	3	15	10
B	10-36	0.6	0.6	10	83.5	40
F	5	3	3	5	$5^{(1)}$	$5^{(1)}$
D	50	6.85	13	40	45	50

In Fig. 2, we sketch the reader range obtained by using the reported values as a function of the reader power. The symbols represent each of the articles listed in Table 1. There is a bold line in this figure which represents the relationship between power and range which uses the values from [13], [15] but with varying power. It is plotted as a reference.

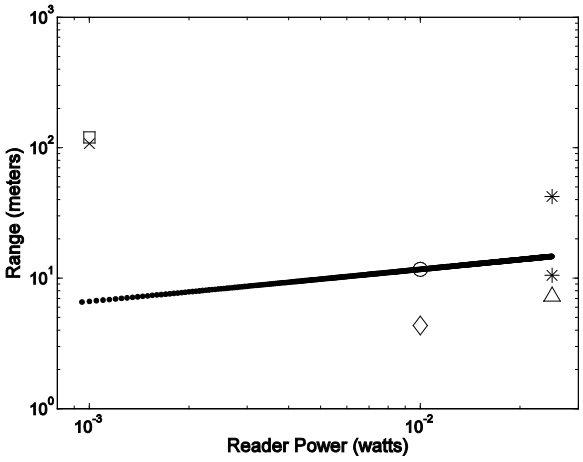


Fig. 2. Power vs. Range for examples presented in Table 1.

I. Conclusion

We have presented the derivation and analysis of the Radar Equation related to the SAW tag systems and compared it to several values taken from diverse sources. We show that using the radar equation developed one can predict the system range as a function of the system power. It is worth noticing the range obtained by [9] and [10] which is attributable to two parameters: the operational frequency and an extremely low loss by the tag device.

We continue work in this field and expect optimizing the SAW tag system performance in diverse environments. Specifically, we expect finding the system parameters which will guarantee the maximum possible range for the minimum possible interrogator power.

References

[1] V. Plessky and L. Reindl, "Review on SAW RFID tags" *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 57, no. 3, pp. 654–668, 2010.

[2] L. Reindl, A. Pohl, G. Scholl, and R. Weigel, "SAW-based radio sensor systems"*IEEE Sensor Journal*, vol. 1, no. 1, pp. 69–78, 2001.

[3] W. E. Bulst, G. Fischerauer, and L. Reindl, "State of the Art in Wireless Sensing with Surface Acoustic Waves" *IEEE Transactions on Industrial Electronics*, vol. 48, no. 2, pp. 265-271, 2001.

[4] L. M. Rodriguez, D. R. Gallagher, M. W. Gallagher, B. H. Fisher, J. R. Humphries, and D. C. Malocha, "Wireless SAW Sensor Temperature Extraction Precision" *IEEE Sensors Journal*, vol. 14, no. 11, pp. 3830-3837, 2014.

[5] F. Nawaz and V. Jeoti, "SAW sensor read range limitations and perspectives", *Wireless Networks*, vol. 20, no. 8, pp. 2581-2587, 2014.

[6] G. Cerda-Villafana and Y. S. Shmaliy, "Threshold-based identification of wireless SAW RFID-tags with pulse position encoding, " *Measurement*, vol. 44, no. 4, pp. 730-737, Apr. 2011.

[7] M. Penza, F. Antolini, and M. Vittori Antisari, "Carbon nanotubes as SAW chemical sensors materials", *Sensors and Actuators*, vol. 100, no. 1-2, pp. 47-59, 2004.

[8] S. J. Orfanidis, *Electromagnetic waves and Antennas*, 2008, available: http://eceweb1.rutgers.edu/~orfanidi/ewa/

[9] R. W. Brocato, *Passive Microwave Tags*, Sandia National Laboratories, 2004.

[10] R.W. Brocato G. A. Wouters, E. Heller, J. Blaich, and D. W. Palmer, "Re-configurable Completely Unpowered Wireless Sensors" in *Proc. 57th Electron. Comp. Techn. Conf. (ECTC'07)*, pp. 179-183, 2007.

[11] S. Scheiblhofer, C. Pfeffer, R. Feger, and A. Stelzer, "An S-FSCW based multi-channel reader system for beamforming applications using surface acoustic wave sensors" in *Proc. ICECom Conf.*, 2010, pp. 1-4.

[12] F. Schmidt, O. Sczesny, C.C.W. Ruppel, and V. Magori, "Wireless interrogator system for SAW-identification-marks and SAW-sensor components" in *Proc. 50th IEEE Int. Freq. Contr. Symp.*, 1996, pp. 208-215.

[13] Y.S. Shmaliy, V. Plessky, G. Cerda-Villafana, and O. Ibarra-Manzano, "Error probability for RFID SAW tags with pulse position coding and peak-pulse detection", *IEEE Trans. on Ultrason., Ferroel. Freq. Contr.*, vol. 59, no. 11, pp. 2528-2536, 2012.

[14] G. Cerda-Villafana, O. Ibarra-Manzano, Y.S. Shmaliy, and V. Plessky, "Peak-pulse detection error probability for RFID SAW-tags with pulse position coding, " in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Process. (ICASSP 2012)*, Kyoto, Japan, March 25-30, 2012, pp. 1697-1700.

[15] G. Cerda-Vallafaña, Y. S. Shmaliy, "Range vs. Error probability relation for passive wireless SAW tags", in *Recent Advances in Circuits, Systems, Telecom. and Control. Proc. WSEAS Multi Conf.*, Paris, France, October 29-31, 2013, pp. 62-66.

Evolving Optimal Digital Circuits Using Cartesian Genetic Programming With Solution Repair Methods

Spyros A. Kazarlis, John Kalomiros, Anastasios Balouktsis and Vassilios Kalaitzis

***Abstract*—**In this work, a new approach is presented for automatically evolving optimal digital circuits using a specific implementation of Evolutionary Algorithms called Cartesian Genetic Programming (CGP). This CGP approach uses an MxN grid of forward interconnected logic gates that is genetically evolved and evaluated through a custom digital circuit emulator developed especially for this purpose. The CGP approach uses a special repairing technique for transforming invalid evolved circuits into valid ones. The CGP algorithm is tested on the evolution of a number of different digital circuits, used as benchmarks. The simulation results are promising and show the effectiveness of the proposed technique.

***Keywords*—**Cartesian Genetic Programming, Digital Circuits, Evolvable Hardware, Digital Circuit Emulation.

I. Introduction

EVOLUTIONARY Algorithms (EAs) [1], [2], are stochastic global optimization algorithms inspired from the principles of natural evolution, and have already been established as powerful optimization tools for solving real world optimization problems. One of EAs greater paradigms is Genetic Programming (GP) [3], [6], a technique initially conceived for evolving optimal forms of software, using tree encodings. Later GP has been successfully applied to other real world problems where tree encodings could describe their potential solutions, like the design of analog and digital circuits [4], [5]. Moreover, the evolution of simulated circuits can be seen as an extrinsic implementation of another EC sub-domain called Evolvable Hardware (EH) [7], [8].

Miller and Thomson, proposed a variation of GP called Cartesian Genetic Programming (CGP) [9] where circuits are encoded as directed graphs instead of trees. In CGP implementations on digital circuits [20], [21] usually an MxN grid is constructed where each node represents a specific gate. Nodes are interconnected with each other, but usually only forward connections are allowed, of a certain depth "d", in order to avoid feedbacks.

In order to apply such a GP, CGP or generally EC algorithm, a Fitness Function [10] has to be constructed that will evaluate each genetically produced solution and provide a measure of its quality or fitness. Through this fitness function the EC algorithm can sample the quality of points in the search space and guide its way through the fitness landscape and towards the optimal solution. For this reason, a special Digital Circuit Emulator [11] has been constructed to be especially used in this work as a fitness function for the CGP evolutionary algorithm.

Genetically produced solutions can be valid or invalid, according to the specific problem constraints, and the encoding used. Invalid solutions form forbidden areas through the whole fitness landscape that complicate the search procedure and makes the optimization task more difficult [12], [13]. In order to overcome this, a special "repair" mechanism [14] is proposed for the evolved digital circuits in the form of MxN grids that do not correspond to valid solutions, and which could be fed to the Digital Circuit Emulator for evaluation and produce erratic results. This "repair" technique transforms the invalid solutions into their nearest valid ones and thus it smoothens the fitness landscape, by eliminating the forbidden areas of invalid solutions. Thus, from the EAs point of view the whole landscape is valid and searchable.

The proposed CGP implementation, that features the "repair" technique and uses the custom Digital Circuit Emulator, is applied on a number of six (6) benchmark problems. In each problem it is asked to evolve a well-known elementary digital circuit given a specific MxN grid of gates. The optimization goal is twofold: a) to evolve a digital circuit that will match the desired truth table, and b) to achieve this with the least possible number of gates. Thus, this problem is defined as a Multi-Objective optimization problem [15].

In order to cope with the two objectives, a consolidated

This work has been co-financed by the European Union (European Social Fund – ESF) and Greek national funds through the Operational Program "Education and Lifelong Learning" of the National Strategic Reference Framework (NSRF) - Research Funding Program: ARCHIMEDES III. Investing in knowledge society through the European Social Fund..

S. A. Kazarlis is with the Technological and Educational Institute of Central Macedonia, Serres, Greece, Dept. of Informatics Engineering, Terma Magnesias St., 62124, Serres, Greece, (phone: +30-23210-49343; fax: +30-23210-49128; e-mail: kazarlis@teicm.gr).

J. Kalomiros is also with the Technological and Educational Institute of Central Macedonia, Serres, Greece, Dept. of Informatics Engineering, Terma Magnesias St., 62124, Serres, Greece, (e-mail: ikalom@teicm.gr).

A. Balouktsis is also with the Technological and Educational Institute of Central Macedonia, Serres, Greece, Dept. of Informatics Engineering, Terma Magnesias St., 62124, Serres, Greece, (e-mail: tasosb@teicm.gr).

V. Kalaitzis is also with the Technological and Educational Institute of Central Macedonia, Serres, Greece, Dept. of Informatics Engineering, Terma Magnesias St., 62124, Serres, Greece, (e-mail: kalaitz@teicm.gr).

fitness function is formed that contains terms for both objectives. Thus, by optimizing the consolidated fitness function, the EA optimizes both objectives simultaneously.

Extensive simulation runs have been conducted in order to test the efficiency of the proposed CGP algorithm. Many different parameter values have been tested in order to find the optimal parameter set that provides the best performance for the CGP algorithm.

The organization of the paper is as follows: in Section II the Cartesian Genetic Programming implementation is described. The digital circuit simulator is described in Section III. Section IV presents the repair method used for transforming invalid solutions to valid ones. Section V describes the test set used for testing the CGP scheme. The simulation results are presented in Section VI. Conclusions and possible future work are discussed in Section VII.

II. The Cartesian Genetic Programming Implementation

A. The Cartesian Grid Structure

In order to encode digital circuits in a standard form that can be easily transformed into a suitable genotype for genetic optimization, the CGP method is adopted. According to this method a MxN grid is formed that represents a potential digital circuit. Each grid node represents a digital gate whose Boolean function is genetically chosen from a set of available Boolean functions or logic gates. Grid nodes can be genetically interconnected with each other in an arbitrary manner. However, in this implementation, only forward connections are allowed of depth 1, which means that a node at column i can only connect to another node in column i+1. The form of a 3x3 such grid is depicted in fig. 1

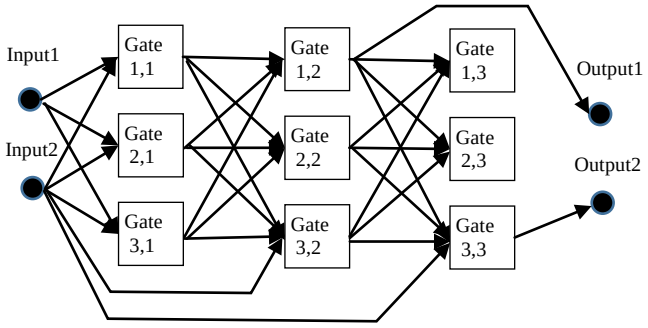


Fig. 1. A 3x3 Cartesian Grid for digital circuit evolution

As can be seen in Fig. 1 circuit inputs are allowed to be connected not only to the gates of the first column, but can be arbitrarily connected to any gate in the grid. Moreover, the circuit's outputs can be taken not only from the nodes of the last column but from any node in the grid. Specifically for our implementation we have adopted the following rules:

Rule 1: Each node in the grid can be configured to represent one of the following functions: 0. Non-existent, 1. AND, 2. OR, 3. NOT, 4. NAND, 5. NOR, 6, XOR, 7, XNOR. When a node is configured as "non-existent", it means that it does not implement a logic gate and it is left unused. This allows for a circuit to be implemented with less gates than the total number of gates in the grid (less than MxN). Thus, the Evolutionary Algorithm can also optimize the number of gates used to achieve a specific truth table.

Rule 2: All implemented gates have two inputs and one output except for the NOT gate that has a single input.

Rule 3: The inputs of each gate can be connected either to an input of the whole circuit (circuit input) or to an output of a gate located in the previous column. Also a gate's input can be connected to a constant logical 1, or a logical 0. It is clear that gates of column 1 have their inputs connected only to the circuit's inputs, or to logical 1's or 0's. It is also clear that even gates of the last column can have inputs driven directly from the circuit's inputs.

Rule 4: The outputs of the circuit can be drawn from the output of any gate in the grid and not only the nodes of the last (right most) column. This gives an extra degree of freedom to the optimization algorithm to evolve the optimal circuit.

B. The genotype encoding scheme

In order for the Evolutionary Algorithm to be applied, an encoding scheme should be adopted in order to encode potential solutions into strings of symbols that resemble DNA. For this reason the following encoding scheme is implemented: each genotype consists of a number of chromosomes equal to the total number of the grid nodes. Thus for a 3x3 grid the number of chromosomes would be 9.

Each chromosome consists of 4 sections:

Section 1 (gate type): this section consists of 3 bits that are adequate to encode the 8 different node configurations mentioned earlier, including the "non-existent" case.

Section 2 (gate input 1): this section encodes the source of the gate's first input and consists of a number of bits equal to:

$$\text{ceil}\left(\log_2\left(\text{NoOfGridRows}+\text{NoOfCircuitInputs}+2\right)\right) \quad (1)$$

where ceil() is a function that returns the smallest integer greater or equal than its argument, NoOfGridRows is the number of rows in the grid, NoOfCircuitInputs is the number of inputs of the circuit and the "+2" term is for including the cases of logical 1 and 0.

Section 3 (gate input 2): this section encodes the source of the gate's second input and consists of a number of bits equal to those of section 2.

Section 4 (gate output): this section encodes the nature of the gate's output, i.e. if it enrolls as an output of the circuit and which output, and consists of a number of bits equal to:

$$\text{ceil}\left(\log_2\left(\text{NoOfOutputs}+1\right)\right) \quad (2)$$

For example, for a 3x3 grid employed to find a circuit with 3 inputs and 2 outputs (like the full-adder) each chromosome consists of 3 gate-type bits, 4 gate-input-1 bits, 4 gate-input-2 bits, and 2 gate-output bits, i.e. a total of 13 bits. Thus the whole genotype has a length of 9x13=117 bits.

The genotype encoding scheme can be seen in Fig. 2.

ISBN: 978-1-61804-290-3

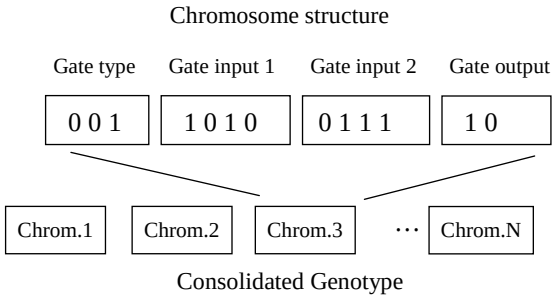


Fig.2. The genotype encoding scheme

C. The Evolutionary Algorithm

For the evolutionary discovery of optimal digital circuits using the Cartesian grid and the solution encoding described earlier, a Genetic Algorithm [16], [17] was used with the following parameters: a constant population of 500 or 1000 genotypes, randomly initialized at the beginning, roulette wheel parent selection [16], uniform crossover [18], binary mutation with a low per-bit probability, replacement of all parents with offspring at each generation, and a generation limit of 10,000 generations. Moreover, we have used the elitism mechanism [16], and an automatic scheme for adaptive operator probabilities described in [19], that adapts the probabilities of the crossover and mutation operators throughout the optimization run, according to the degree of population diversity and convergence.

III. The Digital Circuit Simulator

A. The Simulator

In order for the GA to work, a fitness function needs to be defined in order to provide quality metrics for every genetically produced solution. For the evolution of digital circuits, the fitness function should be either a digital circuit simulator or a platform for real implementation and testing of every proposed circuit, possibly on an FPGA device. In this work, the first solution was chosen.

The digital circuit simulator used in this work was described in [11]. This simulator uses a special alphanumeric encoding for the description of the digital circuit that is passed to the simulator as an input. This encoding can describe any digital circuit up to a given number of gates. The simulator also accepts an array of binary input vectors that must be tested on the given circuit and produce the corresponding binary output vectors. This array usually contains all possible combinations of the input values for all circuit inputs. For example, for the full adder circuit that has 3 inputs, this input array contains 8 lines of 3 bits each, containing all possible combinations of the values of the three inputs.

When invoked, the simulator first parses the input string that describes the circuit and builds an internal array structure with the circuit topology and parameters. Then it loops over all input vectors given and simulates the circuit's response for each such vector, producing and recording its output. The simulation is performed using a second internal array that keeps the state of binary signals for each gate and interconnection of the circuit. The simulation is based on a discrete-time simulation technique that is inspired by the principles of the propagation of digital signals through the logic gates at discrete time quanta. This signal propagation is simulated step by step, using these discrete time quanta, that effectively are iterations of an internal simulation loop. These time steps play the role of consequent periods of an informal internal state clock. The period of this ideal clock coincides with the delay of a single gate, whilst the signal propagation through the interconnections is considered to occur instantaneously.

Finally, after all input vectors are presented to the circuit and all corresponding output vectors have been calculated, the complete array of output vectors is returned to the Genetic Algorithm as the outcome of the simulation.

B. The GA Fitness Function

After receiving the output array from the simulator, the GA fitness function compares this array of outputs to the array of desired outputs for the evolving circuit. From this comparison a hamming distance metric is calculated as an integer that counts the number of bits that differ between these two arrays. This hamming distance has to be minimized by the GA and constitutes the primary goal of the optimization process. However there is also another optimization goal: to minimize the number of gates of the circuit. Thus, the problem is classified as a multi-objective optimization problem [15]. In order to cope with these two optimization goals, the most common solution is to build a consolidated optimization fitness function containing both optimization terms in the following form:

$$\text{Fitness}(S) = w_1 \times F_1(S) + w_2 \times F_2(S) \quad (3)$$

$$F_1(S) = \text{HammingDistance}(O_s, D) \quad (4)$$

$$F_2(S) = \text{NoOfGates}(S) \quad (5)$$

where S is a solution (circuit) under evaluation, $F_1(S)$ and $F_2(S)$ are the two individual optimization functions, w_1 and w_2 are weight factors, O_s is the output array for solution S, and D is the array of desired outputs, according to the desired truth table for the evolving circuit.

Preliminary runs have shown that even with the w_1 weight selected to be significantly larger than w_2, the GA tends to converge into a solution that does not exhibit zero hamming distance between the desired and real outputs, but has a minimal number of gates that in many cases are even less than the gates needed in original well-known designs of the circuit. When this happens it is observed that the GA locks itself in these local optima and it is thereafter extremely difficult to bypass these local optima and find the global optimum solution. These sub-optimal solutions, with minimal number of gates but nonzero hamming distance are of course unacceptable.

In order to make the GA to avoid locking itself in such local optima, another form of fitness function is used that firstly qualifies each proposed circuit only by $F_1(S)$, and if and when

ISBN: 978-1-61804-290-3

this objective function reaches the value of 0, then the fitness function changes and thereafter starts to qualify each successive solution with a consolidate function similar to (3), with the values of 100 and 1 for the w_1 and w_2 weights respectively.

Thus the optimization is performed in two stages: in the first stage the GA tries to find a digital circuit that will satisfy the complete truth table, disregarding of the size of the circuit (i.e. the number of gates), and in the second stage, where it has already satisfied the truth table, it tries to minimize the size of the circuit.

The new fitness function can be described in an algorithmic form as follows:

If ($F_1(S) > 0$) Fitness(S) = 100 x $F_1(S)$

Else Fitness(S) = 100 x $F_1(S)$ + 1 x $F_2(S)$ (6)

Of course in the else clause the term 100 x $F_1(S)$ can be omitted as it values to 0. This fitness function has proved to give much better optimization results over all test cases.

IV. The "Solution Repair" Method

Using the encoding method proposed in Section II a fitness landscape is formed that contains both feasible and unfeasible solutions. For example, if a specific node is genetically selected to be "non-existent" then if there exist any connections from other nodes to the non-existent node, then this represents an unfeasible solution.

Although the GA does not feature an embedded mechanism to handle unfeasible solutions, there are generally a lot of methods proposed in the literature that can handle such constraints [12], [13], [14], like the well-known penalty methods [10]. However, in this work many of these methods cannot be implemented because when a solution is unfeasible, then the circuit cannot even be passed to the simulator and cannot even be tested, in order to produce an acceptable fitness value.

For this reason, a special repair mechanism is implemented that transforms invalid circuits to valid ones and then sends them to the simulator for evaluation. This repair mechanism works as follows:

Step1: Firstly, it decodes the genetically produced genotype to its phenotypic vector of values, comprising the gate type, the connections for input 1 and 2 of the gate and whether its output coincides with a circuit output.

Step 2: It checks whether all phenotypic values are within nominal ranges. This is essential because if, for example, a parameter takes 6 values (0..5) and this parameter is encoded with 3 digital bits, then there exist 6 valid values (0..5) and two invalid values (6,7) that could be encoded. If an invalid value is found it is transformed to a random valid one.

Step 3: It checks for the existence of NOT gates within the grid. If such a gate exists then its second input is invalidated as NOT gates have a single input.

Step 4: It ensures that all gates of the first column (the left-most column) take their inputs either from the circuit inputs or the constant values of logic 0 or 1, and not from the gates of the previous column that simply does not exist.

Step 5: It checks if there are any "non-existent" gates within the grid and if other gates are interconnected to them. If so, it rearranges those connections so that they connect to valid gates only.

Step 6: It also checks if there exists a whole column of invalid gates. This is an abnormal situation where due to the interconnection limitations, the resulting circuit is split in two unconnected parts. To avoid this, the repair mechanism ensures that each grid column has at least one valid gate, to ensure forward connectivity.

Step 7: Finally all repair transformations are re-encoded into the original genotype and the genotype is injected back to the population after it has been "repaired".

This repair technique that transforms all invalid solutions to valid ones is beneficial for the GA, as the latter moves in a landscape with solutions that are effectively all feasible, a fact that facilitates the search.

V. The Simulation Test Set

For testing the CGP implementation proposed in this work, a set of six (6) elementary and well known digital circuits of increasing complexity and number of gates have been employed. The complete test set is shown in Table I.

TABLE I – The Complete Test Set

Circuit	No of 2-Input Gates	No of Inputs	No of Input Combinations	No of Outputs	No of Output Bits
Half Adder	2	2	4	2	8
Decoder 2 to 4	6	2	4	4	16
Full Adder	5	3	8	2	16
2-bit Multiplier	8	4	16	4	64
Decoder 3 to 8	19	3	8	8	64
2-bit Comparator	15	4	16	3	48

In Table I, the "No of Output Bits" column contains the product of the "No of Input Combinations" and the "No of Outputs", thus expressing the size of the output vector that has to match the desired one from the circuit's truth table.

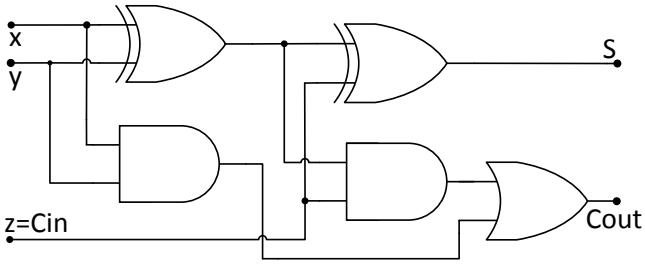


Fig. 3. The Full Adder circuit used in the test set

The typical synthesis of such circuits as well as their complete truth tables are well described in the literature [22].

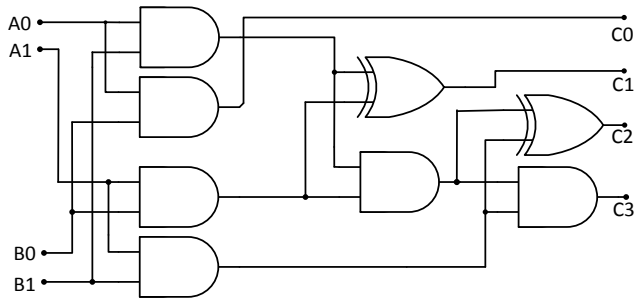


Fig. 4. The 2-bit multiplier used in the test set

For better comprehension the pictures of a "Full Adder", a "2-bit Multiplier" and a "2-bit Comparator" circuits are depicted in Fig. 3, 4 and 5 respectively.

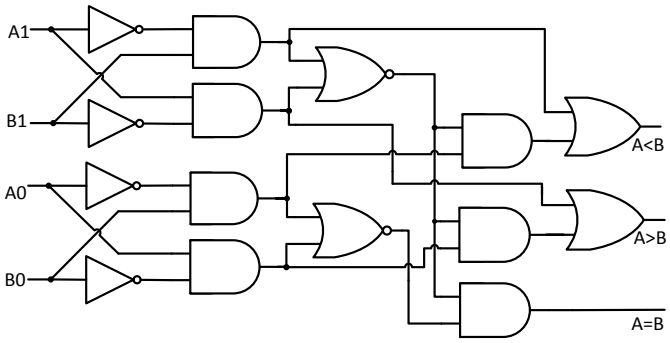


Fig. 5. The 2-bit comparator used in the test set

VI. Simulation Results

Simulation results were carried out with the same set of GA parameters for all cases in the test set. The only exception to this rule was the GA population size, which varied among two distinct values: 500 and 1000 genotypes at each generation. So two tests have been conducted for each case: one with a population of 500, and one with a population of 1000 genotypes. The complete set of GA parameters is shown in Table II.

TABLE II – The GA Parameters

GA Parameter	Value	GA Parameter	Value
Population	500 or 1000	Crossover Probability	0.4 to 0.9
Selection	Roulette Wheel	Mutation Probability	0.001 to 0.1
Crossover	Uniform	Elitism	Yes
Mutation	Binary Mutation	Population Replacement	Whole Population
Operator Probabilities	Automatically adapted	Termination	10,000 generations

As Evolutionary Algorithms are stochastic algorithms, and in order to avoid statistical errors, twenty (20) runs have been made for each circuit case and each population size. Thus for each test case several statistical figures have been calculated in order to judge the performance of the proposed implementation.

The results for the population size of 500 genotypes are shown in Table III, while corresponding results for the population value of 1000 are shown in Table IV. A test case is considered successful if it finds a solution that satisfies the circuit's truth table that is when the hamming distance between the output vector of the genetically produced solution and the desired one is zero (0).

For all test cases a 5x5 Grid has been used, allowing for a maximum of 25 gates per circuit.

The results were obtained on an Intel Core-i7 workstation with 8GB RAM, running Windows 8.1, and the software was developed using native C++.

TABLE III – Simulation Results for Population 500

Circuit	Success Rate	Avg gates on success	Min gates on success	Max gates on success	Avg gener. to find optimum	Avg exec. time per task
Half Adder	**100%**	2,3	2	3	710	5 min
Decoder 2 to 4	**100%**	5,1	4	7	850	6 min
Full Adder	**100%**	6	5	7	890	8 min
2-bit Multiplier	**30%**	12	11	13	5500	36 min
Decoder 3 to 8	**20%**	17,5	17	19	6150	30 min
2-bit Comparator	**10%**	14	14	14	8450	48 min

TABLE IV – Simulation Results for Population 1000

Circuit	Success Rate	Avg gates on success	Min gates on success	Max gates on success	Avg gener. to find optimum	Avg exec. time per task
Half Adder	**100%**	2,7	2	4	380	9 min
Decoder 2 to 4	**100%**	5,1	4	6	640	12 min
Full Adder	**100%**	6,5	5	10	650	15 min
2-bit Multiplier	**80%**	9,9	8	13	4300	64 min
Decoder 3 to 8	**40%**	15,9	14	19	5230	60 min
2-bit Comparator	**20%**	11,8	10	14	6570	85 min

As can be seen from Tables III, and IV, the runs with a population of 1000 genotypes clearly outperform the ones with a population of 500. This was expected as the runs with population 1000 perform twice as much fitness evaluations as compared to the runs with population 500. The runs with population 1000 sample the search space using double points at each generation than those with population 500, resulting thus in a much more efficient search procedure.

As can also be seen from the two Tables, the problems of the Half Adder, The Decoder 2-to-4 and the Full Adder have a 100% success with both population 500 and 1000. However, as can be seen from Tables III and IV the tasks with

ISBN: 978-1-61804-290-3

population 1000 find the optimum in less generations on average.

The success rate seems to drop as the complexity of the circuits increases, and seems to depend not only on the size of the output vector (total number of output bits) but also on the internal complexity of the circuit. For example in Table IV, the Decoder 3-to-8 case exhibits half the success rate of the 2-bit multiplier case, although they have the same output vector size (64 bits). This may be attributed to the fact that the decoder 3-to-8 needs at least 14 gates compared to the 8 gates needed by the 2-bit multiplier. Also the case of the 2-bit comparator exhibits half the success rate of the decoder 3-to-8, although it features a 48 bit output vector compared to the 64-bit vector of the decoder, and although it needs only 10 gates compared to the 14 of the decoder.

It is also worth mentioning that in most cases the CGP algorithm has produced many unconventional solutions with the same minimal number of gates as the conventional ones, and in some cases even less. In the case of the decoder 3-to-8 the CGP scheme has found solutions with only 14 gates, compared to the 19 gates of a typical design [22]. Moreover, in the case of the 2-bit comparator, the CGP scheme finds solutions with a minimum number of 10 gates, and a maximum number of 14 gates, while the typical design shown in Fig. 5, that can be found in [22], uses a total number of 15 gates. The analytical study of these produced solutions could be included in future work. The ability of the CGP scheme to discover unconventional solutions with the same or even lower number of gates than the typical designs, is justifying its characterization as an “invention machine”.

VII. Conclusions and Future Work

In this work a Cartesian Genetic Programming implementation is proposed for discovering optimal digital circuits. The implementation uses a forward connected MxN grid to represent digital circuits. This grid is encoded in bit-string genotypes in order for the GA to be applied. The GA uses a digital circuit simulator as a fitness function for the evaluation of the produced solutions. The fitness function also uses some repair techniques to transform invalid solutions to valid ones, and make the whole fitness landscape feasible for the GA. The proposed implementation is tested on a set of 6 elementary and well-known digital circuits of increasing complexity. The simulation results are promising. The CGP scheme manages to find the best solution for each benchmark case, although the probability for success diminishes as the complexity of the circuit increases. A number of very interesting unconventional solutions with even less gates than typical designs are produced, that, together with several interesting sub-optimal solutions, are worth studying. Finally, in a possible future work, the proposed CGP scheme could be tested using some problem oriented reproductive or mutation operators. It could also be tested against more complex circuits.

References

[1] Back, T., Fogel, D., Michalewicz, Z., Handbook of Evolutionary Computation, Oxford Univ. Press, 1997.

[2] Holland, J. H., Adaptation in Natural and Artificial Systems, The University of Michigan Press, Ann Arbor, 1975.

[3] Koza, J.R.: Genetic Programming: On the Programming of Computers by Means of Natural Selection. MIT Press, Cambridge (1992)

[4] Koza, J. R., Bennett III, F. H., Andre, D., Keane, M. A., and Dunlap, F. "Automated synthesis of analog electrical circuits by means of genetic programming." Evolutionary Computation, IEEE Transactions on 1.2 (1997): 109-128.

[5] Miller J.F., Job D., Vassilev V.K., “Principles in the Evolutionary Design of Digital Circuits – Part I,” Genetic Programming and Evolvable Machines 1(1), 8-35, (2000), Kluwer Academic Publishers

[6] Banzhaf, W., Nordin, P., Keller, R. E., and Francone, F. D., Genetic programming: an introduction (Vol. 1). San Francisco: Morgan Kaufmann, 1998.

[7] Higuchi, Tetsuya, et al. "Evolvable hardware with genetic learning." Circuits and Systems, 1996. ISCAS'96., Connecting the World., 1996 IEEE International Symposium on. Vol. 4. IEEE, 1996.

[8] Higuchi, Tetsuya, and Xin Yao. Evolvable hardware. Vol. 11. Springer Science & Business Media, 2006.

[9] Miller, J.F., Thomson, P., “Cartesian Genetic Programming,” in: Poli, R., Banzhaf, W., Langdon, W.B., Miller, J., Nordin, P., Fogarty, T.C. (eds.) EuroGP 2000, LNCS, vol. 1802, pp. 121–132. Springer, Heidelberg (2000)

[10] Kazarlis S. and Petridis V., “Varying Fitness Functions in Genetic Algorithms: Studying the Rate of Increase of the Dynamic Penalty Terms,” Proceedings of the 5th International Conference on Parallel Problem Solving from Nature (PPSN-V), Amsterdam, 27-30 September 1998, pp. 211-220.

[11] S. Kazarlis, J. Kalomiros, P. Mastorocostas, V. Petridis, A. Balouktsis, V. Kalaitzis, A. Valais, “A Method for Simulating Digital Circuits for Evolutionary Optimization,” Proceedings of the 10th Annual International Joint Conferences on Computer, Information, and Systems Sciences, and Engineering (CISSE 2014).

[12] Michalewicz, Zbigniew. "A Survey of Constraint Handling Techniques in Evolutionary Computation Methods." Evolutionary Programming 4 (1995): 135-155.

[13] Spyros A. Kazarlis, “Constraint Handling Methods in Genetic Algorithms”, Proceedings of the 11th Panhellenic Conference in Informatics (PCI 2007), May 18-20, 2007, University of Patras, Patras, Greece, Vol. A, pp. 591-606.

[14] Orvosh D. and Davis L.(1994), Using a Genetic Algorithm to Optimize Problems with Feasibility Constraints in Proceedings of the 1st IEEE Conference on Evolutionary Computation. Piscataway, NJ: IEEE Press, pp. 548-553.

[15] Deb, Kalyanmoy. Multi-objective optimization using evolutionary algorithms. Vol. 16. John Wiley & Sons, 2001.

[16] Davis, Lawrence, ed. Handbook of genetic algorithms. Vol. 115. New York: Van Nostrand Reinhold, 1991.

[17] Golberg, David E. "Genetic algorithms in search, optimization, and machine learning." Addion wesley 1989 (1989).

[18] Sywerda, Gilbert. "Uniform crossover in genetic algorithms." Proceedings of the third international conference on Genetic algorithms. Morgan Kaufmann Publishers Inc., 1989.

[19] V. Petridis and S. Kazarlis, “Varying Quality Function in Genetic Algorithms and the Cutting Problem,” Proceedings of the First IEEE Conference on Evolutionary Computation, IEEE Service Center, 1994, Vol. 1, pp. 166-169.

[20] Miller, Julian F. Cartesian genetic programming. Springer Berlin Heidelberg, 2011.

[21] Miller, Julian F., and Stephen L. Smith. "Redundancy and computational efficiency in cartesian genetic programming." Evolutionary Computation, IEEE Transactions on 10.2 (2006): 167-174.

[22] Morris Mano M., Ciletti M.D., Digital Design, 5th Edition, Prentice Hall, 2013.

New Speech Enhancement Method based on Wavelet Transform and Tracking of Non Stationary Noise Algorithm

Riadh AJGOU[(1,2)], Salim SBAA[(1)], Said GHENDIR[(1,2)], Ali CHEMSA[(1,2)] and A. TALEB-AHMED[(3)]

***Abstract*—**In this work, we have developed an efficient approach for enhancing speech by combining tracking of non stationary noise algorithm and Continues Wavelet Transform (CWT). Tracking of non stationary noise method that is based on data-driven recursive noise power estimation was proposed by Jan S. Erkelens and Richard Heusdens. The Continues Wavelet decomposition of speech signal uses adaptive level with Harr mother wavelet. In this paper, our novel method was evaluated in presence of different kind of noise using the NOIZEUS noisy speech corpus developed in Hu and Loizou laboratory that is suitable for evaluation of speech enhancement algorithms. The noisy database contains 30 IEEE sentences (produced by three male and three female speakers) corrupted by eight different real -world noises at different SNRs. The noise was taken from the AURORA database and includes suburban train noise, babble, car, exhibition hall , restaurant , street , airport and train-station noise. For evaluating the performance of speech enhancement methods we have used Perceptual Evaluation of Speech Quality scores (PESQ, ITU-T P.862). Simulation results demonstrate that the proposed approach offers an improved performance of speech enhancement in comparison with state-of-the-art methods in terms of PESQ measure.

***Keywords*—** Speech enhancement, Tracking of non stationary noise method, Wavelet Transform; PESQ.

I. INTRODUCTION

The paper addressed the problem of suppressing the background noise in noisy speech. Speech signal can be corrupted by noise in various situations, such as trains, cars, airport, babble, factory, street ..etc. The problem of enhancing speech degraded by the noise is largely open to research, although many significant techniques have been introduced over the past decade because there are many areas where it is necessary to enhance the quality of speech that has been degraded by background noise. Some of these areas include automobile interiors for hands free cellular, aircraft cockpits, voice communications using mobile telephone, automatic speech recognition (ASR) and speech coders[1]. speech enhancement has become more important than ever before. A speech enhancement system helps in increasing the quality of noisy speech [2].

We propose a novel approach to improve the performance of speech enhancement systems by combining tracking of non stationary noise algorithm [3] and De-noising Speech Signals by Wavelet Transform [4].

The problem of de-noising consists of removing noise from corrupted signal without altering it. Thus, we have evaluated our approach by evaluating speech quality. Reconstructed speech quality is measured with Perceptual Evaluation of Speech Quality (PESQ) score [5]. The PESQ measure was not generally intended to assess speech enhancement algorithms. However, it has been used in the past years in several speech enhancement algorithms. It converts the disturbance parameters in speech to a MOS-like listening quality score in a very wide range of conditions that may include codec distortions, errors, filtering, and variable signal delay. The higher score means better perceptual speech quality [6].

The simulation results show that the proposed speech enhancement method provide better speech quality compared to the traditional state-of-the-art methods using PESQ evaluation method.

In this paper various methods for speech enhancement methods have been introduced.

II. STATE-OF-THE ART OF SPEECH ENHANCEMENT ALGORITHMS

In this section we introduce seven of the most famous speech enhancement methods.

A. Tracking of Non-stationary Noise Based on Data-Driven Recursive Noise Power Estimation

We have to describe this method that was proposed by Jan S. Erkelens and Richard Heusdens [3]. The authors considers estimation of the noise spectral variance from speech signals contaminated by highly non-stationary noise sources. The method can accurately track fast changes in noise power level (up to about 10 dB/s). The enhancement algorithm is based on the minimum mean-square error (MMSE) [7]-[8] estimation in

Riadh AJGOU, Salim SBAA, Said GHENDIR and Ali CHEMSA Authors are with:

(1)LI3CUB Laboratory, Electric engineering department, University of Biskra . B.P 145 R.P, 07000 Biskra ALGERIA. Email: ajgou2007@yahoo.fr/ riadh-ajgou@univ-eloued.dz,. s.sbaa@univ-biskra.dz, said-ghendir@univ-eloued.dz, chemsadoct@yahoo.fr

(2) Department of sciences and technologyEl-oued UniversityPO Box 789 39000 El-oued ALGERIA

(3) A. TALEB-AHMED Author is with LAMIH Laboratory University of UVHC Mont Houy - 59313 Valenciennes Cedex 9 FRANCE. Email: abdelmalik.taleb-ahmed@univ-valenciennes.fr

the DFT (Discrete Fourier Transform) domain of speech spectral amplitudes. MMSE estimation of the noise power is to update the noise spectrum estimates with a reduced risk of speech leakage [3]. The MMSE estimates are obtained with the standard method of multiplying the noisy powers by a spectral gain function. This removes most of the speech contribution from the noisy spectrum, allowing for fast and accurate tracking of changing noise levels

1) Prior SNR Estimator $\hat{\xi}_{SE}$ for Speech Enhancement

For speech estimation, "decision-directed" estimator was used[3]:

$$\hat{\xi}_{SE}(k,m)=\max\left[\begin{array}{l}\alpha_{SE}\dfrac{\hat{A}^2(k,m-1)}{\hat{\lambda}_D(k,m)}+\\ (1-\alpha_{SE})\left[\dfrac{R^2(k,m)}{\hat{\lambda}_D(k,m)}-1\right],\xi_{\min}\end{array}\right] \quad (1)$$

Where: k is frequency index in signal

m is frame index.

$\hat{\lambda}_D$ is the noise variance.

$\hat{A}^2$ is the speech power estimate.

R^2 is the noisy power.

α_{SE} is speech enhancement factor between 0 and 1.

$\xi_{\min}$ is a small value larger than 0 in [dB].

Where the latest available estimate of the noise variance $\hat{\lambda}_D(k,m)$ was used[3]:

$$\hat{\lambda}_D(k,m)=\alpha_s(k,m)\hat{\lambda}_D(k,m-1)+(1-\alpha_s(k,m))\hat{D}^2(k,m) \quad (2)$$

Where: $\hat{D}^2(k,m)$ is the noise power.

$\alpha_s(k,m)$ is the smoothing parameter (equation 8).

Note that the speech power $\hat{A}^2$ estimate is used in the first term instead of the square of the amplitude estimate $\hat{A}^2$ (the standard definition) The standard "decision-directed" estimator is the most commonly used estimator of prior SNR[3]:

$$\hat{\xi}(k,m)=\max\left[\begin{array}{l}\alpha\dfrac{\hat{A}^2(k,m-1)}{\hat{\lambda}_D(k,m)}+\\ (1-\alpha)\left[\dfrac{R^2(k,m)}{\hat{\lambda}_D(k,m)}-1\right],\xi_{\min}\end{array}\right] \quad (3)$$

An advantage of the alternative definition in equation (1) is that the estimate of prior SNR does not depend on the final amplitude estimate used for speech reconstruction. This prevents the prior SNR estimator from changing its behavior when another estimator for the speech amplitude is preferred, for example the log-spectral amplitude estimator [3], or any other perceptually relevant amplitude estimator [3]. Another advantage of using $\hat{A}^2$ is that it reduces a bias that leads to the underestimation of prior SNR when α_{SE} is near 1and the SNR is low [9]. An experimental comparison done in [3] with the standard definition showed that, for parameter settings for which both definitions have the same tradeoff between noise reduction and speech distortion, the definition of equation (1) leaded to less musical noise [3].

2) Amplitude Gain Functions: The gain functions for $\hat{A}$ and $\hat{A}^2$ are based on a generalized-Gamma speech amplitude prior [3]. The generalized-Gamma prior is given by:

$$f_A(a)=\frac{\gamma\beta^{\nu}}{\Gamma(\nu)}a^{\gamma\nu-1}\exp\left(-\beta a^{\gamma}\right),\beta>0,\gamma>0,\nu>0,a\geq 0 \quad (4)$$

where $\Gamma(.)$ is the gamma function, and β depends on γ, ν and λ_s. The random variable A represents the DFT magnitude. The MMSE gain functions for $\gamma=1,\nu=1$ and for which the expressions can be found in [8]. For these parameter values, we have: $\beta=\sqrt{2/\lambda_s}$

(λ_s is the speech spectral variance that is the expectation of the speech power $\hat{A}$).

3) Noise Tracking

The steps taken in the noise tracking algorithm:

First, the prior SNR parameter $\hat{\xi}_{NT}(k,m)$ and the posterior SNR $\hat{\zeta}(k,m)$ are estimated, using the latest available noise variance estimate $\hat{\lambda}_D(k,m)$ [3]:

$$\hat{\xi}_{NT}(k,m)=\max\left[\begin{array}{l}\alpha_{NT}\dfrac{R^2(k,m-1)}{\hat{\lambda}_D(k,m)}+\\ (1-\alpha_{NT})\left[\dfrac{R^2(k,m)}{\hat{\lambda}_D(k,m)}-1\right],\xi_{\min}\end{array}\right] \quad (5)$$

α_{NT} : is a factor of noise tracking between 0 and 1.

Next, the speech presence probability estimate $\hat{p}$ is updated using[3]:

$$\zeta(k,m)=\sum_{i=-w}^{w}b(i)\zeta(k-i,m),\text{ with }\sum_{i=-w}^{w}b(i)=1 \quad (6)$$

A rectangular window with $w=1$ is used for $b(i)$. Then a hard decision about speech presence is made:

ISBN: 978-1-61804-290-3

if $\zeta(k,m) > T(k,m)$

$I(k,m) = 1$ speech present (7)

else

$I(k,m) = 0$ speech absent

end

Where: $T(k,m)$ is a threshold.

Otherwise, the speech presence probability determines the smoothing parameter, that is estimated by[3]:

$$\alpha_s(k,m) = \alpha_d + (1-\alpha_d)\hat{p}(k,m) \quad (8)$$

The speech presence probability estimate is updated with a first order recursion [3]:

$$\hat{p}(k,m) = \alpha_p \hat{p}(k,m-1) + (1-\alpha_p) I(k,m) \quad (9)$$

where α_p lies between 0 and 1. This estimate is used in equation (8) to find the smoothing parameter in (2).

The noise variance estimate is now updated using equation (2), where $\hat{D}^2$ is computed with a gain function found in [3].

Finally, for the speech spectral amplitude estimation, we compute prior SNR $\hat{\xi}_{SE}(k,m)$ from equation (1) and recompute posterior SNR $\zeta(k,m)$ from equation 10) using the new noise variance estimate [3]:

$$\zeta(k,m) = \frac{R^2(k,m)}{\lambda_D(k,m)} \quad (10)$$

4) Safety Net

The method will react quite slowly to sudden, large jumps in the noise level. For these cases, the safety net ensures that the algorithms continue to work properly:

The idea is to push the noise variance estimate into the right direction when we detect that its value is much too low.

As a reference value, we use the minima $P_{\min}(k,m)$ of the smoothed values $\overline{P}(k,m)$ of the noisy power $R^2(k,m)$ in a short window of length $w_{\min}$, where $\overline{P}(k,m)$ is given by[3]:

$$\overline{P}(k,m) = \eta \overline{P}(k,m-1) + (1-\eta) R^2(k,m) \quad (11)$$

Where η is a small smoothing parameter. After updating with $\hat{\lambda}_D$ (equation (2)), we check whether it fulfills the following condition[3]:

$$B.P_{\min}(k,m) < \hat{\lambda}_D(k,m) \quad (12)$$

where $B > 1$ is a correction factor. In case of a large increase in noise level that the algorithm cannot follow $B.P_{\min}(k,m)$, will become larger than $\hat{\lambda}_D(k,m)$ after a time of the order of the window length. If that happens, we reset the $\hat{\lambda}_D(k,m)$ values that violated (14) to $\max\left[B.P_{\min}(k,m), \hat{D}^2(k,m)\right]$, and the corresponding $\hat{P}(k,m)$ to 0. The factor B is taken larger than 1, but much smaller than the bias correction that would apply if the window $w_{\min}$ would contain only noise. This ensures that the safety net will not unintentionally come into action when some speech energy leaks into $P_{\min}$. We use very little smoothing of R^2 values (small η) to compute the minimum $P_{\min}$, because that allows us to keep the window $w_{\min}$ short. We have observed that the value of B and the window length are not very critical for good performance, but a window length of at least 0.5 s is required[3].

B. Speech Enhancement Based on a Priori Signal to Noise Estimation

This method was proposed by P. Scalart, and J. Vieira Filho (1996) [10]. Because The a Priori SNR estimation leads to the best subjective results. According to this conclusions, an approach was developed [10].

C. Geometric Approach (GA)

This recent method was proposed by Yang Lu, Philipos C. Loizou (2008) [11] that is A geometric approach to spectral subtraction Abstract. Yang Lu, Philipos C. Loizou presented a Geometric Algorithm (GA) to spectral subtraction based on geometric principles [11]. Unlike the conventional power spectral subtraction algorithm which assumes that the cross terms involving the phase difference between the signal and noise are zero, the algorithm makes no such assumptions. This was supported by error analysis that indicated that while it is safe to ignore the cross terms when the spectral SNR is either extremely high or extremely low, it is not safe to do so when the spectral SNR falls near 0 dB [11]. A method for incorporating the cross terms involving phase differences between the noisy (and clean) signals and noise was proposed [11]. Analysis of the suppression curves of the GA algorithm indicated that it possesses similar properties as the traditional MMSE algorithm (Ephraim and Malah, 1984) [11]. Objective evaluation of the GA algorithm showed that it performed significantly better than the traditional spectral subtraction algorithm in all conditions.

D. Harmonic Regeneration Noise Reduction (HRNR)

This method was proposed by Cyril Plapous, Claude Marro, and Pascal Scalart (2006) [12]. This approach addressed the

problem of single microphone speech enhancement in noisy environments. The well-known decision-directed (DD) approach drastically limits the level of musical noise but the estimated a priori SNR is biased since it depends on the speech spectrum estimation in the previous frame[12]. Therefore, the gain function matches the previous frame rather than the current one which degrades the noise reduction performance[12]. The consequence of this bias is an annoying reverberation effect. The authors proposed a method called Two-Step Noise Reduction (TSNR) technique which solves this problem while maintaining the benefits of the decision-directed approach. The estimation of the a priori SNR is refined by a second step to remove the bias of the DD approach, thus removing the reverberation effect. However, classic short-time noise reduction techniques, including TSNR, introduce harmonic distortion in enhanced speech because of the unreliability of estimators for small signal-tonoise ratios. This is mainly due to the difficult task of noise PSD estimation in single microphone schemes. To overcome this problem, a method called Harmonic Regeneration Noise Reduction (HRNR) was proposed. A non-linearity is used to regenerate the degraded harmonics of the distorted signal in an efficient way.. These methods are analyzed and objective and formal subjective test results between HRNR and TSNR techniques are provided. A significant improvement is brought by HRNR compared to TSNR thanks to the preservation of harmonics[12].

E. Phase Spectrum Compensation (PSC)

This work was proposed by Anthony P. Stark, Kamil K (2008) [13]. In this paper a novel approach for speech enhancement has been presented, where the noisy magnitude spectrum is recombined with a phase spectrum compensated for additive noise distortion to produce a modified complex spectrum. Noise estimates are incorporated into the phase spectrum compensation procedure. During synthesis the low energy components of the modified complex spectrum cancel out more than the high energy components, thus reducing background noise [13].

F. Speech enhancement using a priori SNR estimator

This method was proposed by I. Cohen (2004) [14], where it based on a priori SNR estimator, minimum mean-square error (MMSE). The author proposed a non causal estimator for the a priori signal-to-noise ratio (SNR), and a corresponding non causal speech enhancement algorithm. In contrast to the decision directed estimator of Ephraim and Malah [15], the non causal estimator is capable of discriminating between speech onsets and noise irregularities [14]. Onsets of speech are better preserved, while a further reduction of musical noise is achieved. Experimental results show that the non causal estimator yields a higher improvement in the segmental SNR, lower log-spectral distortion, and better Perceptual Evaluation of Speech Quality scores (PESQ, ITU-T P.862) [14].

G. Unbiased MMSE-Based Noise Power Estimation with Low Complexity and Low Tracking Delay.

This method was proposed by T. Gerkmann and C. Richard (2012) [16]. It has been proposed to estimate the noise power spectral density by means of minimum mean-square error (MMSE) optimal estimation[16]. Otherwise, the resulting estimator can be interpreted as a voice activity detector (VAD)-based noise power estimator, where the noise power is updated only when speech absence is signaled, compensated with a required bias compensation[16].The bias compensation is unnecessary when we replace the VAD by a soft speech presence probability (SPP) with fixed priors [16]. Choosing fixed priors also has the benefit of decoupling the noise power estimator from subsequent steps in a speech enhancement framework, such as the estimation of the speech power and the estimation of the clean speech[16]. In addition, the proposed SPP approach maintains the quick noise tracking performance of the bias compensated MMSE-based approach while exhibiting less overestimation of the spectral noise power and an even lower computational complexity[16].

III. PROPOSED APPROACH

Our approach to enhance speech is based on two speech enhancement methods. First method is Continues Wavelet Transform (CWT). Second method is Tracking of Non-stationary Noise Based on Data-Driven Recursive Noise Power Estimation[3] that is developed by Jan S. Erkelens and Richard Heusdens[3].

The performance of the proposed speech enhancement is evaluated in presence of different kind of noise using the NOIZEUS noisy speech corpus developed in Hu and Loizou laboratory[5] that is suitable for evaluation of speech enhancement algorithms.

A. Continues Wavelet Transform (CWT)method :

The motivation to use wavelet to achieve better noise reduction performance [4]. In our work, the Continues Wavelet decomposition of speech signal S(t) uses adaptive level with Harr mother wavelet. Decomposition process produces 'N' vectors of wavelet coefficients according to adaptive threshold.

Wavelet transform is based on the idea of filtering a signal $S(t)$ with a dilated and translated versions of a prototype function $\psi_{a,\tau}(t)$. This function is called the mother wavelet and it has to satisfy certain requirements [8]. The Continuous Wavelet Transform(CWT) for $S(t)$, is defined as [17]:

$$CWT(S,a,\tau) = \int_{-\infty}^{+\infty} S(t) \times \Psi_{a,\tau}(t)\,dt \tag{13}$$

Where:

$$\Psi_{a,\tau}(t) = \frac{1}{\sqrt{a}} \Psi\left(\frac{t-\tau}{a}\right), a \in \Re_{+}^{*} \tag{14}$$

where: a is the scale parameter and $\tau \in \Re$ is the translation parameter. In addition to its simple interpretation, the CWT

satisfies some other useful properties such as linearity and conservation of energy [17].

1) level decomposition with Adaptive threshold

The number of level (scales) decomposition to be considered is according to the formula:

$$p = 3 \times \left(\frac{\log(n)}{\log(2)} \right) \quad (15)$$

Where: n is samples number. (We keep the integer number of p). In the analysis of speech signals, we calculate wavelet coefficients corresponding for each scale $(a = 1.....p)$.

1) Which scale to be considered?

After wavelet coefficient calculation for each scale (equation 11),we assume it's sufficient to consider wavelet coefficients corresponding to a maximum energy of scale a :

$$E(a) = \sum_{i}^{n} |c_i|^2 \quad (16)$$

$E(a)$: Energy corresponding to scale a .

n : samples number of speech signal.

c : wavelet coefficient.

So, we adopt wavelet coefficient that concentrate more signal energy. it provides better reconstruction quality and introduce less distortion into processed speech. The speech signal to be reconstructed using these wavelet coefficients and passed through Jan S. Erkelens and Richard Heusdens algorithm (Tracking of Non-stationary Noise Based on Data-Driven Recursive Noise Power Estimation). Fig.1 shows wavelet coefficients energy of a speech signal taken from TIMIT database[18] as function of scale (a) where the wavelet coefficients to be used that are with a scale of $a = 18$.

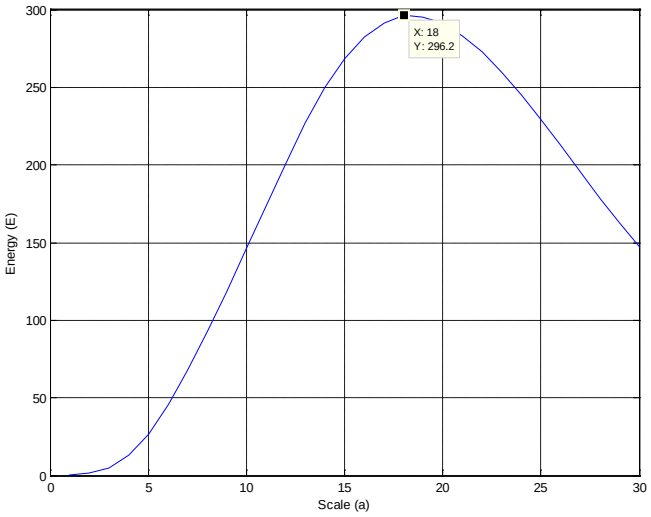


Fig.1. Wavelet coefficients Energy versus scale a (the maximum of wavelet coefficients energy matching with scale $a = 18$).

B. De-noising procedure using CWT and Tracking of Non-Stationary Noise Based on Data-Driven Recursive Noise Power Estimation algorithm

The general de-noising procedure involves six steps. The procedure follows the steps described below:

1. Choose a wavelet (Haar wavelet).
2. Compute of level (scales) decomposition to be considered according to formula (15).
3. Calculate wavelet coefficient for each scale (formula 11).
4. Compute wavelet coefficient energy related to each scale (formula 16) .
5. We consider wavelet coefficients matching to a maximum energy.
6. Passed wavelet coefficient through Tracking of Non stationary Noise Based on Data-Driven Recursive Noise Power Estimation algorithm.

IV. SPEECH QUALITY ASSESSMENT

The perceptual speech quality was objectively measured using Perceptual Evaluation of Speech Quality method (PESQ)[19]-[5].The PESQ method evaluates the quality of the speech signal by comparing the reference signal with the degraded signal. The PESQ algorithm models the human perception of the speech signal and thus enables the prediction of speech quality comparable to the subjective assessment as it would be performed by the human audience[18]. In this work we have adopted Loizou's PESQ implementation [5].

V. RESULTS AND DISCUSSION

A. Experimental Setup

To evaluate the proposed method, we have used speech signals taken from TIMIT database[18] in presence of white Gaussian Noise and NOIZEUS noisy speech corpus developed in Hu and Loizou laboratory[5]. The NOIZEUS corpus is suitable for evaluation of speech enhancement algorithms. The noisy database contains 30 IEEE sentences (produced by three male and three female speakers) corrupted by eight different real -world noises at different SNRs. The noise was taken from the AURORA[5] database and includes suburban train noise, babble, car, exhibition hall , restaurant , street , airport and train-station noise.

Parameter Settings: For the wavelet we have chosen Harr wavelet. For Tracking of Non-Stationary Noise algorithm (as the authors in [3] used), the following parameter settings are used in the experiments: α_d in (8) is set to 0.85, $\alpha_p = 0.1$ in (9), and $T(k,m) = 4$ in (7) independent of time and frequency. We have used $w = 1$ and $b(i) = 1/(2w+1)$ in (6). The same value 0.98 is used for the smoothing parameters α_{NT} in (5) and α_{SE} in (1), and $\xi_{\min}$ is set to -19 dB. We use $\eta = 0.1$ in (13), $B = 1.5$ in (14), and the length of $w_{\min}$ spans 0.8 s.

ISBN: 978-1-61804-290-3

B. *Performance evaluation using TIMIT database*

The proposed approach is objectively evaluated against several popular speech enhancement methods under noise conditions. We compare the proposed approach (CWT+ Tracking of Non-Stationary Noise algorithm) with seven methods of state of the art (Tracking of Non-stationary Noise Based on Data Driven Recursive Noise Power Estimation, Speech Enhancement Based on a Priori Signal to Noise Estimation (P. Scalart1996), Geometric Approach (GA), Harmonic Regeneration Noise Reduction (HRNR), Phase Spectrum Compensation (PSC), Speech enhancement using a priori SNR estimator (I. Cohen 2004) and Unbiased MMSE-Based Noise Power Estimation with Low Complexity and Low Tracking Delay).

Fig.2 Illustrates PESQ for various noise levels, obtained using proposed method, and various seven methods of state of the art. Where, white noise has been added to speech signal at several SNRs, from -5 to 30 dB in steps of 5 dB.

From fig.2, it can be seen that the proposed method score higher than the other methods in terms of the PESQ measure in presence of additive White Gaussian noise. Also, PSC and GA have the same PESQ scores (Curves are superposed).

C. *Performance evaluation using NOIZEUS corpus*

Otherwise the proposed approach was also evaluated in presence of several kinds of noise that are :Babble, Airport, Cart, Street, Restaurant. Where, The noise level varies between SNRs of -5 dB and 30 dB in steps of 5dB. The proposed approach was objectively evaluated against famous speech enhancement methods. (seven methods of state of the art).

The fig.3 represents PESQ measure for the proposed approach and the seven state-of-the-art methods in presence of Babble noise. From this figure we can conclude the efficiency of our approach but HRNR method is the worst under 10 dB. PSC and GA methods have nearly the same PESQ scores. Otherwise, Unbiased MMSE-Based Noise Power and Tracking noise methods have almost the same PESQ.

The fig.4 represents PESQ measure for the proposed approach and the seven state-of-the-art methods in presence of Airport noise. From this figure we can conclude the efficiency of our approach. PSC and GA methods have always nearly the same PESQ.

The fig.5 represents PESQ measure for the proposed approach and the seven state-of-the-art methods in presence of Car noise. From this figure we can conclude the efficiency of our approach. PSC and GA methods have always nearly the same PESQ, but the P. Scalart(1996) method was the worst.

The fig.6 represents PESQ measure for the proposed approach and the seven state-of-the-art methods in presence of Street noise. From this figure we can conclude the efficiency of our approach. PSC and GA methods have always nearly the same PESQ, but Tracking method was the second one after our approach in term of PESQ..

The fig7 represents PESQ measure for the proposed approach and the seven state-of-the-art methods in presence of Restaurant noise. From this figure we can conclude the efficiency of our approach. PSC and GA methods have always nearly the same PESQ. Unbiased MMSE-Based Noise Power and Tracking noise methods have nearly the same PESQ in this type of noise.

D. *Performance evaluation against Run Time*

As a third test we compare our approach with seven stat of the art methods in terms of runtime. Table. I shows simulation results in terms of runtime, where we can observe that Geometric approach has less run time than the other algorithms, but our approach has more run time than four methods (not the best). In our simulation we have used a Laptop that is Intel (R) core (TM) i5-3210M CPU @ 2.5GHZ 2.50GHZ.

TABLE I. RUN TIME OF: PROPOSED METHOD AND SEVEN STATE-OF-THE-ART METHODS

Speech enhancement methods	Elapsed time [sec]
Proposed method	0.4194
Noise tracking method	0.1972
Geometric Approach (GA)	**0.0944**
Phase Spectrum Compensation(PSC)	0.0981
Speech Enhancement Based on a Priori Signal to Noise Estimation (P. Scalart1996)	0.7808
Unbiased MMSE-Based Noise Power Estimation with Low Complexity and Low Tracking Delay (2012)	0.2058
Harmonic Regeneration Noise Reduction (HRNR)	0.6113
Speech enhancement using a priori SNR estimator (I. Cohen 2004)	1.3023

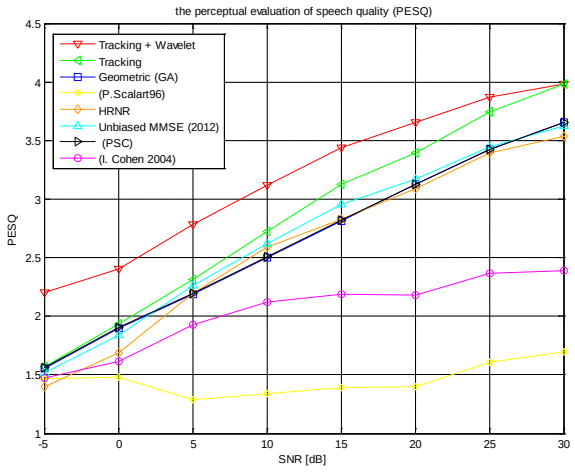


Fig.2. PESQ measure for the proposed approach and seven state-of-the-art methods in presence of additive White Gaussian noise. The noise level varies between SNRs of -5 dB and 30 dB in steps of 5dB.

ISBN: 978-1-61804-290-3

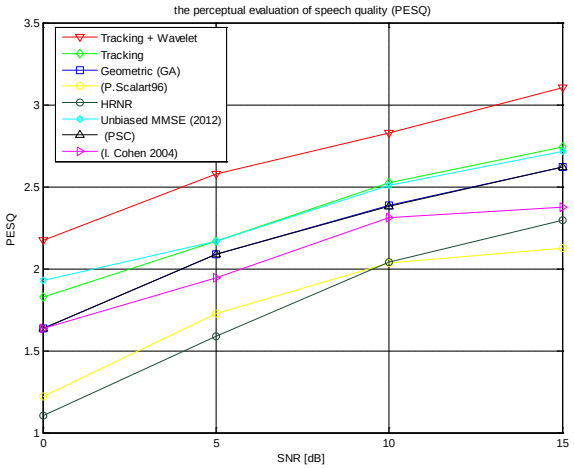


Fig.3. PESQ measure for the proposed approach and seven state-of-the-art methods in presence of Babble noise. The noise level varies between SNRs of 0 dB and 15 dB in steps of 5dB.

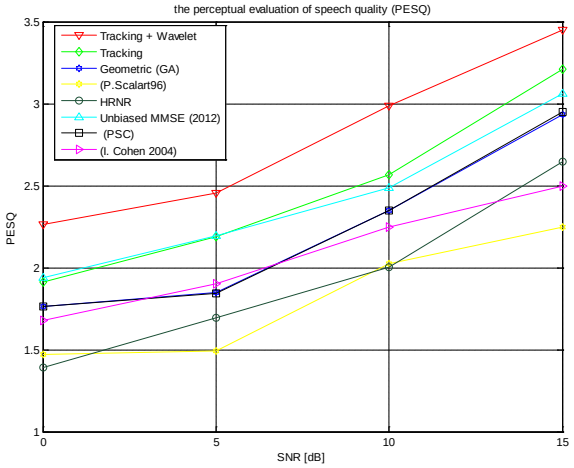


Fig.4. PESQ measure for the proposed approach and seven state-of-the-art methods in presence of Airport noise. The noise level varies between SNRs of 0 dB and 15 dB in steps of 5dB.

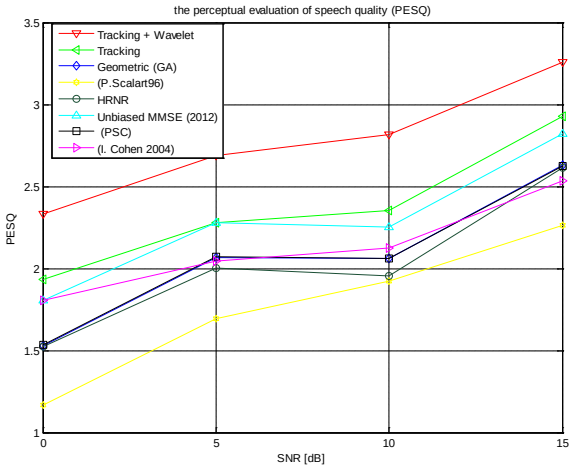


Fig.5. PESQ measure for the proposed approach and seven state-of-the-art methods in presence of Cart noise. The noise level varies between SNRs of 0 dB and 15 dB in steps of 5dB.

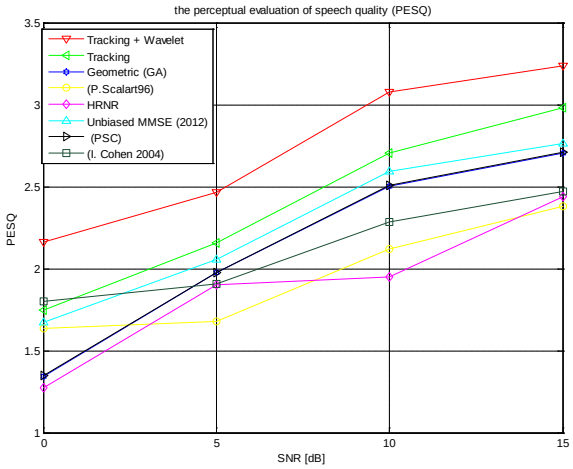


Fig.6. PESQ measure for the proposed approach and seven state-of-the-art methods in presence of Street noise. The noise level varies between SNRs of 0 dB and 15 dB in steps of 5dB.

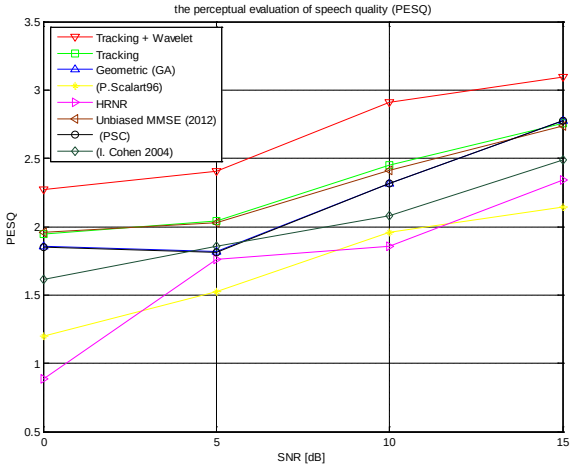


Fig.7. PESQ measure for the proposed approach and seven state-of-the-art methods in presence of Restaurant noise. The noise level varies between SNRs of 0 dB and 15 dB in steps of 5dB.

VI. Conclusion

In this paper, we have provided a novel speech enhancement method which robust to several kind of noise (White Gaussian noise, car, babble, street, airport, and restaurant) and in comparison with seven stat-of-the-art methods. For evaluating the performance of speech enhancement methods we have used Perceptual Evaluation of Speech Quality scores (PESQ, ITU-T P.862).

Our approach is based on two methods, the first is Tracking of Non-stationary Noise Based on Data Driven Recursive Noise Power Estimation and the second is Continuous wavelet Coefficients where the wavelet coefficients to be used are from which the Energy is maximum. Otherwise, we have evaluated our approach in terms of runtime where it has more run time than four methods thus, not the best in view of runtime.

The usefulness of the proposed algorithm for some applications needs to be verified.

ISBN: 978-1-61804-290-3

REFERENCES

[1] Ergun Erc elebi, "Speech enhancement based on the discrete Gabor transform and multi-notch adaptive digital filters" Applied Acoustics 65 (2004) 739–762.

[2] Mohsen Rahmani, Ahmad Akbari, Beghdad Ayad,"An iterative noise cross-PSD estimation for two-microphone speech enhancement" Applied Acoustics70 (2009) 514–521.

[3] J.S. Erkelens and R. Heusdens, "Tracking of nonstationary noise based on data-driven recursive noise power estimation", IEEE Trans. Audio,Speech & Lang. Proc., Vol. 16, No. 6, pp. 1112-1123, August 2008.

[4] Mahesh S. Chavan, Nikos Mastorakis "Studies on Implementation of Harr and daubechies Wavelet for Denoising of Speech Signal". *international journal of circuits, systems and signal processing*. Issue 3, Volume 4, 2010.

[5] Yi Hu and Philipos C. Loizou, "Evaluation of Objective Quality Measures for Speech Enhancement". IEEE transactions on audio, speech, and language processing, vol. 16, no. 1, January 2008.

[6] Atanu Saha, and Tetsuya Shimamura,"Perceptually Motivated Bayesian Estimators With Generalized Gamma Distribution Under Speech Presence Probability". *international journal of circuits, systems and signal processing*. Issue 1, Volume 6, 2012.

[7] Ephraim, Y., & Malah, D. (1984). Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, *32*(6), 1109-1121.

[8] Farsi, H., Mozaffarian, M. A., & Rahmani, H. Adapting Correction Factors in Probability Distribution Function for VAD Improvement. *NAUN International Journal of Circuits, Systems and Signal Processing*, Issue 1, Volume 3, 2009.

[9] Erkelens, J., Jensen, J., & Heusdens, R. (2007). A data-driven approach to optimizing spectral speech enhancement methods for various error criteria. *Speech communication*, *49*(7), 530-541. Elsevier.

[10] P. Scalart, and J. Vieira Filho, "Speech Enhancement Based on a Priori Signal to Noise Estimation," IEEE Intl. Conf. Acoust., Speech, Signal Processing , Atlanta, GA, USA, Vol. 2, pp. 629–632, May 1996.

[11] LU, Yang et LOIZOU, Philipos C. A geometric approach to spectral subtraction. *Speech communication*, 2008, vol. 50, no 6, p. 453-466.

[12] Plapous, C.; Marro, C.; Scalart, P., "Improved Signal-to-Noise Ratio Estimation for Speech Enhancement", IEEE Transactions on Audio, Speech, and Language Processing, Vol. 14, Issue 6, pp. 2098 - 2108, Nov. 2006.

[13] [1] A.P. Stark, K.K. Wojcicki, J.G. Lyons and K.K. Paliwal,"Noise driven short time phase spectrum compensation procedure for speech enhancement",Proc. INTERSPEECH 2008, Brisbane, Australia, pp. 549-552, Sep. 2008.

[14] I. Cohen .Speech Enhancement Using a Noncausal A PrioriSNR Estimator. IEEE, signal processing letters, vol. 11, no. 9, september 2004.

[15] Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-square error log-spectral amplitude estimator," IEEE Trans. Acoust., Speech, Signal Processing, vol. ASSP-33, pp. 443–445, Apr. 1985.

[16] Gerkmann, T. & Hendriks, R. C. Unbiased MMSE-Based Noise Power Estimation With Low Complexity and Low Tracking Delay. IEEE Trans Audio, Speech, Language Processing, 2012, 20, 1383-1393.

[17] W.Shabana and J.Fitch « a wavelet-based pitch detector for musical signals ». Department of Mathematical Sciences, University of Bath,Bath BA2 7AY, UK.2000.

[18] arofolo, J. S., Lamel, L. F., Fisher, W. M., Fiscus, J. G., Pallett, D. S, and Dahlgren, N. L., "DARPA TIMIT Acoustic Phonetic Continuous Speech Corpus CDROM," *NIST*, 1993.

[19] Robert Blatnik, Gorazd Kandus, and Tomaž Šef . Influence of the perceptual speech quality on the performance of the text-independent speaker recognition system. international journal of circuits, systems and signal processing. Issue 4, Volume 5, 2011.

Pose Estimation Methodology For Target Identification And Tracking

Part I. Target Signatures and Hypothesis Testing

Migdat I. Hodzic, Tarik Namas

International University of Sarajevo, FENS, Sarajevo, Bosnia and Herzegovina
mhodzic@ius.edu.ba, tnamas@ius.edu.ba

Abstract— **Ground Moving Target Indicator (GMTI) and High Resolution Radar (HRR) can track position and velocity of ground moving target. Pose, angle between position and velocity, can be derived from estimates of position and velocity and it is often used to reduce the search space and hence increase likelihood of target identification (ID) and Automatic Target Recognition (ATR) algorithms. Due to low resolution in some radar systems, the GMTI estimated pose may exhibit large errors contributing to a faulty identification of potential targets. Our goal in this paper is to define better methodology to improve pose estimate, using real time target signature versus stored signatures. Besides applications in target tracking, there are numerous commercial applications in machine learning, augmented reality and body tracking.**

Keywords – Target Identification and Tracking, Pose Estimation, Target Signature, Statistical Hypothesis Testing

I. INTRODUCTION

With the recent developments in the areas of HRR (High Resolution Radar) and SAR (Synthetic Aperture Radar), ground targets reflection signatures become much richer and indicate various geometrical features of these targets. Hence feature and signature aided tracking and ATR applications benefit from HRR radar processing. Successful simultaneous tracking and identification applications exploit feature information to determine the target type and dynamics. This has enabled target classification and identification (ID), as well as ATR. As a "by-product" of the target ID/ATR process, the pose angle estimates are available as well. Figure 1 shows geometry of pose angle information. The depression angle relates to sensor position and aspect angle could be deduced from HRR signature. For ground targets, which are constrained to move on the earth surface, their velocity vector direction is aligned most of time along the body's longitudinal axis. As a result, the pose angles carry kinematic information that can be used to aid target tracking particularly during the maneuvering periods. See more in [1],[2],[4],[5] and [6]. In this paper we present a methodology for a better pose estimate algorithms, which can in turn accomplish improved target tracking and identification, as well as reduce target miss-association probability (MAP). In a typical application an airborne system first detects ground targets from a distance, initiates the tracks, and then continues with tracks "maintenance". One of the practical issues which arise is when there are closely spaced targets, which then calls for further signature analysis to distinguish the targets. Some of the signature features supplied by HRR or SAR are range, aspect angle, peak amplitude, and other characteristics. Typical method applied is to store a large number of signatures in an on board data base and then compare real time signature of interest with the stored ones. Various statistical and other methods are then applied to seek the match between stored and real time signatures. In this paper we propose to use new statistical comparison approach based on a novel idea of (i) treating the signatures as portions of some stochastic processes, and (ii) combining several statistical methods to test for an independence measure between the signatures. We use off line generated signature template against real target signature data. We exploit time, frequency and correlation features of these signatures. The algorithm then aligns real time signatures to the library templates of targets and determines the best correlation value for the aligned features, or put it differently, eliminates highly independent signatures. In doing so we propose classic methods of Pearson and Spearman coefficients as well as chi square testing. Also, a new methodology via Haar Transform [24] and recently introduced statistical method of Brownian Distance Correlation [27] are considered as well. Based on these methods, various statistics of the outputs are compared and cross-correlation among two sets of data (stored and real) is calculated. In doing so we will obtain an estimate for the pose angle as well. In Part II of this paper, we will determine and predict performance of coupled target tracker and target identifier as a function of pose estimate. Target tracking will be implemented using multiple Extended Kalman Filters (M-EKF) or Unscented Kalman Filters (M-UKF), in the spirit of similar approach described in [6]. Each filter can handle one target type and it produces target trajectory estimates plus associated probability of being used, which would determine target probability as well. Our emphasis in Part II will be on pose sensitivity assessment to parameter changes as well as its effect on tracking performance. Besides defense applications, results in this paper can be used in Intelligent Vehicle Highways and Intersection Traffic Control where the benefits are (i) improved traffic flow, (ii) increased safety and (iii) improved gas mileage. Additional applications are in the area of Augmented Reality, Facial Features estimation and Machine Learning algorithms development.

II. TECHNICAL OBJECTIVES SUMMARY

The overall objective of this paper is to introduce a methodology to develop better pose estimation technique, improve coupled tracking-identification process, and reduce target miss-association. We will use a case - study of well known USA AirForce airborne platform of JSTAR [4], [5]. The original platform was based on old Boeing 707 aircraft that has been fitted with many sophisticated target identification and tracking equipment. JSTAR has seen

ISBN: 978-1-61804-290-3

deployments in Bosnia to enforce Dayton Peace Agreement, and Kosovo, among other places, in mid and late 1990s. There is a current JSTAR modernization effort under way.

The objectives of this paper are listed bellow. In Part I of this paper, we cover first three objectives, while remaining three will be covered in Part II.

1. Target Signature Profiles. Present new approach for generating spatial and frequency target data templates obtained from raw and digitized HRR (SAR) data. The data can be derived from public Moving and Stationary Target Acquisition and Recognition (MSTAR) program database. The MSTAR data consists of SAR data (X-band 1 x 1 foot resolution). Once real time target raw signature is obtained, it will be discretized and stored in a form as any of the template entries, Table 1 and Table 2.

2. **Target Signature Statistics**. Define effective and less computationally intensive ID/ATR search and registration process which uses statistical comparison of off line analyzed HRR template data, step 1 above, and HRR target data obtained in real time. The key requirement here is that the process is fast and computationally simple so it can be done in real time, when many 1000s of templates are scanned and compared. We propose to use new techniques in spatial and frequency domains, as well as some new and classic correlation methods, all aimed at identifying weak or strong target signature correlations.

3. Target Hypothesis Testing. Once various target signature statistics are generated in Objective 2 above, we proceed and perform a variety of hypothesis testing cases, to identify strong and weak correlations between real time signature and template signatures stored ahead of any real time operation. This is very important step and it results in an estimate of the most likely pose angle, based on real time vs. template match. Note that the best match would correspond to a rough pose angle estimate from Table 1, or maybe a narrow range of pose angles. This all leads us to Objective 4 bellow.

4. Pose Angle Estimate. Good pose estimate (depression and aspect angles, φ and ϕ) accomplishes reduction in probability of real time target miss - association, which allows for the capability to discern relevant targets and reject non-plausible targets. Radar tracking assumes that after receiving the energy return from the target, the approximate coarse position of the target results. Since a finite number of range bins are collected, the center bin is assumed to be the position of the target (see Fig. 2 and 3 bellow). Additionally, the radar data has an associated depression and azimuth angle to the target, see Fig. 1 hence further pose estimate fine tuning will be done with the help of the results of Objective 5.

5. Target Tracking. With a rough pose angle estimate we define a bank of parallel target tracking filters, each one for a particular target type. The radar measurements provide the measurements of Range (r), Range rate (dr/dt), Azimuth (α) and Elevation angle (ε), from the aircraft to the target (Figure 1). The HRR provides a target range profile from which we can deduce the pose estimate, i.e. depression and aspect angles. Typically this reduces to just aspect angle because the depression angle is kept fixed, as it will be assumed throughout the work here. Hence we will have two rough pose estimates to work with, and to improve on.

6. Sensitivity Analysis. Predict the performance (better or worse) of target identifier and tracker due to pose estimate quality. This step corresponds to calculating sensitivity of identification (or, in turn, target miss-association) and tracking performance with respect to pose estimate. We will calculate the sensitivity of various statistical parameters to probability of miss – association and identify less sensitive and more sensitive parameters from the steps above. This corresponds to breaking down pose sensitivity calculation, into individual calculations of statistical parameters and signature profiles.

III. Target Signature Profiles

Our methodology consists of six steps (3 in Part I, 3 in Part II), for six technical objectives of Section II. We start with:

STEP 1. Develop target signature profiles as in [6].

1.1 Obtain row and digitized data for all template targets of interest. We can use public domain MSTAR, or other sources. Table 1 shows N targets and M pose angles. To each entry we add various features, Space Features (SF), Space Features Statistics (SFS), Frequency Features (FF) and Frequency Features Statistics (FFS), Table 2.

1.2 Get row and digitized signature for a real time target. It is in the same form as any Table 2 entry. We assume this is done in real time on JSTAR or similar platform.

Table 1. Signature Template (T) Profile

Pose / Target	Pose 1	Pose 2	..	Pose M	Features / Statistics
Target 1	T(1,1)	T(1,2)	..	T(1,M)	F/S Set 1
Target 2	T(2,1)	T(2,2)	..	T(2,M)	F/S Set 2
…..	…..	…..	..	…..	…..
Target N	T(N,1)	T(N,2)	..	T(N,M)	F/S Set N

The MSTAR data consists of SAR data in X band, 1 x1 foot resolution. Images are recorded at 15° and 17° depression angles with aspect angle 360° range, at around 1° spacing in azimuth. The methodology used to convert the SAR imagery to HRR is discussed in [15]. While the MSTAR data consists of all aspect data, for each measurement pose estimate, viewing aspect angles considered are $-5^\circ < \phi < 5^\circ$ in azimuth and 15° and 17° in depression angle. Each HRR image results in 101 HRR profiles, which represent approximately a 3° variation in azimuth. If we have 50 target types, that is a total of 5,050 signature profiles, plus additional statistical data (last column in Table 1 and Table 2) as described in Section IV.

As stated earlier in this paper we will assume K = 16 points and will calculate various statistics in Section IV. For an extended paper we are working on K = 64where better statistics will be obtained.

Table 2. Signature Template T(I,J), Target I and Pose J

Spatial Features SF(I,J)	SF Statistics SFS(I,J)
	Max, Min, Average, Median, Variance, Energy, Ratios, etc.
Frequency Features FF(I,J)	FF Statistics FFS(I,J)
	Max, Min, Average, Median, Variance, Energy, Ratios, etc.

ISBN: 978-1-61804-290-3

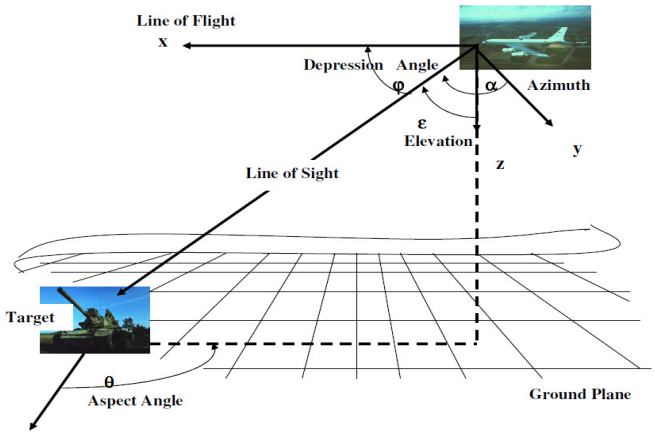


Figure 1. Pose Geometry

In Steps 1.1 and 1.2, we form specific Spatial, Amplitude and Frequency Signature Feature vectors.

Spatial Features (SF) consist of:

- Row HRR data (MSTAR or other source), See Fig. 2.
- Vector of K digitized raw data, Fig. 3. Here K = 16 for the purposes of this paper. In the extended version of the paper we will present the results for K = 64. In real situations, digitization is done by dedicated equipment and the number of points could be 128 or more, and it could be case specific. For the purposes of our work we also assume that the number of points is $K = 2^m$,and m = 1,2,3,… .

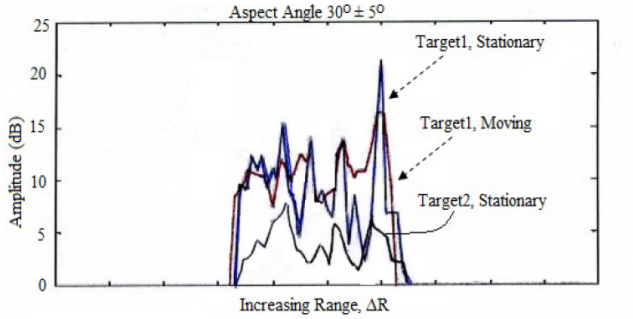


Figure 2. Continuous Range Data Over ΔR
Stationary Target 1 and 2, and Moving Target 1 Signatures

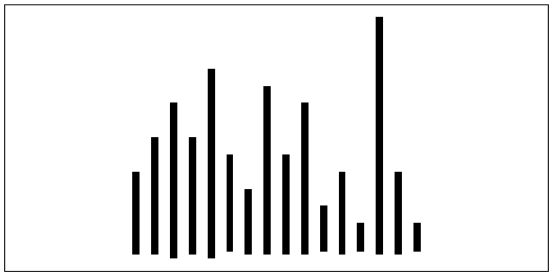


Figure 3. Digitized Data, Stationary Target 1 in Fig. 2
Each vertical line is one digitized range bin

Besides Spatial Features, we also define additional signal characteristics. These include various amplitude features and corresponding statistics SFS. All statistics are localized to the total range span ΔR indicated in Fig. 2.

Spatial Features Statistics (SFS) formed of the statistics:

- Highest amplitude (A_{max}) (or two highest ones)
- Lowest amplitude (A_{min}) (or two lowest ones)
- Average signal amplitude (A_{av})
- Median amplitude value (A_{med})
- Standard deviation (A_{sd})

plus such indicators as:

- Ratio (A_{max}/A_{min})
- Total energy (A_e), as the sum of amplitudes squared
- Number of discretized peaks and valleys (A_p), (A_v)
  A_p = Number of Low-High-Low cases
  A_v = Number of High-Low-High cases = A_p-1

Note that in order to obtain better statistics, we would need more than 16 sampling points, which will be presented in the extended paper for 64 sampling points. In addition to SF we also consider Frequency Features (FF).

Frequency Features (FF) vector is generated by using Haar Transform Matrix operating on SF vector, such as in Equation 5 and Tables 3, 4 and 5 in Section 6.

The Haar transform is very useful in signal processing applications where real-time implementation is essential. The Haar transform is based on Haar functions which are periodic and orthogonal. The Haar functions become increasingly localized as their number increases [24], which provides frequency domain in which signature energy is concentrated in localized regions. This property is very useful in various applications.

The Haar transform matrix is an orthogonal one, hence the inverse Haar transform can be derived from:

$$HH^T = I, \quad H^{-1} = H^T \tag{1}$$

where I is the identity matrix. The 1st order Haar matrix is:

$$H(1) = 1/\sqrt{2^1} \text{ times} \tag{2}$$

1	1
1	-1

The recursive equation for higher order Haar matrices is:

$$H(k+1) = 1/\sqrt{2^{k+1}} \text{ times} \tag{3}$$

H(k)*[I I]
$2^{k/2}$ I(2^k)*[I –I]

where "*" is Kronecker product, I(2^k) is Identity matrix of order 2^k, such that H(2) is 4x4, H(3) is 8x8, and H(4) is 16x16 matrix, and so on. Take the 8x8 Haar matrix H(3):

$$H(3) = 1/\sqrt{2^3} \text{ times} \tag{4}$$

1	1	1	1	1	1	1	1
1	1	1	1	-1	-1	-1	-1
√2	√2	-√2	-√2	0	0	0	0
0	0	0	0	√2	√2	-√2	-√2
2	-2	0	0	0	0	0	0
0	0	2	-2	0	0	0	0
0	0	0	0	2	-2	0	0
0	0	0	0	0	0	2	-2

ISBN: 978-1-61804-290-3

Unlike Fourier transform, Haar transformation involves only real numbers. We propose to use Haar matrix not because it is the best choice for frequency transform, but because of its simplicity for target signature features.

The Haar transform Y(K) of an K - input vector X(K) is:

$$Y(K) = H(m)X(K),\ X(K)= H^T(m)Y(K) \qquad (5)$$

where $K = 2^m$. To obtain proper averages, for H(4), first 2 rows are scaled by $1/\sqrt{16}$, next 2 by $1/\sqrt{8}$, next 4 by $1/\sqrt{4}$, and the last 8 by $1/\sqrt{2}$. Note that this operation will scale orthogonality of the matrix. We note the following:

- The first element in Y is the average (DC) value of X
- The second element is a low frequency component of the input vector X
- The next two components of Y correspond to moderate frequencies in input X
- The next four elements correspond to moderate-to-high frequency components in X, and
- The last eight elements correspond to high X frequencies.

Note that we did not specify what "low" and "high" frequencies are. We just want to indicate relative sizes of groups of frequencies which can be calculated as inverses of spatial differences.

All X(K) and Y(K) vectors can be stored into Target Template (Table 1) for each entry as in Table 2. During real time operation, as JSTAR plane is scanning an area for ground targets, real time X(K) and Y(K) would be calculated for any target detected, and compared against the stored ones of Table 1, to determine specific target presence. Note that **FF Statistics (FFS)** similar to SFS are also calculated, Section IV

IV. Target Signature Statistics

Our aim is to devise effective statistical correlation methods.

STEP 2. Produce various statistical measures and correlations

2.1 Calculate template standard statistics and correlations for SFs, AFs and FFs (last column in Table 1)
This is all done ahead of any real time operation
2.2 Calculate real time target statistics and correlations, same type as in Step 2.1. This is done as real target is acquired
2.3 Produce a variety of cross statistics and compare them in real time (all or selected templates vs real time signature).

In all the Steps above we treat the template and real time data as two sequences of either "independent" or "dependent" stochastic processes, on which we perform statistical testing as described bellow. Specifically, in Steps 2.1 and 2.2 we form:

Standard Statistical Features (SSF) which is equivalent to individual components of AF vector. In this embodiment, we will use SSFs to perform individual hypothesis testing in Section 5, and the full SF, FF and AF vectors are used for correlation features, Step 2.3.

Correlation Features (CF):

- Pearson sample correlation coefficient (p)
- Spearman rank correlation coefficient (ρ)
- Chi square test and P-value (χ^2), (P)
- Skewness measure (s)
- Sample distance (Brownian) covariance and coefficient (V, r), two new concepts introduced recently in [27].

Standard Pearson coefficient recovers a linear relationship that may exist among two sets of data sequences, per:

$$p_{XY}=[\Sigma(X(i)-X_a)(Y(i)-Y_a)]/[\Sigma(X(i)-X_a)^2\,\Sigma(Y(i)-Y_a))^2]^{1/2} \quad (13)$$

where the sums are evaluated from 1 to K, and X_a and Y_a are mean (average) values of X(K) and Y(K) sequences. The R_{XY} coefficient ranges from -1 to 1. It has some serious limitation as far as capturing non linear and non stationary correlations, plus $p_{XY}=0$ does not imply independence in general (only for normal distributions). On the positive, the method is simple.

Spearman rank correlation tests how relationship between two variables can be described using a monotonic function (increasing or decreasing). With no repeated values, a perfect correlation +/-1 occurs when each of the variable is a perfect monotone function of the other. Ranking refers to data transformation in which numerical values are ranked by their size. The Spearman coefficient is Pearson's for the ranked entries. For a K-sample size, K X(k) row values and Y(K) are converted to ranked $X_r(K)$ and $Y_r(K)$ and ρ is computed from:

$$\rho =1- 6\Sigma d^2(i)/n(n^2-1) \qquad (14)$$

where $d(i) = X_r(i)-Y_r(i)$ is the rank difference, and the sum goes from 1 to K. We can also use (13) above with $X_r(K)$ and $Y_r(K)$ instead of original samples.

Chi square test is used to test independency of the two sequences. The test returns the value from chi-squared distribution for the statistic and the degrees of freedom number. In the context of our paper, we will test independency of real time (an experiment) and template signature entries (hypothesized results). The test is defined as:

$$\chi^2=\Sigma\ (A_i-E_i)^2/E_i \qquad (15)$$

where A_i=actual sample value, and E_i=expected sample value, and the sum is from 1 to K. A low value of χ^2 is an indicator of independence. As can be seen from the formula, χ^2 is always positive or 0 (only if $A_i = E_i$ for every i). Once χ^2 is calculated, an appropriate program, for example Excel CHITEST which returns "the probability P that a value of the χ^2 statistic at least as high as the value calculated by (15) could have happened by chance under the assumption of independence". [30]. In computing P-value, program uses χ^2 distribution with an appropriate number of degrees of freedom, df = K–1. The test is most appropriate when E_i's are ≥5.

The skewness is based upon the sample formula:

$$s=(1/K)\ \Sigma(X(i)-X_a(i))^3/[(1/K-1)\ \Sigma(X(i)-X_a(i))^{3/2}] \qquad (16)$$

with average value $X_a(i)$. Another formula is often used, i.e. the adjusted Fisher-Pearson standardized moment coefficient:

$$s_1=s(K^2)/(K-1)(K-2) \qquad (17)$$

(SKEW in Excel). The variance from a normal distribution is:

$$Var(s_1)=6K(K-1)/(K-2)(K+1)(K+3) \qquad (18)$$

Skewenss is obviously zero for any symmetric distribution.

Finally, we use the newest form of correlation [27], i.e. distance correlation and distance covariance (equivalent to Brownian distance covariance and coefficient). The key advantage is that zero correlation implies independence, plus the coefficient captures non stationary and non linear correlations as well. Here is a brief summary taken from [27].

For a random sample (X,Y)=[(X(k),Y(k), k=1,…,K] of K i.i.d. (independent identically distributed) vectors (X,Y) from the joint distribution of the random vectors X in R^p and Y in R^q, compute the Euclidean distance matrices:

$$[a(k,l)] = (|X(k)-X(l)|_p), \quad [b(k,l)] = (|Y(k)-Y(l)|_q) \qquad (19)$$

and then define:

$$A(k,l) = a(k,l) - a(k) - a(l) + a$$
$$B(k,l) = b(k,l) - b(k) - b(l) + b, \quad k,l = 1,2,\ldots,K \qquad (20)$$

where:

$$a(k) = (1/n)\Sigma_l\, a(k,l)$$
$$a(l) = (1/n)\Sigma_k\, a(k,l)$$
$$a(k) = (1/n^2)\Sigma_k\Sigma_l\, a(k,l) \qquad (21)$$

and the sums go from 1 to K. Then, the non negative sample distance covariance is defined as:

$$V_n^2(X,Y) = (1/n^2)\Sigma_k\Sigma_l A(k,l)B(k,l), \quad k,l = 1,2,\ldots K \qquad (22)$$

and the corresponding sample distance correlation as:

$$R_n^2(X,Y) = V_n^2(X,Y)/[V_n^2(X)V_n^2(Y)]^{1/2} \qquad (23)$$

whenever $V_n^2(X)V_n^2(Y)] > 0$, and $R_n^2(X,Y) = 0$, when we have $V_n^2(X)V_n^2(Y)] = 0$. Obviously, we have sample distance variance as:

$$V_n^2(X) = V_n^2(X,X) = (1/n^2)\,\Sigma_k\Sigma_l A_{kl}^2, \quad k,l = 1,2,\ldots K \qquad (24)$$

Another interesting result in is that:

$$V_n^2(X,Y) = \|f^n_{XY}(t,s) - f^n_X(t)\, f^n_Y(s)\|^2 \qquad (25)$$

where f's are corresponding characteristic functions. The results in (22) and (23) turn out to be equal to Brownian (Wiener Process) distance covariance and correlation coefficient. Again consult [27]. These are very useful results, and we believe they can be applied successfully in ATR/ID environment for hypothesis testing.

V. Target Hypothesis Testing

STEP 3. Execute target signature hypothesis testing.

3.1 Hypothesis testing for Steps 2.1 and 2.2
3.2 Hypothesis testing for Step 2.3

We treat real time signature as "an experiment" and template data as "test" signatures to compare against. The Null Hypothesis H_0 refers to a default "no match" position that there is no relationship between two phenomena, i.e. they are independent. The $\mathbf{H_0}$ is assumed true until evidence indicates an alternative "match" $\mathbf{H}_1$ hypothesis.

Statistical hypothesis (see Figure 4) testing is defined as:

- Compute real time and template statistics and compare to produce the value $O(T_k)$ of some test T_k statistic. This could be simple variance comparison, chi square, or any other of the methods discussed earlier. Comparison may be as simple as a quotient of the two statistics or more sophisticated calculation of so called p-value. Under the H_0, this is the probability of sampling a test statistic at least as extreme as the observed one

- Reject the null hypothesis, in favor of the alternative hypothesis, if and only if the compared or p-value is less than the significance level (selected probability) threshold.

In general, we will define statistical variable $\mathbf{S^{ij}}$ (target "i" and pose "j") as a weighted average of the outputs of all the statistical tests, denoted by $O(T_k)$ in Figure 4, i.e.

$$\mathbf{S^{ij}} = \sum_{k=1}^{n} a_k\, O(T_k) \qquad (26)$$

where a_k is TBD weight determined by experiment or extensive Monte Carlo simulation of the observed and template signatures. Hence variable $\mathbf{S^{ij}}$ determines the threshold for the H_0 acceptance or rejection. We are interested in whether the observed data is significantly different from what would be expected if the null hypothesis is true. So the main goal is to make a proper trade-off between the probability of Type I error, i.e.:

$$\text{False Alarm} = \text{Declaring } H_1 \text{ when } H_0 \text{ holds} \qquad (27)$$

and Type II error, the probability of:

$$\text{Missed Alarm} = \text{Declaring (keeping) } H_0 \text{ when } H_1 \text{ holds} \qquad (28)$$

In the case of (27) False Alarm may result in a wrong pose angle estimate, which is not desirable. This would correspond to a target miss-association possibility. In the case of (28) Missed Alarm would result in missing to identify a right pose angle estimate. The best possible pose estimate is determined by properly choosing $\mathbf{S^{ij}}$ threshold, based on several statistical tests listed in Task 2.

If the null hypothesis is true, then we can evaluate the probability of a Type I error and this value represents our tolerance for Type I errors, i.e. of rejecting when in fact it is true. This probability provides an important design criterion for testing. Specifically, the rejection region is chosen so that the probability of Type I error is no greater than a specified level, for example 1% and 5%. An alternative approach is to ask the question: Assuming H_0 is true, what is p-value of the test statistic. If it is close to one, then there is no reason to reject the null hypothesis, but if it is small, then there is reason to reject H_0. Which way to test H_0 will be decided by Monte Carlo simulation which is outside the scope of this paper. This can be done for a specific target class case at hand.

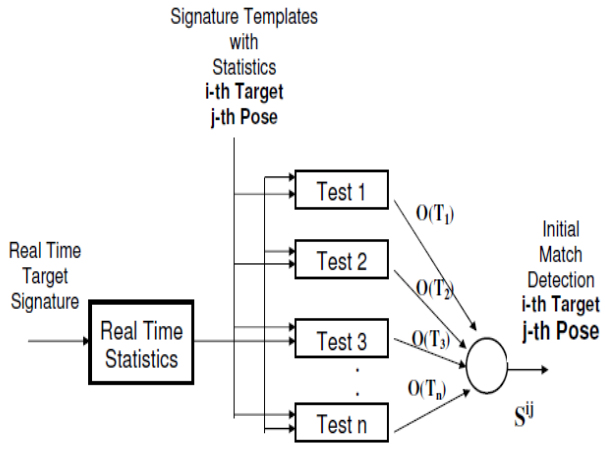


Figure 4. Forming S^{ij} Statistical Threshold

ISBN: 978-1-61804-290-3

VI. Target Signature Example

We use Figures 2 and 3 and form specific signature features.

Spatial Features (SF). We have the following Tables 3, 4 and 5, with N=16 discretized signature amplitudes (SF) going from left to right, such as in Fig. 3.

TABLE 3. STATIONARY TARGET 1, SFs VECTOR $X^1_s(16)$

7	10	12	10	15	8	5	14	8	13	4	7	3	22	7	4

TABLE 4. MOVING TARGET 1, SFs VECTOR $X^1_m(16)$

7	10	11	8	12	10	12	13	8	13	11	12	13	17	7	4

TABLE 5. STATIONARY TARGET 2, SFs VECTOR $X^2_s(16)$

3	5	4	7	8	6	3	5	7	5	3	5	8	6	3	2

Spatial Features Statistics (SFS). Based on Tables 3, 4 and 5, and earlier defined SFS, we have three signature results:

TABLE 6. STATIONARY TARGET 1, AFs VECTOR $A^1_s(8)$

A_{max}	A_{min}	A_{max}/A_{min}	A_{av}	A_{med}	A_{sd}	A_p	A_e
22	4	5.5	9.3125	8	4.98	6	14.9

TABLE 7. MOVING TARGET 1, AF VECTOR $A^1_m(8)$

A_{max}	A_{min}	A_{max}/A_{min}	A_{av}	A_{med}	A_{sd}	A_p	A_e
17	4	4.25	10.5	11	3.14	5	16.8

TABLE 8. STATIONARY TARGET 2, AFs VECTOR $A^2_s(8)$

A_{max}	A_{min}	A_{max}/A_{min}	A_{av}	A_{med}	A_{sd}	A_p	A_e
8	2	4	5	5	1.9	5	8.1

Frequency Features (FF). In our example, n=16, m=4, hence we need to calculate 16x16 H(4) Haar matrix which can be deduced from (3) and H(3). We have:

$$H(4) = (1/\sqrt{2^4}) \text{ times} \quad (28)$$

1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	-1	-1	-1	-1	-1	-1	-1	-1
√2	√2	√2	√2	-√2	-√2	-√2	-√2	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	√2	√2	√2	√2	-√2	-√2	-√2	-√2
2	2	-2	-2	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	2	2	-2	-2	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	2	2	-2	-2	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	2	2	-2	-2
√8	-√8	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	√8	-√8	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	√8	-√8	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	√8	-√8	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	√8	-√8	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	√8	-√8	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	√8	-√8	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	√8	-√8

From Table 3 (stationary Target 1) we obtain corresponding frequency vector $Y^1_s(16)$ by way of H(4) and:

$$X^1_s(16) = [7,10,12,10,15,8,5,14,8,13,4,7,3,22,7,4]^T \quad (29)$$

producing:

$$Y^1_s(16) = H(4)X^1_s(16) = \quad (30)$$
$$[9.3125,0.8125,-0.375,-0.5,-1.25,1.0,2.5,3.5,-1.5,1,3.5,-4.5,-2.5,-1.5,-9.5,1.5]^T$$

If the Target 1 is moving (Table 4), we obtain the vectors:

$$X^1_m(16)= [7,10,11,8,12,10,12,13,8,13,11,12,13,17,7,4]^T \quad (31)$$

and:

$$Y^1_m(16) = H(4)X^1_m(16) = \quad (32)$$
$$[10.5,-0.125,-1.375,0.375,-0.5,-0.75,-0.5,4.75,-1.5,1.5,1,-0.5,-2.5,-0.5,-2,1.5]^T$$

Finally, the vectors for the stationary Target 2 (Table 5) are:

$$X^2_s(16) = [3,5,4,7,8,6,3,5,7,5,3,5,8,6,3,2]^T \quad (33)$$

and

$$Y^2_s(16) = H(4)X^2_s(16) = \quad (34)$$
$$[5,0.125,-0.375,0.125,-0.75,1.5,1,2.25,-1,-1.5,1,-1,1,-1,1,0.5]^T$$

Based on the above data we form various statistics summarized in the Tables 9 and 10 bellow.

Target Scenarios / Statistical Tests		St T1	Mv T1	St T2
SFs	Mean	9.31	10.5	5
16 samples	Median	8	11	5
	St Deviation	4.9762	3.1411	1.8974
	Variance	24.7625	9.8667	3.6
	Energy	1759	1912	454
	Skewness	1.0674	-0.1475	0.1338
FFs	Mean	0.0625	0.5859	0.6406
16 samples	Median	0.0625	-0.5	0.625
	St Deviation	4.0396	3.151	1.5103
	Variance	16.3182	9.9286	2.281
	Energy	1759	1912	454
AFs	Mean	9.4585	8.315	4.3167
5 samples with median	Median	7.6667	7.375	4.5
	St Deviation	7.4365	5.4651	2.2719
	Variance	55.3107	29.8674	5.1617
	Skewness	1.6315	0.3334	0.4911

Table 9. Statistics for Step 2.1. and 2.2

Target Scenarios / Statistical Tests		St T1/St T2	St T1/Mv T1	Mv T1/St T2
SFs	Pearson Correlation	0.353	0.5864	0.4139
16 samples	Spearmen Correlation	0.925	0.936	0.9396
	Brownian Correlation	0.2653	0.354	0.2308
	Chi Square Test Prob.	2.77E-18	0.1187	9.73E-23
	Chi Square Test Value	89	33.375	87.43
	Sample Covariance	3.125	8.594	2.3125
	Brownian Covariance	1.1505	2.238	0.6382
FFs	Pearson Correlation	0.6036	0.7898	0.759
16 samples	Spearmen Correlation	N/A	N/A	N/A
	Brownian Correlation	0.4795	0.6303	0.5331
	Chi Square Test Prob.	N/A	N/A	N/A
	Chi Square Test Value	N/A	N/A	N/A
	Sample Covariance	3.453	9.425	3.386
	Brownian Covariance 1	1.034	2.371	0.8783
AFs	Pearson Correlation	0.9391	0.8968	0.9521
4 samples no median	Spearmen Correlation	0.9429	0.934	0.998
	Brownian Correlation	0.9524	0.9517	0.9997
	Chi Square Test Prob.	3.30E-07	0.3074	8.16E-05
	Chi Square Test Value	49	4.7	30.6
	Sample Covariance	11.93	27.4	9.85
	Brownian Covariance	11.6	22.6	9.62

Table 10. Statistics for Step 2.3

Some statistics, at least for this example, are more useful than others. That may translate into larger or smaller coefficients a_k in (26) or we may opt not to use some of the statistics at all. Case in point, Table 9 entry for Spearman correlation coefficient does not look useful because all three numbers are very close to each other. On the other hand, Table 9 shows that Brownian Correlation and Chi Square test give good resolution between the targets, i.e. indicating which signatures appear to be less dependent on each other and less correlated, which would support H_0 hypothesis. As the template is spanned for all targets and all pose angles, there will be a strong correlation at one point indicating matching or H_1. In our limited size example we just wanted to indicate a possible

ISBN: 978-1-61804-290-3

method to find the best match between real time signature against an entry in template Table 1. As far as specific tests for defining H_0, we would look for good discrimination in offered by ratio of variances, means, energy and skewness (Table 9) and Chi Square, Brownian correlation, plus Pearson correlation, sample and Brownian covariance (Table 10). Each of these would correspond to one specific entry in S^{ij} of (26). As stated earlier we would use Monte Carlo simulation to determine specific weights a_k. For example, based on Tables 9 and 10, it appears as if more credence should be given to Chi Square and Brownian correlation compared to other methods.

At the end to summarize the results of this section, based on the statistics in Tables 9 and 10, we conclude that for the three signatures of Figure 2 both stationary and moving Target 1 show less correlation with the stationary Target 2 than to each other, which is what we would expect. Both Tables 9 and 10 use limited size K=16 signature samples. We are currently working on K-64 which would give better overall statistics, plus larger template in Table 1 and 2. Also suitable Monte Carlo simulation will be used to determine a_k coefficients and $O(T_k)$ in (26).

In any case, the outcome of all of the above is a rough estimate of the pose angle, based on the best fit between real time signature and one of the entries in signature template table. This outcome will feed into the next step, i.e. Tracking Filter which we will describe in Part II of the paper.

7. CONCLUSION

In this paper we presented an improved methodology for estimating pose angle of a target with HRR or similarly generated signature. We employ a variety of local and cross statistics in spatial, frequency and amplitude domains, and then compare real time against stored template signatures. The net result is either a match (Hypothesis H_1) or a miss (Hypothesis H_0) based on statistics comparison which produces the best estimate for the pose angle. We plan to extend the results in this paper to more discretized points, such as K = 64 where the statistics would be more effective, in particular Haar transform frequency data.

8. REFERENCES

[1] B. Kahler and E. Blasch, Robust Multi-Look HRR ATR Investigation Through Decision - Level Fusion Evaluation", Proc. 11th International Conference On Information Fusion, July 2008.

[2] B. Kahler, J. Querns, G. Arnold, An ATR Challenge Problem Using HRR Data, Proc. SPIE, Vol. 6970, 2008.

[3] Peter S. Maybeck, Stochastic Models, Estimation, and Control, Vol. 1-3, Academic Press, 1979-1982.

[4] JSTARS- Joint Surveillance and Target Attack Radar System, USA, from The Website for Defense Industries – Air Force, December 2001.

[5] Joint STARS/JSTARS, Intelligence Resource Program, FAS Website, December 2001.

[6] J. Layne and D. Simon, A Multiple Model Estimator for a Tightly Coupled HRR Automatic Target Recognition and MTI Tracking System, SPIE , Orlando, Florida, April 1999.

[7] E. Blasch, Derivation of A Belief Filter for High Range Resolution Radar Simultaneous Target Tracking and Identification, Ph.D. Diss., Wright State University, 1999.

[8] E Blasch and C. Yang, Ten methods to Fuse GMTI and HRRR Measurements for Joint Tracking and ID", Fusion 04, July 2004.

[9] M. I. Hodzic, Estimation Algorithms for Real-Time Airborne Target Tracking, MSc. Thesis, UB, 1985

[10] G. Minkler and J. Minkler, Theory and Application of Kalman Filtering, Magellan Book Company, 1993.

[11] M. S. Grewal, Angus P. Andrews, Kalman Filtering: Theory and Practice Using Matlab, Willey, 2001.

[12] M. I. Hodzic, Monte - Carlo Simulation of a Real Time Target Tracking Filter Operation, Informatika, 1985.

[13] M. I. Hodzic and Radovan Krtolica, Robustness of the Maneuvering Target Tracking Filters, ETAN, 1987.

[14] Y. Bar-Shalom & X. Li, Multitarget - Multisensor Tracking: Principles and Techniques, YBS, New York, 1995.

[15] E. Blasch and M. Bryant, SAR Information Exploitation Using an Information Filter Metric, IEEE, 1998.

[16] S. G. Nikolov, E. Fernandez Canga, J. J. Lewis, A. Loza, D. R. Bull, and C. N. Canagarajah, Adaptive Image Fusion Using Wavelets: Algorithms and System Design, in Multisensor Data and Information Processing for Rapid and Robust Situation and Threat Assessment, Eds. E. Lefebvre, P Valin, IOS Press, 2006.

[17] D. Gross, M. Oppenheimer, B. Kahler, B. Keaffaber, and R. Williams, Preliminary Comparison of HRR Signatures of Moving and Stationary Ground Vehicles, Proc. SPIE, Vol. 4727, 2002.

[18] H-C. Chiang, R.L. Moses, and L.C Potter, Model based classification of radar images, IEEE Transactions on Information Theory, 46, 5 (2000), 1842-1854.

[19] R. Williams, J. Westerkamp, D. Gross, and A. Palomino, Automatic Target Recognition of Time Critical Moving Targets Using 1D High Range Resolution (HRR) Radar, IEEE AES Systems Magazine, April 2000.

[20] R. Wu, Q. Gao, J. Liu, and H. Gu, ATR Scheme Based On 1-D HRR Profiles, Electronic Letters, Vol. 38, Issue 24, Nov. 2002.

[21] S. Paul, A. K. Shaw, K. Das, and A. K. Mitra, Improved HRRATR Using Hybridization Of HMM and Eigen-Template-Matched Filtering IEEE Conf. on Aco., Speech, and Sig. Proc., 2003.

[22] R.A. Mitchell and J.J. Westerkamp, Robust statistical feature based aircraft identification, IEEE Trans. Aerospace & Electronic Systems, 35, 3, 1999.

[23] E. Blasch, J.J. Westerkamp, J.R. Layne, L. Hong, F. D. Garber and A. Shaw, Identifying moving HRR signatures with an ATR Belief Filter, SPIE 2000.

[24] Chethan Parameswariah, Wavelet analysis and filters for engineering Applications, PhD Thesis, LTU, 2003

[25] E. Blasch and S. Huang, Multilevel Feature-based fuzzy fusion for target recognition, Proc. SPIE 2000.

[26] W. Snyder, G. Ettinger and S. Laprise, Modeling Performance and Image Collection Utility for Multiple Look ATR, Proc SPIE 2003.

[27] Gabor J. Szekelyi and Maria L. Rizzo, Brownian Distance Covariance, The Annals of Applied Statistics, Vol. 3, No. 4, 1236–1265, 2009

[28] F. Dicander and R. Jonsson, Comparison of Some HRR Classification Algorithms, Proc. SPIE, Vol. 4382, 2001.

[29] E. Blasch and L. Hong, Simultaneous Tracking and Identification, Conference on Decision Control, Tampa, FL, December 1998

[30] Microsoft web site, www.microsoft.com

Bayesian channel estimation in chaos based DS-CDMA system

Meher Krishna Patel, Stevan M. Berber, *Senior Member, IEEE,* and Kevin W. Sowerby, *Senior Member, IEEE*

***Abstract*—This paper proposes maximum a priori (MAP) channel estimation technique in chaos based code division multiple access (CDMA) system. Two different cases are considered for estimating the fading channel. In the first case, channel coefficients are estimated with the help of chaotic sequences. In the second case estimation is performed without including the chaotic sequence in the estimation algorithm. Simulation results shows that the MAP estimation algorithms performance is better for the first case.**

***Index Terms*—Channel estimation, CDMA, Chaotic sequence, Bayesian estimation, MAP**

I. Introduction

Fading is the phenomena which makes wireless communication more difficult as compare to other communication systems e.g. optical fiber communication and wired communication etc. For many wireless systems, independent of whether time division multiple access (TDMA) or code division multiple access (CDMA) is employed, estimation of channel fading coefficient is necessary for high speed communication. Channel estimates can be updated frame by frame for slower fading rate as compare to frame rate. If channel coefficients changes significantly within the frame then it is necessary to update coefficients iteratively based on symbol by symbol basis [1], [2].

Various estimation methods have been studied in last few decades and each method has its own advantages and disadvantages. Minimum mean square estimators (MMSE) [3], [4], [5] are easy to implement and perform well in flat fading environment. But these estimators require correlation computation and have poor performance for time varying channel estimation. Bayesian estimators [6], [7], [8], [9] used prior knowledge of data to generate posterior analysis. Therefore performance extensively depends on prior informations. On the other hand, neural networks [10], [11], [12] do not require prior knowledge of channel statistic, but there is huge computational burden for training process. Finally, particle filters [13], [14], [15], [16] use the sequential Monte Carlo sampling method to implement recursive Bayesian filter. But these filters have very high computational load for correcting each particle, which results in higher energy consumption. Therefore hardware implementation of these filters are difficult.

The chaotic signals generated from the same chaotic map has high auto correlation and very low cross correlation values. Further, these signals are very sensitive to initial conditions, therefore infinite number of chaotic sequences can be generated from a chaotic map. Hence, chaos based CDMA system gains significant interest among the researchers in last decade [17], [18], [12], [19], [20], [21], [22], [23], [24], [25]. Each user in CDMA system is distinguish by it's spreading code. Bayesian estimators i.e. MAP and maximum likelihood (ML) are extensively studied for CDMA systems with binary spreading codes [26], [27], [28], [29], [30], [31]. However, to our best knowledge, performance of these estimators never studied for chaos based CDMA system.

Objective of this research work is to study the Bayesian channel estimator for chaos based CDMA system for downlink communication. MAP estimator equation is derived for these systems, which needs a prior knowledge of channel statistics. Further, we have also derived the ML estimation equation for considering the case where the mean and variance of the channel is unknown at the receiver. Two algorithms are derived to consider the multiplexed pilot-data case and added pilot-data case. In multiplexed pilot-data case, after demultiplexing, channle estimation can be performed directly on the extracted pilot signal. Whereas for added pilot-data case, pilot needs to be extracted by multiplying corresponding chaotic sequence, before channel estimation process. Performance difference in these two methods have been shown using simulation results.

This paper is organized as follows. In section II chaos based CDMA system with Bayesian estimator is shown. MAP and ML estimation algorithms are derived in section III. Simulation results are shown in section IV. Finally some concluding remarks are given in section V.

II. System model

Fig. 1 shows the baseband representation of the chaos based CDMA system with Bayesian channel estimator. In this figure channel estimation is performed after multiplying the chaotic signal to received signal. The wireless channel is assume to be quasi-static fading channel i.e. path gains are constant over a symbol duration. Then the received signal at user can be described as:

$$y(n) = \left(\sum_{k=1}^{N} \mathbf{s}_k^T(n) \mathbf{C}_k(n) \right) \mathbf{h}(n) + w(n) \qquad (1)$$

where $\mathbf{s}(n) = [s(n), s(n-1), \cdots, s(n-L+1)]^T$ is the transmitted signal, $\mathbf{h}(n) = [h_0(n), h_1(n), \cdots, h_{L-1}(n)]^T$ is the quasi-static time varying channel for k^{th} user and $w(n)$ is the zero mean White Gaussian noise with variance of σ_w^2. L and N represents the total number of paths and users respectively. $\mathbf{C}(n) = diag[\mathbf{c}(n), \mathbf{c}(n-1), \cdots, \mathbf{c}(n-L+1)]$ is the diagonal matrix with elements $\mathbf{c}(\cdot)$ of length 2β known as

ISBN: 978-1-61804-290-3

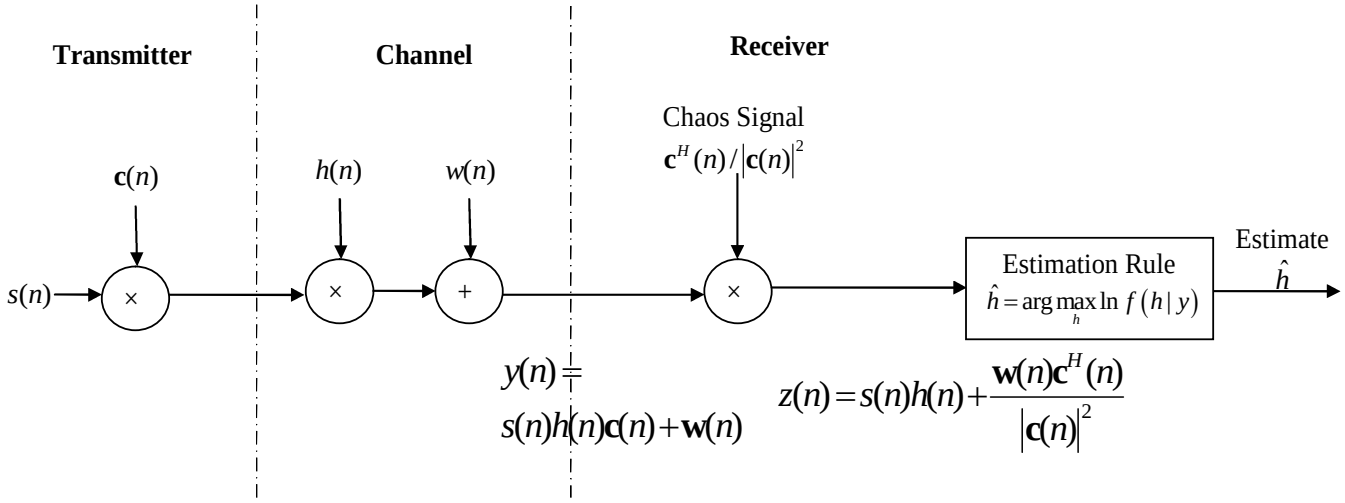


Fig. 1. Block Diagram of Proposed System

spreading factor. Subscript k denotes that the symbol is related to k^{th} user. Channel coefficients are assumed to be Gaussian distributed [27] i.e. $\mathbf{h} \sim N(\mathbf{m_h}, \sigma_\mathbf{h}^2)$ where $\mathbf{m_h}$ and $\sigma_\mathbf{h}^2$ are the mean and variance of the channel respectively.

III. MAP AND ML ESTIMATOR

In this section we have derived Bayesian estimator equations for two cases i.e. for multiplexed pilot-data and add pilot-data case.

A. MAP and ML estimation with multiplexed pilot and user data

If pilot is multiplexed with user data then it can be extracted using demultiplexer at the receiver and can be processed by channel estimator. Here we have to assume that fading and nosie have same effect on pilot and user data symbols. In this case, the condition distribution function $p\left(y(n)|\mathbf{h}(n)\right)$ for k^{th} user is defined as

$$\begin{aligned} &p\left(y(n)|\mathbf{h}(n)\right) \\ &= \frac{1}{\sqrt{2\pi\sigma_w^2}} \exp\left(-\frac{\left(y(n)-\mathbf{s}_k^T(n)\mathbf{C}_k(n)\mathbf{h}(n)\right)^2}{2\sigma_w^2}\right) \end{aligned} \tag{2}$$

Since mean $\mathbf{m_h}$ and variance $\sigma_\mathbf{h}^\mathbf{2}$ of Gaussian distributed channel is known at receiver, therefore MAP estimation algorithm is given by

$$\begin{aligned} &\nabla_\mathbf{h}\left(-\frac{\left(y(n)-\mathbf{s}_k^T(n)\mathbf{C}_k(n)\mathbf{h}(n)\right)^2}{2\sigma_w^2}\right. \\ &\left.-\frac{(\mathbf{h}-\mathbf{m_h})\sigma_\mathbf{h}^{-\mathbf{2}}(\mathbf{h}-\mathbf{m_h})^T}{2} + \text{constants}\Big|_{\mathbf{h}=\hat{\mathbf{h}}}\right) = 0 \end{aligned} \tag{3}$$

Above derivative reduces to following equation (see appendix A for derivation)

$$\begin{aligned} &\hat{\mathbf{h}}_{MAP}(n) = \\ &\mathbf{m_h} + \tfrac{1}{\sigma_w^2}\left(\sigma_\mathbf{h}^{-\mathbf{2}} + \tfrac{1}{\sigma_w^2}\mathbf{C}_k^T(n)\mathbf{s}_k(n)\mathbf{s}_k^T(n)\mathbf{C}_k(n)\right)^{-1} \\ &\times\mathbf{C}_k^T(n)\mathbf{s}_k\left(y(n) - \mathbf{s}_k^T(n)\mathbf{C}_k(n)\mathbf{m_h}\right) \end{aligned} \tag{4}$$

If we do not have a prior knowledge of the channel statistic, then we remove the second term in equation (3) and resultant algorithm is known as ML estimation i.e.

$$\nabla_\mathbf{h}\left(-\frac{\left(y(n)-\mathbf{s}_k^T(n)\mathbf{C}_k(n)\mathbf{h}(n)\right)^2}{2\sigma_w^2} + \text{constants}\Big|_{\mathbf{h}=\hat{\mathbf{h}}}\right) = 0 \tag{5}$$

After solving derivative, we have following ML estimation equation

$$\hat{\mathbf{h}}_{ML}(n) = \left(\mathbf{s}_k^T(n)\mathbf{C}_k(n)\right)^{-1} y(n) \tag{6}$$

B. MAP and ML algorithm for added pilot and user data

For multiplexed pilot and user data, we have to assume same fading effects on both the signals. If we add pilot symbols to user symbols then fading have same effect on both the signals. Therefore same fading and noise effect assumptions can be removed. Further, in this case pilot has to be extracted from data for the channel estimation process. Since the chaotic sequences are orthogonal to each other therefore pilot symbols can be extracted by multiplying the received signal with chaotic sequence of pilot symbols. Multiplying received signal i.e. equation (1) with chaotic signal of k^{th} user we have

$$z(n) = y(n)\frac{\mathbf{C}_k^H(n)}{\mathbf{C}_k(n)\mathbf{C}_k^H(n)} \tag{7}$$

In this case MAP and ML estimation equations are given by (see appendix B for derivation)

$$\begin{aligned} &\hat{\mathbf{h}}_{MAP}(n) = \mathbf{m_h} + \tfrac{1}{\sigma_w^2}\left(\sigma_\mathbf{h}^{-\mathbf{2}} + \tfrac{1}{\sigma_w^2}\mathbf{s}_k(n)\mathbf{s}_k^T(n)\right)^{-1} \times \\ &\mathbf{s}_k\left(z(n) - \mathbf{s}_k^T(n)\mathbf{m_h}\right) \end{aligned} \tag{8}$$

and

$$\hat{\mathbf{h}}_{ML}(n) = \left(\mathbf{s}_k(n)\mathbf{s}_k^T(n)\right)^{-1}\mathbf{s}_k(n)z(n) \tag{9}$$

ISBN: 978-1-61804-290-3

IV. Simulation Results

In the simulation we compare the performance of estimators for three cases. In first case, the pilot is multiplexed with data without multiplying with chaotic sequences at the transmitter. We represent this case as 'Without Chaotic Multiplication' in the simulation results. Similarly other two cases i.e. multiplexed pilot-data and added pilot-data with chaotic sequences multiplication at transmitter, are denoted by 'Before Chaotic Multiplication' and 'After Chaotic Multiplication' respectively in the results.

The value of the spreading factor 2β is 50. Following Chebyshev polynomial function i is used to generate the chaotic sequence [32].

$$x_k = 1 - 2(x_{k-1})^2 \tag{10}$$

where x_k denotes the k^{th} chip value chaotic sequence.

Fig. 2 and Fig. 3 show the channel tracking performance of the three estimators at $0dB$ and $20dB$ SNR conditions respectively. Chaotic sequences spread the data over entire bandwidth during transmission and despreading takes place during reception. Further noise is spread by the chaotic sequence multiplication at the receiver. Due to this spreading of noise, performance of the estimator in the presence of chaotic sequence is better than the without chaotic spreading case, as shown in Fig. 2 and Fig. 3.

From these figures it is clear that performance of all the estimators are improved with increase in the SNR. Since chaotic sequences are directly used for channel estimation as well as noise spreading in multiplexed pilot-data case therefore its performance is better than the added pilot-data case, for lower SNR conditions. For higher SNR conditions performance of both the chaotic estimators are same as shown in Fig. 3.

Finally in Fig. 4, the BER performances are shown. $20dB$ performance improvement can be seen at $SNR = 10dB$ with chaotic spreading sequence over without spreading case. Further, performance improvement can be seen in 'Before Multiplication Case' over 'After Multiplication case' at lower SNR conditions.

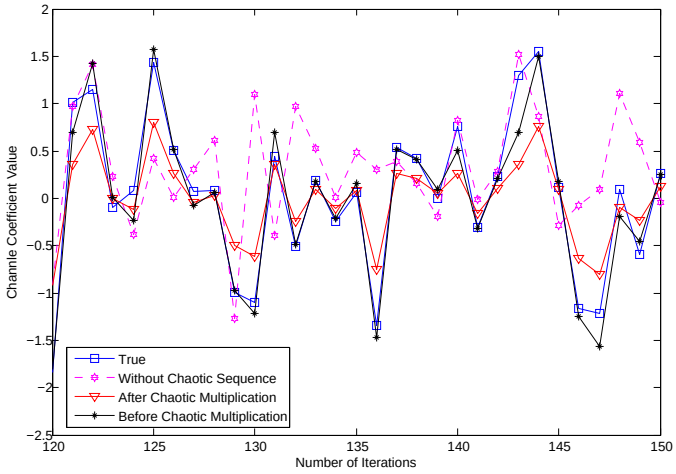


Fig. 2. MAP channel estimators performance, SNR = 0dB

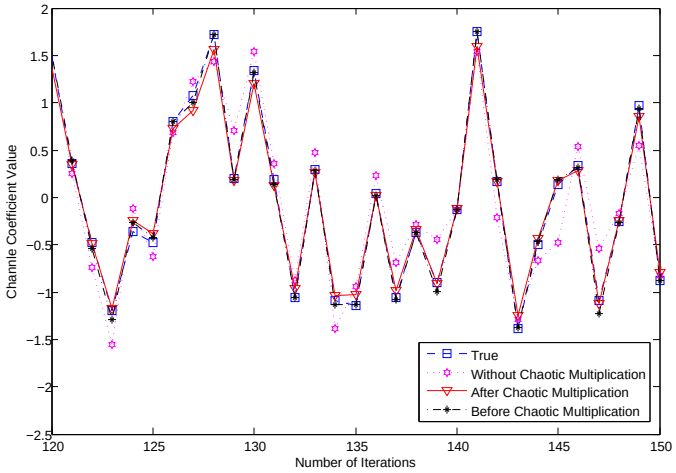


Fig. 3. MAP channel estimators performance, SNR = 20dB

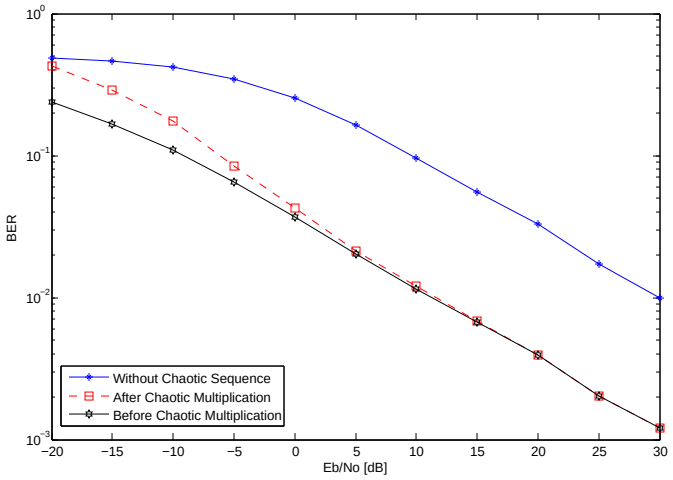


Fig. 4. BER performance of chaos based CDMA system using MAP, $2\beta = 50$

V. Conclusion

In this paper, we propose MAP channel estimation algorithms for chaos based CDMA system. In simulation, channel estimators are compared for various cases. In the first case, channel estimation is performed on the received signal directly i.e. in the presence of chaotic sequence. In second case chaotic sequence is not available for channel estimation because estimation is done after multiplying the received signal with chaotic sequence. The MAP estimators work better in the presence of chaotic sequence because in this case chaotic sequences are used for channel estimation as well as noise spreading. However, for this case pilot and data are multiplexed with each other, therefore we have to assume that pilot and data have same fading and noise effect. Further, if data and pilot have different fading effect then second estimation algorithm can be used.

Appendix A
Derivation of Equation (4)

Rewriting equation (3) after solving derivative, we have

ISBN: 978-1-61804-290-3

$$\hat{\mathbf{h}}(n) = \left(\sigma_{\mathbf{h}}^{-\mathbf{2}} + \frac{1}{\sigma_w^2}\mathbf{C}_k^T(n)\mathbf{s}_k(n)\mathbf{s}_k^T(n)\mathbf{C}_k(n)\right)^{-1} \times \left(\mathbf{m_h}\sigma_{\mathbf{h}}^{-\mathbf{2}} + \frac{\mathbf{C}_k^T(n)\mathbf{s}_k(n)y(n)}{\sigma_w^2}\right) \tag{11}$$

Let

$$\sigma_{\mathbf{h}}^{-\mathbf{2}} + \frac{1}{\sigma_w^2}\mathbf{C}_k^T(n)\mathbf{s}_k(n)\mathbf{s}_k^T(n)\mathbf{C}_k(n) = \mathbf{T} \tag{12}$$

Hence equation (11) becomes

$$\hat{\mathbf{h}}(n) = \mathbf{T}^{-1}\left(\mathbf{m_h}\sigma_{\mathbf{h}}^{-\mathbf{2}} + \frac{\mathbf{C}_k^T(n)\mathbf{s}_k(n)y(n)}{\sigma_w^2}\right) \tag{13}$$

Put the value of $\sigma_{\mathbf{h}}^{-\mathbf{2}}$ from equation (12) to (13), we get equation (4)

APPENDIX B
DERIVATION OF EQUATION (8) AND (9)

Putting the value of $y(n)$ from equation (1) in equation (7), we have

$$z(n) = \mathbf{s}_k^T(n)\mathbf{h}(n) + \left(\sum_{j=1,k\neq j}^{N}\mathbf{s}_k^T(n)\mathbf{C}_k(n)\right)\mathbf{h}(n)\times \frac{\mathbf{C}_k^T(n)}{\mathbf{C}_k(n)\mathbf{C}_k^T(n)} + \frac{w(n)\mathbf{C}_k^T(n)}{\mathbf{C}_k(n)\mathbf{C}_k^T(n)} \tag{14}$$

Since the cross-correlation of two different chaotic signals is very small, hence we can neglect the second term i.e.

$$z(n) \approx \mathbf{s}_k^T(n)\mathbf{h}(n) + \frac{w(n)\mathbf{C}_k^T(n)}{\mathbf{C}_k(n)\mathbf{C}_k^T(n)} \tag{15}$$

Now, following the same steps as in section III-A, we get equation (8) and equation (9).

REFERENCES

[1] L.-M. Chen and B.-S. Chen, "A robust adaptive DFE receiver for DS-CDMA systems under multipath fading channels," *IEEE Transactions on Signal Processing*, vol. 49, no. 7, pp. 1523–1532, 2001.

[2] M.-A. Baissas and A. M. Sayeed, "Pilot-based estimation of time-varying multipath channels for coherent CDMA receivers," *IEEE Transactions on Signal Processing*, vol. 50, no. 8, pp. 2037–2049, 2002.

[3] G. Rice, D. Garcia-Alfs, L. Stirling, S. Weiss, and R. Stewart, "An adaptive MMSE RAKE receiver," in *Conference Record of the Thirty-Fourth Asilomar Conference on Signals, Systems and Computers, 2000.*, vol. 1. IEEE, 2000, pp. 808–812.

[4] H. Cheng and S. C. Chan, "Blind linear MMSE receivers for MC-CDMA systems," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 54, no. 2, pp. 367–376, 2007.

[5] S. Haykin, "Adaptive filter theory, 4 ed." *Prentice Hall*, 2002.

[6] H. L. V. Trees and K. L. Bell, *Detection estimation and modulation theory, pt. I*. Wiley, 2013.

[7] R. Jose and K. V. S. Hari, "Bayesian approach for joint estimation of phase noise and channel in orthogonal frequency division multiplexing system," *IET Signal Processing,*, vol. 8, no. 1, pp. 10–20, 2014, iD: 1.

[8] A. Vempaty, H. He, B. Chen, and P. K. Varshney, "On quantizer design for distributed Bayesian estimation in sensor networks," *IEEE Transactions on Signal Processing,*, vol. 62, no. 20, pp. 5359–5369, 2014, iD: 1.

[9] K. Zhong and S. Li, "On symbol-wise variational Bayesian CSI estimation and detection for distributed antenna systems subjected to multiple unknown jammers," *IEEE Signal Processing Letters,*, vol. 21, no. 7, pp. 782–786, 2014, iD: 1.

[10] S. Guarnieri, F. Piazza, and A. Uncini, "Multilayer feedforward networks with adaptive spline activation function," *IEEE Transactions on Neural Networks*, vol. 10, no. 3, pp. 672–683, 1999.

[11] H. H. Yang and S. ichi Amari, "Adaptive online learning algorithms for blind separation: maximum entropy and minimum mutual information," *Neural computation*, vol. 9, no. 7, pp. 1457–1482, 1997.

[12] J.-H. Kao, S. M. Berber, and V. Kecman, "Blind multiuser detector for chaos-based CDMA using support vector machine," *IEEE Transactions on Neural Networks*, vol. 21, no. 8, pp. 1221–1231, 2010.

[13] S. Wang, L. Cui, L. Stankovic, V. Stankovic, and S. Cheng, "Adaptive correlation estimation with particle filtering for distributed video coding," *IEEE Transactions on Circuits and Systems for Video Technology,*, vol. 22, no. 5, pp. 649–658, 2012, iD: 1.

[14] L. Anping and X. Nan, "Blind multiple frequency offsets and channel estimation using particle filter in cooperative transmission wireless networks," in *IEEE International Conference on Computer and Information Technology (CIT), 2014*, 2014, pp. 837–841, iD: 1.

[15] S. A. Banani and R. G. Vaughan, "Blind channel estimation and discrete speed tracking in wireless systems using independent component analysis with particle filtering," *IET Communications,*, vol. 6, no. 2, pp. 224–234, 2012, iD: 1.

[16] H. Hu, S. Zhang, and H. Li, "Individual channel tracking for one-way relay networks with particle filtering," in *IEEE Global Communications Conference (GLOBECOM), 2014*, 2014, pp. 3198–3202, iD: 1.

[17] F. C. Lau and K. T. Chi, *Chaos-based digital communication systems: Operating principles, analysis methods, and performance evaluation*. Springer, 2003.

[18] R. Rovatti, G. Mazzini, and G. Setti, "Enhanced RAKE receivers for chaos-based DS-CDMA," *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, vol. 48, no. 7, pp. 818–829, 2001.

[19] S. Berber, "Probability of error derivatives for binary and chaos-based CDMA systems in wide-band channels," *IEEE Transactions on Wireless Communications*, 2013.

[20] M. Coulon and D. Roviras, "Multi-user receivers for synchronous and asynchronous transmissions for chaos-based multiple-access systems," *Signal Processing*, vol. 89, no. 4, pp. 583–598, 2009.

[21] G. Kaddoum, D. Roviras, P. Charg, and D. Fournier-Prunaret, "Accurate bit error rate calculation for asynchronous chaos-based DS-CDMA over multipath channel," *EURASIP Journal on Advances in Signal Processing*, vol. 2009, p. 48, 2009.

[22] G. Kaddoum, P. Charg, D. Roviras, and D. Fournier-Prunaret, "Performance analysis of differential chaos shift keying over an AWGN channel," in *International Conference on Advances in Computational Tools for Engineering Applications, 2009. ACTEA'09.* IEEE, 2009, pp. 255–258.

[23] R. Rovatti, G. Mazzini, and G. Setti, "Enhanced RAKE receivers for chaos-based DS-CDMA," *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications,*, vol. 48, no. 7, pp. 818–829, 2001.

[24] S. Vitali, R. Rovatti, and G. Setti, "Improving PA efficiency by chaos-based spreading in multicarrier DS-CDMA systems," in *IEEE International Symposium on Circuits and Systems, 2006. ISCAS 2006. Proceedings. 2006*, 2006, pp. 4 pp.–1198, iD: 1.

[25] M. G. Zia, "Performance of chaos-based MC-MC-CDMA in frequency selective fading channel," *International Journal of Scientific and Engineering Research*, vol. 4, no. 1, pp. 1–4, 2013.

[26] B. Hu, I. Land, R. Piton, and B. H. Fleury, "A Bayesian framework for iterative channel estimation and multiuser decoding in coded ds-cdma," in *IEEE Global Telecommunications Conference, 2007. GLOBECOM'07.* IEEE, 2007, pp. 1582–1586.

[27] E. Aydin and H. A. irpan, "Bayesian-based iterative blind joint data detection, code delay and channel estimation for DS-CDMA systems in multipath environments," in *7th International Wireless Communications and Mobile Computing Conference (IWCMC), 2011*. IEEE, 2011, pp. 1413–1417.

[28] L. Wu, G. Liao, C. Wang, and Y. Shang, "Bayesian multiuser detection for CDMA system with unknown interference," in *IEEE International Conference on Communications, 2003. ICC '03.*, vol. 4, 2003, pp. 2490–2493 vol.4, iD: 1.

[29] Z. Yang, B. Lu, and X. Wang, "Blind Bayesian multiuser receiver for space-time coded MC-CDMA system over frequency-selective fading channel," in *IEEE Global Telecommunications Conference, 2001. GLOBECOM '01.*, vol. 2, 2001, pp. 781–785 vol.2, iD: 1.

[30] Q. Yu, G. Bi, and L. Zhang, "Bayesian blind multiuser detection for long code multipath DS-CDMA systems," in *International Conference*

on Communications, Circuits and Systems, 2004. ICCCAS 2004. 2004, vol. 1, 2004, pp. 84–88 Vol.1, iD: 1.

[31] A. Vosoughi and A. Scaglione, "Optimal training designs for Bayesian channel estimators with application in CDMA systems," in *IEEE/SP 13th Workshop on Statistical Signal Processing, 2005*, 2005, pp. 1348–1353, iD: 1.

[32] G. Kaddoum, P. Charg, D. Roviras, and D. Fournier-Prunaret, "Comparison of chaotic sequences in a chaos based DS-CDMA system," in *Proceedings of the International Symposium on Nonlinear Theory and its Applications, Vancouver, Canada*, 2007.

DDS On Top Of FlexRay Driver: Simulink Blockset Implementation of FlexRay Driver for SAE Application using the DDS Middelware

Zouhaira Abdellaoui, Rim Bouhouch, Houda Jaouaini, Salem Hasnaoui
University of Tunis El Manar
National Engineering School of Tunis LR-99-ES21
Communications Systems Research Laboratory SYSCOM, Tunis
1002, Tunis, Tunisie
Zouhairaabdellaoui@yahoo.fr , rim.bouhouch@yahoo.fr, jouani_houda@yahoo.fr, Salem.Hasnaoui@enit.rnu.tn

Abstract**—**Thanks to its several features such as flexibility, fault-tolerance, and determinism and high-speed, the FlexRay networks was known as one of the newest x-by wire communication systems [1] known for their speed and performance insuring communication over a shared medium. It offers reliable and real-time capable high-speed data transmission between electrical and mechatronic components. In the same context, the real-time middleware Data Distribution Service (DDS) is an appropriate alternative for the standard vehicular middleware. In this paper, we have implemented a Simuink Blockset of FlexRay's Driver in order to validate its performances and design and to provide current and future innovative functions into distributed systems within automotive applications.

The proposed blockset is dedicated for the SAE application. The Society of Automotive Engineers SAE benchmark model is normally connected by the CAN bus we extended it to the FlexRay Bus. We have applied the real-time middleware Data Distribution Service (DDS) in order to guarantee the QoS of SAE benchmark using the FlexRay protocol.

Keywords**—**DDS; Embedded MATLAB; FlexRay; Suspension; SAE Benchmark ;Simulink Blockset

.

I. INTRODUCTION

Vehicles real-time Networks are based on the interaction between the driver, reflecting the working process of the network, and the application in need of communication [1]. In order to guarantee the intra-vehicular confidentiality Networks and to improve security, safety luxury and reliability in the automotive sector, we chose the FlexRay Networks instead the CAN.

FlexRay is a new communication system that offers reliable and real-time capable high-speed data transmission. This protocol protocol is meeting safety critical applications performance requirements (flexibility, fault-tolerance, determinism, high-speed…) [2][3] .

We have exploited the platform of a vehicular network based on the Society of Automotive Engineers (SAE).

There for, to validate the vehicle system design and to guarantee the arrival of the right data on the right time we have chosen to work within the bus FlexRay.

In this paper, we have implemented a Simulink blockset of FlexRay driver, we have used the SAE Benchmark extended Model as application.

We have applied the studied approach on a new vehicle benchmark developed in [4]. The author added to the original benchmark a number of nodes and messages to better represent the complexity of today's vehicles and to model some options responsible for improving vehicle safety, reliability, cost, and luxury. We have adapted it to the FlexRay protocol instead of CAN network. FlexRay protocol is known for its speed and performances which uses ranges from planes to cars networks to insure communication over a shared medium [5].

We applied the real-time middleware Data Distribution Service (DDS) to this model in order to connect the designed sub-blocks [6].

In fact, DDS is an appropriate alternative for the standard vehicular middleware considering its ability to handle Quality of Service (QoS) parameters.

Since the association between the DDS middleware and real-time network such as FlexRay enable automotive applications to run in a hardware and software environment meeting their timing requirements, thanks in part to the rapid access technique provided by the network and secondly to the temporal QoS guarantees by DDS.

II. RELATED WORK

A. Data Distribution Service DDS

The Data Distribution Service (DDS) is an open standard managed by the Object Management Group (OMG) and representing the first general-purpose middleware standard that addresses challenging real-time requirements. In fact, DDS handles a wide range of data flows, from extremely high performance combat management or flight control to slower command sequences. This specification describes two levels of interfaces: A lower level Data-Centric Publish-Subscribe (DCPS) that is targeted towards the efficient delivery of the proper information to the proper recipients and an optional

ISBN: 978-1-61804-290-3

higher-level Data-Local Reconstruction Layer (DLRL), which allows for a simpler integration into the application layer . It is based on publish-subscribe communication model illustrated in fig1, and supports both messaging and data-object centric data models [7]. Also, DDS provides the DEADLINE QoS Policy, LATENCY_BUDGET Qos Policy, TRANSPORT_PRIORITY QoS Policy and other policies specifically targeted to minimum latency, predictable real-time operation in high-performance distributed data critical systems [8].

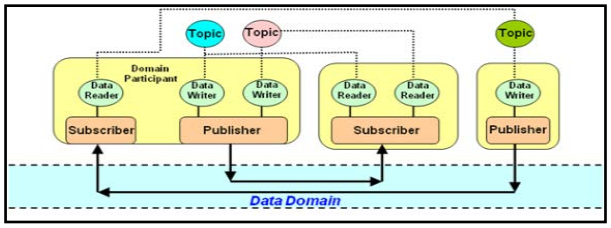


Fig.1.DDS Communication Model : Publication/souscription

B. FlexRay Bus

FlexRay Networks are one of the newest x-by wire communication systems .It is a real-time communication bus designed to operate at speeds of up to 10 M bits/s. He is developed by a consortium that includes automobile builder. It provides time-triggered and an event triggered architecture. Data are transmitted in payload segment containing between 0 and 254 bytes of data, 5 bytes for the Header segment and 3 bytes for the trailer segment. The topology may be linear bus, star or hybrid topologies. This bus contains two channels; each node could be connected to either one or both channels [9].

This bus contains a static segment for time triggered messages and a dynamic segment for event triggered messages. In time triggered networks, nodes only obtain network access at specific time periods, also called time slots. In event triggered networks, nodes may obtain network access at any time instant. The static (ST) segment and the dynamic (DYN) segment lengths can differ, but are fixed over the cycles. Both the ST and DYN segments are composed of several slots. The first two bytes of the payload segment are called message ID, this is used only in dynamic segment. The message ID can be used as a filterable data.

In order to guarantee safety and performance required for time-critical applications (flexibility, fault tolerance, determinism, speed FlexRay), we have to develop its driver (API and Services) and implement its Simulink Blockset.

III. Simulink Blockset Implementation of Flexary Driver

In this paper, we focus our interest on studying a new methodology of development of automotive Nteworks's driver based in Model Based design approach. It is the implementation of Simulink Blockset.This implementation will be a good and reliable tool regarded as a solution to resolve software development process. It facilitates the modeling that is required as design basis. It guarantees the complex systems validation.

A .Application Model

We have chosen to work within a platform of a vehicular network based on the extended SAE benchmark .In this system a set of network processors subsystems produces routing data. It must be distributed along the vehicular network. It is based on the SAE Benchmark presented in fig1**.** The author has added to the original benchmark a number of nodes and messages to better represent the complexity of today's vehicles and to model some added options responsible for improving vehicle safety, reliability, cost, and luxury. The resulting architecture is composed of 15 nodes connected by the FlexRay bus [10] [11].

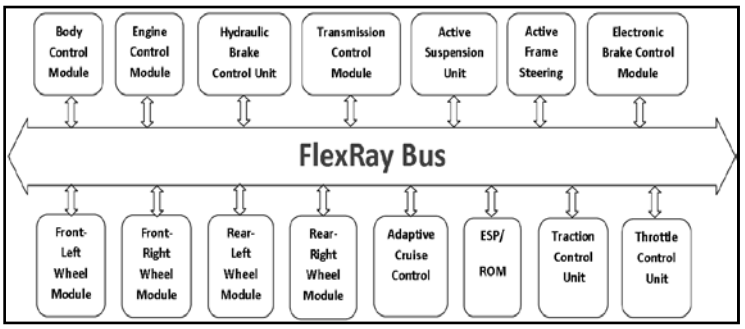


Fig.2. The Application Model

B .Development of FlexRay Driver

In this section, we focus our interest on developing the driver of our automotive Networks. It serves communication between its Network controller MB88121B and the SAE Benchmark application with the incorporation of the real time operating system µC/OS-III [12]**.**

We have begun with the development of the API (Application Programming Interface) of our protocol. The API is a program interface between our SAE Benchmark application and the FlexRay Driver which consist of procedures and functions that give access to the various fields composing the network screen (data, ID,...).

The protocolar treatement present the first step of Driver development.It consist of set of services identified by functions **Get** and **Set**.

This protocol processing requires the assignment to each field of the header of the FlexRay protocol the two function **Get** and **Set** as shown in (1) and (2).

Get_Nom_of_feild () (1)

Set_Nom_of_feild () (2)

Also, we must provide both services **Send ()** and **Receive** ().Thus the number of API function that we have developed will include all function Set, Get,Send and Receive (3).

The number of function of API≈

2x [Number of feild of header +1] (3)

Let's take the example of WRHS1 (Write Header Section Register 1) register .It is a Read/ Write(R/W) register consisting of 7 groups such as the [bit 27] PPIT: Payload

ISBN: 978-1-61804-290-3

Preamble Indicator Transmit. This service is responsible for monitoring the payload preamble indicator status of the transmitted frame. Get PPIT () and Set PPIT ().

The second step of developing of services according to the specificity of the FlexRay Network Controller. The FlexRay protocol has 453 registers and 1248 services that we have developed.We have to affect to ech group of bit of all controller register the two function set and get.

If the register is Read/Write,we have to affect him two function Set and Get as expressed in (4) and (5).

Get_Nom_of_groupment () (4)

Set_Nom_of_groupment () (5)

However, if the register is Read Only we affect him only the Read function. We have to add the three services:Read(),Write() and ImmediateWrite.Ther for we have:

The number of services of driver ∈ [Number of groupment, 2x Number of groupment] +3 (6)

C .Simulink Blockset Implementation of FlexRay Driver

In order to exploit the different services of the FlexRay Driver in Simulink models we have developed a new methodology of implementation according to the Model-Based Design.

Since we find in Matlab library the Simlink Blockset only of CAN protocol.This approach presents a solution conceived for software development process. Thus it facilitates the modeling that is required as design basis. We have used the S-Function tool to implement this Simulink Blokset and specially the LCT (Legacy Code Tool).The following figure illustrates the process of this tool.

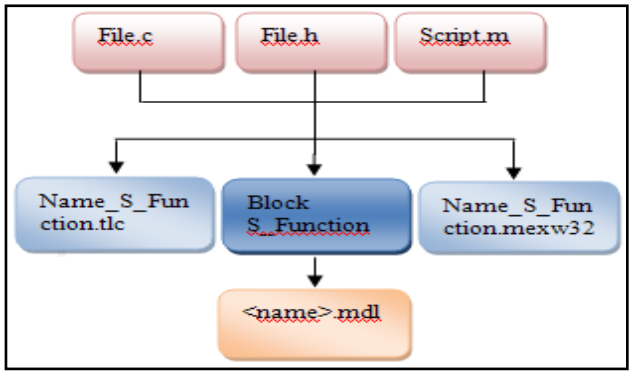


Fig.3. Process generation of Blok S-Function

We have used this tool to implement the Blockset Simulink of MB88121B Driver to be present in the Matlab Library.As shown in the Fig.4.

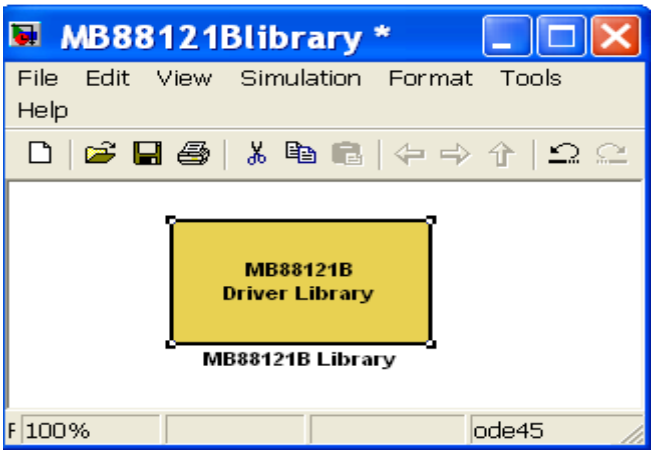


Fig.4. Simulink Blockset Implementation of FlexRay Driver

This Blockset consist of FlexRay Interface,Interruption Handing blocks and Execution Profiling blocks .Fig5.

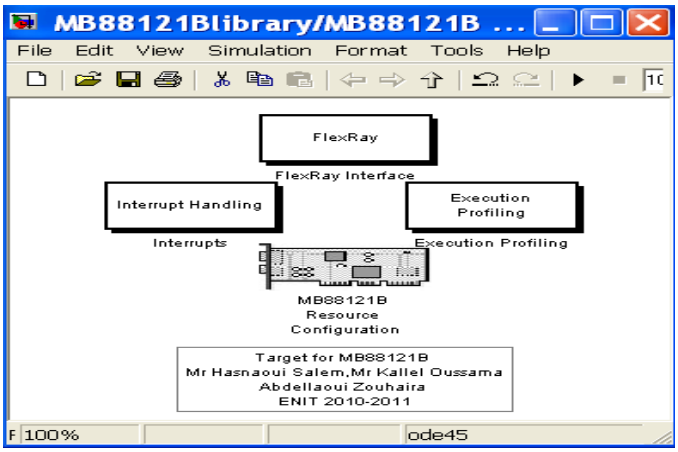


Fig.5. Blockset of MB88121B Driver

The following figure shows the contents of FlexRay Interface.

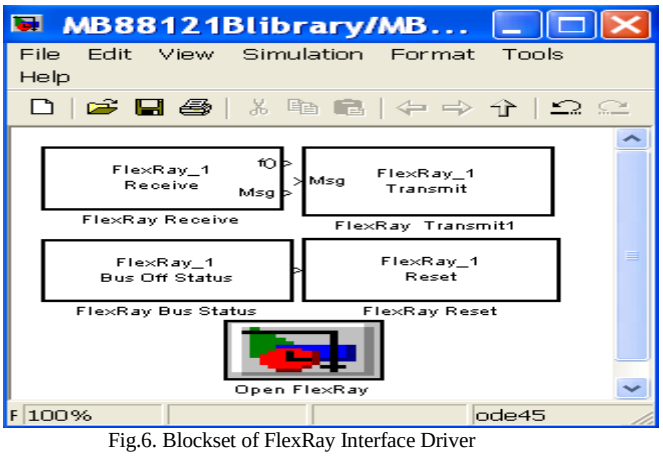


Fig.6. Blockset of FlexRay Interface Driver

V.TEST AND RESULT

In this section, we have tested the FlexRay Driver that we have developed and the blockset that we have implemented in Simulink.We have used the **IAR** (Integrated ARM) Embedded

ISBN: 978-1-61804-290-3

Workbench as tool pack designed for the development of the ARM environment to validate our development of FlexRay Driver.Take for example the function Set of VIEW register that belong to OBCR (Output Buffer Command Request Register). As shown in fig.7.

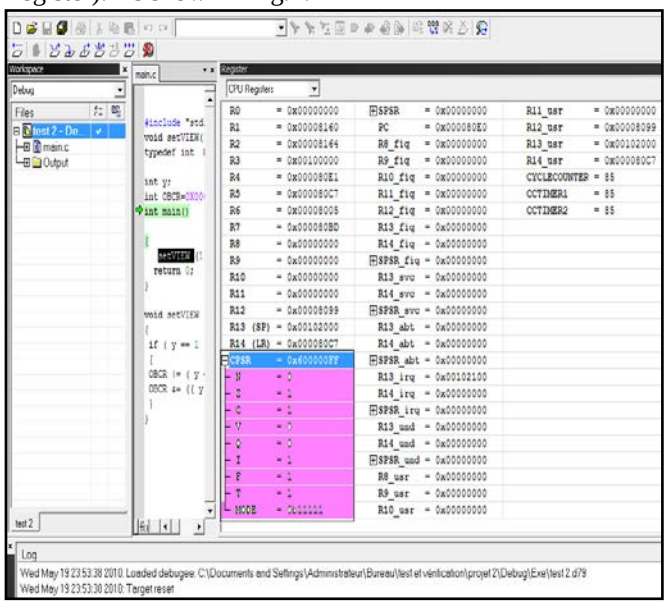


Fig.7. the behavior of the CPU registers of SETVIEW () Function

Also, we have tested the Blockset Simulink that we have implemented, let's take the example of FlexRay-Receive block.Fig.8.

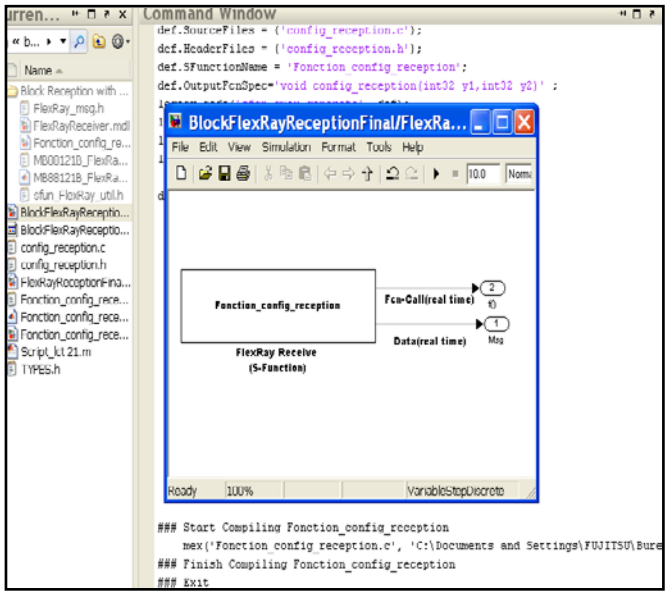


Fig.8. Compilation and execution of FlexRay receive code

VI.CONCLUSION

This paper presents a new methodology of development of automotive Networks driver. It guarantees the achievement of reliability and luxury in the developed Vehicle model.In my present paper, we focused our interest on the development of FlexRay driver and its controller Simulink Blockset implementation using the SAE Benchmark model with its different sub-blocks. We have proposed to use DDS on top of the real-time network FlexRay to take advantage of its high speed and to profit of the DDS QoS management in an automotive context

In my future work, I will integrate the Simulink blockset of other blocks of the SAE Benchmark model extended.

ACKNOWLEDGMENT

The researchers presented in my paper are thanks to the support of many people. We wish to express our gratitude to the SYSCOM ENIT members for their help and assistance.

REFERENCES

[1] N. Navet (editor) and F. Simonot-Lion, "Automotive embedded systems handbook". Industrial Information Technology Series. CRC Press, 2009.

[2] FlexRay Consortium, FlexRay Communications System-Protocol Specification, Version 2.1, Revision A, 2005.

[3] I. Broster. "Flexibility in dependable real-time communication". PhD Thesis, Department of Computer Science, University of York, August 2003.

[4] M. Utayba Mohammad, N. Al-Holou, "Development of an Automotive Communication Benchmark", Canadian Journal on Electrical and Electronics Engineering, Vol. 1, No. 5, August 2010.)

[5] D. Millinger, R. Nossal, « FlexRay CommunicationTechnology », The Industrial Communication Technology Handbook, CRC Press, Taylor & Francis, éditeur R. Zurawski, ISBN 0-8493-3077-7, janvier 2005.

[6] « Data Distribution Service for Real-time Systems», Version 1.2. OMG Available Specification formal/07-01-01.

[7] R. Bouhouch, H. Jaouani, Amel Ben Ncira3, S. Hasnaoui, "DDS on Top of FlexRay Vehicle Networks: Scheduling Analysis", International Journal of Computer Science and Artificial Intelligence, Vol. 3 Iss. 1, PP. 10-26, Mar. 2013

[8] Teton SNA Core Team, "DDS vs DDS4CCM", Northrop Grumman, the Teton Project, July 13 2011.

[9] D. Millinger, R. Nossal, « FlexRay CommunicationTechnology », The Industrial Communication Technology Handbook, CRC Press, Taylor & Francis, éditeur R. Zurawski, ISBN 0-8493-3077-7, janvier 2005.

[10] Christopher A. Lupini, "Vehicle Multiplex Communication - Serial Data Networking Applied to Vehicular Engineering", SAE, April 2004.

[11] Tindel, K., Burns, A., "Guaranteeing Message Latencies On Control Area Network (CAN)", Real-Time Systems Research Group, Department of Computer Science, University of York, England, 1994. [Online] available.

[12] A J. J. Labrosse. "MicroC/OS-II The Real Time Kernel". Miller Freeman, Inc, United States of America, 1999.

Cooperative guidance of multi-missile system based on extreme learning machine

Xing Wei, Yongji Wang, Shuai Dong, Lei Liu

***Abstract*—The cooperative guidance problem of multi-missile system, aiming at simultaneous attack on a static target, is considered. A distributed guidance strategy is proposed based on classic proportional guidance law. A distributed protocol is designed to coordinate the time-to-go commands of all missiles. Then an expert system, consisted of two artificial neural networks (ANNs) using extreme learning machine (ELM), can regulate the local proportional guidance coefficient of each missile according to the command. All missiles will arrive at the target at the same time with the assumption that the network is connected. An example is given to show the validity of the proposed method.**

***Keywords*—multi-missile, cooperative proportional guidance, consensus, extreme learning machine, saturation attack**

I. Introduction

Saturation attack, which involves simultaneous attack from different missiles in a communication network, is an important combat manner to penetrate the missile defence system. In fact, a group of well-organized missiles of low-cost and poor-performance may yield better results than a single excellent one. The key to cooperative guidance of multi-missile system is that all missiles reach the target at the same time when saturation attacking.

Cooperative control theories have been researched broadly with respect to different agents, such as unmanned aerial vehicle [1, 2], satellite [3] and some abstract objects [4, 5]. A cooperative control strategy for achieving cooperative timing among teams of vehicles which is based on coordination variables and functions, was developed [6]. However, there are only a few existing literatures considering the cooperative timing problem of missiles. In [7], with a combination of the well-known proportional navigation guidance (PNG) law and the feedback of the impact time error, an impact time control guidance (ITCG) law for salvo attack of anti-ship missiles was presented and could be used to guide multiple missiles to hit a stationary target simultaneously at a desirable impact time. Based on this law, a cooperative PNG law was proposed by introducing a new concept of the variance of time-to-go of multiple missiles in [8]. With the weighting average consensus algorithm [9], a cooperative guidance scheme based on the ITCG law, where distributed coordination algorithms and local guidance laws were combined together, was developed in [10].

Different from the unmanned aerial vehicle or satellite, the maneuver flight of missile is mainly based on aerodynamic force and the guidance of it doesn't include task assignment, cooperative path planning and path tracking. This makes it particularity hard to design a cooperative strategy for the multi-missile system. In the case of a group of missiles intercepting a single maneuvering target, an optimal cooperative guidance law based on comprehensive cost function of missiles, was derived with the constraint of a relative intercept angle [11]. A time-cooperative guidance architecture which is a centralized coordination control form, was proposed based on leader- follower strategy [12]. Using the dynamic surface control theory and disturbance observation technology, [13] developed a novel integrated guidance and control law, which under the condition of variable missile velocity, unknown uncertainties and limited actuator deflection angle can realize cooperation of impact time and flight position for multiple missiles during their cooperative attack. All these methods adopt a centralized control manner that one missile must exchange information with all the other missiles.

In order to achieve the simultaneous attack of missiles, the consensus of time-to-go (remaining flight time, i.e. arriving time) is considered in this paper. However, the relationship between the t_{go} command and the local proportional guidance law is hard to obtain directly and it needs to turn to the fitting methods. An artificial neural network (ANN), with a fixed number of numeric inputs and outputs, can be regarded as a complex nonlinear function. By learning iteratively large numbers of samples, the ANN stores the mapping relation between inputs and outputs. Due to this ability of self-learning and self-adaption, ANN is widely used in expert systems. Despite ANN is of high precision, the learning speed of traditional ANN is so slow that can't be used in on-line cooperative guidance. ANN with ELM is introduced in the paper to overcome this problem.

In this paper, we design a cooperative guidance strategy to achieve simultaneous attack based on expert system using ELM, which just requires that the communication network is connected. Via a distributed protocol through the connected network, which aims at asymptotical consensus of time-to-go (remaining flight time, i.e. arriving time) commands, the t_{go} commands of all missiles are coordinated. Then the local

This work was supported in part by the National Nature Science Foundation of China (NO. 61203081 and 61174079), Doctoral Fund of Ministry of Education of China (NO. 20120142120091), Fundamental Research Funds for the Central Universities of HUST (NO. 2013054), and Precision Manufacturing Technology and Equipment for Metal Parts (NO. 2012DFG70640).

Xing Wei is with the School of Automation, Key Laboratory of Ministry of Education for Image Processing and Intelligent Control, Huazhong University of Science and Technology, Wuhan, China. (e-mail: weixingkuaile@hust.edu.cn).

Corresponding author, YongJi Wang (Invited-Dimitrova) is a professor at the School of Automation, Huazhong University of Science and Technology, Wuhan, China. (e-mail: wangyjch@mail.hust.edu.cn).

Shuai Dong is with the School of Automation, Huazhong University of Science and Technology, Wuhan, China. (e-mail: hustacds@hust.edu.cn).

proportional guidance law of each missile is regulated by the expert system according to the t_{go} commands.

The remainder of this paper is organized as follows. Section II describes the cooperative guidance problem of multiple missiles. Section III proposes a new cooperative guidance (CG) strategy based on the expert system with ANNs using ELM. Section IV shows a simulation result to illustrate the validity of the proposed CG. Finally, some concluding remarks to this paper are presented in Section V.

II. Cooperative Guidance Problem Formulation

Consider the cooperative guidance of n similar missiles which are denoted as $1,2,\cdots,n$. The objective is to make all missiles arrive at a static target at the same time. Assume that the missiles can only change the direction of the velocity, which means that the speeds of all missiles are an equal constant v and always perpendicular to the accelerations. In general, the flight of missile is decomposed into the movement in longitudinal and lateral plane when designing the terminal guidance law. Furthermore, when missiles approach the static target, the range of lateral motion is quite small. Hence, only consider the longitudinal motion, and ignore the influence of gravity. Without loss of generality, all n missiles are assumed to move in the same plane and normalize the speed v to 1, meanwhile the missiles collision is ignored. The relative motion of missile i is depicted in Fig. 1, and the corresponding equation is written as

$$\begin{aligned}
&\begin{bmatrix}\dot{x}_i\\ \dot{y}_i\end{bmatrix}=\begin{bmatrix}v_{xi}\\ v_{yi}\end{bmatrix}\\
&\begin{bmatrix}\dot{v}_{xi}\\ \dot{v}_{yi}\end{bmatrix}=\begin{bmatrix}a_{xi}\\ a_{yi}\end{bmatrix}=a_i\begin{bmatrix}v_{xi}\Big/\sqrt{v_{xi}^2+v_{yi}^2}\\ -v_{yi}\Big/\sqrt{v_{xi}^2+v_{yi}^2}\end{bmatrix}\\
&r_i=\sqrt{x_i^2+y_i^2}\\
&a_i=\sqrt{a_{xi}^2+a_{yi}^2}\\
&v=1=\sqrt{v_{xi}^2+v_{yi}^2}
\end{aligned}\tag{1}$$

where $[x_i \quad y_i]^{\mathrm{T}}$ is the position vector, $[v_{xi} \quad v_{yi}]^{\mathrm{T}}$ is the velocity vector, $[a_{xi} \quad a_{yi}]^{\mathrm{T}}$ is the acceleration vector, v is the velocity scalar, a_i is the acceleration scalar and r_i is the distance from missile i to the target. Define λ_i and θ_i as the anticlockwise angles from the My_i to the vector $[x_i \quad y_i]^{\mathrm{T}}$ and $[v_{xi} \quad v_{yi}]^{\mathrm{T}}$ respectively, and denote them as

$$\lambda_i=\angle(x_i,y_i),\ \theta_i=\angle(v_{xi},v_{yi}) \tag{2}$$

Because the sensor assembled on the missile can only measure r_i,λ_i,θ_i, it is difficult to design a guidance law a_i based on the system (1) which is established on the XOY rectangular coordinate system. We often use the model expressed in a polar coordinate system below

$$\begin{aligned}
&\dot{r}_i=-\cos(\lambda_i-\theta_i)\\
&\dot{\lambda}_i=\sin(\lambda_i-\theta_i)/r_i\\
&\dot{\theta}_i=-a_i
\end{aligned}\tag{3}$$

The classic proportional guidance law is

$$a_i=-k_i\dot{\lambda}_i \tag{4}$$

where, k_i is the proportional coefficient. Since the overload imposed on the missile is limited, usually $k_i\le 6$.

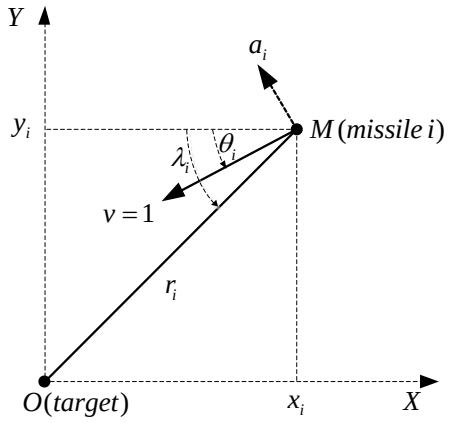


Fig. 1 The relative motion of missile i

Substitute (4) into the closed loop system of (3), we obtain

$$\begin{aligned}
&\dot{r}_i=-\cos(\lambda_i-\theta_i)\\
&\dot{\lambda}_i-\dot{\theta}_i=(1-k_i)\sin(\lambda_i-\theta_i)/r_i
\end{aligned}\tag{5}$$

How multi-missiles cooperatively attack the ground static target is shown in Fig. 2. n missiles fly towards the target from different distances and directions respectively, besides, the communication topology of them is undirected and strongly connected. Namely every missile can only exchange information with its neighbors.

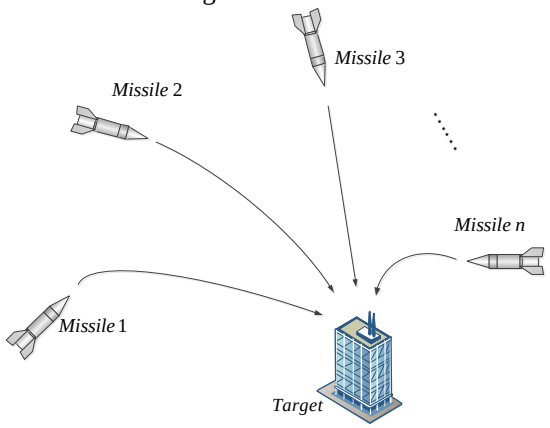


Fig. 2 The concept map of multi-missiles cooperative saturation attack

Then the objective of cooperative guidance is to find the appropriate acceleration command a_i and make all missiles reach the target simultaneously, i.e., $r_1(t_f)=r_2(t_f)=\cdots=0$. The time-to-go of missile i is a monotonic decreasing function of k_i, so we can regulate k_i to adjust the arrival time.

In order to achieve simultaneous attack, an indirect cooperative guidance strategy is adopted as shown in Fig. 3. The protocol makes the expected t_{go} of all missiles achieve consensus asymptotically. Then, consisted of ANNs trained offline, these local expert systems transform the t_{go} command to corresponding proportional coefficient k_i. The details will be introduced in the next section.

III. Algorithm Description

Denote t_{go} of missile i as T_i for convenience. Since the

closed loop system of (5) is autonomous, if $r_i(t_0)$ and $\lambda_i(t_0)-\theta_i(t_0)$ are known and $k_i(t)=k_i(t_0)$ when $t\geq t_0$, then $T_i(t_0)$ is determined by the function

$$T_i(t_0)=f(r_i(t_0),\lambda_i(t_0)-\theta_i(t_0),k_i(t_0)) \quad (6)$$

Similarly, when the required time-to-go $T_i(t_0)$ is given, there exists a corresponding $k_i(t_0)$ that can be obtained by

$$k_i(t_0)=f^{-1}(r_i(t_0),\lambda_i(t_0)-\theta_i(t_0),T_i(t_0)) \quad (7)$$

Fig. 4 shows some curves of $T_i(t_0)$ with respect to $k_i(t_0)$. Although it is hard to derive the analytic form of the function $f(\cdot)$ and its inverse $f^{-1}(\cdot)$, fortunately it can resort to the ANNs.

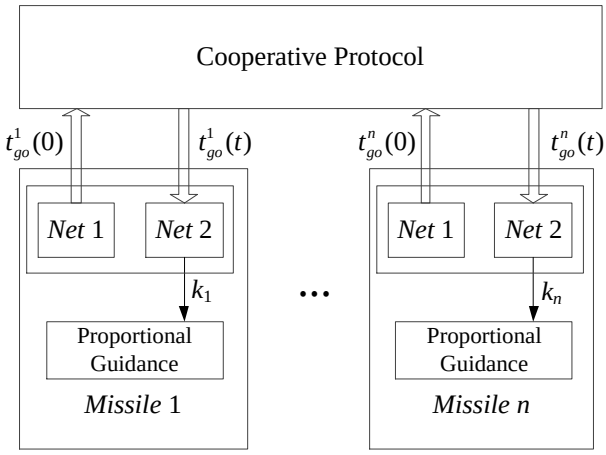


Fig. 3 The framework of cooperative guidance strategy

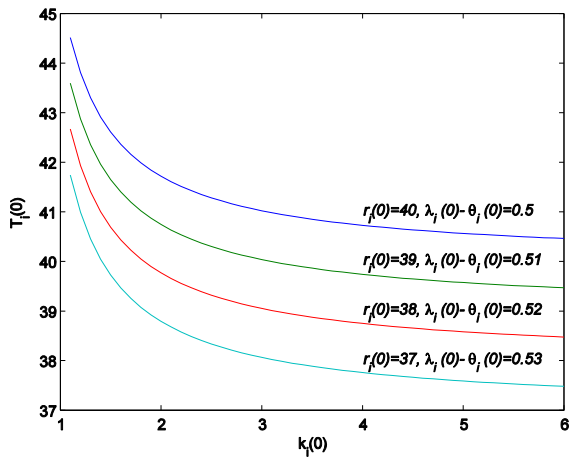


Fig. 4 The mapping relation between proportional guidance coefficient and time-to-go

A. *Extreme learning machine*

ELM is a simple learning algorithm for single-hidden layer feedforward neural networks (SLFNs) which achieves fast learning through increasing the number of hidden nodes and obtains good generalization performance [14]. In the training process with ELM, the input weights (linking the input layer to the hidden layer) and hidden layer biases of SLFNs can be assigned arbitrarily and don't need to adjust. After chosen randomly, SLFNs can be simply considered as a linear system and the output weights (linking the hidden layer to the output layer) can be analytically determined through simple generalized inverse operation of the hidden layer output matrices. Only by setting the number of hidden nodes, it's easy to get the single optimal solution of SLFNs with ELM which has a learning speed much faster than traditional feedforward neural network learning algorithms like back-propagation (BP) algorithm while obtaining better generalization performance. ELM not only tends to reach the smallest training error but also obtains the smallest norm of weights.

The structure of standard SLFNs is described in Fig. 5. The network has n input nodes, l hidden nodes and m output nodes. Let $\boldsymbol{w}_i=[w_{i1},w_{i2},\ldots,w_{in}]^T$ denote the weight vector connecting the ith hidden node and the input nodes, b_i denote the threshold of the ith hidden node, and $\boldsymbol{u}_i=[u_{i1},u_{i2},\ldots,u_{im}]^T$ denote the weight vector connecting the ith hidden node and the output nodes. There exists a training set with s arbitrary distinct samples, whose input vectors and output vectors are $\boldsymbol{x}_i=[x_{i1},x_{i2},\ldots,x_{in}]^T\in\boldsymbol{R}^n$ and $\boldsymbol{o}_i=[o_{i1},o_{i2},\ldots,o_{im}]^T\in\boldsymbol{R}^m$. The standard SLFNs with l hidden nodes and activation function $g(x)$ can approximate these s samples with zero error means that $\sum_{j=1}^{s}\left\|\boldsymbol{y}_j-\boldsymbol{o}_j\right\|=0$, i.e., there exist $\boldsymbol{w}_i$, $\boldsymbol{u}_i$ and b_i such that

$$\sum_{i=1}^{l}\boldsymbol{u}_i g(\boldsymbol{w}_i\cdot\boldsymbol{x}_j+b_i)=\boldsymbol{o}_j,\quad j=1,\ldots,s. \quad (8)$$

where $y_i=[y_{i1},y_{i2},\ldots,y_{im}]^T$ is the output vector of SLFNs.

The above s equations can be written compactly as $\boldsymbol{HU}=\boldsymbol{O}$, where

$$\boldsymbol{H}=\begin{bmatrix} g(\boldsymbol{w}_1\boldsymbol{x}_1+b_1) & g(\boldsymbol{w}_2\boldsymbol{x}_1+b_2) & \cdots & g(\boldsymbol{w}_l\boldsymbol{x}_1+b_l) \\ g(\boldsymbol{w}_1\boldsymbol{x}_2+b_1) & g(\boldsymbol{w}_2\boldsymbol{x}_2+b_2) & \cdots & g(\boldsymbol{w}_l\boldsymbol{x}_2+b_l) \\ \vdots & \vdots & \ddots & \vdots \\ g(\boldsymbol{w}_1\boldsymbol{x}_s+b_1) & g(\boldsymbol{w}_2\boldsymbol{x}_s+b_2) & \cdots & g(\boldsymbol{w}_l\boldsymbol{x}_s+b_l) \end{bmatrix}$$

$\boldsymbol{U}=[\boldsymbol{u}_1^T,\boldsymbol{u}_2^T,\ldots,\boldsymbol{u}_l^T]^T$ and $\boldsymbol{O}=[\boldsymbol{o}_1^T,\boldsymbol{o}_2^T,\ldots,\boldsymbol{o}_s^T]^T$. $\boldsymbol{H}$ is the hidden layer output matrix of the neural network.

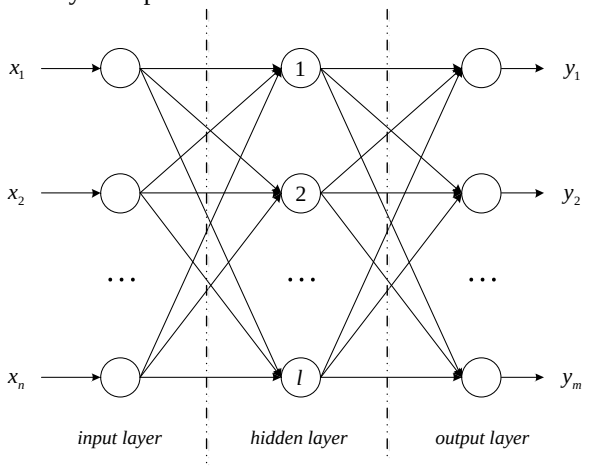


Fig. 5 The structure of standard SLFNs

Form the theorems of [14], it is known that the input weights and hidden layer biases of SLFNs can be randomly chosen if the activation function is infinitely differentiable. Unlike the common understanding that all the parameters of SLFNs need to be adjusted, the input weights $\boldsymbol{w}_i$ and the hidden layer biases b_i are in fact not necessarily tuned and

ISBN: 978-1-61804-290-3

the hidden layer output matrix $\boldsymbol{H}$ can remain unchanged once random values have been assigned to these parameters in the beginning of learning. Then training an SLFN is equivalent to finding a least-squares solution $\hat{\boldsymbol{U}}$ of the linear system $\boldsymbol{HU}=\boldsymbol{O}$:

$$\left\|\boldsymbol{H}\hat{\boldsymbol{U}}-\boldsymbol{O}\right\|=\min_{\boldsymbol{U}}\left\|\boldsymbol{HU}-\boldsymbol{O}\right\| \tag{9}$$

In most cases, the amount of distinct training samples is quite large. In order to reduce the calculation cost, the number of hidden nodes is chosen much less than number of samples. Therefore, the minimum norm least-square solution of the above linear system is

$$\hat{\boldsymbol{U}}=\boldsymbol{H}^{+}\boldsymbol{O} \tag{10}$$

where $\boldsymbol{H}^{+}$ is the Moore-Penrose generalized inverse of matrix $\boldsymbol{H}$ and the smallest solution is unique.

Thus, the procedure of ELM for SLFNs is as follows:

Step 1: Obtain the training set $\boldsymbol{x}_i,\boldsymbol{o}_i,i=1,2,\ldots,s$.

Step 2: Adjust the number of hidden nodes l ($l<s$).

Step 3: Randomly assign the input weight $\boldsymbol{w}_i$ and hidden layer bias b_i, $i=1,2,\ldots,l$.

Step 4: Calculate the hidden layer output matrix $\boldsymbol{H}$.

Step 5: Calculate the output weight matrix $\boldsymbol{U}=\boldsymbol{H}^{+}\boldsymbol{O}$.

Compared with BP, ELM only needs to adjust the number of hidden nodes and avoids the multiple iterations, hence, it is absence of the problem of local minimum or infinite training iteration and can reach the minimum training error.

B. Cooperative guidance strategy

In order to approximate $f(\cdot)$ and $f^{-1}(\cdot)$, we build two SLFNs and train them with plentiful simulation data beforehand under ELM. Then introduce T_i as coordination variables, the structure of cooperative guidance strategy is built as shown in Fig. 3. These two SLFNs get assembled into an expert system as shown in Fig. 6. *Net* 1 is the approximation of $f(\cdot)$ and used to evaluate the $T_i(0)$ as the initial state under the state of missile at the beginning of cooperative guidance; *Net* 2 is the approximation of $f^{-1}(\cdot)$ and used to calculate $k_i(t)$ according to $T_i(t)$ in real time.

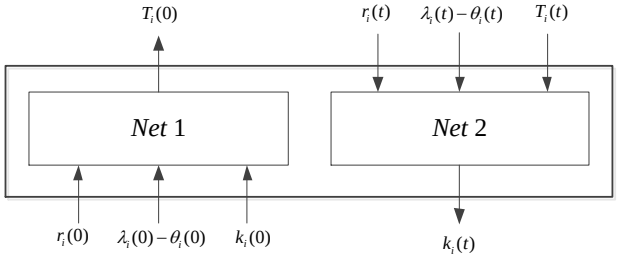


Fig. 6 Expert system with two SLFNs

The time-variant undirected graph [15] of the multi-missile system is defined as $G(t)=\{V,E(t)\}$, where $V=\{1,2,\cdots,n\}$ is the set of all missiles and $E(t)=\{(i,j)\in V\times V:i\sim j\}$ is a set of edges, in which the edge $i\sim j$ means that missile i and j can exchange with each other. The Laplacian matrix $L(t)$ is defined as

$$\begin{aligned}&L(t)=D(t)-A(t)=D(t)-(a_{ij}(t))_{n\times n}\\&a_{ij}(t)=\begin{cases}1, & if\ (i,j)\in E(t)\\0, & otherwise\end{cases}\\&D(t)=diag\left\{\sum_{j=1}^{n}a_{1j}(t),\sum_{j=1}^{n}a_{2j}(t),\cdots,\sum_{j=1}^{n}a_{nj}(t)\right\}\end{aligned} \tag{11}$$

where $D(t)$ is the degree matrix and $A(t)$ is the adjacency matrix of the graph. If $G(t)$ is connected, there exists $L(t)=L^{\mathrm{T}}(t)\geq 0$.

If there is no cooperation among missiles, the original $T_i(t)$ satisfies

$$\begin{aligned}&\dot{T}_i(t)=-1\\&T_i(t)=T_i(0)-t\end{aligned} \tag{12}$$

In order to make $T_i(t),i=1,2,\cdots,n$ achieve consensus, we choose the cooperative protocol below for missile i

$$\dot{T}_i(t)=-1-c_i\sum_{j\in N_i}[T_i(t)-T_j(t)] \tag{13}$$

where N_i is the neighborhood (the set of missiles that can exchange information with i) of missile i, and c_i is specified as

$$\begin{cases}c_i>0, & if\ \sum_{j\in N_i}[T_i(t)-T_j(t)]<0\\c_i=0, & otherwise\end{cases} \tag{14}$$

The purpose of c_i is that all T_i converge to $\max_i\{T_i\}$ and that

$$T_i(t)\geq T_i(0)-t \tag{15}$$

which means $k_i(t)\leq k_i(0)$.

From Fig. 4, we can see that T_i decreases with k_i increasing. The range of traditional proportional guidance coefficient is [3,6]. But it is worth noting that $\partial T_i/\partial k_i$ is close to 0 when $3<k_i\leq 6$, and that T_i is too sensitive to k_i when $k_i<1$. So the ideal interval of k_i to time-to-go is [1,3]. If $k_i(0)$ is specified as 3, then $k_i(t)\leq 3$.

Theorem 1 The system (13) will achieve consensus asymptotically.

Proof

Let $T=[T_1\ \ T_2\ \ \cdots\ \ T_n]^{\mathrm{T}}$ and $C=diag\{c_1,c_2,\cdots,c_n\}$, then (13) can be rewritten as

$$\dot{T}=-\mathbf{1}_{n\times 1}-CLT \tag{16}$$

Define energy function

$$Q(T)=\frac{1}{4}\sum_{i=1}^{n}\sum_{j\in N_i}(T_i-T_j)^2 \tag{17}$$

then, the derivate is

$$\begin{aligned}\dot{Q}&=\frac{1}{2}\sum_{i=1}^{n}\sum_{j\in N_i}(T_i-T_j)(\dot{T}_i-\dot{T}_j)\\&=T^{\mathrm{T}}L\dot{T}\end{aligned} \tag{18}$$

By substituting (16) into (18), we can obtain

$$\dot{Q}=-T^{\mathrm{T}}L\mathbf{1}_{n\times 1}-T^{\mathrm{T}}LCLT \tag{19}$$

Because $\mathbf{1}_{n\times 1}$ can be regarded as the eigenvector of matrix L under the eigenvalue of zero, $L\mathbf{1}_{n\times 1}=\mathbf{0}$. Besides, $LCL\geq 0$, then we have

$$\dot{Q} = -T^{\mathrm{T}} LCLT \le 0 \qquad (20)$$

where the condition of $LCL = 0$ is $T_1 = T_2 = \cdots = T_n$, hence

$$\lim_{t\to\infty}(T_i - T_j) = 0,\ \forall i, j \in V$$

which means that the system (13) will achieve consensus asymptotically. □

Usually, we should choose a big c_i to accelerate the convergence rate of system (13). Furthermore, when max $\{|T_i - T_j|\}$ decreases into the allowable tolerance scope at t^*, we can stop cooperative guidance and fix $k_i = k_i(t^*)$. The reason is that a longer convergence process may require a wider varying range of $r_i(t_0)$ and $\lambda_i(t_0) - \theta_i(t_0)$ of the samples for these two SLFNs. This will lead to longer training time and more complicated topological structure.

IV. Simulation and Results

In this section, an example is given to illustrate the proposed cooperative strategy. Consider the saturation attack of 5 missiles on a static target. The initial states of these missiles are listed as below

$$r_1 = 30 \quad \lambda_1 - \theta_1 = 0.7$$
$$r_2 = 33 \quad \lambda_2 - \theta_2 = 0.9$$
$$r_3 = 32 \quad \lambda_3 - \theta_3 = 0.8$$
$$r_4 = 35 \quad \lambda_4 - \theta_4 = 0.77$$
$$r_5 = 38 \quad \lambda_5 - \theta_5 = 1$$

The topological structure of the communication network among them is shown in Fig. 7 and the corresponding Laplacian matrix is

$$L = \begin{bmatrix} 1 & -1 & 0 & 0 & 0 \\ -1 & 2 & -1 & 0 & 0 \\ 0 & -1 & 2 & -1 & 0 \\ 0 & 0 & -1 & 2 & -1 \\ 0 & 0 & 0 & -1 & 1 \end{bmatrix}$$

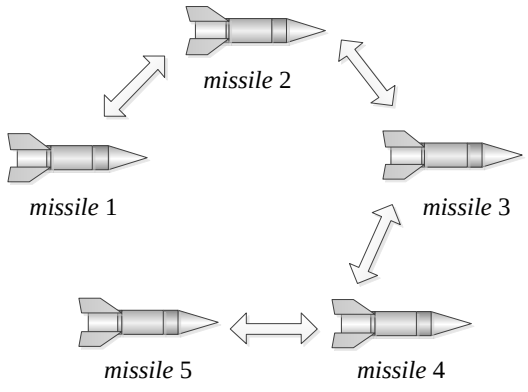


Fig. 7 Topological structure of the network

With the algorithm of ELM and Neural Network Toolbox of Matlab, these two SLFNs can be built and trained conveniently. Each SLFN includes 250 neurons in the hidden layer and adopts activation functions 'tansig' and 'purelin' for the hidden layer and output layer respectively. In order to generate enough samples, simulate the guidance of a single missile with fixed proportional guidance coefficient iteratively, with the initial conditions $r(0)$, $\lambda(0) - \theta(0)$, and $k(0)$ varying at points

$$r(0) = 40:-1:10$$
$$\lambda(0) - \theta(0) = 0.5:0.01:1$$
$$k(0) = 1.1:0.1:6$$

The training results of the SLFN with 250 hidden neurons using ELM are quite good. The error of mean square is 0.0029, which is close to the result of BP. However, the training time is only 2.3 seconds, which is much less than BP of 41.1 seconds. When increasing the number of hidden neurons to 500, the training time with ELM still needs only 4.5 seconds but the training results improve significantly.

With the two trained SLFNs, setting the parameters $c_i = 1$, the simulation is conducted and results are depicted in Fig. 8 ~Fig. 11. When $t = 5s$, the time-to-go commands have already converged to $T_5(t)$. So we stop the cooperative guidance, and fix the proportional guidance coefficients. And almost at the same time $t = 42s$, all missiles arrive the target. It is worth noting that $k_i(t) \le 3$ always.

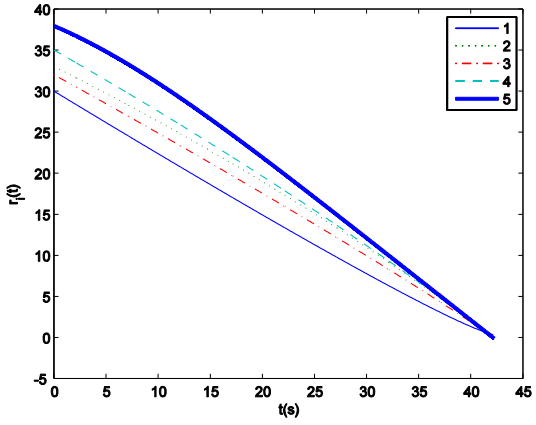


Fig. 8 The trajectories of $r_i(t)$

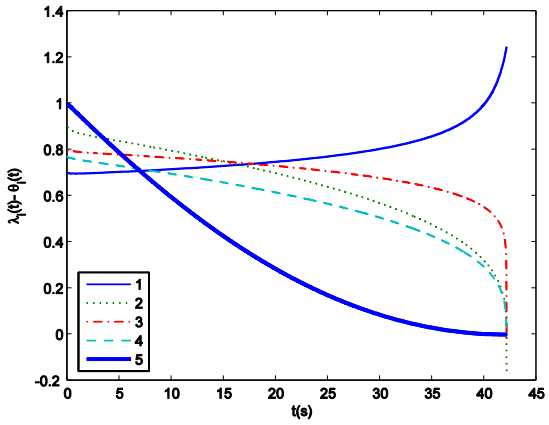


Fig. 9 The trajectories of $\lambda_i(t) - \theta_i(t)$

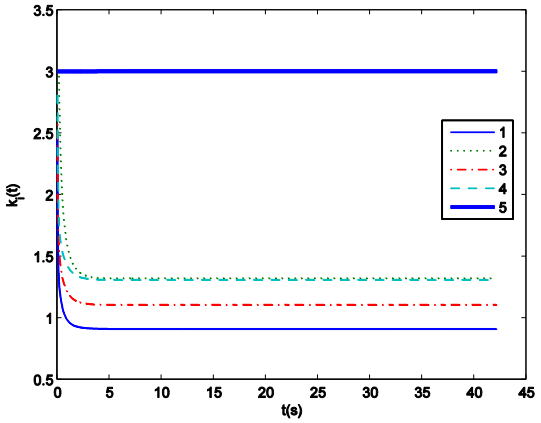


Fig. 10 The trajectories of $k_i(t)$

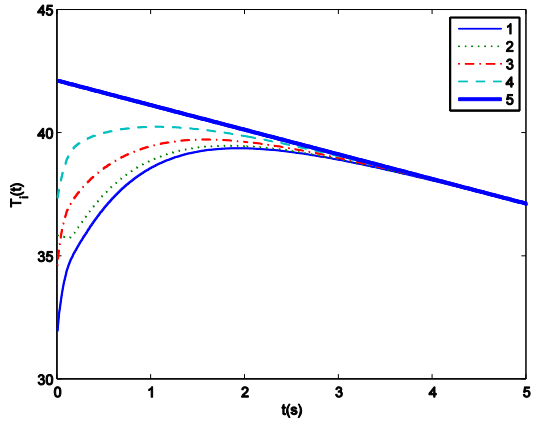


Fig. 11 The converging process of $T_i(t)$

V. Conclusion

In this paper, the cooperative guidance for simultaneous attack on a static target of multi-missile system is investigated. An indirect and distributed cooperative proportional guidance strategy, based on expert systems and first-order consensus protocol, is proposed. This strategy can guarantee that the overload satisfies limitation of the missile body since the proportional guidance coefficient is always less than 3 in the simulation. Additionally, it is demonstrated that all missiles can arrive at the static target simultaneously with this strategy. Future studies will be conducted on the cooperative guidance of simultaneous attack on a maneuvering target within finite time and control saturation.

References

[1] Z. T. Dydek, A. M. Annaswamy, and E. Lavretsky, "Adaptive configuration control of multiple UAVs," *Control Engineering Practice*, Vol. 21, No. 8, Aug. 2013, pp. 1043-1052.

[2] W. Ren, R. W. Beard, and E. M. Atkins, "Information consensus in multivehicle cooperative control," *IEEE Control Systems Magazine*, Vol. 27, No. 2, Apr. 2007, pp. 71-82.

[3] M. Bando and A. Ichikawa, "Active Formation Flying Along an Elliptic Orbit," *Journal of Guidance, Control, and Dynamics*, Vol. 36, No. 1, Jan.-Feb. 2013, pp. 324-332.

[4] H. S. Su, G. R. Chen, X. F. Wang, and Z. L. Lin, "Adaptive second-order consensus of networked mobile agents with nonlinear dynamics," *Automatica*, Vol. 47, No. 2, Feb. 2011, pp. 368-375.

[5] F. Pasqualetti, D. Borra, and F. Bullo, "Consensus networks over finite fields," *Automatica*, Vol. 50, No. 2, Feb. 2014, pp. 349-358.

[6] T. W. McLain and R. W. Beard, "Coordination Variables, Coordination Functions, and Cooperative Timing Missions," *Journal of Guidance, Control, and Dynamics*, Vol. 28, No. 1, Jan.-Feb. 2005, pp. 150-161.

[7] I. S. Jeon, J. I. Lee, and M. J. Tahk, "Impact-time-control guidance law for anti-ship missiles," *IEEE Transactions on Control Systems Technology*, Vol. 14, No. 2, Mar. 2006, pp. 260-266.

[8] I. S. Jeon, J. I. Lee, and M. J. Tahk, "Homing Guidance Law for Cooperative Attack of Multiple Missiles," *Journal of Guidance Control and Dynamics*, Vol. 33, No. 1, Jan.-Feb. 2010, pp. 275-280.

[9] R. Olfati-Saber and R. M. Murray, "Consensus problems in networks of agents with switching topology and time-delays," *IEEE Transactions on Automatic Control*, Vol. 49, No. 9, Sep. 2004, pp. 1520-1533.

[10] Z. Shiyu and Z. Rui, "Cooperative Guidance for Multimissile Salvo Attack," *Chinese Journal of Aeronautics*, Vol. 21, No. 6, Dec. 2008, pp. 533-539.

[11] S. Vitaly and S. Tal, "Cooperative Optimal Guidance Laws for Imposing a Relative Intercept Angle," in *AIAA Guidance, Navigation, and Control Conference*, Minneapolis, Minneasota, Aug. 2012.

[12] Z. Enjiao, W. Songyan, C. Tao, and Y. Ming, "Multiple missiles cooperative guidance based on leader-follower strategy," in *2014 IEEE Chinese Guidance, Navigation and Control Conference (CGNCC2014)*, Yantai, China, 8-10 Aug. 2014, pp. 1163-1167.

[13] X. Wang, Y. Zheng, and H. Lin, "Integrated guidance and control law for cooperative attack of multiple missiles," *Aerospace Science and Technology*, Vol. 42, Apr.-May. 2015, pp. 1-11.

[14] G. B. Huang, Q. Y. Zhu, and C. K. Siew, "Extreme learning machine: Theory and applications," *Neurocomputing*, Vol. 70, No. 1-3, Dec. 2006, pp. 489-501.

[15] C. Godsil and G. Royle, *Algebraic graph theory*. New York: Springer-Verlag, 2001.

A neural network framework for face recognition by elastic bunch graph matching

Francisco A. Pujol López, Higinio Mora Mora*, José A. Girona Selva

***Abstract*—** Automated biometric systems are being widely used in many applications. Face recognition is one of the most promising methods due to its good acceptance by users. In this paper, we have explored including neural networks models for face recognition combined with other methods. As a result, a new recognition algorithm based on Elastic Bunch Graph Matching and Self-Organizing Maps is introduced. In this algorithm a combination of global and local techniques are applied to construct graphs whose nodes encode the facial features. A formal framework for specifying the functions involved is defined. The experiments performed were aimed at calibrating the map and evaluating the performance. The experimental results show the effectiveness of the proposal when compared to other well-known methods.

***Keywords*—** Pattern recognition, Face Recognition, Neural Networks, Self-Organizing Maps.

I. Introduction

IN recent years, there has been an intensive research to develop complex security systems involving a new kind of *'key'*: the biometric features. Automated biometric systems are being widely used in many applications such as surveillance, digital libraries, forensic work, law enforcement, human computer intelligent interaction, and banking, among others. For applications requiring high levels of security, biometrics can be integrated with other authentication means such as smart cards and passwords. In relation to this, face recognition is an emerging research area and, in the next few years, it is supposed to be extensively used for automatic human recognition systems in many of the applications mentioned before.

One of the most popular methods for face recognition is the Elastic Graph Bunch Matching (EBGM), proposed in [1], and it is an evolution of the method known as Dynamic Link Architecture (DLA) [2]. The main idea in the elastic graph matching is to represent a face starting from a set of reference or fiducial points known as landmarks. These fiducial points have a spatial coherence, as these points are connected using a graph structure. Considering these nodes, geometric information can be extracted and both distance and angle metrics can be defined accordingly.

* H. Mora-Mora, MT. Signes-Pont, J. Azorín-López and L. Corral-Sánchez are with the Department of Computer Technology and Computation, University of Alicante, Spain, 03690, San Vicente del Raspeig, Alicante, Spain. e-mail: ({fpujol, hmora, jags20}@ua.es).

This algorithm takes into account that actual facial images have many nonlinear features (variations in lighting, pose and expression) that are not generally considered in linear analysis methods, such as LDA or PCA. Moreover, it is particularly robust when out-of-plane rotations appear. However, the main drawback of this method is that it requires of an accurate location of the fiducial points.

Artificial Neural Networks (ANN) is one of the most used paradigms to address problems in artificial intelligence. Among the techniques and architectures proposed by the scientific community in this field, the Self Organizing Map (SOM) has special features for association and pattern classification [3], and it is one of the most popular neural network models. The key aspect that characterizes the family of problems in which it is desirable to apply this technique is the inaccuracy or lack of formalization of problems. In these cases, there is not a precise mathematical formulation of the relationship of input patterns [4].

Although there have been many scientific advances in the field of artificial intelligence and facial recognition, it is still not clear enough how the human brain recognizes different faces. This reasoning motivates the application in this work of neural network techniques to the face recognition problem with the goal of improving existing approaches.

Consequently, in this paper we will use ANNs to improve the efficiency of the EBGM algorithm. To do this, a SOM is applied in the construction of the database of facial graphs in an adaptive learning process. First, the fiducial points will be extracted automatedly and, after that, known faces will be grouped (or clustered) into M classes, each class corresponding to a different person.

This paper is organized as follows: Section II describes the EBGM method and summarizes the related work in the domain of using that method for face recognition; Section III explains the proposal of an EBGM-based face recognition method and the formal framework to define it; Section IV introduces the neural network approach with a Self Organizing Map for recognition; Section V describes the experiments carried out; and finally, conclusions and some future work are discussed in Section VI.

ISBN: 978-1-61804-290-3

II. EBGM ALGORITHM AND RELATED WORK

In this section, the EBGM algorithm is described and, afterwards, some recent, related works are discussed.

A. The Elastic Bunch Graph Matching

Elastic Bunch Graph Matching is a *feature-based* face identification method. EBGM derives a bunch of jets for each training image and uses the jets to represent the graph node. To form a bunch graph, a collection of facial images is marked with node locations at defined positions of the head. These node locations are called landmarks and are obtained by a semiautomatic process. When matching a bunch graph to an image, the jet extracted from the image is compared to all jets in the corresponding bunch attached to the bunch graph and the best matching one is selected.

Jets are defined as *Gabor coefficients* in a landmark location computed by convoluting a set of Gabor wavelet filters around each landmark location. The jets of all training images are collected in a data structure called a bunch graph. The bunch graph has a node for every landmark on the face and every node is a collection of jets for the corresponding landmark. The main steps for face recognition by EBGM are outlined below [5]:

• *Step 1*: Select the landmarks on the training face images to create the face models. The selection is performed manually.

• *Step 2*: Convolve these points with a Gabor wavelet to construct the Gabor jets *J*. The local appearance around a fiducial point $\vec{x}$ will be coded by using the convolution of the input image $I(\vec{x})$ with a Gabor filter $\psi_m(\vec{x})$, so that:

$$J_j(\vec{x}) = \int I(\vec{x}')\, y_j(\vec{x}-\vec{x}')\, d^2\vec{x}' \tag{1}$$

where

$$\psi_m(\vec{x}) = \frac{\left\|\overrightarrow{k_m}\right\|}{\sigma^2}\exp\left(-\frac{\left\|\overrightarrow{k_m}\right\|^2\left\|\vec{x}\right\|^2}{2\sigma^2}\right) \cdot\left[\exp(i\,k_m\cdot\vec{x})-\exp(-0.5\sigma^2)\right] \tag{2}$$

and $\overrightarrow{k_m}$ is the wave vector:

$$k_j=\begin{pmatrix}k_{jx}\\k_{jy}\end{pmatrix}=\begin{pmatrix}k_\nu\cos\varphi_\mu\\k_\nu\sin\varphi_\mu\end{pmatrix},\ k_\nu=2^{-\frac{\nu+2}{2}}\pi,\ \varphi_\mu=\mu\frac{\pi}{8}. \tag{3}$$

Consequently, 5 different frequencies (ν = 0, 1, …, 4) and 8 different orientations (μ = 0, 1, …, 7) are used and, as a result, a jet will have 40 coefficients.

• *Step 3*: Create a data structure called bunch graph corresponding to facial landmarks that contains a bunch of model jets extracted from the face model.

• *Step 4*: Then for every new image to be recognized:

(a) Estimate and locate the landmark positions with the use of the bunch graph. (b) Calculate the new jets displacement from the actual position by comparing it to the most similar model jet. (c) Create a new face graph containing each landmark position and jet values for that landmark position.

• *Step 5*: Similarly, for each new image, estimate and locate the landmarks using bunch graph. Then the features are extracted by convoluting with the number of instances of Gabor filters followed by the creation of face graph. The matching score is calculated on the basis of similarity between face graphs of images in the database and the one in a new input image.

B. Related work

EBGM has been used for recognition in the last few years. Most of the methods based on EBGM use Gabor wavelets for feature extraction [6]. These features are represented by a grid of points geometrically adjusted to the features extracted. The recognition is based on the wavelet coefficients, which are calculated for the nodes of a 2D elastic graph representing the grid containing the landmarks. This method combines a local and a global representation through the processing of a Gabor filter with several scales and several directions (jets), of a point set –called fiducial points– located in specific regions of the face. The location of the fiducial points is the most complex task of this method. These points depend on lighting conditions, the expression and the pose of the face.

An alternative method proposed in [7] is the application of the histogram of orientation gradients (Histogram of Oriented Gradients, HOG) instead of using Gabor filters to locate features. This algorithm provides invariance in terms of location and orientation. To do this, the reference points are extracted in the space representation and the gradients of the image with respect to the dominant orientation around each reference point are calculated. Finally, the HOG descriptor, which is a statistic measure where the orientations of all the image gradients around a reference point are taken into account, is calculated.

Recently, a combination of EBGM with PCA and soft biometrics is used to make a study on the influence of age variations in face recognition [8]. Additionally, some new versions of EBGM focus on fast versions of the algorithm in order to make it feasible for real conditions; thus, a parallel version of EBGM for fast face recognition using MPI (Message Passing Interface) is presented in [9]; the authors divide the training process into *p* processors, then the recognition process is made simultaneously. Khatun and Bhuiyan [10] presented a neural network based face recognition system using Gabor filter coefficients, where the recognition used a hybrid neural network with a two networks, a Bidirectional Associative Memory (BAM) for dimensional reduction of the feature matrix to make the recognition faster and a Multilayer Perceptron with backpropagation algorithm for training the network.

In [11] a data mining approach to improve the performance

ISBN: 978-1-61804-290-3

of EBGM in case of using a large database was proposed, based on an entropy decision tree with the most important features in the face recognition process. Sarkar [12] combines skin detection with EBGM so as to obtain an accurate recognition, since skin segmented images remove background noises and reduce errors in identifying Gabor features. Finally, Li and Wachs [13] applied EBGM to hand gesture recognition, where a hierarchy is assigned to each node of the graph and the classification of hand gestures is performed against complex backgrounds. The location of the fiducial points is one of the most complex tasks of the method. Their positions highly depend on the lighting conditions, facial expressions and pose. In the original EBGM algorithm, a fixed number of features were established. These features corresponded to specific face characteristics, such as the pupils or the corners of the mouth. As a result, a facial model graph is obtained and the fiducial points are manually selected for each image in the database.

Another way to locate the features is based on a uniformly distributed grid of points that deforms and conforms to a pattern, such as the contours identified by an edge detector (Canny, Sobel, MLSEC, etc.) [14, 15].

Some recent advances have been made for the detection of the fiducial points in faces. Thus, Belhumeur et al. [16] used a Bayesian model that combines the output of local detectors with a non-parametric set of global models for the part locations based on a set of hand-labeled face images. The experiments were performed both using the BioID database and the new Labeled Face Parts in the Wild (LFPW) database, with very accurate results in any case. The Labeled Faces in the Wild (LFW) database was used instead in [17], where a method based on regression forests that detects 2D facial feature points in real-time is presented.

Other recent relevant works include: Baltrusaitis et al. [18] proposed a probabilistic patch expert (landmark detector) that can learn non-linear and spatial relationships between the input pixels and the probability of a landmark being aligned. Then, the 2-D fiducial detection method proposed in EBGM is extended in [19] to independently detect fiducial points by restricting the search range corresponding to each target fiducial, thereby removing the computationally expensive iterative scheme present in the original EBGM. Finally, Jin et al. [20] developed a Hough voting-based method to improve the efficiency and accuracy of fiducial points localization.

To sum up, from this revision two conclusions emerge: first of all, there is still a great interest from many research groups in order to use and improve the original EBGM method for face recognition; moreover, most of these investigations are focused on adapting EBGM to be used in real-time conditions with an accurate location of the landmarks or fiducial points for faces. It is clear that there is still much work to do in this field, and no previous works on the application of neural networks to EBGM have been found.

III. A PROPOSAL OF AN EBGM-BASED FACE RECOGNITION METHOD

To facilitate the work of connectionist models, in this work an adaption of basic EBGM method described in the previously section is performed.

Therefore, the faces are represented using a facial graph that includes geometric and textural information. The facial graph is defined as a pair $\{V, A\}$, where V refers to the set of vertices or nodes and A to the set of edges. Each vertex corresponds to a fiducial point and encodes the corresponding vector of jets and its location, that is, $V_i = \{J_i, P_i(x,y)\}$. Each edge A_{ij} encodes information on the distance and angle between the two nodes it connects, so that $A_{ij} = \{d_{ij}, \theta_{ij}\}$.

For each node, a 2-dimensional histogram $hist_i$ is constructed. In this histogram, the information about the distance $D = \{d_{i1}, d_{i2}, \ldots, d_{in}\}$ and the angle $\theta = \{\theta_{i1}, \theta_{i2}, \ldots, \theta_{in}\}$ from node i to the other nodes in the graph will be stored. Therefore, the histogram $hist_i$ consist of k bins corresponding to x distance-intervals by y angle-intervals. Thus, the k bins in histogram $hist_i$ are uniformly constructed in a *log-polar* space. Each pair $(\log(d_{ij}), \theta_{ij})$ increases the corresponding histogram bin. The algorithm followed to obtain the fiducial points is represented by commented pseudo-code below:

Algorithm 1. Obtaining the fiducial points from face

1. Normalize image sizes.
2. Apply an edge detector. In this work, the well-known *Canny edge detector* [26] is used.
3. Create a grid of *Nx x Ny* points, where nodes are uniformly distributed.
4. Each node adjusts its position to the nearest point in the edges obtained in Step 2.
5. The distances and angles from each final node to the rest of nodes are calculated.

A Gabor jet J is now constructed. Following Wiskott's approach [1], a vector of 40 complex components will be obtained. A jet J is then obtained considering the magnitude parts only. The position of each of the nodes in both facial graphs is known, as each vertex V_i encodes this information: $V_1 = \{J_1, P\}$, $V_2 = \{J_2, Q\}$, where $P = \{p_1, p_2, \ldots, p_n\}$ and $Q = \{q_1, q_2, \ldots, q_n\}$ are the vectors with the positions of each of the fiducial points for both faces.

So that, just as basic EBGM method, in order to match two facial graphs, $G_1 = \{V_1, A_1\}$ and $G_2 = \{V_2, A_2\}$, both geometric and texture information will be used.

Three functions of similarity are proposed in this work: the *Match Cost Function* (MCF), the *Norm Vector Function* (NVF) and *Gabor Feature Match Function* (GFMF).

Taking into account the histograms previously computed with geometric information of the nodes, MCF is calculated adding the matching costs for each node in the input facial graph G_1 with its corresponding node in the stored facial graph G_2:

ISBN: 978-1-61804-290-3

$$\mathrm{MCF}(G_1, G_2) = \mathrm{MCF}(P, Q) = \frac{\sum_{i=1}^{n} \sum_{k} \frac{[hist_{p_i}(k) - hist_{q_i}(k)]^2}{h_{p_i}(k) + h_{qi}(k)}}{\|P\| \cdot \|Q\|} \quad (4)$$

where $||P||$, $||Q||$ refer to the norm of vectors P and Q.

The NVF is calculated by adding the norm of the vector of differences among the matched nodes:

$$\mathrm{NVF}(G_1, G_2) = \mathrm{NVF}(P, Q) = \sum_{i=1}^{n} \left\| \overrightarrow{p_i c_P} - \overrightarrow{q_i c_Q} \right\| \quad (5)$$

where

$$c_P = \frac{1}{n} \sum_{i=1}^{n} p_i \text{ and } c_Q = \frac{1}{n} \sum_{i=1}^{n} q_i$$

The texture information given by the Gabor jets from each node will be used to define the third similarity function: *Gabor Feature Match Function* (GFMF); thus, for each node $pi \in P$, a jet J_{pi} is calculated. Let R contain the Gabor jets of all the nodes in a facial graph, $R = \{J_{p1}, J_{p2}, \ldots, J_{pn}\}$. The function GFMF between two facial graphs is calculated as follows:

$$\mathrm{GFMF}(G_1, G_2) = \mathrm{GFMF}(R_1, R_2) = \frac{1}{n} \sum_{i=1}^{n} < R_{1_i}, R_{2_i} > \quad (6)$$

where $< R_{1_i}, R_{2_i} >$ is the normalized dot product between the i-th jet in R_1 and the i-th jet in R_2 . As mentioned before, only the magnitude of the Gabor coefficients in the jets is considered.

Finally, from the expressions defined in (4) to (6), the final similarity function called *Global Distortion function* (GD) is defined, which combines the results from each of them:

$$\mathrm{GD} = \lambda_1 \mathrm{MCF} + \lambda_2 \mathrm{NVF} + \lambda_3 \mathrm{GFMF} \quad (7)$$

where λ_1, λ_2 and λ_3 are coefficients to be obtained experimentally and $\lambda_1 + \lambda_2 + \lambda_3 = 1$.

The functions that make up GD are normalized to the range [0, 1]. This normalization is performed with the maximum values for each component function using the distance between the input graph facial and facial graphs stored in neurons. Finally, the GD function is as follows:

$$\mathrm{GD} = \frac{\lambda_1 \mathrm{MCF}}{\max(\mathrm{MCF})} + \frac{\lambda_2 \mathrm{NVF}}{\max(\mathrm{NVF})} + \frac{\lambda_3 \mathrm{GFMF}}{\max(\mathrm{GFMF})} \quad (8)$$

The values of GD $\in$ [0, 1], where facial images belonging to the same person will give results close to 0, and facial images of different people will have a high GD value. The correct acceptance threshold value will be calculated to perform a recognition rates according contextual application features.

IV. Improving EBGM Algorithm with a Neural Network Approach

A. Self Organizing Map formulation

The Self Organizing Map (SOM) is a Neural Network technique that implements a nonlinear projection from a high-dimensional space onto a low-dimensional array of neurons. That mapping tends to preserve the topological relationship of the inputs, so, the visual image of this map depicts clusters of input information and their neighbour relationships on the map [3], [21].

The utility of the Self Organizing Maps (SOMs) for recognition problems has been proven in numerous studies. Next, we describe the SOM formalization to the proposed facial identification process.

A SOM is defined at any time by a collection of neurons, their position on the map and the weight of each. The neurons are connected to adjacent neurons by a neighbourhood relation. This set up the topology or the structure, of the map. The topological configuration of neurons are generally rectangular or hexagonal grid [21]. For a SOM of M neurons, the set,

$$\mathrm{W}(\mathrm{t}) = \{\mathrm{w}_1, \mathrm{w}_2, \ldots, \mathrm{w}_\mathrm{M}\} \quad (9)$$

has the weight information as a whole, where w_i is the weight vector associated to neuron i and is a vector of the same dimension of the input. The set W(t) evolves according the Self-Organizing Map algorithm. The neurons position, defined by their weight vector, are configuring a topological mapping of the input space.

Let $X \in \mathbb{R}^k$ be the input vector of the SOM. For application to face recognition, this vector consists of k features extracted from the face to identify. We define,

$$\Psi: \mathscr{V} \rightarrow \mathbb{R}^k \quad (10)$$

as the function that obtains the characteristics of the face to make up the vector X. So that,

$$\forall\, I \in \mathscr{V},\ \Psi(I) = X \in \mathbb{R}^k \quad (11)$$

A classification of the face images is obtained when running the SOM algorithm on feature vectors calculated with Ψ from images of faces. On this classification can be applied a clustering process according to a method known. There are numerous methods of clustering [22]. However, the representation clusters in a 2D region is usually not a simple problem because the input data is usually of a high dimensionality. Let,

$$G_\Psi = \{g_1, g_2, \ldots, g_m\} \quad (12)$$

be the set of groups obtained when clustering process is made. So that,

$$\forall I \in \mathcal{Y} \Rightarrow SOM(\Psi(I)) \in G_\Psi \quad (13)$$

where, SOM function is called to the classification function of a feature vector, from input of the map to one of the groups generated.

The working hypothesis of this research is to consider that the classification with SOM network, suitably trained with images of individuals ($\mathcal{Y}$), is correct. That is, a bijective function between groups and individuals can be established.

The main research described in this paper is to determine the extent to which the above hypothesis is true and under what conditions. For this, SOM operation and feature extraction functions from a face are analyzed under various configurations.

In this way, aspects of the facial identification made through SOM classification network are defined, suitable feature extraction function Ψ is identified, and also, the characteristics of the similarity function and acceptance threshold will be established under the premise that this function should provide the minimum value for a face when it is classified within the cluster corresponding to its individual.

The above information will enable to calibrate the method of face recognition under different scenarios and conditions.

B. Self Organizing Map configuration

As mencioned previously, in this paper we propose a neural network as a function of face recognition. The self-organizing map is responsible for building itself the database of facial graphs from the training images. Specifically, a two-dimensional $N_x x N_y$ SOM neural network is used. The number of neurons of the map will be determined experimentally in order to establish the minimum size that maximizes the efficiency in identification.

To analyze and extract features from each image, let's use one of the techniques that provides better results. The EBGM method will obtain the data of each face to be used as inputs for the training phase of the network and subsequent recognition of the method. From these data, the SOM uses the features extracted as inputs. The identification threshold t consist of the maximum distance that characterizes the clusters organized into the SOM in the training process. In this case, SOM network applies a classification process from the set of face graphs obtained from training images and generates G_Ψ clusters where each of them corresponding to one of the individuals to be identified.

The input data to the SOM network come from the EBGM method output's face graphs. A face graph is the structure used to represent the face through EBGM method. This graph has the set of nodes corresponding to the set of landmarks of the face and, each of them contains both geometric and texture characteristics. The following figure shows the face graph representation as input matrix array.

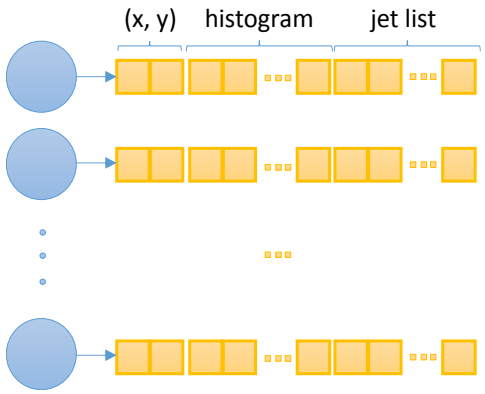


Fig. 1 SOM input data from facial graph

For practical implementation, the matrix arrangement of the entry represented in above figure can be transformed into one-dimensional vector placing continuously columns into memory.

The following algorithm shows the learning algorithm of the SOM.

Algorithm 2 Learning Algorithm Overview

1. Randomize the map's neurons' weight vectors.
2. Obtain face graph bunch using EBGM function of input face.
3. Every neuron is examined to calculate which one's weights are most like the input data.
4. The radius of the neighborhood of the winning neuron is now calculated.
5. The radius of the winning neuron is now updated according a neighborhood function.
6. Each neighboring node's (found in previous step) weights are adjusted to make them more like the input vector according a learning function.
7. Repeat step 2 for N iterations.

The calculation of the winning neuron are performed using for this similarity function defined in (13). Then, the neuron which minimizes the result of GD function is the winning neuron.

After this training process, there is a map composed of $N_x x N_y$ neurons organized according to the similarity between the input data. That is, the neurons which are near each other and are in the same cluster (g_i) have information about the face of the same user.

Once the self-organizing map has built and trained, it can be used as recognition function (F) to identify user by his/her face.

V. Results

A. Experimental setup

The recognition scheme has been tested with the FERET database [23]. The version used in this work is the Color FERET Database. It contains 11338 pictures of 994 different

individuals. For our experiments we used the sets of images fa and fb, corresponding to 843 individuals with frontal images only. There is an image of each of the 843 individuals in fa and another one in fb. Images stored in fb were taken a few moments of time after the ones in fa, so in most cases some changes in the expression of the model can be noticed. All FERET images have a size of 512 x 768 pixels.

For the experiments, a set of 20 users were used, with 5 training images and 2 test images per user, with a resolution 128 x 192 pixels. The tests have been performed using Matlab® with a 2 GHz Intel Core i5 and 8 GB memory.

B. Experimental results

The experimentation made has a dual objective: in first place, it seeks to configure the SOM network to provide the best results, and secondly, it tests the effectiveness of the method in the process of facial recognition. The two types of experiments are combined in a set of tests on the input images. The following subsections focus on testing one aspect leaving the other aspects as invariant. In all cases, the number of training iterations made is 200.

1) Type and size of map

The following table shows the recognition rate, where N (20) is the number of different individuals used for the experimentation (and 5 different images for person).

TABLE I. RECOGNITION RATE (%): TYPE AND SIZE OF MAP

SOM size	rectangular grid (gridtop)	hexagonal grid (hextop)
N-10 x N-10	77.6	77.61
N-5 x N-5	80.9	80.9
N x N	88.6	89.7
N+5 x N+5	88.5	88.9
N+10 x N+10	87.3	88.5

The above table shows that the optimal size of the map is around the number of individuals to identify. As regard type, a hexagonal distribution of neurons provides slightly better results.

2) Size of facial graph

In this experiment, the connection between the size of facial graph obtained from EBGM method and recognition rate is analysed. The results are shown in the next table:

TABLE II. RECOGNITION RATE (%): SIZE OF FACIAL GRAPH

SOM size	Size of facial graph			
	6x6	8x8	10x10	12x12
N x N	86.1	87.4	89.6	89.7

It is observed that, at first seems, there are a correlation between the size of the graphs and the recognition accuracy. However, in this respect (as in the above), we must find a compromise between speed and accuracy of training, as training cost increases considerably with the size of the graph and the map.

3) Weighting λ coefficients of similarity function

The following table shows the results obtained according to various configurations of λ.

TABLE III. RECOGNITION RATE (%): SIZE OF FACIAL GRAPH

SOM size	λ coefficients weight		
	$\lambda_1=0.2, \lambda_2=0.7, \lambda_3=0.1.$	$\lambda_1=0.7, \lambda_2=0.2, \lambda_3=0.1.$	$\lambda_1=0.2, \lambda_2=0.2, \lambda_3=0.6.$
N x N	89.7	89.7	81.6

As seen in the above table, greater weighting metric GFMF not give good results, whereas the other two metrics shows that the best correct identifications is reached. Generally, it appears that giving high weight to the NVF and MCF functions the best results are obtained for experimentation.

C. Comparison with other methods

A set of experiments have been performed in order for our system to be compared with some other existing algorithms for face recognition. In particular, the following methods have been chosen: Wiskott's Elastic Bunch Graph Matching (EBGM) [1], eigenfaces (PCA) [24], and Ahonen's Local Binary Patterns (LBP) [25]. The results are shown in Table IV.

TABLE IV. COMPARISON BETWEEN METHODS

Algorithm	Accuracy (%)
EBGM	80.9
PCA	66.4
LBP	74.5
Our proposal	89.7

From these results, we can deduce that connectionist proposals provides a promising method to compete with other well-known methods in face recognition applications.

VI. CONCLUSIONS & FUTURE WORK

This paper has carried out a study on the state of current knowledge on the problem of face recognition focused on the use of the EBGM method. Applying connectionist techniques to build the database of knowledge that make up the collection of faces to recognition has improved the results.

Taking into account that the system is applied in a controlled environment and with a small number of individuals, both the size of the information required and the time spent searching to identify an individual from an image is optimized.

We can highlight here a key feature: Applying a SOM obtains a database more compact than traditional methods. In EBGM based algorithms, the bunch graph on which perform the matching process contains information of each of the graphs obtained for each input image of the training phase,

ISBN: 978-1-61804-290-3

therefore, for more images of individuals larger size of the database. In the proposed method, the information is compacted. It is the map size which determines the number of graphs to be used in the matching phase. Thus, training can be applied to an extensive battery of images without affecting the amount of memory required to store the entire database in memory during execution of the matching algorithm.

Base on the current outcomes, our future work will be unfolded along two directions: one is to make more exhaustive experiments with a great number of images and individuals. The other direction is to prove with other recognition methods in combining with neural network techniques to explore the potential of this approach.

Finally, as demonstrated in the experimental section, the classification efficiencies above 89% can be achieved, leading to optimism on the implementation of the proposed work in a real environment.

REFERENCES

[1] L. Wiskott, J-M. Fellous, N. Krüger, and C. von der Malsburg. "Face Recognition by Elastic Bunch Graph Matching". IEEE Transactions on Pattern Analysis and Machine Intelligence. vol. 19, n. 7, pp. 775-789. Jul. 1997.

[2] C. Kotropoulos, and I. Pitas "Rule-based face detection in frontal views", IEEE International Conference on Acoustics, Speech, and Signal Processing, vol. 4, pp. 2537-2540, 1997.

[3] T. Kohonen, "Self-Organising Maps", 2nd ed., Springer-Verlag, Berlin, 1997.

[4] J. Azorín-López, M. Saval-Calvo, A. Fuster-Guilló, A. Oliver-Albert: A Predictive Model for Recognizing Human Behaviour based on Trajectory Representation. International Joint Conference on Neural Networks, 2014.

[5] A. Rattani, N. Agarwal, H. Mehrotra, and P. Gupta, "An efficient fusion-base.d classifier". In Workshop on Computer Vision, Graphics and Image Processing (WCVGIP), pp. 104-109, 2006.

[6] L. L. Shen, and L. Bai, "A review on Gabor wavelets for face recognition", Pattern Anal. Appl., vol. 9, pp. 273 -292, 2006.

[7] D. Monzo, A. Albiol, and J. M. Mossi, "A comparative study of facial landmark localization methods for face recognition using HOG descriptors". In Proceedings of the 20th International Conference on Pattern Recognition (ICPR), pp. 1330-1333, 2010.

[8] G. Guo, G. Mu, and K. Ricanek, "Cross-Age Face Recognition on a Very Large Database: The Performance versus Age Intervals and Improvement Using Soft Biometric Traits", 20th International Conference on Pattern Recognition, ICPR, pp. 3392-3395, 2010.

[9] X. Chen, C. Zhang, F. Dong, and Z. Zhou, "Parallelization of elastic bunch graph matching (EBGM) algorithm for fast face recognition". In Proceedings of the 2013 IEEE China Summit & International Conference on Signal and Information Processing (ChinaSIP) pp. 201-205.

[10] A. Khatun, and M. A. A. Bhuiyan, "Neural network based face recognition with Gabor filters". International Journal of Computer Science and Network Security, vol. 11, pp. 71-74, 2011.

[11] S. Mitra, S. Parua, A. Das, and D. Mazumdar, "A Novel Datamining Approach for performance improvement of EBGM based face recognition system to handle large database", In Advances in Computer Science and Information Technology, pp. 532-541, 2011.

[12] S. Sarkar, "Skin segmentation based elastic bunch graph matching for efficient multiple face recognition". In Advances in Computer Science, Engineering & Applications, pp. 31-40, 2012.

[13] Y. T. Li, and J. P. Wachs, "Hierarchical elastic graph matching for hand gesture recognition". In Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications, pp. 308-315, 2012.

[14] R. Espí, F. A Pujol, H. Mora, J. Mora, Development of a Distributed Facial Recognition System Based on Graph-Matching, International Symposium on Distributed Computing and Artificial Intelligence, pp. 498, 502, 2008.

[15] D. González-Jiménez, and J. L. Alba-Castro, "Shape-Driven Gabor Jets for Face Description and Authentication," IEEE Trans. Information Forensics and Security, vol. 2, n. 4, pp. 769-780, 2007.

[16] P. N. Belhumeur, D. W. Jacobs, D. J. Kriegman, and N. Kumar, "Localizing Parts of Faces Using a Consensus of Exemplars", Proc. IEEE Conf. Computer Vision and Pattern Recognition, pp. 545-552, 2011.

[17] M. Dantone, J. Gall, G. Fanelli, and L. Van Gool, "Real-time facial feature detection using conditional regression forests". In Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2578-2585.

[18] T. Baltrusaitis, P. Robinson, and L. Morency, "3D constrained local model for rigid and non-rigid facial tracking". In Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2610-2617.

[19] S. Jahanbin, H. Choi, A. C. Bovik, "Passive Multimodal 2-D+3-D Face Recognition Using Gabor Features and Landmark Distances," Transactions on Information Forensics and Security,IEEE, vol. 6, n. 4, pp. 1287-1304, December 2011.

[20] X. Jin, X. Tan and L. Zhou. "Face Alignment Using Local Hough Voting". In Proceedings of the 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG), pp. 1-8, 2013.

[21] H. Yin, "The self-organizing maps: Background, theories, extensions and applications". In Computational intelligence: A compendium, pp. 715-762, 2008.

[22] J. A. F. Costa, "Clustering and visualizing SOM results". In Intelligent Data Engineering and Automated Learning–IDEAL 2010, pp. 334-343.

[23] P. J. Phillips, H. Moon, P. J. Rauss, and S. Rizvi, "The FERET evaluation methodology for face recognition algorithms", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 22 (10), 2000.

[24] M. Turk, and A. Pentland, "Eigenfaces for recognition". J Cognitive Neurosci, vol. 3, pp. 71-86, 1991.

[25] T. Ahonen, A. Hadid, and M. Pietikäinen, "Face recognition with local binary patterns". In Pajdla T, Matas J (eds.) Proceedings of the 8th European Conference on Computer Vision, ECCV 2004, Part I, Springer Berlin/Heidelberg, pp. 469-481.

[26] J. Canny, "A Computational Approach to Edge Detection", IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 8, pp. 679-698, 1986.

Improved ESPRIT-TLS Algorithm for Wind Turbine Fault Discrimination

Saad Chakkor, Mostafa Baghouri, Abderrahmane Hajraoui

***Abstract*—** ESPRIT-TLS method appears a good choice for high resolution fault detection in induction machines. It has a very high effectiveness in the frequency and amplitude identification. Contrariwise, it presents a high computation complexity which affects its implementation in real time fault diagnosis. To avoid this problem, a Fast-ESPRIT algorithm that combined the IIR band-pass filtering technique, the decimation technique and the original ESPRIT-TLS method was employed to enhance extracting accurately frequencies and their magnitudes from the wind stator current with less computation cost. The proposed algorithm has been applied to verify the wind turbine machine need in the implementation of an on-line, fast, and proactive condition monitoring. This type of remote and periodic maintenance provides an acceptable machine lifetime, minimize its downtimes and maximize its productivity. The developed technique has evaluated by computer simulations under many fault scenarios. Study results prove the performance of Fast-ESPRIT offering rapid and high resolution harmonics recognizing with minimum computation time and less memory cost.

***Keywords*—**Spectral Estimation, ESPRIT-TLS, Real Time, Diagnosis, Wind Turbine Faults, Band-Pass Filtering, Decimation.

I. Introduction

WIND energy has become one of the popular renewable powers all over the world in electricity generation capacity. Wind turbines contain a complex electromecanical system which is prone to defects. Consequently, there is an increase need to implement a predictive monitoring scheme of wind turbines, allowing an early detection of electromechanical faults, in order to avoid catastrophic damage, to reduce maintenance costs, to ensure continuity of production and to minimize downtime. It means that stopping a wind installation for unexpected failures could lead to expensive repair. These faults cause a modulation impact in the magnetic field of the wind generator, which is characterized by the appearance of a significant harmonics (peaks) in the stator current spectrum. For this reason, most of the recent researches have been oriented their interest toward electrical monitoring, with focus on frequency analysis of stator current (CSA). This technique is more practical and less costly [1]-[4]. Furthermore, with recent digital signal processor (DSP) technology developments, motor and generator fault diagnosis can now be done in real-time [1]. ESPRIT is one high resolution or subspace method (HRM) which is widely adopted in electromechanical machine diagnosis. It can be used for spectral estimation [3], [5], [6]. This algorithm allows very high spectral detection accuracy and a high resistance to noise compared to others methods like MUSIC and Root-MUSIC. Contrariwise, it require long computation time to find more frequency estimates when the autocorrelation matrix is large and the order of sampled data dimension increase. This fact makes its application in real time detection very limited despite its high precision. This article presents an ameliorated version of ESPRIT-TLS method for fast wind turbine faults detection and diagnosis based on a band pass filtering technique. The proposed improvement allows many advantages: reduction of computation time, saving of memory space and accuracy increase in a specified frequency bandwidth. The paper is organized as follows: the problem is formulated in Section II, the stator current signal is presented in Section III, and then Section IV describes wind turbine fault models. While Section V focuses on ESPRIT method theory, Section VI explains in details the proposed approach to enhance original ESPRIT algorithm. Simulation results are presented and discussed in Section VII. Finally, conclusions with future work are drawn in the last section.

Saad Chakkor, Mostafa Baghouri, and Abderrahmane Hajraoui are working with the Dept. of Physics, Abdelmalek Essaâdi University, Communication and Detection Systems Laboratory, Faculty of Sciences, BP. 2121 M'Hannech II, 93030, Tetouan, Morocco.

(e-mail: [1]saadchakkor@gmail.com, [2]baghouri.mostafa@gmail.com, [3]ad_hajraoui@hotmail.com).

II. Related Work

Many research studies applying enhanced and advanced signal processing techniques have been used in the motor and generator stator current to monitor and to diagnose prospective electromechanical faults. The classical methods like periodogram and its extensions which are evaluated through a Fast Fourier Transform (FFT) are not a consistent estimator of the PSD because its variance does not tend to zero as the data length tends to infinity. Despite this drawback, the periodogram has been used extensively for failure detection in research works [4], [6]. The (FFT) does not give any information on the time at which a frequency component occurs. Therefore, the Short Time Fourier Transform approach (STFT) is used to remove this shortcoming. A disadvantage of this approach is the increased sampling time for a good frequency resolution [7]. The discrimination of the frequency components contained within the signal, is limited by the length of the window relative to the duration of the signal [8]. To overcome this problem, in [9] and [10] Discrete Wavelet Transform (DWT) is used to diagnose failures under transient conditions for wind energy conversion systems by analyzing frequencies with different resolutions. This method facilitates signal interpretation because it operates with all information contained in the signal by time-frequency redistribution. One

limitation of this technique that it gives a good time resolution and poor frequency resolution at high frequencies, and it provides a good frequency resolution and poor time resolution at low frequencies [4], [11]. Recently, high resolution methods (HRM) are applied to detect more frequencies with low SNR. In fact, MUSIC and ESPRIT techniques with its zooming extensions are conjugated to improve the identification of a large number of frequencies in a given range [12], [13]. In [14] a comparative performance analysis of (HRM) is made. This study has demonstrated that ESPRIT method has a high accuracy which exceeds all other algorithms even with the existence of an annoying noise. Moreover, these algorithms are based on an eigenanalysis of the autocorrelation matrix of a signal corrupted by noise. This decomposition requires a long computation time mainly when the size of the autocorrelation matrix and the number of data samples increase. In [15] a rank reduced ESPRIT technique is proposed to transform it into simplified low-complexity algorithm. However, this method presents performance deterioration mainly with the SNR decreasing and lowers harmonic amplitudes. Moreover, it has not focused on the minimization of the computational time execution for real applications. This work proposes a solution to overcome the complexity cost of ESPRIT in the purpose of its use in a real time wind turbine monitoring.

III. Stator Current Model

The application of CSA technique for diagnosis of wind turbine machine requires a well previously knowledge of various frequency and amplitudes components in stator current spectrum stemmed from the wind turbine generator in both healthy and faulty states. In fact, to build a correct detection of the wind turbine fault modulations and signatures in the stator current, it is necessary to construct a complex signal associated with the real one. This analytical signal model describes precisely the behavior and the evolution of the real stator current. It contains relevant fault information. For these reasons it is often used for command purposes. The studied wind generator stator current will be denoted by the discrete signal $i[n]$. This signal is considered as a sum of L complex sinusoids and white noise. It is obtained by sampling the continuous time current every $T_s=1/F_s$ seconds. The induction generator stator current $i[n]$ in presence of mechanical and/or electrical faults has a data model which can be expressed as follows [10]:

$$i[n]=\sum_{k=1}^{L} I_k e^{j\left(2\pi.\frac{f_k}{F_s}.n+\varphi_k\right)}+b[n] \quad (1)$$

where $i[n]$ corresponds to the n^{th} stator current sample with n=0, 1, 2...N_s-1. I_k, f_k, and φ_k are the amplitude, the frequency and the phase of the k^{th} complex sinusoid (harmonic components) respectively, b[n] is a gaussian white noise. F_s is the sampling frequency and N_s is the number of data samples. L represents the number of researched harmonics.

IV. Wind Turbine Fault Models

The wind machine is subject to diverse electro-mechanical anomalies that involve mostly five components: the stator, the rotor, the bearings, gearbox and/or the air gap (eccentricity) [16]. These defects require a predictive recognition to avert any side effect provoking a breakdown or a fatal spoilage. Because it contains the totally relevant fault information, the stator current spectrum is examined to withdraw the sideband frequency components inserted by the fault. These fault frequencies are located around the fundamental line frequency and are called lower sideband and upper sideband components. This detection technique is used in collaboration with one bit vibration sensors for an early identifying of prospective electromechanical failures which can occurs in any time. A synopsis of wind turbine faults and their related frequencies formulas are presented in Table I.

TABLE I
Wind Turbines Faults Signatures

Failure	Harmonic Frequencies	Parameters
Broken rotor bars (brb)	$f_{brb}=f_0\left[k\left(\frac{1-s}{P}\right)\pm s\right]$	$k=1,3,5,...$
Bearing damage (bng)	$f_{bng}=\lvert f_0 \pm k\, f_{i,o}\rvert$	$k=1,3,5,...$ $f_{i,o}=\begin{cases}0.4\ n_b\ f_r\\ 0.6\ n_b\ f_r\end{cases}$
Misalignment (mis)	$f_{mis}=\lvert f_0 \pm k\, f_r\rvert$	$k=1,3,5,...$
Air gap eccentricity (ecc)	$f_{ecc}=f_0\left[1\pm m\left(\frac{1-s}{P}\right)\right]$	$m=1,2,3,...$

f_0 is the electrical supply frequency, s is the per-unit slip, P is the number of poles, f_r is the rotor frequency, n_b is the bearing balls number, $f_{i,o}$ is the inner and the outer frequencies depending on the bearing characteristics, and m, $k\in\mathbb{N}$ are the harmonic frequency index [4], [9], [10]. Slip s is defined as:

$$s=\frac{\omega_s-\omega_r}{\omega_s} \quad (2)$$

$$\omega_s=\frac{120\,f_0}{P} \quad (3)$$

ω_s is the generator synchronous speed,
ω_r is the relative mechanical speed of the generator.
These harmonics are extensively used as diagnostic measures in the CSA approach.

V. ESPRIT Method Theory

High resolution methods are recently used for fault diagnosis. They can detect and identify the faulty element based on its frequency. The most accurate and efficient technique is ESPRIT which belongs to the subspace parametric spectrum estimation methods. It is based on eigenvector decomposition which aims to separate the observation space in a signal subspace, containing only useful information, and its orthogonal complement, called noise subspace. The rotational invariance between both subspaces

allows extracting of the parameters of spectral components present within the investigated waveform [17], [18], [20].

A. Autocorrelation Matrix Estimation

Based on the stator current model defined by (1), the autocorrelation matrix can be then estimated as [19]:

$$R_i = E\left[i(n).i^H(n)\right] = R_s + R_b = S.P.S^H + \sigma_b^2.I \tag{4}$$

It is composed by the sum of signal and noise autocorrelation matrices. Where H is the Hermitian transpose, σ_b^2 is the variance of the white noise, I is the identity matrix of size (N_s x N_s) and P is the power matrix of the harmonics:

$$P = diag\left[I_1^2\, I_2^2 \cdots I_L^2\right] \tag{5}$$

S is the Vandermonde matrix defined by:

$$S = \left[s_1 \cdots s_i \cdots s_L\right] \tag{6}$$

$$S_k = \left[1\; e^{j\left(2\pi.\frac{f_k}{F_s}\right)}\; e^{j\left(4\pi.\frac{f_k}{F_s}\right)} \cdots e^{j\left(2\pi.(N_s-1).\frac{f_k}{F_s}\right)}\right]^T \tag{7}$$

The finite data length of the signal makes the computation of the autocorrelation matrix R_i inaccurate. For real purpose, this matrix is unknown and it must be singular. For effective detection, it is necessary to reduce the statistical fluctuations present in estimating the autocorrelation matrix by the averaging [7], [19]. In addition, the accuracy of ESPRIT depends on the dimension ($M \le N_s$) of R_i. It is possible to estimate it from the acquired data samples by [7], [19]:

$$\hat{R}_i = \frac{1}{N_s - M + 1} D.D^H \tag{8}$$

Where M is the data matrix order and D is a Hankel data matrix defined by:

$$D = \begin{bmatrix} i(0) & \cdots & i(N_s - M) \\ \vdots & \cdots & \vdots \\ i(M-1) & \cdots & i(N_s - 1) \end{bmatrix} \tag{9}$$

The dimension of R_i should be high enough to have more eigenvalues for noise space and should be low enough to minimize the computation time cost. When the value of M decreases below $N_s/3$, it can be seen the increase of the frequency detection error. Contrariwise, if M increases beyond $N_s/2$, calculation time increases. So, there is a trade-off for the right choice of M. Empirically, the value of M is chosen to be bounded as shown in (10) to give a good performance:

$$\frac{N_s}{3} < M < \frac{N_s}{2} \tag{10}$$

In this paper, the autocorrelation matrix dimension M is taken rounded down as:

$$\hat{M} = Round\left(\frac{N_s - 1}{2}\right) \tag{11}$$

Evidently, the number of frequencies L is not a priori known. The frequency signal dimension order (FSDO) L must to be estimated by the minimization of a cost function $MDL(k)$ named minimum description length. In order to obtain a robust estimate, (MDL) criterion is used as shown in the following formula [18] for $k=1, 2,\ldots, L$:

$$MDL(k) = -\log\left(\frac{\prod_{i=k+1}^{L} \lambda_i^{\frac{1}{L-k}}}{\frac{1}{L-k}\sum_{i=k+1}^{L}\lambda_i}\right)^{\rho(L-k)} + \frac{1}{2}k(2L-k)\log(\rho) \tag{12}$$

$$\rho = N_s - L - 2 \tag{13}$$

where λ_i are eigenvalues autocorrelation matrix R_i. Analytically, the estimate of L can then be expressed in the form:

$$\hat{L} = \arg_k \min\left(MDL(k)\right) \tag{14}$$

However, ESPRIT performances are completely degraded by choosing a wrong FSDO value.

B. Eigendecomposition of Autocorrelation Matrix

The eigendecomposition of the autocorrelation matrix R_i is given by exploiting the eigenvalues $\{\lambda_1, \lambda_2, \ldots, \lambda_M\}$ and their corresponding signal eigenvectors $\{v_1, v_2, \ldots, v_M\}$ [17]:

$$R_i = \sum_{k=1}^{N_s} \lambda_k v_k v_k^H = \underbrace{U_s.E_s.U_s^H}_{R_s} + \underbrace{U_b.E_b.U_b^H}_{R_b} \tag{15}$$

Where:

$$U_s = \left[v_1 \cdots v_L\right],\; E_s = diag\left[\lambda_1 \cdots \lambda_L\right] \tag{16}$$

$$U_b = \left[v_{L+1} \cdots v_{N_s}\right],\; E_b = \sigma_b^2 I_{N_s - L} \tag{17}$$

U_s represents the eigenvectors matrix of the signal space related to the L largest eigenvalues arranged in descending order. Whereas, U_b represents the eigenvectors matrix of the noise space related to the N_s-L eigenvectors that, ideally, have eigenvalues equal to the variance noise σ_b^2. Diagonal matrices E_S and E_b contain eigenvalues λ_i corresponding to eigenvectors v_i.

C. Abbreviations Frequency Estimation

ESPRIT-TLS method is based on the study of the signal subspace E_s. It uses some rotational invariance properties founded naturally in the case of exponential. A decomposition of the matrix S into two matrices S_1 and S_2 is considered as follows:

$$S=\begin{bmatrix} 1 & 1 & \cdots & 1 \\ e^{j\left(2\pi.\frac{f_1}{F_s}\right)} & e^{j\left(2\pi.\frac{f_2}{F_s}\right)} & \cdots & e^{j\left(2\pi.\frac{f_L}{F_s}\right)} \\ \vdots & \vdots & \cdots & \vdots \\ e^{j\left(2\pi.(N_s-1).\frac{f_1}{F_s}\right)} & e^{j\left(2\pi.(N_s-1).\frac{f_2}{F_s}\right)} & \cdots & e^{j\left(2\pi.(N_s-1).\frac{f_L}{F_s}\right)} \end{bmatrix} \begin{matrix} \} S_1 \\ \} S_2 \end{matrix} \quad (18)$$

S_1 represents the first N_s-1 rows of the matrix S, S_2 represents the last N_s-1 rows of the matrix S. The rotational invariance between both subspaces leads to:

$$S_1 = \Phi S_2 \quad (19)$$

The matrix Φ contains all information about L components frequencies. Nevertheless, the estimated matrices S can contain errors. Thereafter, the ESPRIT-TLS (total least-squares) algorithm finds the matrix Φ by minimization of matrix error given by (20) and (21). The determination of this matrix can lead to obtain the frequency estimates defined by [20]:

$$f_k = \frac{Arg\left(\Phi_{k,k}\right)}{2\pi} F_s, \ k=1,2,...,L \quad (20)$$

$$\Phi=\begin{bmatrix} e^{j\left(2\pi.\frac{f_1}{F_s}\right)} & 0 & \cdots & 0 \\ 0 & e^{j\left(2\pi.\frac{f_2}{F_s}\right)} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & e^{j\left(2\pi.\frac{f_L}{F_s}\right)} \end{bmatrix} \quad (21)$$

D.Harmonics Powers Estimation

Once the searched frequencies components of the signal are estimated by ESPRIT, the values of their amplitudes and then their powers can be estimated. By using the eigendecomposition of the subspace signal [17], [19]:

$$R_s = S.P.S^H = \sum_{k=1}^{L}\left(\lambda_k + \sigma_b^2\right).v_k.v_k^H \quad (22)$$

It is assumed that the eigenvectors of the signal subspace are normalized as follows:

$$v_k^H.v_k = 1 \quad (23)$$

Thus, for $k=1, 2,\ldots, L$:

$$R_i.v_k = \lambda_k.v_k \quad (24)$$

Multiplying both sides of this equation by v_k^H:

$$v_k^H.R_i.v_k = \lambda_k.v_k^H.v_k \quad (25)$$

According to (4), (11) and (21):

$$v_k^H.R_i.v_k = v_k^H.\left[\sum_{k=1}^{L} P_k s_k s_k^H + \sigma_b^2.I\right].v_k = \lambda_k \quad (26)$$

This equation can be simplified as follows:

$$\sum_{k=1}^{L} P_k.\left|s_k^H.v_k\right|^2 = \lambda_k - \sigma_b^2 \quad (27)$$

Using:

$$\left|s_k^H.v_k\right|^2 = \left|Q_k\left(e^{j2\pi f_k}\right)\right|^2 \quad (28)$$

Equation (22) can be written in:

$$\sum_{k=1}^{L} P_k.\left|Q_k\left(e^{j2\pi f_k}\right)\right|^2 = \lambda_k - \sigma_b^2 \quad (29)$$

This equation is a set of L linear equations with a number L of unknown harmonics powers. It is very easy to extract the harmonics powers vector P from (25) by simple resolution.

VI. Improved ESPRIT Method

The discrimination of all small amplitude frequency components around f_0 by ESPRIT method is difficult. This is mainly due to the significant computation time elapsed by this algorithm to find harmonic sideband components correctly.

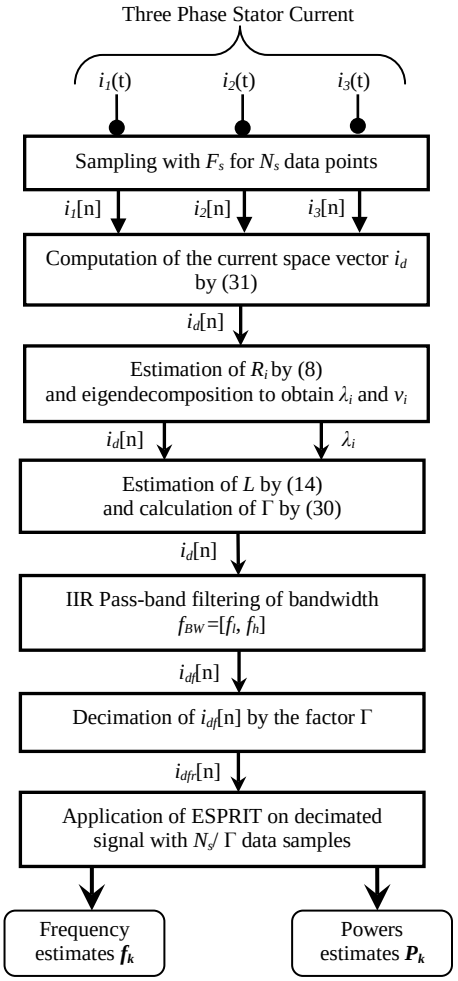


Fig. 1 Block diagram scheme of the Fast-ESPRIT algorithm

Furthermore, ESPRIT calculation cost increases when the size of the autocorrelation matrix and the number of data samples increase. It depends on the complexity of N_s^3. This delay forms a major drawback that can cause a catastrophic evolution of a wind turbine fault which may lead to greatest damages. In order to apply a proactive, robust and real time wind turbine condition monitoring, an improved version of ESPRIT algorithm entitled Fast-ESPRIT was used. Fig. 1 shows the block diagram scheme of different stages that Fast-ESPRIT algorithm must execute to identify the fault harmonic frequencies and their powers. The ameliorated algorithm is based on both a band-pass IIR filtering and decimation technique in the fault frequency bandwidth [f_l, f_h], where f_l, f_h are the low cut-off and high cut-off frequency of the band-pass filter. This process provides a remarkable reduction in computation time and in data memory size. The decimation factor used in this research is computed with respect to the Nyquist criteria as [21]:

$$\Gamma = \begin{cases} \dfrac{F_{Nyquist}}{2f_0} = \dfrac{F_s}{4f_0} \;\; if \;\; f_h < 95\,Hz \\ \dfrac{F_{Nyquist}}{6f_0} = \dfrac{F_s}{12f_0} \;\; if \;\; 95\,Hz \le f_h < 500\,Hz \end{cases} \quad (30)$$

Fig. 2 shows the variation of Γ according to f_h. The decimation factor decreases with the increase of the maximum harmonics frequency detected in the signal.

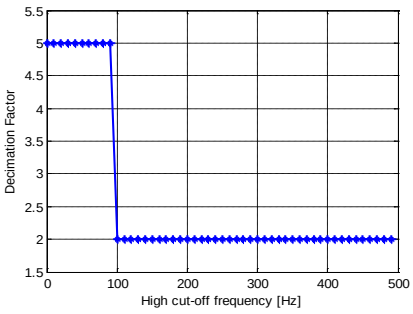


Fig. 2 Evolution of decimation factor with fault frequency

In the first time, the acquired sequences $i_{1,2,3}[n]$ of the three phase stator current sampled at the frequency F_s, are used to calculate the stator current space vector as [22]:

$$i_d = \frac{i_1 + a.i_2 + a^2.i_3}{3}, \quad a = e^{j\frac{2\pi}{3}} \quad (31)$$

Where a, a^2 are the spatial operators. This vector allows a fault diagnosis on all phase stator current instead of examining fault signature on each ones. With this method computation time will be minimized. In the second step, an estimation of the autocorrelation matrix R_i is realized and therefore the eigenvalues λ_i are extracted to estimate the number of researched harmonics L in the stator current signal with respect to *MDL* criterion seen in (14). Then, the signal sequence $i_d[n]$ is filtered via a recursive Infinite Impulse Response (IIR) digital band-pass filter based on a least squares fit in the frequency range [f_l, f_h] characterizing the fault. This filter has a flat response in the desired bandwidth and its use is justified by the fact that it will be helpful to extract just the informations contained in the signal which are useful in the fault recognition which can occurs at any time. In the third stage, the received sequence of the stator current space vector $i_{df}[n]$ is decimated by a factor Γ shown in (30). In addition, the applied decimation uses low-pass filter to ensure anti-aliasing. The motivation for this decimation is to reduce the cost processing and memory required for a cheaper implementation. Finally, the ESPRIT algorithm is applied on the decimated signal sequence having N_s/Γ data samples to identify all frequency components and their powers contained in the signal.

VII. Simulation Results and Analysis

The developed approach seen in the previous section has been applied and simulated under different scenarios of wind turbine fault types shown in Table I. To evaluate its performance in real time fault detection, Fast-ESPRIT algorithm has been integrated with a fault diagnosis controller which coordinates with vibration sensors localized in specific wind turbine mechanical components to monitor vibration levels. The controller decides and classifies the existence of a fault depending on vibration measurements collected by the sensors and the harmonic frequencies with their powers estimated by the Fast-ESPRIT method. Fig. 3 illustrates the explained technique. Besides, the applied diagnosis algorithm is based on the use of a fault frequency band switching which sweeps any prospective faults that may occur and subsequently classify them by type according to their frequencies. Thus, the diagnosis is made by the intervals of the spectrum reflecting the signature of a possible default [23], [25]. This means that the Fast-ESPRIT method will not be applied to the entire signal but only on a part that contains the target information to be extracted for analysis. In case of fault detection, a system alarm is triggered to alert monitoring and maintenance staff for an emergency intervention repair.

This procedure provides many benefits because it allows high recognizing and classification of faults with economic and real time implementation [24]. Computer simulations are realized in Matlab for a faulty wind turbine generator using 2 pair poles, 4kW/50Hz, 230/400V. The induction generator stator current, is simulated by using the signal model described in (1) for the different failure kinds described in Table I. The parameters of the simulations are illustrated in Table II.

TABLE II
Parameters used in the simulations

Parameter	Value
s	0,033
P	2
f_0	50 Hz
f_r	29,01 Hz
n_b	12
N_s	1024
F_s	1000 Hz
Fundamental Stator Current Amplitude	10 A
Computer Processor	Intel Core2 Duo T6570 2,1 GHz

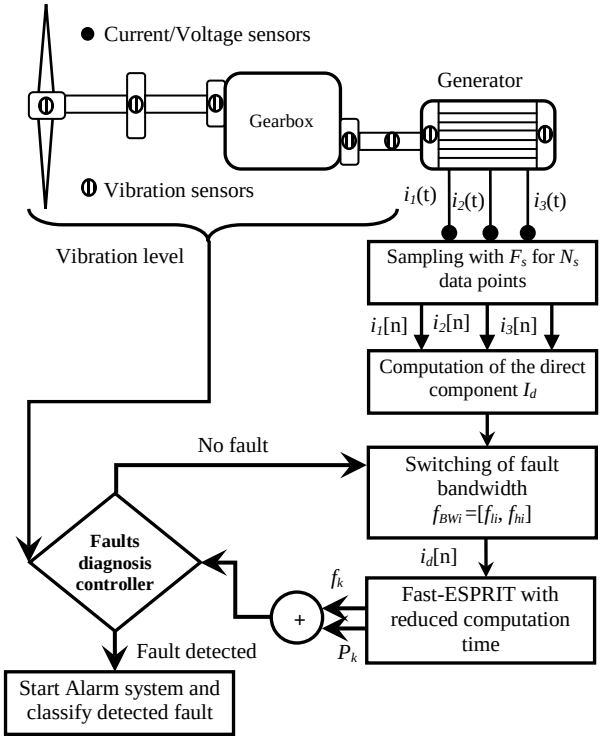


Fig. 3 Intelligent wind turbine faults diagnosis by Fast-ESPRIT

To simplify the simulation, a single phase of the generator stator current has been studied. The power of each fault is calculated based on its amplitude as follows:

$$P_k = 10\log_{10}\left(\frac{I_k}{2}\right) \tag{32}$$

Before examining the stator current signal, it must be filtered to obtain in the output a composite signal having a totally negligible noise compared to the fundamental and its harmonics.

A. Air Gap Eccentricity Detection

Table IV shows the simulation results for identifying wind turbine generator air gap eccentricity fault signature in the goal to compare the performance of the original ESPRIT-TLS with the proposed Fast-ESPRIT. The harmonics characterizing this fault are given by Table III.

TABLE III
AIR GAP ECCENTRICITY FAULT PARAMETERS

f_{ecc} (Hz)	I_{ecc} (A)	P_{ecc} (dB)	N_h	SNR (dB)
25.825	0.4	-10.97	3	80
74.175	0.3	-13.46		

This experiment was done with a high signal to noise ratio to determine the computing time and the required memory size in both algorithms.

TABLE IV
COMPUTATION PERFORMANCE COMPARISON

Method	Data samples	Harmonics f_k/P_k	Signal Memory size (KB)	M	Time (s)
Original ESPRIT	1024	50.00Hz/ 16.99dB 25.82 Hz/ -10.97dB 74.17Hz/ -13.47dB	16	511	4.3471
Fast ESPRIT	205	49.99Hz/ 16.96dB 25.81Hz/ -12.10dB 74.17Hz/ -14.12dB	3.2	102	0.03046

It is very clear from Table IV that both original and fast ESPRIT algorithms provide satisfactory accuracy, and they correctly identify the $L=3$ harmonics despite with smallest powers case. The little performance difference observed in the Fast-ESPRIT is justified by the attenuations caused by the IIR band pass filter used. Furthermore, the obtained results confirm the important reduction of the computational time with 142.7 times, the memory size required for processing with 5 times and complexity has been changed from N_s^3 to $(N_s/\Gamma)^3$.

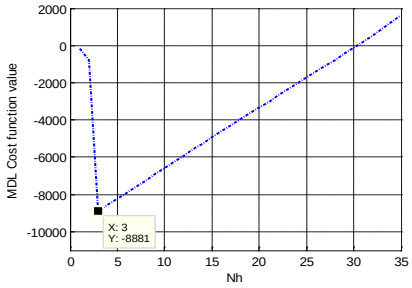


Fig. 4 Estimation of signal harmonics number by MDL criterion

In addition, a negligible performance loss is observed in the power and frequency estimation caused by the band pass filter attenuations. Fig. 4 illustrates the estimation of the signal subspace dimension by means of the Rissanen criteria based on *MDL* function cost shown in (12) and (14). Fig. 5 shows the frequency response gain of the Yule-Walk IIR band pass filter used in the Fast-ESPRIT algorithm having an order h=25. Obviously, the filter has a flat response in the bandwidth target. Whereas, Fig. 6 illustrates graphically the power and frequency estimates given by the proposed method. It seems obviously that Fast-ESPRIT has detects all harmonics existents in the eccentricity fault range [20, 80]Hz with high precision.

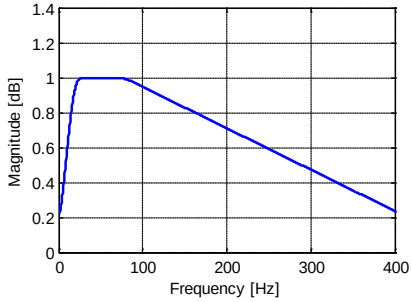


Fig. 5 IIR filter Yule-Walk frequency response gain

ISBN: 978-1-61804-290-3

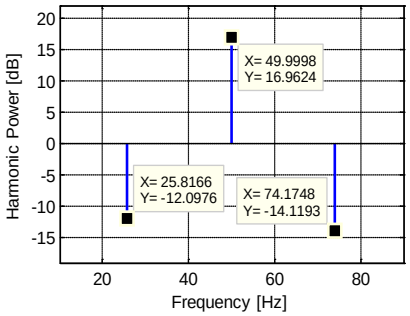


Fig. 6 Power and Frequency estimates by Fast-ESPRIT for eccentricity fault

B. Broken Rotor Bars Detection

The proposed Fast-ESPRIT method has been applied for a wind turbine generator stator current to detect broken rotor bars fault signature in the frequency range [15, 80]Hz. The characteristics of this fault harmonics are shown in Table V. The obtained results are averaged over fifty realizations.

TABLE V
BROKEN ROTOR BARS FAULT PARAMETERS

f_{brb} (Hz)	I_{brb} (A)	P_{brb} (dB)	N_h	SNR (dB)
22.525	0.3	-13.47	5	20
25.825	0.45	-9.95		
70.875	0.35	-12.13		
74.175	0.4	-10.97		

From Fig. 7, the Fast ESPRIT method flows a large calculation time with a large average estimation error rate for discriminating the fault harmonics and their powers contained in a very noisy stator current signal having SNR value less than or equal to 15dB. This can be interpreted by the delays caused by the algorithm search to find the eigenvalues and eigenvectors of the autocorrelation matrices in both signal and noise spaces. Contrariwise, the calculation time in question decreases gradually for a slightly noisy signal with SNR>15dB. The average estimation error declines also to fall to a minimum value for large values having SNR>55dB. The method finds difficulty in identifying faults in a very noisy environment. As shown in Fig. 8, for a stator current signal with a high annoying noise SNR<10dB, detection fault powers presents a considerable error and a remarkable instability level.

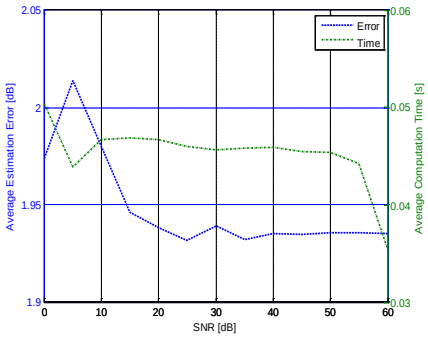


Fig. 7 Evolution of fault powers estimation error and computation time with SNR

This error rate decreases when the order of the Yule-Walk IIR band pass filter increases. However, the identification performance improves when signal to noise ratio SNR exceeds 10dB. In this case, the average error rate estimation stabilizes gradually to reach an asymptotic value. Although the standard deviation of the average estimation error is low, thus by increasing the order of the Yule-Walk IIR band pass filter, the accuracy of the method is improved proportionately. By analyzing Fig. 9, the fault harmonics discrimination having low amplitudes is so difficult because the power average estimation error reaches a maximum value especially when the SNR decreases.

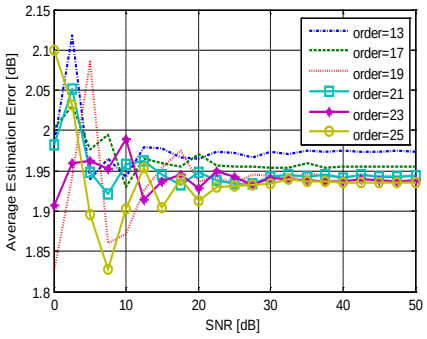


Fig. 8 Variation of fault powers estimation error depending on IIR filter order and SNR

Thus a satisfactory recognition results requires an SNR greater than 15dB.

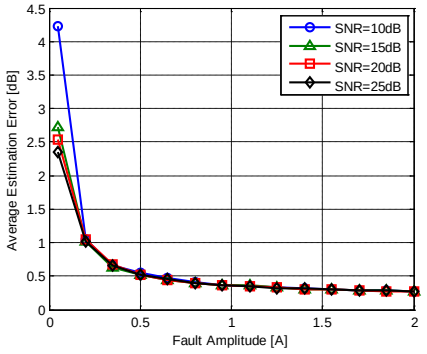


Fig. 9 Variation of fault powers estimation error according to fault amplitude and SNR

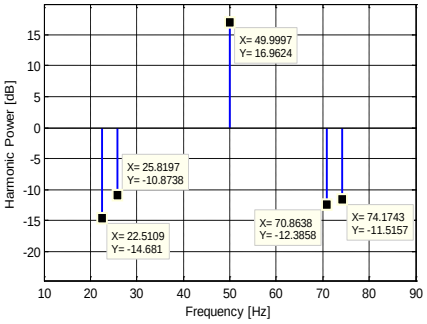


Fig. 10 Power and Frequency estimates by Fast-ESPRIT for broken rotor bars fault

ISBN: 978-1-61804-290-3

Against, when the fault magnitude increases the algorithm becomes able to track automatically the harmonics more accurately. Fig. 10 provides the achieved detection by the proposed method of broken rotor bars fault harmonics. It is noted that the Fast-ESPRIT algorithm was able to separate spectral components much closer and lower accurately in an optimal computation time which equal to 0.03021 second.

C. Bearing Damage Detection

In the third test, the proposed algorithm has been used to detect bearing damage fault signature in the frequency range [40, 200]Hz. Table VI gives the simulated fault parameters.

TABLE VI
BEARING DAMAGE FAULT PARAMETERS

f_{bng} (Hz)	I_{bng} (A)	P_{bng} (dB)	N_h	SNR (dB)
89.248	0.2	-16.99	3	20
155.248	0.25	-15.05		

Fig. 11 shows that the proposed approach provides a satisfactory result with high accuracy with a minimum computation cost which reach 0.07093 second even if the frequency range is wide. It is noted that this computation time is the twice compared to the time required to detect the previous faults.

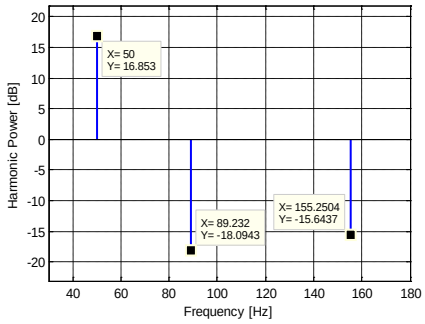


Fig. 11 Power and Frequency estimates by Fast-ESPRIT for bearing damage fault

D. Misalignment Detection

In this simulation, Fast-ESPRIT method has been evaluated to identify an important number of harmonics charactering misalignment fault signature in the frequency bandwidth [10, 210]Hz as showed in Table VII.

TABLE VII
MISALIGNMENT FAULT PARAMETERS

f_{mis} (Hz)	I_{mis} (A)	P_{mis} (dB)	N_h	SNR (dB)
21	0.22	-16.16	7	20
37.03	0.33	-12.64		
79	0.27	-14.38		
95.05	0.37	-11.65		
137.03	0.18	-17.90		
195.05	0.15	-19.49		

The detection results are given in the following figure.

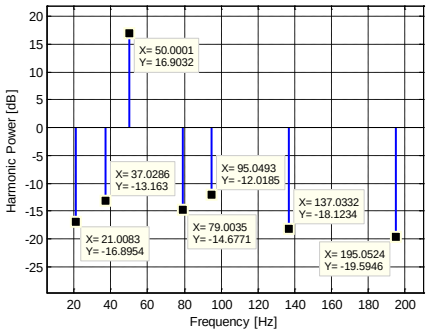


Fig. 12 Power and Frequency estimates by Fast-ESPRIT for misalignment fault

Referring to Fig. 12, the applied method offers good estimation ability with a very good computation cost equal to 0.07178 second. As illustrated in Fig. 13, when the frequency bandwidth [f_l, f_h] characterizing a fault contains an upper bound f_h which is increased and approaching or exceeding the value $F_{Nyquist}/2=F_s/4$, the decimation factor Γ decreases and thereafter the signal data samples increases. This causes the increase of the signal autocorrelation matrix dimension. Consequently this leads to a large calculation time. On the other side, if $f_h < F_{Nyquist}/2$ the computation time required by the Fast ESPRIT algorithm becomes minimal and it is almost without a big change despite the increase of the fault harmonics number contained in the stator current signal. This increase influences slightly on the calculation time which can be calculated as:

$$T_c = \Delta t \frac{N_s}{\Gamma} \tag{33}$$

Where Δt is the time required to process one data sample.

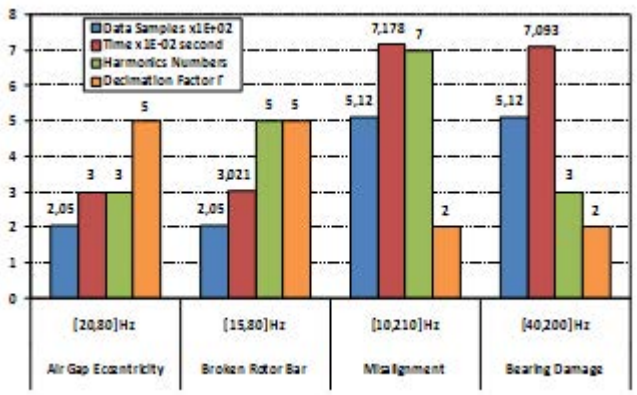


Fig. 13 Computation time depending on Fault Frequency range and decimation factor

Another remark that can be added to this interpretation is that the increase of F_s leads to increase the size of the signal data samples. This causes the augmentation of the autocorrelation matrix dimension. As result, the Fast-ESPRIT method takes a long time to discriminate all the frequencies contained in the processed signal with an acceptable accuracy.

ISBN: 978-1-61804-290-3

So in order to adapt accurately the detection algorithm to a real-time application, there is a trade-off between the choice of an optimal sampling frequency in a side and the computation time in the other side.

VIII. Conclusion

ESPRIT method has better performances than others super resolution algorithms for identifying frequencies from a short data signal acquisition drowned in a noise. However, the main drawback of this technique is the high computational time especially when the size of the signal autocorrelation matrix increases. A new version of the ESPRIT algorithm is proposed in this paper entitled Fast-ESPRIT. An improvement is realized with an adequate IIR band pass filtering and an optimal decimation technique. This enhancement leads to low complexity, to satisfactory accuracy and to memory storage reduction algorithm. The proposed technique was applied under different wind turbine faults to evaluate its resolution ability. Analysis of the simulation results shows that estimator achieves remarkable performance estimation in extracting frequencies and amplitudes in a specified bandwidth than the original ESPRIT. Moreover, due to its advantages, Fast-ESPRIT method can be implementable for real time fault diagnosis. The future work will be focused to ameliorate Fast-ESPRIT detection efficiency in the case of low amplitudes harmonics.

References

[1] Hamid A. Toliyat et al., "Electric Machines Modeling, Condition Monitoring, and Fault Diagnosis", CRC Press Taylor & Francis Group NW 2013

[2] M. L. Sin, W. L. Soong and N. Ertugrul, "On-Line Condition Monitoring and Fault Diagnosis – A Survey" Australian Universities Power Engineering Conference, New Zealand, 2003

[3] K. K. Pandey et al, "Review on Fault Diagnosis in Three-Phase Induction Motor", MEDHA – 2012, Proceedings published by International Journal of Computer Applications (IJCA)

[4] E. Al Ahmar et al, "Advanced Signal Processing Techniques for Fault Detection and Diagnosis in a Wind Turbine Induction Generator Drive Train: A Comparative Study", IEEE Energy Conversion Congress and Exposition ECCE 2010, Atlanta United States 2010

[5] John L. Semmlow, "Biosignal and Biomedical Matlab-Based Applications", Marcel Dekker, Inc New York 2004

[6] Neelam Mehala et al, "Condition monitoring methods, failure identification and analysis for Induction machines", International Journal of Circuits, Systems and Signal Processing, Issue 1, Volume 3, 2009, pages 10-17

[7] Gérard Blanchet and Maurice Charbit, "Digital Signal and Image Processing using Matlab", ISTE USA 2006

[8] Yassine Amirat et al, "Wind Turbine Bearing Failure Detection Using Generator Stator Current Homopolar Component Ensemble Empirical Mode Decomposition", IECON 2012 - 38th Annual Conference on IEEE Industrial Electronics Society

[9] Elie Al-Ahmar et al, "Wind Energy Conversion Systems Fault Diagnosis Using Wavelet Analysis", International Review of Electrical Engineering Volume 3, No 4 2008, pages: 646-652, http://hal.univ-brest.fr/docs/00/52/65/07/PDF/IREE_2008_AL-AHMAR.pdf

[10] El Houssin El Bouchikhi, Vincent Choqueuse, M.E.H. Benbouzid, "Non-stationary spectral estimation for wind turbine induction generator faults detection", Industrial Electronics Society, IECON 2013-39th Annual Conference of the IEEE 2013, pp 7376-738

[11] Ioannis Tsoumas et al, "A Comparative Study of Induction Motor Current Signature Analysis Techniques for Mechanical Faults Detection, SDEMPED 2005 - International Symposium on Diagnostics for Electric Machines", Power Electronics and Drives Vienna, Austria, 7-9, September 2005

[12] Yong-Hwa Kim, "High-Resolution Parameter Estimation Method to Identify Broken Rotor Bar Faults in Induction Motors, IEEE Transactions on Industrial Electronics, Vol. 60, Issue 9, pages 4103 – 4117, September 2013

[13] Shahin Hedayati Kia et al, "A High-Resolution Frequency Estimation Method for Three-Phase Induction Machine Fault Detection", IEEE Transactions on Industrial Electronics, Vol. 54, No. 4, AUGUST 2007

[14] Saad Chakkor et al., "Performance Analysis of Faults Detection in Wind Turbine Generator Based on High-Resolution Frequency Estimation Methods", International Journal of Advanced Computer Science and Applications, SAI Publisher, Volume 5 No 4, May 2014, pages 139-148

[15] Jian Zhang et al, "Rank Reduced ESPRIT Techniques in the Estimation of Principle Signal Components", Proceedings 5th Australian Communications Theory Workshop, Australian National University, 2004

[16] Shawn Sheng and Jon Keller et al, "Gearbox Reliability Collaborative Update", NREL U.S. Department of Energy, http://www.nrel.gov/docs/fy14osti/60141.pdf

[17] J. Proakis and D. Manolakis, "Digital Signal Processing : Principles, Algorithms, and Applications", New York: Macmillan Publishing Company, 1992

[18] André Quinquis, "Digital Signal Processing using MATLAB", ISTE Ltd, London UK, 2008

[19] Monson H. Hayes, "Statistical Digital signal processing and modeling", John Wiley & Sons, New York, 1996

[20] R. Roy and T. Kailath, "ESPRIT-estimation of signal parameters via rotational invariance techniques," IEEE Trans. Acoust., Speech, Signal Processing, vol. 37(7), pp. 984 –995, July 1989

[21] Fredric J. Harris, "Multirate Signal Processing for Communication Systems", Prentice Hall, Mai 2004

[22] Joao Paulo C. L. da Costa et al, "Comparison of Model Order Selection Techniques For High-Resolution Parameter Estimation Algorithms", 54th Internationales Wissenschaftliches Kolloquium, Technische Universität Ilmenau, Germany, 2009

[23] Janos J. Gertler, "Fault Detection and Diagnosis in Engineering Systems Basic concepts with simple examples", Marcel Dekker Inc., New York, 1998

[24] Saad Chakkor et al., "Wind Turbine Fault Detection System in Real Time Remote Monitoring", International Journal of Electrical and Computer Engineering (IJECE), IAES Publisher, Volume 4 No 6, December 2014

[25] H. Vincent Poor, "An Introduction to Signal Detection and Estimation", Second Edition, Springer-Verlag texts in electrical engineering, Virginia USA 1994

Saad Chakkor was born in Tangier Morocco. He's a member in the Physics department, Communication and detection Systems laboratory, Faculty of sciences, University of Abdelmalek Essaâdi, Tetouan Morocco, and his research area is: wireless intelligent sensors and theirs applications, frequency estimation algorithms for faults detection and diagnosis system in electromecanical machines. He obtained the Master's degree in Electrical and Computer Engineering from the Faculty of Sciences and Techniques of Tangier, Morocco in 2002. He graduated enabling teaching computer science for secondary qualifying school in 2003. In 2006, he graduated from DESA in Automatics and information processing at the same faculty. He works as teacher of computer science in the high school.

Mostafa Baghouri was born in Tangier Morocco. He's a member in the Physics department, Communication and detection Systems laboratory, Faculty of sciences, University of Abdelmalek Essaâdi, Tetouan Morocco, his research area is: routing and real time protocols for energy optimization in wireless sensors networks. He obtained a Master's degree in Electrical and Computer Engineering from the Faculty of Science and Technology of Tangier in Morocco in 2002. He graduated enabling teaching computer science for secondary qualifying school in 2004. In 2006, he graduated from DESA in Automatics and information processing at the same faculty. He work teacher of computer science in the high school.

Abderrahmane Hajraoui is a professor of the Higher Education at University of Abdelmalek Essaâdi. He's a director thesis in the Physics department, Communication and detection Systems laboratory, Faculty of sciences, University of Abdelmalek Essaâdi, Tetouan, Morocco. His research areas are: Signal and image processing, automation systems, simulation systems, antennas and radiation, microwave devices and intelligent wireless sensors networks.

ISBN: 978-1-61804-290-3

Modeling Security Risks for Smart Grid Networks

Suleyman Kondakci
Izmir University of Economics,
Faculty of Engineering & Computer Sciences, Izmir–Turkey
suleyman.kondakci@ieu.edu.tr

***Abstract*—A set of models is presented here for analyzing risks to smart grid networks and for the determination of joint risks caused by multiple threat sources. Regarding heterogeneous communication environments, it is still an open issue to define justifiable models that can associate a risk assessment and its decision–making process on a solid ground. This numerically astute model proposes a novel concept that can help a security evaluator to quantitatively determine dependence and causality within a network of interconnected systems and their applications.**

***Index Terms*—Security analysis, quantitative risk modeling, joint causality.**

I. Introduction

AN electrical (power) grid network is composed of a set of interconnected networks for producing electricity and delivering it to consumers. Obviously, it serves as a critical infrastructure. Due to the geographical dispersion of the power plants, functions for the management and control of power grids are distributed over wide are networks (WAN). One may think of a smart grid network as a three–layered complex structure consisting of (i) power generation and distribution, (ii) data exchange, and (iii) management (control and command) layers. Such complex structures bearing various intelligence and functionality for managing all aspects of electrical grids are called smart grid networks (SGN). Further details on the structure and functionality of SGNs can be found in [1]–[4], and [5].

The community of power grids still needs appropriate techniques and tools to develop adequate analysis and assessment approaches that can quantitatively determine causal and joint risks and other adversarial implications. Because such techniques and tools can be used to build more accurate and balanced safety measures for dynamically growing heterogeneous environments. A SGNsas a critical infrastructure requires balanced security and reliability mechanisms being the most fundamental requirements. Such infrastructures should provide economically and technically feasible non-overlapping functionality, while being protected by the vital security functions (e.g., periphery protection, secure authentication, authorization, effective load-shedding, and malware protection). To shed some light on this issue, we propose here a fundamental concept based on deterministic and stochastic causality by focusing on a model based analysis of risks.

Methods applicable regardless of the protection states of environments can be a valuable means to build concepts for the analyses of generic risk factors and parameters. Because security analyses independent of the protection state of environments can constitute the most fundamental approaches for generic risk assessment and protection methods that can be easily adapted to a wide range of engineering fields. In light of this, we focus here on the definition and use of some fundamental models that are needed for the quantification of risks that may be caused by various threat sources, including human failures and deliberate attacks. Threats emanating from different sources may lead to impacts of a joint character. In such cases, the evaluation of security breeches should not be considered independently, unless a careful analysis of the dependency structure has been performed prior to the evaluation.

II. Related Work

As also documented by several other organizations, NIST Framework and Roadmap for Smart Grid Interoperability Standards clearly specify Cyber threats to grids, [5]. As stated by [6], it is important to consider the vulnerability analyses of SGNs from both structural and functional perspective. Amongst others, the intended contribution of this paper is mainly to define a generic model for emphasize risks caused by various threat sources. Some related models considering similar problems dealing with IP-networks are presented in [7] and [8]. As also stated by several SGN forums, Cyber security must address not only deliberate attacks, such as from disgruntled employees, industrial espionage, and terrorists, but inadvertent compromises of the information infrastructure due to user errors, equipment failures, and natural disasters. Vulnerabilities might allow an attacker to penetrate a network, gain access to control software, and alter load conditions to destabilize the grid in unpredictable ways. Our work considers risk assessment rather than detailing the Cyber security procedures. An approach dealing with Layer 2 security for smart grid networks is presented in [9]. More security stuff can be found in [10]–[14], and [15].

In a wider perspective risk modeling can be applied to solving of various types of problems, in particular fault localizations and prediction of component errors, e.g., the work given in [16] applies a risk modeling strategy to fault localization in IP–networks. The concept presented in this paper introduces also joint effect analyses in order to familiarize the security engineer with the simplicity of the concept, [17]. As an introductory example, a joint vulnerability analysis of security and routing protocols is also considered by [18].

III. Modeling the Threat Classes

A smart grid network consists of non-homogeneous domains (networks and information pathways) that enable an

ISBN: 978-1-61804-290-3

application in a particular domain to communicate with an application in any other domain within the interconnected networks, see Fig. 1. Also, devices in different network segments manage and control each other for the provision of stable and high-quality power to consumers. Securely authorized accesses to various resources need to be established with proper management having control over who and where applications can be interconnected. That is, authentication and authorization of principals (users, devices, code-pieces, utility applications, etc) that occur through the infrastructure of SGNs differ significantly from the traditional homogeneous business networks. This means that alongside with threats and protection mechanisms risk assessment approaches and techniques will significantly be different compared to that of the homogeneous environments.

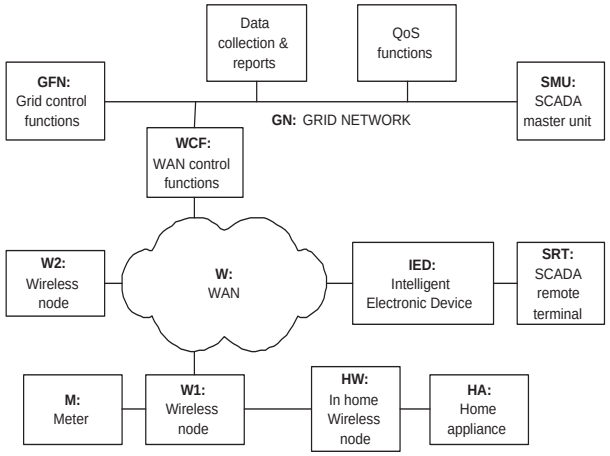


Fig. 1. An Architectural overview of smart grid networks

As shown in Fig. 1, we have a collection of networks that exchange information and control data among the services, devices, and functions they possess in an SGN segment. Prior to a security assessment, actual threat classes and associated targets within these networks are identified often by use of a validated threat taxonomy, e.g., some of those to mention, if not older, are discussed in [19]–[25]. Usually, based on some penetration scenarios, a data mining and preprocessing of existing data need to be performed. That is, success probabilities to every penetration action and its associated attack scenario are calculated and inserted into the risk calculation process. Further, eventual impacts of future attacks will be evaluated by use of these probabilities and quantitative risk values. If appropriate, some existing data can be used, e.g. the DARPA repository. Reader can refer to [26] for some useful critiques about DARPA data sets.

A. Adversary Threat Sources

Supervisory Control and Data Acquisition (SCADA) represents a wide range of protocols and technologies for monitoring and managing equipment and machinery in various sectors of critical infrastructure and industry, [5]. This includes power generation, manufacturing, oil and gas, water treatment, and waste management. Therefore, the security of SCADA technologies and protocols is a serious concern mostly related to infrastructure security. Because attacks (or threats) to such crucial services can lead to serious failures in the infrastructure that can result in potential losses of services for consumers of a wide spectrum. SCADA systems are often purchase as bundled packages, which are mostly opaque to end-users, who are not often able to know what is inside and what precautions needed to keep the system safe from threats and vulnerabilities. SCADA uses central hosts to receive monitoring information from remote locations, send commands for managing remote instruments and hosts, analyze and control information on the operator screens.

SCADA networks were earlier isolated from public networks. Along side the rapid growth of industry and operational complexity, the connectivity demand to other external systems has also increased. SCADA systems are nowadays connected to other networks to increase the scope of functionality that are required for remote access and control operations. This will naturally give rise to immense security threats to the SCADA networks, [6]. Since the SCADA networks are connected to public networks they will confront threats that also exist for the Internet. This makes the SCADA systems more vulnerable. Especially, the use of the TCP/IP protocol suite connecting SCADA systems together will also bear the implementation specific deficiencies found in the TCP/IP protocol suite. This will facilitate attackers to easily gain access to in-depth knowledge about the SCADA networks.

We usually encounter three distinct classes of external–adversary threats: (i) vulnerable applications, (ii) user–activated exploits, and (iii) organized (dedicated) intensive attacks. For further analyses, a brief description of the three major threat sources considered here are given below. The first category is due to random user activities that may lead to exploits in vulnerable applications. Some of these threats are Trojans, SQL-injection, cross-site scripts, password phishing, pop-ups, and many others that give the attacker the opportunity to activate attacks on the victim machine by a vulnerable application. Since these attacks make use of user interventions, they are directly associated with the human-related threat types, which should be analyzed under a broader focus. Because there are several manually initiated legacy functions that perform critical data exchange among various SGN networks, management, maintenance, and control operations within a given SGN and within their substations, remotely as well.

1) Model 1: Messaging and Data Sharing: Some control messages may unintentionally contain worms and Trojans that can interrupt a service, spread to other applications, cause hazards to devices of interest. We call such messages as malicious codes or malmessages. Malmessages regarding network management functions, home appliances, instant communications, intelligent meters, manually activated grid control data packets, and peer-to-peer (P2P) SCADA communications belong to the category of user–activated threats.

General IP-traffic, [27], can also be considered as a prominent threat source for SGNs. Consider a set of virus-infected messages accessed by a user or by an application running within an SGN segment (or central host). Let the probability that each of the fixed number of contacts of a node with the

ISBN: 978-1-61804-290-3

viral messages causing k_i infections of type i is denoted by $p_k(t)$. Probability of an infection transmission is a random variable ξ with parameters p_i and n, where p_i denotes the probability of a user action among n activities causing faults on grid control functions. This is analogous to incidents caused by malicious codes, email messages, instant messengers, and P2P fileshares on a computer node with small to medium number of user activities (i.e., viral message traffic) that exhibit binomially distributed probabilities, [28], expressed as

$$P_\xi(k_i) = P\{\xi = k_i \mid n, p_i\} \quad (1)$$
$$= \frac{n!}{k_i!(n-k_i)!} p_i^{k_i}(1-p_i)^{(n-k_i)}, k_i = 0, 1, \ldots, n$$

This gives precisely k_i successes of the infection and $n - k_i$ attack failures (no infection) where each single success has probability p_i. However, generally, larger malicious message traffics satisfy the condition of fitting an approximate Poisson distribution similar to that of dedicated (adversary) attacks such as the DoS attack. This is true, if the victim user receives infectious messages from a large number of computers throughout the Internet. This is a typical case similar to that of a Distributed DoS attack (DDoS), where hundreds or thousands of attack engines are involved in attacking a single target, [29].

B. Model 2: Intentional Dedicated Attacks

There are many reasons for launching intentional attacks against a given target, e.g., for the purpose of sabotage, espionage, and reconnaissance. Some attacks are launched intensively in a sweep mode in order to quickly achieve the desired goal. We classify these attacks as dedicated intensive attacks, which have the distribution of the threat probability approximated to a Poisson process,

$$P_\xi(k_i) = P\{\xi = k_i \mid \lambda_i\} \approx \frac{\lambda_i^{k_i}}{k_i!} e^{-\lambda_i}, \quad (2)$$

with parameter $\lambda_i = np_i$, if the number of attacks, n, is large and the probability p_i of success of an attack is small. This is the probability of encountering k_i incidents of type i. As known, a Poisson process is a stochastic process which counts the number of events and the time that these events occur in a given time-sweep. The time between each pair of consecutive events has an exponential distribution with parameter λ and each of these inter-arrival sweeps is assumed to be independent of other inter-arrival sweeps. An example to dedicated massive attacks could be such that the adversary can make use of a compromised SCADA terminal to launch diffusion attacks to grid control systems, which in turn, can delay or disrupt power generation and distribution tasks. WAN control functions, AMIs, and EDIs can also be considered as victims of this category of threats, e.g., DoS attacks. Reader can refer to reports from SANS and Symantec for threats to SCADA and vulnerabilities found in SCADA systems, see also [30]–[32] and [33] for some details on the SCADA security.

C. Model 3: Catastrophic situations

Suppose that attacks were composed of a common threat type launched via a (SCADA) system remote terminal, e.g., denial of service attack (DoS), that took advantage over vulnerabilities inherent in N different SCADA master units having applications with identical vulnerabilities. There are several channels of attacks that can cause s complete crash-down effect. For example, the UDP protocol is widely used in network servers and applications, e.g., DNS, NTP, SNMP, and NetBIOS. A Distributed Reflective Denial of Service (DRDoS) attack is an emerging form of Distributed Denial of Service (DDoS) attack that relies on the use of publicly accessible UDP servers. Let such an attack event occur with probability $p(t)$, and let the probability that k successful attacks occur as a result be conditionally binomially distributed with parameter θ. Let also $a_{i,k}(t)$ be the probability that at time t we discover attacks with success size $k = \{0, 1, 2, \ldots\}$, given that i nodes (computers that control operations of an SGN) are already compromised. Then,

$$a_{i,k}(t) = p(t)\binom{N-i}{k}\theta^k\Big[1-\theta\Big]^{N-i-k} \quad (3)$$

can be used to form one-step transition probabilities of a Markov chain, which can be used to calculate state transitions at any time slot. Here, *state* defines the number of elements found in a specific condition in a given time slot, e.g., healthy or failed (compromised). This model can be used to describe a catastrophic situation, i.e., if $\theta = 1$ then all machines will go into the down state. That is, $I - i$ failures will occur simultaneously.

IV. Overall System State: State Transition Model for the Compromised Systems

At any given time t, the number of failed nodes is specified by the current state regardless of the behavior of the system (set of compromised/attacked computers) before the time t. State s_x means that x computers are in state s (healthy or problematic). Starting with some initial state at time $t_0 = 0$, the system under attack may change its state randomly at subsequent times. Thus, at time t, the state of the system will be denoted by the random variable $\xi(t)$. Based on this assumption, the evolution of the system in time is described by the Markov chain $\xi(0), \xi(1), \ldots, \xi(n)$, giving the state of the system at time $t = 0, t = 1, \ldots, t = n$.

Let

$$p_j(n) = \mathbf{P}\{\xi(n) = s_j\}, \quad j = 1, 2, \ldots \quad (4)$$

be the probability that the system (the SGN network) will be in state s_j after n steps of transitions. To justify this, first we assume that after $n - 1$ steps the network will be in one of the s_k states, i.e., the events $\{\xi(n) = s_k\}, k = 1, 2, \ldots$ form a complete set of mutually exclusive events (not everyone in action at the same time) in the sense that one and only one of the k events always occurs. Assuming that at time $t_0 = 0$, the network is in the initial state s_i with the initial probability of

$$p_i^0 = \mathbf{P}\{\xi(0) = s_i\}, \quad i = 1, 2, \ldots, \quad (5)$$

then, the probability that the system goes into the state s_j from state s_i at the next step is given by

$$p_{i,j} = \mathbf{P}\{\xi(n+1) = s_j \mid \xi(n) = s_i\}, \quad i, j = 1, 2, \ldots, \quad (6)$$

ISBN: 978-1-61804-290-3

regardless of its behavior before the time t_i, due to the Markovian property: the memoryless property of a stochastic process. Therefore, the numbers corresponding to $p_{i,j}(t)$ do not depend on the time (or step) n. Let

$$p_j(n) = \mathbf{P}\{\xi(n) = s_j\}$$

be the probability that the system will be in state s_j after n steps, and

$$p_k(n-1) = \mathbf{P}\{\xi(n-1) = s_k\}$$

be the probability that the system was in state s_k before the last step, i.e., $n-1$. Then, we can find the probability of the system state at any step by referring to the *total probability formula*, i.e., at step n,

$$\mathbf{P}\{\xi(n) = s_j\} = \sum_k \mathbf{P}\{\xi(n) = s_j \mid \xi(n-1) = s_k\} p_k(n-1). \tag{7}$$

Rewriting Eq. (7) in terms of (4), (5), and (6) we get to the following recursive formulae, [34],

$$\begin{aligned} p_j(0) &= p_j^0, \\ p_j(n) &= \sum_k p_k(n-1) p_{k,j}, \quad n = 1, 2, \ldots \end{aligned} \tag{8}$$

Naturally, assuming that the system is in a certain state s_i, initially, at time $t = 0$, then the initial probability distribution can be defined as

$$p_i^0 = 1, p_k^0 = 0, \quad k \neq i.$$

Thus, the probability $p_j(n)$ at time n can be expressed as

$$p_{i,j}(n) = \mathbf{P}\{\xi(n) = s_j \mid \xi(0) = s_i\}, \quad i, j = 1, 2, \ldots$$

Using *Chapman–Kolmogorov equations* we can estimate the system status after $n-$ transitions denoting the probability that the system will be compromised due to attacks with the Poisson distribution. Hence, given the initial distribution

$$p_{i,j}(0) = \begin{cases} 1 & \text{if} \quad j = i, \\ 0 & \text{if} \quad j \neq i, \end{cases}$$

the system will go from state s_i to s_j in n steps obeying

$$p_{i,j}(n) = \sum_k p_{i,k}(v) p_{k,j}(n-v), \forall_{(i,j,n)}, 0 \leq v \leq n. \tag{9}$$

This equation states that the system first goes from state i into the intermediate state k after exactly v steps, $0 \leq v \leq n$. Thus, $p_{i,k}(v) p_{k,j}(n-v)$ denotes the conditional probability that the system goes to state k (starting in state i) after v steps and then to state j in $n-v$ steps. Assuming the special cases of $v = 1$ and $v = n-1$ we can define the state transition probabilities

$$p_{i,j}(n) = \sum_k p_{i,k}(1) p_{k,j}(n-1) = \sum_k p_{i,k} p_{k,j}(n-1)$$

and

$$p_{i,j}(n) = \sum_k p_{i,k}(n-1) p_{k,j}.$$

From these definitions, we can recursively obtain the n-step transition probabilities, $\left(p_{i,j}\right)^{(n)}$, expressed in a *transition probability matrix* given as

$$\mathbf{P}^{(n)} = \begin{pmatrix} p_{0,0}(n) & p_{0,1}(n) & \ddots & p_{0,N}(n) \\ p_{1,0}(n) & p_{1,1}(n) & \ddots & p_{1,N}(n) \\ p_{2,0}(n) & p_{2,1}(n) & \ddots & p_{2,N}(n) \\ \vdots & \vdots & \ddots & \vdots \\ p_{M,0}(n) & p_{M,1}(n) & \ddots & p_{M,N}(n) \end{pmatrix}.$$

Furthermore, we need to determine *steady–state transition probabilities* after a long–run of transitions. In order to compute the n-step steady–state (stationary) transition probabilities, consecutive probabilities at each step are multiplied, i.e.,

$$\mathbf{P}^{(0)} = \mathbf{P}, \mathbf{P}^{(2)} = \mathbf{P}.\mathbf{P} = \mathbf{P}^2, \ldots$$

In short, given the initial probability matrix $\mathbf{P}(0)$, a general form for the n-step stationary transition probability can be obtained by the product of n matrices,

$$\mathbf{P}^{(n)} = \mathbf{P}^n, \quad n = 1, 2, \ldots$$

For example, given the empirically determined initial probability matrix, e.g.,

$$\mathbf{P}^{(0)} = \begin{pmatrix} 0.000 & 0.202 & 0.124 & 0.200 & 0.474 \\ 0.130 & 0.320 & 0.200 & 0.129 & 0.221 \\ 0.180 & 0.084 & 0.100 & 0.152 & 0.484 \\ 0.000 & 0.110 & 0.219 & 0.320 & 0.351 \\ 0.140 & 0.026 & 0.024 & 0.310 & 0.500 \end{pmatrix}, \tag{10}$$

and the distribution function of its n–state (long–run) transitions, we can easily compute the corresponding steady–sate probabilities

$$\vec{\pi} = [\pi_0, \pi_1, \ldots \pi_n]. \tag{11}$$

Using the initial state probabilities (e.g., Eq. (10)), we can also determine the stationary probabilities $\pi_0, \pi_1, \ldots, \pi_n$ by solving the equation set

$$\begin{aligned} \pi_0 &= p_{0,0}\pi_0 + p_{1,0}\pi_1 + p_{2,0}\pi_2 + \ldots \\ \pi_1 &= p_{0,1}\pi_0 + p_{1,1}\pi_1 + p_{2,1}\pi_2 + \ldots \\ \pi_2 &= p_{0,2}\pi_0 + p_{1,2}\pi_1 + p_{2,2}\pi_2 + \ldots \\ \pi_3 &= p_{0,3}\pi_0 + p_{1,3}\pi_1 + p_{2,3}\pi_2 + \ldots \\ &\vdots \quad \vdots \quad \ddots \quad \vdots \\ \pi_n &= p_{0,n}\pi_0 + p_{1,n}\pi_1 + p_{2,n}\pi_2 + \ldots \end{aligned} \tag{12}$$

where the stationary probabilities sum to 1, i.e.,

$$1 = \pi_0 + \pi_1 + \ldots + \pi_n.$$

For example, considering the sample matrix Eq. (10), a vector of stationary probabilities after 16 consecutive transitions are computed as

$$\vec{\pi} = [\pi_0, \pi_1, \pi_2, \pi_3, \pi_4] = [0.093, 0.100, 0.112, 0.267, 0.428].$$

Thus, after 16 time units (e.g., minutes), the probability of observing zero, one, two, three, and four successful attacks resulting in system failures of a specific type tend to be $\pi_0 = 0.093, \pi_1 = 0.100, \pi_2 = 0.112, \pi_3 = 0.267$, and $\pi_4 = 0.428$.

ISBN: 978-1-61804-290-3

V. Determining the Joint Risk

The quantitative/probabilistic models described above can now be incorporated into the joint risk computation for any part undergoing the risk assessment. We define first a causal model for incorporating the quantitative threat levels, which will be used to compute the risk for the related asset (or system unit) shown in Fig. 2. As a reference, we use the

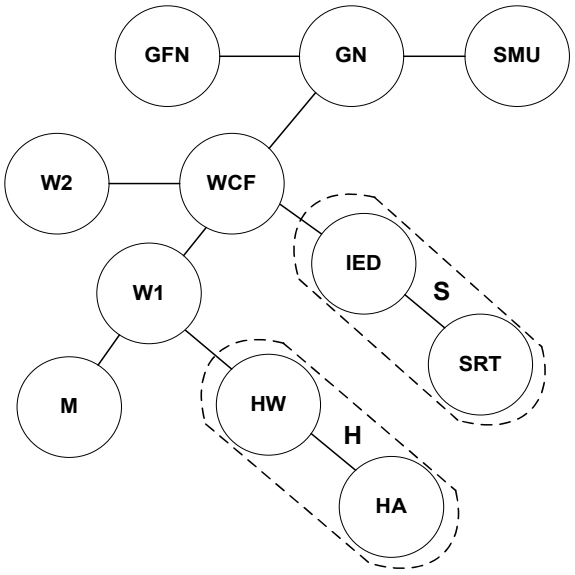


Fig. 2. Causal dependence diagram showing the nodes of the threat flow path

architectural model shown in Fig. 1 to build a causal model for the determination of causal risks. Therefore, the abbreviations from Fig. 1 are directly copied into a risk-flow diagram to represent the risky nodes of the entire architecture. The risky nodes represent the critical path that threats use for propagating to other nodes. The causal model will then be converted to a simplified model given as a directed acyclic graph (DAG) in order to simplify the computation of the overall risk, Fig. 3. As already noticed, Fig. 3 shows a simplified version of the risk-flow diagram, so that the overall risk computation can be facilitated by use of a Bayesian belief networks (BBN) tool. Results from the BBN–based tool are used for making inferences on the causal (or marginal) risk distribution. A BBN is a compact representation of a probabilistic model, by which dependences (and, implicitly, conditional independences) among a given set of variables can be completely described, [35]. Given a graphical structure (i.e., DAG) of the dependency variable set, the joint probability distribution over that set can be completely described by specifying the appropriate set of marginal and conditional distributions over the variables. For example, given the input probability (prior) distributions for the belief network shown in Figure 3, we can determine posterior probability functions in order to estimate the causal (joint) risk parameters, α, β, γ, which can then be used to make an overall inference on the expected risk level for the entire network. Probably the most important constraint on the use of belief networks (BNs) is the fact that, in general, computation of every nodes's belief of very large networks is NP–hard, [36]. However, there exist

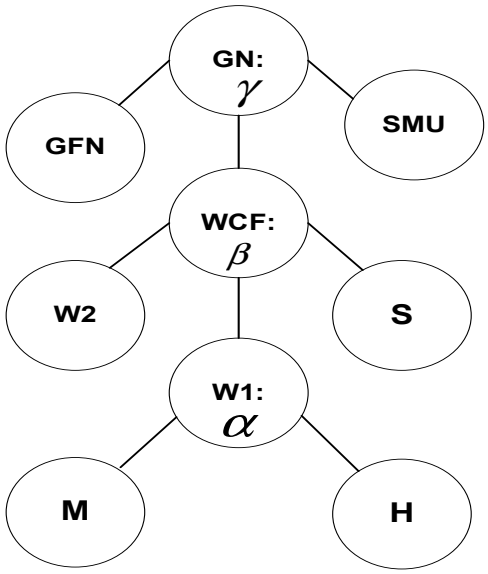


Fig. 3. Risk-flow diagram of the SGN architecture shown in Fig. 1

inference algorithms to manipulate conditional independences in BNs with limited sizes. Some known algorithms from [37] and [38] can be modified to apply to such networks, [35].

Thus, having the threat flow diagram shown in Fig. 2, we can build a simplified risk-flow diagram, which is shown in Fig. 3. In order to determine the major impact parameters α, β,and γ of the risk-flow model shown in Figure 3, joint posterior probabilities of these parameters must be computed first. The events for α, β, and γ are random, which may also occur simultaneously and independently at different time slots.

Considering the threat model shown in Figure 3, the set of conditionally independent model parameters (α, β, and γ) are used to construct the following union:

$$\bigcup_{i=1}^{n} A_i = \{\alpha, \beta, \gamma\}.$$

Then, given the Belief network of the risk-flow shown in Figure 3, the joint probability can be computed by first determining the individual values for the parameters α, β, and γ. Referring to the reliability theory, we can easily compute risk values corresponding to the above probabilistic definitions. Following this, the probability of the combined risk, i.e., the probability of the above union can be obtained by rewriting

$$P_{\mathcal{A}} = P\Big(\bigcup_{i=1}^{n} \mathcal{A}_i\Big) = 1 - \prod_{i=1}^{n} [1 - P(\mathcal{A}_i)]$$

as

$$P_{\mathcal{A}} = P\Big(\bigcup_{i=1}^{4} \mathcal{A}_i\Big) = 1 - \prod_{i=1}^{4} [1 - \alpha][1 - \beta][1 - \gamma]. \quad (13)$$

Thus, per–system risk for any system, say $\mathcal{A}$, is defined as the product of the asset weight $\mathcal{A}_w$ and the probability of the combined risk for system $\mathcal{A}$. Given the observed data set, posterior probabilities of these events are computed and transfered to quantitative values for the estimation of a total

ISBN: 978-1-61804-290-3

risk $\mathbf{R}_{\mathcal{A}}$ for asset $\mathcal{A}$ having weight $\mathcal{A}_w$., i.e., simply

$$\mathbf{R}_{\mathcal{A}} = (P_{\mathcal{A}} - F_{\mathcal{A}})\mathcal{A}_w, \tag{14}$$

where $\{P_{\mathcal{A}}, F_{\mathcal{A}}\} \in [0, \ldots, 1], \{\mathbf{R}_{\mathcal{A}}, \mathcal{A}_w\} \in [0, \ldots, 100]$. Here, $F_{\mathcal{A}}$ denotes the failure probability of a threat incident. $P_{\mathcal{A}}$, Eq. (13), denotes the joint causal probability of a set of threat incidences for system $\mathcal{A}$, which can be easily incorporated into probabilistic inference models, e.g., into an appropriate BBN.

A. Putting it Together

Suppose we have derived the equations for computing the per-system/node risk and have the proper equations for computing the probabilities of attacks and failures, Eq. (1)–(12). Then we can expand the idea of per-system risk computation to contain a finite set of heterogeneous systems (or assets). It is thus obvious that the joint risk of a finite set of systems can be easily computed, since we have the necessary data at the disposition:

- $\mathcal{P}^{(n)} = \{P_1, P_2, \ldots, P_n\}$:= Probabilities of threats,
- $\mathcal{F}^{(n)} = \{F_1, F_2, \ldots, F_n\}$:= Probabilities of attack failure,
- $\mathcal{W}^{(n)} = \{w_1, w_2, \ldots, w_n\}$:= System weights (importance factor).

Hence, the normalized magnitude of the overall risk, $\mathcal{R}^{(n)} = \{R_1, R_2, \ldots R_n\}$, for a network of n systems can be computed by

$$\mathcal{R}^{(n)} = \left(\frac{1}{n} \cdot \sum_{i=1}^{n} R_i^{-1}\right)^{-1} = \frac{n}{\sum_{i=1}^{n} \frac{1}{R_i}}, \quad R_i = (P_i - F_i)w_i, \tag{15}$$

where,

$$\{R^{(n)}, R_i, w_i\} \in [0, \ldots, 100], \{P_i, F_i\} \in [0, \ldots, 1].$$

VI. Conclusions

We have presented a risk assessment approach mainly based on a stochastic model based on an interdependence structure of a set of systems. The model facilitates the determination of the dependence structure of risk parameters and related risk values applicable to smart grid networks. Causal threat analysis producing quantitative risk propagation data is necessary for risk management of critical infrastructures such as the power generation and transmission facilities. The assessment model presented here can easily categorize and analyze data from threat sources and facilitate the assessment of a single-asset (system) and multiple-asset environments. This risk propagation approach makes use of the conditional and stochastic probability methods combined with system weighing scheme, so that risk levels of any component within the network can be easily quantified. There are obvious interdependency factors among different types of operational domains within an SGN, which can affect both the human– and technology–related security risks. Thus, the concept presented here can enable the dissection of various threat paths leading to specific risk-flaw paths within the WAN of SGNs.

References

[1] F. Bouhafs, M. Mackay, and M. Merabti, "Links to the future: Communication requirements and challenges in the smart grid," *Power and Energy Magazine, IEEE*, vol. 10, no. 1, pp. 24–32, Jan 2012.

[2] V. Gungor, D. Sahin, T. Kocak, S. Ergut, C. Buccella, C. Cecati, and G. Hancke, "Smart grid technologies: Communication technologies and standards," *Industrial Informatics, IEEE Transactions on*, vol. 7, no. 4, pp. 529–539, Nov 2011.

[3] C.-H. Lo and N. Ansari, "The progressive smart grid system from both power and communications aspects," *Communications Surveys Tutorials, IEEE*, vol. 14, no. 3, pp. 799–821, Third 2012.

[4] M. Hammoudeh, F. Mancilla-David, J. Selman, and P. Papantoni-Kazakos, "Communication architectures for distribution networks within the smart grid initiative," in *Green Technologies Conference, 2013 IEEE*, April 2013, pp. 65–70.

[5] G. Locke and P. D. Gallagher, "Nist framework and roadmap for smart grid interoperability standards, release 1.0," pp. 1–145, January 2010.

[6] P. Chopade and M. Bikdash, "Structural and functional vulnerability analysis for survivability of smart grid and scada network under severe emergencies and wmd attacks," in *Technologies for Homeland Security (HST), 2013 IEEE International Conference on*, Nov 2013, pp. 99–105.

[7] S. Kondakci, "Intelligent network security assessment with modeling and analysis of attack patterns," *Security and Communication Networks*, vol. 5, no. 12, pp. 1471–1486, 2012.

[8] S. Kondakci, "A causal model for information security risk assessment," in *Information Assurance and Security (IAS), 2010 Sixth International Conference on*, aug. 2010, pp. 143 –148.

[9] N. Indukuri, "Layer 2 security for smart grid networks," in *Advanced Networks and Telecommuncations Systems (ANTS), 2012 IEEE International Conference on*, Dec 2012, pp. 99–104.

[10] A. Metke and R. Ekl, "Security technology for smart grid networks," *Smart Grid, IEEE Transactions on*, vol. 1, no. 1, pp. 99–107, June 2010.

[11] Z. Lu, X. Lu, W. Wang, and C. Wang, "Review and evaluation of security threats on the communication networks in the smart grid," in *MILITARY COMMUNICATIONS CONFERENCE, 2010 - MILCOM 2010*, Oct 2010, pp. 1830–1835.

[12] K. Ahmed, Z. Aung, and D. Svetinovic, "Smart grid wireless network security requirements analysis," in *Green Computing and Communications (GreenCom), 2013 IEEE and Internet of Things (iThings/CPSCom), IEEE International Conference on and IEEE Cyber, Physical and Social Computing*, Aug 2013, pp. 871–878.

[13] A. Hamlyn, H. Cheung, T. Mander, L. Wang, C. Yang, and R. Cheung, "Computer network security management and authentication of smart grids operations," in *Power and Energy Society General Meeting - Conversion and Delivery of Electrical Energy in the 21st Century, 2008 IEEE*, July 2008, pp. 1–7.

[14] C.-M. Yu, C.-Y. Chen, S.-Y. Kuo, and H.-C. Chao, "Privacy-preserving power request in smart grid networks," *Systems Journal, IEEE*, vol. PP, no. 99, pp. 1–9, 2013.

[15] M. Kim, "Protecting delay-sensitive traffic transmission against flooding attacks in smart grid networks," in *ICT Convergence (ICTC), 2013 International Conference on*, Oct 2013, pp. 1121–1124.

[16] R. Kompella, J. Yates, A. Greenberg, and A. Snoeren, "Fault localization via risk modeling," *Dependable and Secure Computing, IEEE Transactions on*, vol. 7, no. 4, pp. 396 –409, oct.-dec. 2010.

[17] S. Kondakci, "Intelligent network security assessment with modeling and analysis of attack patterns," *Security and Communication Networks*, pp. 1–17, 2012.

[18] P. Tague, D. Slater, J. Rogers, and R. Poovendran, "Evaluating the vulnerability of network traffic using joint security and routing analysis," *Dependable and Secure Computing, IEEE Transactions on*, vol. 6, no. 2, pp. 111 –123, april-june 2009.

[19] K. Padayachee, "Taxonomy of compliant information security behavior," *Computers & Security*, vol. 31, no. 5, pp. 673 – 680, 2012.

[20] H. Venter and J. Eloff, "A taxonomy for information security technologies," *Computers & Security*, vol. 22, no. 4, pp. 299 – 307, 2003.

[21] M. Kjaerland, "A taxonomy and comparison of computer security incidents from the commercial and government sectors," *Computers & Security*, vol. 25, no. 7, pp. 522 – 538, 2006.

[22] E. Koutrouli and A. Tsalgatidou, "Taxonomy of attacks and defense mechanisms in p2p reputation systemslessons for reputation system designers," *Computer Science Review*, vol. 6, no. 23, pp. 47 – 70, 2012.

[23] H. Debar, M. Dacier, and A. Wespi, "Towards a taxonomy of intrusion-detection systems," *Computer Networks*, vol. 31, no. 8, pp. 805 – 822, 1999.

ISBN: 978-1-61804-290-3

[24] S. Marti and H. Garcia-Molina, "Taxonomy of trust: Categorizing p2p reputation systems," *Computer Networks*, vol. 50, no. 4, pp. 472 – 484, 2006.
[25] J. M. Estevez-Tapiador, P. Garcia-Teodoro, and J. E. Diaz-Verdejo, "Anomaly detection methods in wired networks: a survey and taxonomy," *Computer Communications*, vol. 27, no. 16, pp. 1569 – 1584, 2004.
[26] J. McHugh, "Testing intrusion detection systems: a critique of the 1998 and 1999 darpa intrusion detection system evaluations as performed by lincoln laboratory," *ACM Trans. Inf. Syst. Secur.*, vol. 3, no. 4, pp. 262–294, Nov. 2000.
[27] S. Kondakci, "A concise cost analysis of Internet malware," *Computers & Security*, vol. 28, no. 7, pp. 648–659, 2009.
[28] S. Kondakci and C. Dincer, "Internet epidemiology: healthy, susceptible, infected, quarantined, and recovered," *Security and Communication Networks*, vol. 4, no. 2, pp. 216–238, 2011.
[29] S. Kondakci, "Analysis of information security reliability: A tutorial," *Reliability Engineering & System Safety*, vol. 133, no. 0, pp. 275 – 299, 2015. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0951832014002373
[30] F. Flammini and F. Flammini, *Critical Infrastructure Security: Assessment, Prevention, Detection, Response*. WIT Press / Computational Mechanics, 2012.
[31] J. Wiles, *Techno Security's Guide to Securing SCADA: A Comprehensive Handbook On Protecting The Critical Infrastructure*. Syngress Publishing, 2008.
[32] C. Alcaraz, G. Fernandez, and F. Carvajal, "Security aspects of scada and dcs environments," in *Critical Infrastructure Protection*, ser. Lecture Notes in Computer Science, J. Lopez, R. Setola, and S. Wolthusen, Eds. Springer Berlin Heidelberg, 2012, vol. 7130, pp. 120–149. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-28920-0_7
[33] R. Derynck, "Scada system security threats, vulnerabilities and solutions," in *Developments in Control in the Water Industry, 2004. The IEE Seminar on (Ref. No. 2004/10729)*, May 2004, pp. 0_49–19/8.
[34] S. Kondakci, "Epidemic state analysis of computers under malware attacks," *Simulation Modelling Practice and Theory*, vol. 16, no. 5, pp. 571 – 584, 2008.
[35] S. Kondakci, "Network security risk assessment using bayesian belief networks," in *Social Computing (SocialCom), 2010 IEEE Second International Conference on*, aug. 2010, pp. 952 –960.
[36] G. F. Cooper, "Probabilistic inference using belief networks is np–hard," in *Knowledge Systems Laboratory*, 1987, pp. 87–27.
[37] S. L. Lauritzen and D. J. Spiegelhalter, *Local computations with probabilities on graphical structures and their application to expert systems*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1990.
[38] P. Dawid, "Application of a general propogation algorithm for probabilistic expert systems," *Statistics and Computing*, vol. 2, no. 1, pp. 25–36, 1992.

ISBN: 978-1-61804-290-3

IP Impairment Testing for LTE Networks

Andrei Rusan, Radu Vasiu

Abstract — IP based networks are on the rise, with a majority of current networks being IP based. This brings an increased interest in the behavior of these networks in various conditions and environments, to be able to understand and even predict their performance in various deployment scenarios. The latest generation of commercially deployed telecommunications networks are full IP, with 4G LTE networks starting commercial operation in most countries and in almost every possible environment. It is thus important to understand IP impairments, their effect on the network and how the network and the network elements react to them. In order to do so, we need to emulate/generate these IP impairments (or the conditions that lead to them) in such a way that allows us to assess their impact on the performance of networks. In this paper we propose a practical and economical solution to achieve this, one that is flexible and easy to use in all types of scenarios, from laboratory networks to live pilot and demo networks – with a focal point on 4G LTE networks and the flexibility needed for testing in real life scenarios.

Keywords— IP impairment testing, LTE IP impairments, 4G LTE network performance.

I. Introduction

NETWORKS built on IP represent a majority of the currently deployed networks, with most new networks being designed and built as full IP networks. Older networks that are still being kept in service are now also being upgraded or reengineered and redesigned for IP. Regardless of whether networks are starting as IP full IP networks or are partially or step-wise migrating to IP, this brings IP networks to almost any application and any industry one can think of. These networks need to operate in any and all conceivable environments, being built upon all possible combinations and permutations of equipments, physical links and interfaces.

Ideally, the IP packets would be transported identically over any such network or network segment. But real networks are far from homogenous, with different specifications for each network element and transport link – and thus different behavior in field deployments. Links and network elements behavior varies also based on capacity and loading/utilization – not to mention technology.

It is also common nowadays to encounter real life scenarios with multiple layers of virtual links over one physical link.Virtual networks, circuits and tunnels are also common in today's architectures and usage scenarios: networks inside networks, on multiple levels, forming a hierarchy of almost unlimited depth. In such scenarios the outer (encapsulating) layers/networks obviously pass their "behavior" down to the layers below, which also add their own specific behavior and response to variable network conditions.

It is clear that the task of understanding the behavior and performance of such networks becomes a very complex one, given the virtually unlimited possible combinations of network elements, links and transport mediums that IP packets need to pass. It is thus vital that engineers and researchers can observe and understand the behavior of such networks, to be able to properly design, engineer and optimize them for the intended applications.

Given the high number of variables in real networks, with many different network elements, active and passive equipments and various topologies involved, it is usually not practical to build real-life field prototypes for each specific deployment. One alternative is to work with simplified laboratory or demo networks and study specific scenarios that are characteristic for a given deployment. One can thus emulate the behavior of network elements and links inside the network or, as needed, the ones interfacing with the network at its border. This way once can effectively build and recreate realistic environments, but with the added advantage of easy access and a controlled environment if needed, and also with a fraction of the effort and cost compared to the final/commercial solution. One can then test and measure the performance of such a simulated network, in one or more points, while running or emulating also the target application(s) or service(s) that will be deployed over this network.

Such scenarios will allow more than simple performance testing of typical applications and services over such networks. They will also allow the simulation of worst case and extreme scenarios, where one can test, measure and even optimize performance in critical conditions. This can provide a better understanding of the limits of a particular technology or network solution. Engineers and researchers can get insights on behavior and performance in such scenarios, enabling further research and optimization work to improve these technologies and solutions.

One obvious advantage of such simulation and testing work is that it can accelerate the time to market in most scenarios and dramatically lower initial costs and effort. It is also highly useful and important for developing expertise and growing experience with new and complex technologies, where these are still limited and actively being developed. It can provide engineers and others involved (including management and marketing) with early hands-on experience and understanding of the technology: on-the-job-training and close to real-life demos, even before these new and complex networks are fully

Andrei Rusan is a PhD student at the Politehnica University of Timisoara, Romania. (phone: +40.749.402503 / +1.347.502.8910; e-mail: arusan@gmx.com).

Radu Vasiu is professor, with the Politehnica University of Timisoara, Department of Communications. (phone: +40.722.516555; e-mail: radu.vasiu@cm.upt.ro).

ISBN: 978-1-61804-290-3

deployed and ready for use – a crucial aspect, as implementing and getting such a network ready for commercial service can take months and years.

In this paper we target exactly such scenarios and use-cases, focusing on the telecommunications industry which sees increased complexity and dynamics for networks. More specifically, this paper discusses IP impairment testing for 4th Generation (4G) Long Term Evolution (LTE) networks – the latest generally available wireless communications technology, with commercial deployments ongoing or about to start in most countries and a rapidly growing number of users worldwide.

II. IP Network Impairments

Real-life commercial networks do not behave in a deterministic way. Impairments in these networks affect traffic packets that are traversing it – and in many instances this is a variable behavior, as these impairments are caused by cumulative random events: busy hours for network activity, high profile events (either attracting a lot of interest or generating peaks due to "panic usage"), localized service disruptions, dynamic routes and paths for packets, faulty or overloaded network elements, etc.

It is out of the scope of this paper to detail each type of impairment, but we'll briefly discuss the top three that have the most impact on quality-sensitive services [1]:

A. Packet Delay or Latency

Packet delay, also known as "latency", is a measurement of how much time it takes for a data packet to get from one point to another. Although a measurement of zero delay is not seen in production networks because it takes some amount of time for the packet to travel from its source to its destination, a low packet delay number is desired for optimum network and application performance.

Propagation, router/switch processing, and storage delays are normal contributors to the expected delay. However, adverse network conditions like queuing delays on the intermediate network elements also contribute to packet delays.

B. Packet Loss

Packet loss is a measure of the number of packets sent over a network that fail to reach their destination. This can result in noticeable performance issues. In addition, a decrease in the throughput is caused by some transport protocols such as TCP, which have a mechanism to ensure reliable delivery of packets, requiring the retransmission of missing packets.

Causes of packet loss may include multi-path fading, channel congestion, in transit rejection of corrupted packets, faulty hardware, drivers, or routing routines.

C. Jitter or Delay Variation

Delay variation, usually referred to as jitter, is the measure of variability of delay values over a period of time. The delay experienced by traffic is not a static value but usually varies due to random events such as fluctuating loads on the network infrastructure. For example, in the morning when a lot of people log on to the network and start accessing network resources versus the evening hours when people retire to bed and network usage goes down. This dynamic nature of network traffic affects the delay that packets experience as they traverse the network.

Packets suffering from delay variance will end up arriving at the destination out of sequence or may even be dropped by the receiving devices. This has a negative impact on the performance of various voice and video applications.

III. Effects of IP Impairments

Even if packets are all being transported over the same network, they carry data used in different applications and for different services. These applications and services are affected in different ways by such impairments on the transport network.

Some are more sensitive than others when it comes to specific IP impairments. This is because of:

- the nature of the services/application,
- higher layer protocols used,
- error correction or recovery algorithms and mechanisms in higher layers.

Packet loss, for example, can be a critical problem for some services, making incomplete final data totally useless. But for the same service things can look different if there are retransmissions and the final data is complete. Retransmissions, however, mean a higher latency (more time needed, to not only transmit but also retransmit data) and might make certain services useless if data arrives too late, whereas for some services it is not that important.

Considering the above examples, the possible effects of the same impairment and the way it branches out as a function of several variables, it is clear that a pure theoretical analysis or prediction of the effects of IP impairments would be very complex and of limited accuracy in real-life scenarios with a much higher number of variables.

Trying to also understand not only how, but also how much an IP impairment (or combination of such impairments) affects a particular service/application, makes this task an almost impossible one.

This is why emulating IP impairments in a real network is a much better solution when trying to observe, measure and understand specific scenarios:

- it is useful to get a feeling about how things work in real life, since even implementations are not perfect (codecs, servers, clients etc.);
- to be able to verify at least some degree of high level reasoning and predictions about functionality;
- to get a feeling for particular and exotic use cases where there is not much prior experience (e.g. connecting a base station or eNodeB on a remote island over a satellite link used for backhaul).

ISBN: 978-1-61804-290-3

IV. Current approaches for IP impairment testing on 4G LTE

IP impairments are usually injected by using network emulators. This is not a new idea – in 1995 a WAN emulator was already used to evaluate TCP Vegas [2].

Based on criteria of interest (a simple, lightweight and cheap way of simulating network impairments), we can classify such emulators into two categories:

- Appliances - integrated, specialized devices and solutions from network test devices manufacturers: e.g. Ixia Anue [3][1], Rugged Tooling Rude [4], etc.
- Software emulators:
  - open source solutions: Dummynet [5][6][7], Netem [8], NISTNet [9]
  - proprietary solutions by network test solutions providers: Spirent INE (and the older IPWave) [10], GL Communications Maps / PacketExpert / IPLinkSim [11], PDS/ZTI NetDisturb [12], etc.

Appliances usually offer the highest accuracy and bandwidth. For specific use cases there can be downsides, though:

- cost: appliances and highly specialized devices usually have a very high cost, even when renting them;
- logistics of getting the device delivered and installed in the testing location; usually a data-center or lab-like location with rack mounting possibilities, cooling and power supply are required;
- range limitations: while very accurate and usually also offering very high density/link speeds (e.g. up to 40GBps for the Ixia tool), some of these devices can be limited for extremely high values of IP impairments (e.g. delay over a few seconds), especially in the lower end versions; while such high values are not commonly emulated, this can be a downside when performing proof-of-concept investigations or just testing the limits.

The open source software emulators given as examples are solid and proven open source solutions for network emulation, with state of the art implementations of scheduling and queuing algorithms (e.g. QFQ [13], ABE [14], WF^2Q [15][16], Round Robin variants like DRR [17], etc.).

Dummynet has been around for more than 15 years, as a standard component of FreeBSD first, which is now also integrated with the FreeBSD firewall IPFW. It is also standard in MAC OS X (since 2006) and available for Windows and Linux, available on many Linux and BSD distributions.

Netem and NISTNet are similar and even share some common code. They are both used mainly under Linux, with netem being distributed with Linux and NISTNet being distributed separately. Netem is also tightly integrated with Linux Traffic Control subsystem (TC).

It is important to note that those three emulators have also been used as building blocks for large-scale emulation platforms like Emulab [18], which uses Dummynet on its FreeBSD nodes and Linux/TC on its Linux nodes. V-em [19] uses NISTNet. There are also other such large scale network emulation platforms, like MicroGrid [20], EMPOWER [21], IMUNES [22], etc. (some based on Dummynet) – but these are outside of the scope of this paper because of their high complexity which usually limits them to the laboratory which developed them.

Proprietary software emulator solutions, like Spirent INE or the GL Communications tools are also interesting solutions, but were not analyzed in detail by the authors, who considered them to be out of the scope of this paper. The main reason for this is the fact that they don't offer the benefits of standalone appliances, as they are not hardware solutions – and also lack the benefits we were looking for in the open source solutions (low cost, detailed understanding of the inner workings, possibility to modify or write our own user interface).

V. Considering an alternative

When looking for a suitable solution for testing IP impairments over 4G LTE networks, the choice comes down to hardware based network emulators versus software based network emulators.

Our target was to find the solution that comes with minimal complexity and cost and can provide the necessary functionality. Another important aspect was the possibility to find a portable solution or one that is easily moveable between test-beds and locations worldwide. This is to encourage IP impairment testing rather than make it a high complexity task to be avoided.

From a cost perspective, it is obvious that integrated appliances and even proprietary software network emulators cannot compete with the open source solutions. This holds true even after a deeper analysis, considering the effort of setting up and running/maintaining the solution.

A few other questions to look at:

- Do open source solutions provide the needed functionality?
- Is their accuracy high enough?
- Can they be used to build portable solutions?
- Can we make it easy to use?

We compared Dummynet, Netem/TC and NISTNet to each other and found that each of them could provide the functionality that we were looking for: simulating packet loss, latency, jitter (and eventual reordering because of jitter). All three can also provide bandwidth (BW) limitation, but that is outside of our interest as in current testing we considered that LTE links are properly engineered in terms of capacity and BW limitation should not be seen in a 4G LTE network.

Given the type of testing we are targeting at this point and the fact that we are looking to generate network impairments rather than model very strict behavior of network elements or complex models, accuracy is also satisfactory with all solutions when optimizing them (e.g. using a 10kHz system clock for Dummynet or high resolution timers for TC/Netem

[23]).

We found that these implementations are similar to some degree, with differences in details. [24] is a good comparative study and investigation into the differences and similarities among the emulators, though we must add to [24] that Dummynet can be used for packet reordering/jitter emulation by using queues with different probabilities and delays.

All 3 emulators are also suitable for building a low profile solution, based on either a desktop machine or even smaller factor PCs that can accept at least 1 (preferably 2) additional Network Interface Cards (NICs) on PCI-eXpress (PCI-X) slots.

In terms of easy to use solutions, we decided that NISTNet is not optimal since it is not actively maintained anymore and porting it to latest Linux versions is not trivial [25]. Since it wasn't clear for us if new versions will be maintained/developed (e.g. NIST Net Next Geneartion [26]), we only shortlisted Dummynet and Netem/TC.

Table 1 offers a synthetic view on our comparison.

Table 1 - Comparison of network emulators

	Dummynet	**Netem / TC**	**NISTNet**
Availability	Included in FreeBSD; available for Windows, Linux, other BSD distros	Included in Linux	Not actively maintained; available for older Linux versions
Accuracy / time resolution	System clock, up to 10kHz on FreeBSD	System clock (max 1 KHz) or high resolution timers	Real time clock, 8192Hz
Latency	Yes	Yes	Yes
Packet Drop	Yes	Yes	Yes
Packet Reordering	Yes, using variable priority and delay queues	Yes	Yes
Packet duplication	No	Yes	Yes
Packet corruption	No	Yes	Yes
Suitable for portable solution	Yes	Yes	Yes
Interception point for packets	Input and output	Output (traffic shaper)	Input

Our final choice was Dummynet on FreeBSD, with the option and intent of also exploring Netem/TC at some point, mainly because of their packet duplication and packet corruption functionalities. Since these functionalities were not our immediate focus, we proceeded with FreeBSD and Dummynet. Factors for our decision:

- extensive past experience with FreeBSD
- Dummynet's capabilities to intercept packets both on the input and output
- FreeBSD's high time resolution with a 10kHz system clock
- Mature, solid and proven solution.

VI. Implementation of the IP Impairment Tool

After finding a suitable network emulator, we had to decide on the details of implementing the IP impairment tool in accordance with our requirement that it be easy to use – both from a user interface and logistics point of view.

The recommendations outline in this section are our original contribution in order to satisfy our requirements:

- simple usage of such an IP impairment solution
- lightweight and simple build: hardware, software, configuration and maintenance
- ideally get to a solution the is cheaper than the ones available already
- easy/quick to move among testing locations.

A. User interface

We decided to write our own user interface, that will allow us to manage and control the network emulation functions of the IP impairment tool interactively. Our requirements stated that is should be possible to:

- change parameters of network impairment functionality on the fly → we wanted to be able to conduct sequential testing with varying parameters, e.g. measure throughput for different values of packet loss over a specific link.
- operate the IP impairment tool remotely, so that testing can be carried out from a different location than the physical location of the machine → this allows us to sit outside of the data center, where the physical machine is connected to the network, while controlling the IP impairment emulation remotely via terminal session over a dedicated maintenance link; the testing engineer can then be located where user side testing is being performed, running tests using the LTE UE, generated the needed traffic and measuring performance with the various IP impairments active.

The management of the IP impairment tool needs to be on a separate interface, so that the management traffic is not impaired itself, or else one could lock himself out of the tool by mistake (e.g. when testing high values of packet loss) or make testing painfully slow when testing high latency.

It is also important to note that even with remote management, the machine running the IP impairment tool is physically connected to the network. Switching links used for IP impairment testing usually also means physically changing cables or ports – it can also be done by proper remote link/port management in most routers (this is outside the scope of this

paper and often not touching the service routers is a better idea).

Given the goal to keep things simple, we chose to implement a console/text-based user interface. This brings the advantage of simple remote access via SSH for fast and secure management of the IP impairment tool and interface..

We used C to develop the user interface, along with the ncurses API – the result is a GUI-like text-based user interface that can run in a terminal emulator. We developed and compiled different small independent applications for each type of impairment testing, to keep things simple and clear.

We mentioned that the goal was to perform sequential testing, covering values for latency and packet loss in a specific range. We thus implemented our tool in such a way that it would start with a given lower bound value for delay or packet loss, then increase this value by a given step value until an upper bound value is reached, returning the system to its initial state in the end (stop any impairment) before finishing on the test sequence. The start, end and step values are all freely configurable for each run.

Figure 1 shows a screenshot of the interface during delay testing on the LTE S1-U interface (the LTE interface between the eNodeB and the Serving Gateway - SGW), with the progress bar at the bottom indicating progress. Figure 2 shows the end of the same test sequence.

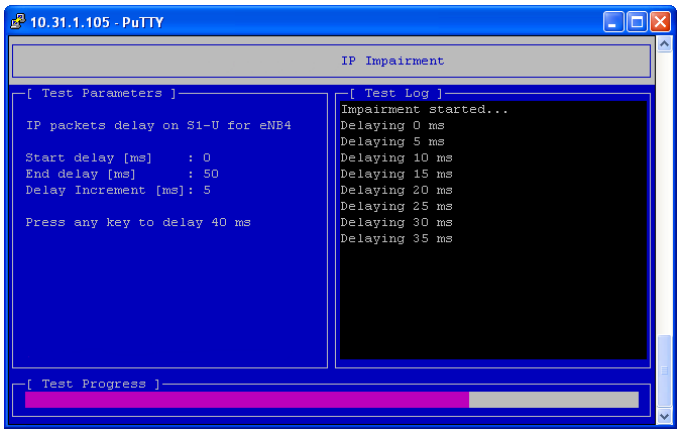


Figure 1 - Delay test on S1-U in progress

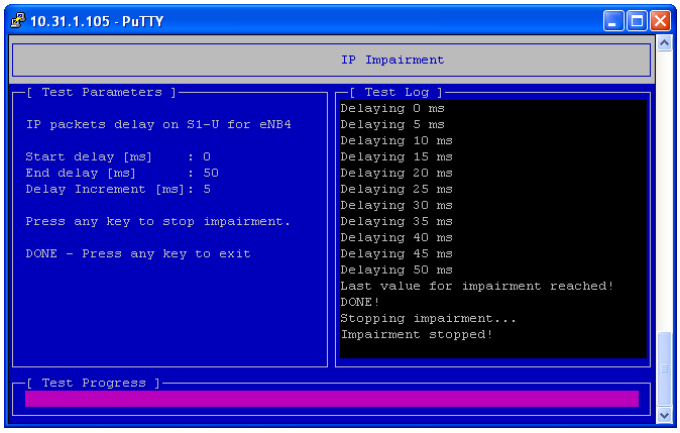


Figure 2 - Delay test on S1-U finalized

B. Hardware

The hardware choice had to satisfy the requirement of mobility or facilitate usage of this tool in various locations with little effort and trouble.

While a laptop machine cannot satisfy the connectivity requirements of such a tool (at least 3 ethernet ports, 2 of them on enterprise-class NIC(s)), we considered desktop/business PCs.

Once again the remote connection requirement is an advantage, as it means that the impairment tool can be a headless hardware implementation (no physical display or input devices are needed locally). This does away with the need for a monitor, keyboard and mouse: those can be used if available, but are not a requirement.

For the machine itself we identified two possible solutions:

- Building the system / tool on a low footprint PC. There are small form factor devices available, that provide an integrated NIC and have at least one or preferably two available PCI-X slots for additional NIC(s). Some of these come with decent hardware configurations and are really small enough to be considered mobile.
- Building the system / tool on a generic PC with generic hardware support built into the OS. Minimizing hardware dependence would allow us to use any available machine with similar (standard) hardware configuration, by just swapping the hard drive (or copying an image) and moving the enterprise-class NIC(s).

While the first option sounds and looks nicer, we decided to go with the second one. This lead to our decision:

- Hardware reliability: a lot of hardware in a very tight enclosure does not go well with constant high usage/load on the same hardware during testing, so failures could happen rather sooner than later
- Easy hardware replacement in case of a failure: one might not easily find the same model of a small form factor PC if the original one fails
- Supply chain considerations: it was easier to get a desktop PC obeying procedures.

Our final choice is an Intel-based desktop PC with integrated gigabit ethernet and graphics. It has 8GB of RAM and a 80GB HDD. We had to use a mechanical HDD but would recommend a SSD, especially if it will be taken out for transport.

We also added 2 server-grade NICs – the essential hardware to work with the network emulation functionality. Server/enterprise-class NICs are selected based on their performance and reliability: adapters in the Intel I350 family [27] or the Intel ET server adapter family [28] are supported under FreeBSD and worked well in our setup. These also have the added benefit of being dual or quad port, so if needed one single add-on NIC and available PCI-X slot are sufficient. We had 2 cards in our setup and used one port on each. These NICs are available with copper or optic fiber connectors, depending on what is needed. In our setup the Intel ET dual port NIC(s) were detected by the em-driver and performed flawlessly for all of our testing.

ISBN: 978-1-61804-290-3

C. OS configuration

We used FreeBSD 8 and fine-tuned the kernel to our liking, by customizing it and recompiling it. As part of this process we set the timer granularity option to 10kHz, along with enabling ipfirewall and dummynet operation. The 3 kernel options to do this are:

```
options IPFIREWALL
options DUMMYNET
options HZ=10000
```

We also streamlined the kernel, exclude unnecessary drivers and options. We included the em driver for the Intel NICs and also support for most common NICs – making sure the OS can detect most on-board adapters, as these will be used for the remote access and control link to the IP impairment tool.

We also left in support for most generic devices, allowing this streamlined FreeBSD kernel to boot on most if not all standard Intel desktop machines with minimal hardware support needed for the IP impairment functionality. This allows to easily move the impairment tool between locations, even independently of the original desktop PC box.

Take only the HDD (or a HDD image) and the NIC(s), and with this small start kit a new impairment tool can be ready in a short time wherever a decent desktop PC is available. This enables quick and easy testing in any location, independent of most logistics that involve hardware shipments, customs etc.

Of course one can always move the desktop PC altogether and avoid any work on hardware.

D. Connecting the IP impairment tool in the LTE core

Dummynet acts as an Ethernet bridge for network impairment emulation, moving packets in between the two ethernet interfaces on the IP impairment machine. This makes it invisible on an IP level: the IP impairment tool is transparent to the network from an IP point of view.

When using the IP impairment tool on a given LTE interface for testing, it will be physically inserted on the physical link used for that specific LTE network interface.

Figure 3 illustrates the placement of the IP impairment tool in the case of testing on the S1-U LTE interface.

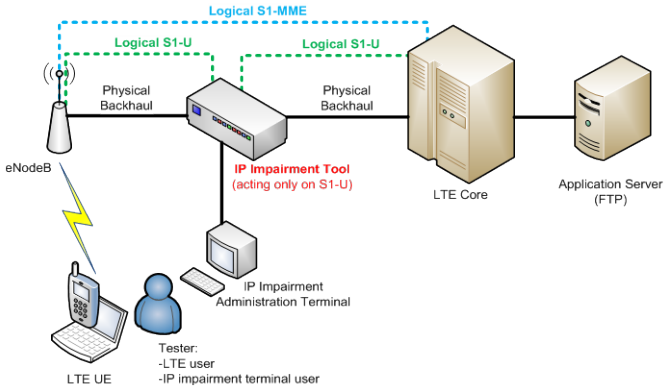


Figure 3 - Connecting the IP impairment tool for testing on the S1-U interface

If testing switches to other LTE links, one simply switches and inserts the IP impairment tool on the target physical link.

It is important to note that packets are matched against a set of rules before they are manipulated and any impairment is applied. These rules are defined so as to filter only the packets for the specific interface that we currently want to test on. We usually filter for IPs/IP ranges, matching endpoint(s) IP addresses for the target link. This way we filter traffic for a specific logical LTE interface on a physical link that might also carry other traffic. IP impairments will then only be applied to this traffic, not touching any other traffic that we do not want to impair – the rest of the untouched packets are simply forwarded through the Ethernet bridge in a transparent manner.

VII. Conclusion

Building on existing blocks, we present a lightweight, cheap, easy to implement and use solution for IP impairment testing on LTE networks.

The actual software platform and functionality is free and readily available. We also presented the key differences between a couple of available network emulator options. The additional needed configuration work is kept at a minimum and should be a one-time task – when following the recommendations that we propose for the implementation of the tool: the same software (OS, interface) can be reused, as needed, across a wide range of hardware in different locations and for different tasks, minimizing logistics and installation effort.

It is also important to note that the proposed solution allows for a single person to perform all testing with minimal effort, as long as he can connect to the LTE network with the LTE UE and also has remote access to the IP impairment tool's administration interface, e.g. over SSH by using the built-in ethernet adapter of his PC/laptop.

When validating our solution, we use one single laptop for all tasks related to testing:

- connecting to the LTE network by using a USB dongle as the LTE UE – and generating traffic by running multiple FTP data transfers in parallel;
- measure end-user throughput on the device connected over LTE;
- control the IP impairment tool (and thus the impairments' parameters), by using the laptop's built-in ethernet adapter to connect via SSH to the IP impairment tool and run its management interface.

Using this setup, one single trained engineer can perform a wide range of testing on different scenarios, in different locations, for laboratory work or even demonstrational purposes. He can quickly move the setup without even moving hardware and minimizing installation effort, and can do so by using cheap and readily available hardware (with the exception of the NICs, that should satisfy a certain higher standard).

This solution is not limited to LTE and can be applied for many other testing scenarios:

- with minimal changes to the software user interface that we developed, to ease user interaction with the tool;
- with no changes to the setup / OS platform.

We have used the setup to test data transfer performance over FTP for a single LTE user. Maximum data rates were in the range of 100Mbps over 4G LTE, an this was far from pushing a decent hardware setup for the IP impairment tool to its limits.

Next steps of interest for our research and that are enabled by this setup include:

- testing of more LTE network interfaces - including S1-MME, S5, S11 – to get a better view on LTE performance in case of IP impairments;
- looking into using TC/Netem, to compare performance over identical test scenarios but also to enable testing of additional impairments;
- simulation of specific real-life scenarios by using IP impairments, in order to understand and measure how LTE performs under specific conditions and in specific deployment scenarios.

REFERENCES

[1] "Predicting and Managing Network Impairments", IXIA Whitepaper, 2012, rev.A, available: http://www.ixiacom.com/sites/default/files/resources/whitepaper/impairment.pdf

[2] J. S. Ahn, P. B. Danzig, Z. Liu, L. Yan, „Evaluation of tcp vegas: emulation and experiment", SIGCOMM Comput. Commun. Rev., 25(4), 1995

[3] Ixia Anue Network Emulators, available: http://www.ixiacom.com/products/ixia-network-emulators

[4] Rugged Tooling Rude, available: http://www.ruggedtooling.com/rude.php

[5] L. Rizzo, "Dummynet: a simple approach to the evaluation of network protocols", ACM Computer Communication Review, 27(1), pp. 31-41, 1997

[6] M. Carbone, L. Rizzo, "Dummynet revisited", ACM SIGCOMM Computer Communication Review, Volume 40 Issue 2, April 2010

[7] The Dummynet Project, available: http://info.iet.unipi.it/~luigi/dummynet/#4f01

[8] S. Hemminger, "Network emulation with NetEm", linux.conf.au 2005

[9] M. Carson, D. Santay, "NIST Net: a Linux-based network emulation tool", SIGCOMM Comput. Commun. Rev., 33(3), pp. 111-126, 2003

[10] Spirent INE, available: http://www.spirent.com/Products/Spirent_INE

[11] GL Communications Network Impairment Simulator, available : http://www.gl.com/telecom-test-solutions/network-impairments-simulation.html

[12] Packet Data Systems/ZTI Telecom NetDisturb, available : http://www.pds-test.co.uk/products/netdisturb.html

[13] F. Checconi, L. Rizzo, P. Valente, "QFQ: Efficient Packet Scheduling with Tight Guarantees", IEEE/ACM Transactions on Networking, Volume 21, Issue 3, October 2012, pp. 802-816

[14] P. Hurley, J. Le Boudec, P. Thiran, M. Kara, "ABE: Providing a low-delay service within best effort.", IEEE Network, Vol. 15, Issue 3, pp. 60–69, 2001

[15] J. C. R. Bennett, H. Zhang, „WF^2Q: Worst-case fair weighted fair queueing", Proceedings of IEEE INFOCOM '96, March 1996, pp. 120–128

[16] D. Stiliadis, A. Varma, "A general methodology for designing efficient traffic scheduling and shaping algorithms", Proceedings of IEEE INFOCOM '97, April 1997, pp. 326–335

[17] M. Shreedhar, G. Varghese, "Efficient fair queuing using deficit round robin", IEEE/ACM Transactions on Networking, Vol. 4, Issue 3, 1996, pp. 375–385

[18] B. White, J. Lepreau, L. Stoller, R. Ricci, S. Guruprasad, M. Newbold, M. Hibler, C. Barb, A. Joglekar, "An integrated experimental environment for distributed systems and networks", SIGOPS Oper. Syst. Rev., 36(SI), 2002

[19] G. Apostolopoulos, C. Hassapis. "V-em: A cluster of virtual machines for robust, detailed, and highperformance network emulation", MASCOTS '06, 2006

[20] H. J. Song, X. Liu, D. Jakobsen, R. Bhagwan, X. Zhang, K. Taura, A. Chien, "The MicroGrid: a scientific tool for modeling computational grids", Supercomputing '00: Proceedings of the 2000 ACM/IEEE conference on Supercomputing, 2000

[21] P. Zheng, L. M. Ni, "Empower: A cluster architecture supporting network emulation", IEEE Trans. Parallel Distrib. Syst., Vol. 15, Issue 7, 2004

[22] M. Zec, M. Mikuc, "Operating system support for integrated network emulation in IMUNES", Proceedings of the 1st Workshop on Operating System and Architectural Support for the on demand IT InfraStructure (OASIS), Boston, MA, 2004

[23] T. Gleixner, D. Niehaus, "HR timers and beyond: Transforming the linux time subsystems", Proceedings of the Ottawa Linux Symposium, 2006

[24] L. Nussbaum, O. Richard, "A Comparative Study of Network Link Emulators", 12th Communications and Networking Simulation Symposium (CNS'09), Mar 2009, San Diego, United States.

[25] NISTNet, available: http://snad.ncsl.nist.gov/nistnet/

[26] NIST Net Next Generation, available: http://sourceforge.net/projects/nistnet/

[27] Intel Ethernet Server Adapter I350 Product Family, available: http://www.intel.com/content/www/us/en/network-adapters/gigabit-network-adapters/ethernet-server-adapter-i350.html

[28] Intel Gigabit ET Dual Port Server Adapter Product Family, available : http://www.intel.com/content/www/us/en/network-adapters/gigabit-network-adapters/ethernet-ef-et.html

Stabilizing lead lag controllers for time delay systems

N. Ben Hassen, K. Saadaoui and M. Benrejeb

***Abstract*—** In this paper, to solve the problem of stabilizing an all poles linear time delay system using lead lag controllers. A new method to design the second order controller is proposed. The complete set of stabilizing parameters is determined by the D-decomposition method. An illustrative example is given to show the effectiveness of the proposed procedure.

***Keywords*—** Second order controller; time delay; D-decomposition; stability; stabilization.

I. Introduction

Recently, the problem of determining all stabilizing fixed order, fixed structure, low order controllers for linear time invariant systems was addressed by several authors, see [1] and [2] and the references therein. The problem is worthwhile as determining this set of all stabilizing controllers is a first and an essential step in calculating optimal fixed order controllers. This line of research was later extended to include time delay systems [3]-[4]-[5]. In fact, the analysis and the study of delay systems is an active research area, that continues to grow in importance [6]-[7]. It is well known that the existence of time delays may deteriorate system's performances and can even cause instability of the closed loop system [7]. This is another reason for the extensive literature on stability and stabilization of time delay processes. These parameterization methods were successfully applied to get stabilizing classical low order controllers, such as PI controller [8], PID controller [9]-[10], and first order controllers [11]- [12].

In this paper, a proposed a method to calculate stabilizing lead lag controllers for delay systems. It consists of determining the admissible values of one of the controller's parameters. Then, this parameter is fixed within the admissible range and the D-decomposition method is used to determine the stabilizing regions in the space of the remaining two parameters. By sweeping over the first parameter the complete set of stabilizing gains can be determined.

N. Ben Hassen Laboratoire de Recherche LARA-Automatique Ecole Nationale d'Ingenieurs de Tunis, BP 37, le Belvédère 1002, Tunis, Tunisia, e-mail : Nidhal.Benhassen@enit.rnu.tn

K. Saadaoui Laboratoire de Recherche LARA-Automatique Ecole Nationale d'Ingenieurs de Tunis, BP 37, le Belvédère 1002, Tunis, Tunisia, e-mail : karim.saadaoui@isa2m.rnu.tn

M. Benrejeb Laboratoire de Recherche LARA-Automatique Ecole Nationale d'Ingenieurs de Tunis, BP 37, le Belvédère 1002, Tunis, Tunisia, e-mail : mohamed.benrejeb@enit.rnu.tn

The paper is organized as follows. In section II, the set of all stabilizing lead lag controllers for time delay systems is calculated. An illustrative example is given in section III. Finally, section IV gives some concluding remarks.

II. Stabilizing second order controllers for time delay systems

In this section, the stabilizing regions in the parameter space of a second order controller are determined. We consider the classical feedback system of Fig. 1, where the system's transfer function is given by

$$G(s) = \frac{e^{-Ls}}{Q(s)} \tag{1}$$

with $L > 0$ the time delay. Many practical systems can be represented by (1). In [6], the dynamic behavior of temperature control in a mix process is represented by (1) and in [7] it is used to model a ship positioning an underwater vehicle through a long cable, to name just few examples. Our objective is to determine the set of all second order controllers given by

$$C(s) = \frac{s^2 + \alpha_3 s + \alpha_1}{s^2 + \alpha_2 s + \alpha_1} \tag{2}$$

that stabilizes the feedback system of figure 1. In fact, the controller given by (2) is a lead-lag controller which combines the effects of phase lead and phase lag in certain frequency ranges and can realize the behavior of a PID controller [13]. Let

$$C(s) = \frac{(1+\tau_2 s)(1+\tau_3 s)}{(1+\tau_1 s)(1+\tau_4 s)} \tag{3}$$

with $\tau_1 > \tau_2 > \tau_3 > \tau_4$. In order to get the same gain for high frequencies and low frequencies, we impose $\tau_2\tau_3 = \tau_1\tau_4$ which leads to the following expression of C(s)

$$C(s) = \frac{\beta_1 s^2 + \beta_3 s + 1}{\beta_1 s^2 + \beta_2 s + 1} \tag{4}$$

or equivalently the controller given by (2). The closed loop characteristic equation is given by

$$\Delta^*(s) = (s^2 + \alpha_2 s + \alpha_1)Q(s) + (s^2 + \alpha_3 s + \alpha_1)e^{-Ls} \quad (5)$$

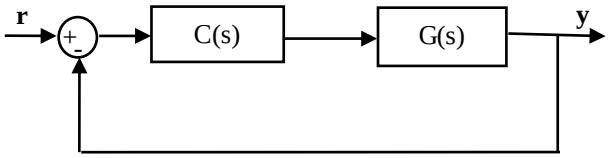


Fig. 1. Classical feedback system

Our aim in this section is to determine the set of all stabilizing regions in the parameter space of the controller. As there are only three parameters $(\alpha_1, \alpha_2, \alpha_3)$, the problem will be solved in two steps. First, we calculate the admissible ranges for one of the controller's parameters. Next, this parameter is fixed within the admissible range and the stabilizing region in the space of the remaining two parameters, if it exists, is determined. By sweeping over the admissible values of the first parameter, the complete set of stabilizing controllers can be obtained.

A. Admissible values of (α_1, α_2)

Let start by determining the admissible values in the parameter space of (α_1, α_2). In order to reduce the number of parameters $(\alpha_1, \alpha_2, \alpha_3)$ in the original stability problem from three into a simpler sub-problem with only two parameters, the following lemma will be used.

Lemma [14]. *Consider the quasi-polynomial*

$$\Delta(s) = \sum_{i=0}^{n} \sum_{l=1}^{r} h_{il} s^{n-1} e^{\tau_1 s}$$

such that $\tau_1 < \tau_1 < \cdots \tau_r$, *with main term* $h_{0r} \neq 0$, *and* $\tau_1 + \tau_r > 0$. *If* $\Delta(s)$ *is stable, then* $\Delta'(s)$ *is also a stable quasi-polynomial, where* $\Delta'(s)$ *is the derivative of* $\Delta(s)$ *compared to s.*

Now, the closed loop characteristic equation of the closed loop system of Fig. 1 is given by (5). Since the term e^{Ls} has no finite roots, the quasi-polynomial $\Delta^*(s)$ and $\Delta(s) = e^{Ls}\Delta^*(s)$ have the same roots, therefore stability of $\Delta(s)$ is equivalent to stability of $\Delta^*(s)$. In the sequel, the quasi-polynomial $\Delta(s)$ will be used to study stability of the closed-loop system of Fig 1, where $\Delta(s)$ is given by

$$\Delta(s, \alpha_1, \alpha_2, \alpha_3) = (s^2 + \alpha_2 s + \alpha_1)e^{Ls}Q(s) + (s^2 + \alpha_3 s + \alpha_1) \quad (6)$$

Using the condition of Lemma, if $\Delta(s)$ is stable then $\Delta'(s)$ is also a stable quasi-polynomial, where $\Delta'(s)$ is given by

$$\Delta'(s, \alpha_1, \alpha_2, \alpha_3) = (2s + \alpha_2)P(s) + (s^2 + \alpha_2 s + \alpha_1)P'(s) + 2s + \alpha_3 \quad (7)$$

where $P(s) = Q(s)e^{Ls}$. Repeating the same reasoning once again, If $\Delta'(s)$ is stable then $\Delta''(s)$ is also stable, where $\Delta''(s)$ is given by

$$\Delta''(s, \alpha_1, \alpha_2) = (s^2 P''(s) + 4sP'(s) + 2P(s) + 2) + \alpha_1 P''(s) + \alpha_2 (sP''(s) + 2P'(s)) \quad (8)$$

At this step, note that the number of parameters is reduced and only two parameters appear in (8). It is possible now to apply the D-decomposition method [15]- [16], and calculate the stabilizing regions in the parameter space of (α_1, α_2). To this end, we evaluate the characteristic polynomials on the imaginary axis by substituting s by jw and equating the real and imaginary parts of (8) to zero. Let

$$P(j\omega) = R(\omega) + jI(\omega)$$
$$P'(j\omega) = R'(\omega) + jI'(\omega)$$

and

$$P''(j\omega) = R''(\omega) + jI''(\omega)$$

then we get the following set of equations represented in matrix form

$$\begin{bmatrix} R'' & -\omega I'' + 2R' \\ I'' & \omega R'' + 2I' \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} = -\begin{bmatrix} -\omega^2 R'' - 4\omega I' + 2R + 2 \\ -\omega^2 I'' + 4\omega R' + 2I \end{bmatrix} \quad (9)$$

Two cases will be considered:

Case1: For $\omega = 0$, we get

$$\alpha_1 = -\frac{2P'(0)}{P''(0)}\alpha_2 - \frac{2(P(0)+1)}{P''(0)} \quad (10)$$

Case2: For $\omega > 0$, the solution of (9) is given by

$$\begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} = -\frac{1}{B_1}\begin{bmatrix} \omega R'' + 2I' & \omega I'' - 2R' \\ -I'' & R'' \end{bmatrix} \begin{bmatrix} -\omega^2 R'' - 4\omega I' + 2R + 2 \\ -\omega^2 I'' + 4\omega R' + 2I \end{bmatrix} \quad (11)$$

where

$$B_1 = \omega R''^2(\omega) + \omega I''^2(\omega) + 2(I'(\omega)R''(\omega) - I''(\omega)R'(\omega)) \quad (12)$$

The (α_1, α_2) plane can be partitioned using equations (10) and (11) into several regions and stability of (8) can be checked by

choosing a point inside a region and applying classical methods for testing stability such as Nyquist criterion or Bode method.

B. Stability regions in (α_1, α_3) plane

Once the admissible values of (α_1, α_2) are determined by the procedure described in the previous sub-section, one parameter is fixed within the admissible range and we determine the stability regions in the space of the remaining two parameters. We choose to fix α_2 and calculate stability regions in (α_1, α_3) plane. Using (6) and replacing $Q(s)e^{Ls}$ by $P(s)$ we get

$$\Delta(s, \alpha_1, \alpha_2, \alpha_3) = (s^2 + \alpha_2 s + \alpha_1)P(s) + (s^2 + \alpha_3 s + \alpha_1) \quad (13)$$

substituting s by $j\omega$ and equating the real and imaginary parts of (13) to zero, we get

Case1: For $\omega = 0$

$$\alpha_1 = 0 \quad (14)$$

Case2 : For $\omega > 0$

$$\alpha_1 = \frac{\omega^2 R(\omega) + \omega^2 + \alpha_2 \omega I(\omega)}{R(\omega) + 1} \quad (15)$$

$$\alpha_3 = \frac{(\omega^2 - \alpha_1) I(\omega) - \alpha_2 \omega R(\omega)}{\omega} \quad (16)$$

where $P(j\omega) = R(\omega) + jI(\omega)$. By the D-decomposition method, using (14), (15) and (16) for $\omega \geq 0$ the (α_1, α_3) plane can be partitioned into several regions and the stability region, if any, can be determined by employing classical methods. By sweeping over admissible values of α_2 the complete set of stabilizing lead lag controller for the linear time delay system given by (1) can be calculated.

III. ILLUSTRATIVE EXAMPLE

Consider stabilizing the third-order plant given by

$$G(s) = \frac{e^{-0.25s}}{s^3 + 2s^2 + 3s + 5}$$

by a second order controller

$$C(s) = \frac{s^2 + \alpha_3 s + \alpha_1}{s^2 + \alpha_2 s + \alpha_1}$$

as described in the previous section, we start by calculating the admissible values of (α_1, α_2). After deriving twice the characteristic polynomial of the closed loop system, the sub-problem to be solved at this step is stabilizing the quasi-polynomial given by

$$\Delta''(s, \alpha_1, \alpha_2) = (0.0625s^5 + 2.625s^4 + 24.1875s^3 + 28.8125s^2 + 23s + 10)e^{0.25s} + \alpha_1(0.0625s^3 + 1.625s^2 + 8.1875s + 5.8125)e^{0.25s} + \alpha_2(0.0625s^4 + 2.125s^3 + 15.1875s^2 + 15.3125s + 8.5)e^{0.25s} + 2$$

Using (10) and (11), the stability region is found as shown in Fig. 2.

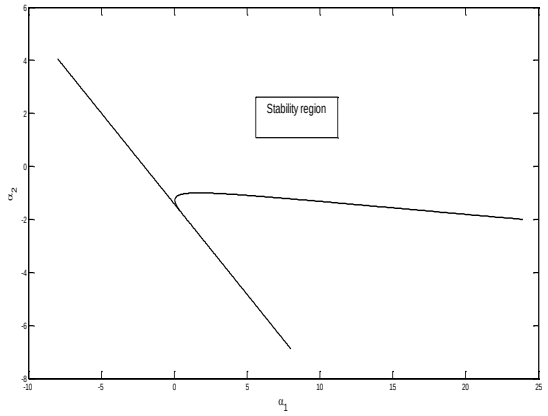


Fig. 2. Stabilizing region in the (α_1, α_2) plane

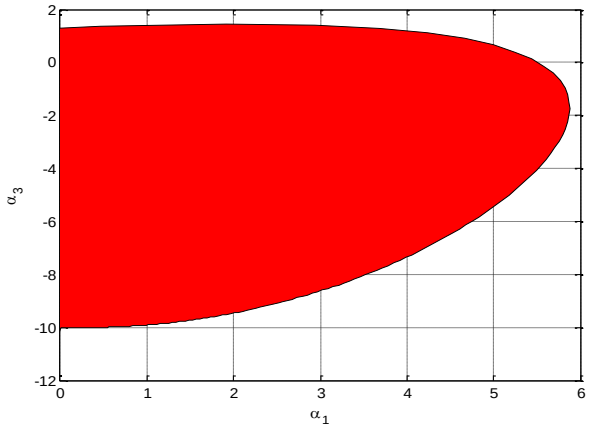


Fig. 3. Stabilizing region in the (α_1, α_3) plane for $\alpha_2 = 2$

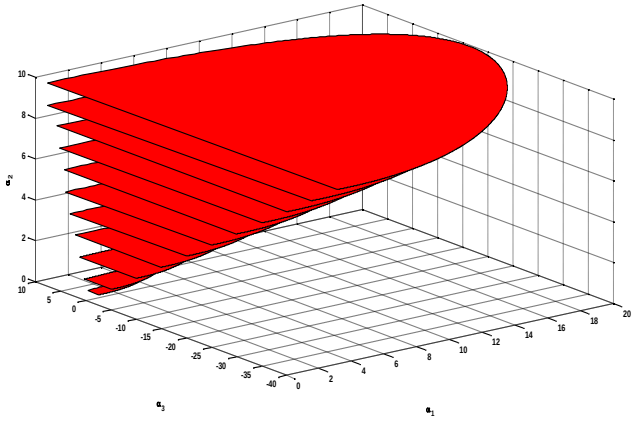


Fig. 4. Stabilizing region in the $(\alpha_1, \alpha_2, \alpha_3)$ plane for $\alpha_2 \in [2\ 10]$

Fixing a value of α_2 within the stability region, for instance $\alpha_2 = 2$, and applying (14), (15) and (16) we get the stability region in the (α_1, α_3) plane as shown in Fig. 3.

By sweeping over admissible values of α_2 the complete stabilizing regions in the parameter space $(\alpha_1, \alpha_2, \alpha_3)$ of the controller can be determined. In Fig. 4, a 3D plot of the stabilizing regions is given for values of α_2 between 2 and 10.

IV. Conclusion

In this paper, the D-decomposition method is used to compute the stability regions of a second order controller applied to an n-th order all poles linear time delay system. The second order controller used to stabilize the feedback system is a lead lag controller. The proposed method is based on determining first the range of one of the controller's parameter, α_2 in our case, and then determining stability regions in the parameter space of the remaining two parameters.

References

[1] A. Datta, M. T. Huo and S. P. Bhattacharyya, "*Structure and synthesis of PID controllers".* Springer: London 2000.

[2] K. Saadaoui and A. B. Ozguler, *"Stabilizing first order controllers with desired stability region"*, Control and Intelligent systems, vol. 49, pp. 31-38, 2009.

[3] K. Saadaoui, S. Elmadssia and M. Benrejeb, *''Stabilizing PID controllers for a class of time delay systems''*, in PID Controller design approaches- theory, tuning and application to frontier areas , ISBN 978-953-51-0405-6, edited by Marialena Vagia, Intech publishing 2012.

[4] H. Alikhani and A. Madady, *"First-Order Controllers Design for Second-Order Integrating Systems with Tme Dlay",* IEEE CCA, Hyderabad, India, August 28-30, 2013

[5] K. Saadaoui, S. Testouri, and M. Benrejeb, "Robust stabilizing first-order controllers for a class of time delay systems", ISA Transactions 49, pp. 277-282, 2010.

[6] Q. C. Zhong, "*Robust control of time delay systems".* Springer: London 2006.

[7] J. E. Normey-Rico, E. F. "Camacho, *Control of dead time processes".* Springer: London 2007.

[8] L. Jinggong, X. Yali and L. Donghai, *"Calculation of PI Controller Stable Region Based on D-Partition Method"*, International conference on control, automation and systems 2010, Korea, October 27-30, 2010.

[9] L. Ou, Y. Tang, D. Gu, and W. Zhang, *"Stability Analysis of PID Controllers for Integral Processes with Time Delay"*, American Control Conference, Portland, OR, USA, June 8-10, 2005

[10] K. Saadaoui, A. Moussa and M. Benrejeb, *"PID controller design for time delay systems using geneti algorithms"*, The Mediterranean Journal of Measurement and Control, vol. 5, pp 31-36, 2009.

[11] S. Testouri, K. Saadaoui and M. Benrejeb, *"Analytical design of first order controllers for the TCP/AQM systems with time delay"*, International Journal of Information technology, Control & Automation, vol. 2, pp 27-37, 2012.

[12] A. Madady and H. Alikhani, *"First-order controllers design employing pole placement"*, 19th Mediterranean Conference on Control and Automation, Corfu, Greece, June 20-23, 2011.

[13] P. Borne, D. F. Geneviève, J. P. Richard, F. Rotella and I. Zambettakis, "*Analyse et régulation des processus industriels. Tome 1 régulation continue*", Editions Technip: 1993.

[14] V. L. Kharitonov, S. Niculescu, J. Moreno and W. Michiels, *"Static output feedback stabilization: necessary conditions for multiple delay controllers"*, IEEE Transactions on Automatic Control, vol. 50, pp 82-86, 2003.

[15] J. Osusky and V. Vesely, *"Modification of Neimark D-partition method for desired phase margin"*, International Conference, Cybernetics and Informatics, Slovak Republic, February 10-13, 2010.

[16] E. N. Gryazina and B. T. Polyak, *"Stability regions in the parameter space d-decomposition revisited"*, Automatica, vol. 42, pp13-26, January, 2006.

Nidhal BEN HASSEN obtained her Master Degree in the Automatic and Industrial Informatics from the National Institute of Applied Science and Technology (INSAT) in 2011. She is a researcher at the research laboratory LARA Automatique of the National Engineering School of Tunis (ENIT), Tunisia. Her research interests are in the areas of time delay systems, stability robustness and applications of robust control to process control problems.

Karim SAADAOUI received his PhD at the Electrical and Electronics Engineering Department of the University of Bilkent of Ankara in 2003. He is a researcher at the research laboratory LARA Automatique of the National Engineering School of Tunis (ENIT), Tunisia. His research interests are in the areas of time delay systems, stability robustness and applications of robust control to process control problems.

Mohamed BENREJEB is a full professor with ENIT since 1985 and received the Ph.D. degree in 1976, from the University of Sciences and Technologies in Lille (France). Prof. BENREJEB also obtained a French State Doctorate in 1980 from the same university. His main scientific interests include analysis and synthesis of complex systems (stability, stabilisability, automatic control by classic and non conventional methods, using fuzzy logic, neural network, neuro-fuzzy) and optimisation with scheduling by evolutionary approaches.

Location-based Application of Secure Coding providing Local Information

Jinyoung Jung, Miyoung Bae, Yangwon Lim, Hankyu Lim(corresponding author)

***Abstract*—** Through developments in IT technology and the wide spread of high-end smartphones, everyone has been given easy access to diverse information. However, users are sometimes exposed to unnecessary information against their will. This paper made use of a location-based service to design a new method of providing local information to users so they can view information more selectively by designating the size of the region from which the information comes, as well as the amount of such information. The Secure Coding technique is very important and it should be considered in developing the applications. If a user's input is entered without processing, it might create a security problem, so coding should be allowed only with valid input data.

***Keywords*—** Location Based Service, Local Information, Smartphone App , Secure Coding

I. Introduction

As users of smart devices have become more experienced, they can now install applications they want more easily than on PCs [1]. Moreover, as smart devices contain cameras, GPS technology, and sensors, such as acceleration sensors, diverse applications that make use of these tools are simultaneously developed and distributed. Among these applications, the growth in the use of living-oriented applications that provide information about food, tourism, traffic, etc. using location-based service (LBS) applications is especially remarkable [2]. In this paper, an LBS application is introduced that provides local information, including tourism, culture, traffic, etc., in real time based on location information about the user in motion. Differing from previous applications, the proposed LBS application provides users with only information that is relevant to their current location. Moreover, users are provided with only desired information by setting the information type and size (amount) themselves.

II. Existing App using LBS

Most applications launched these days provide information in a region of a certain size. Moreover, although they provide information based on location, they usually generate information regarding a single field. Therefore, users must install a number of applications to obtain the desired information. Figure 1 (a) below shows a food information application and it shows there exists a large amount of information in an area of a certain size. Figure 1 (b) shows a bartering service application for a used goods market. It has a disadvantage in that users who want diverse information about different regions must install multiple applications at the same time. Moreover, the size of the field serviced in a metropolitan area should be different from those of small and medium-sized cities.

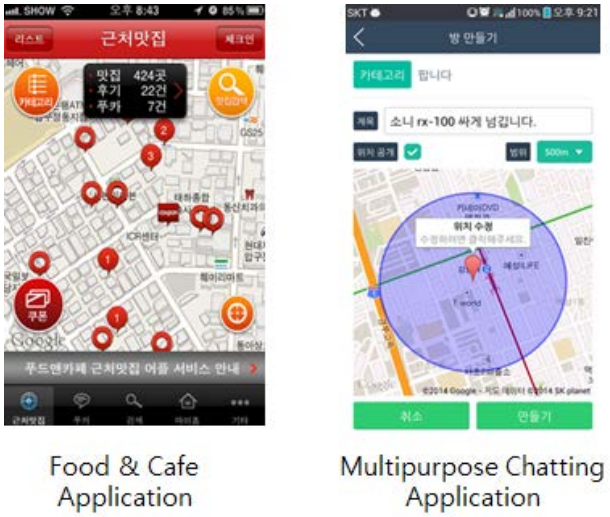


Fig. 1 Applications using LBS

A. Tour API

Tour API is an open API that provides users with tourism information and it is distributed by the Korea Tourism Organization [3]. Tour API supports nine languages, including Korean, English, Chinese, etc., and a user issued a developer key can request data from Tour API along with a certain type of URL to retrieve tourism data. There are two types of data provisions, i.e., XML and JSON, and the developer can make use of data through the parsing procedure [4]. It provides information about tourism, food, cultural facilities, accommodations, etc. and it was used when searching and analyzing information about tourism, food, and cultural facilities in this paper.

III. Design of Local Information using LBS

The LBS application suggested in this paper to provide local

Jinyoung Jung, a student of Andong National University, Korea
Miyoung Bae, a PH.D. student of Andong National University, Korea
Yangwon Lim, a lecturer of Andong National University
Hankyu Lim, received the B.S. degree in Electronics Engineering from the Kyungbook National University in 1981. He received the M.S. degree in Computer Engineering from the Yonsei University in 1984, He received the PH.D. degree in Computer Engineering from the Sung Kyun Kwan University in 1997. He is a professor of Andong National University, Korea.

information was designed with eight category menus, including local community, comprehensive local information, tourism information, local food, recruitment, amenities, cultural facilities, and weather, where the user can select the information they want to view. The menu and setting structures are defined in Figure 2 below, so the location and the amount of information can be controlled.

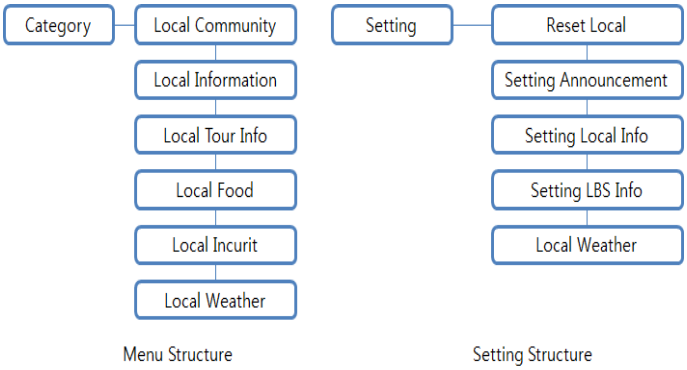


Fig. 2 Menu and Setting Structure

When the application is first launched, it provides only the previously set information about the current location by retrieving the current location's GPS coordinates. However, if the user wants to learn about information in another region beforehand, it is designed so he/she can select the region in the settings. Figure 3 and 4 below show the UI, where a user chooses local information to view information about a certain region.

Fig. 3 UI : Selecting Local Information – Select Area(1)

Fig. 4 UI : Selecting Local Information - Select Area(2)

As previous location-based applications show information within a certain distance from the user's current location, large amounts of information are generated in metropolitan areas, and it is possible the application cannot find information in small and medium-sized cities. The application suggested in this paper designates the amount of displayed information and shows only 10 pieces (default value) of information that are most relevant to the current location. Moreover, for more information, the application allows the user to set the amount of displayed information so he/she can search the desired information more easily without becoming confused. Figure 5 below shows a module structure that locates the user and provides information about the relevant location. The analysis module has a function of recalling only the required information from storage to send to users by capturing the information about the relevant location and checking the amount of generated information

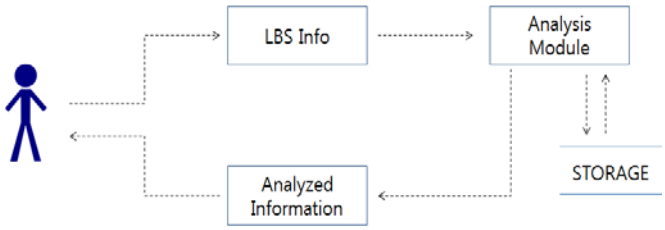


Fig. 5 Module Structure

IV. Implementation

The LBS application suggested in this paper was implemented using Android's Web View. The information search function uses open API, such as Tour API, MAP API, etc., and stored addresses, local codes, user information, local information, etc.

ISBN: 978-1-61804-290-3

were stored in privatized storage. The application was implemented to provide information regarding a user's current location and region using this storage information. Moreover, requests for resources were minimized by requesting data based on server and URL/JSON type. The majority of information was stored in privatized storage to ensure a prompt response to the user's request is possible.

Security features can cause many problems if their use is not considered carefully. The following items are considered in security features: access control, authentication, confidentiality, encryption, and rights management [5].

A. Local Information Update Method

It was composed of a module that revises the user's location, which is implemented when the location revision icon of the action bar on the main screen is touched after initial launch. It retrieves a coordinate by selecting the optimal provider from methods using GPS or base station according to the device settings. The relevant coordinate is converted to an address using the MAP API service and the address is subsequently provided to the user, followed by a series of storing processes that keep coordinate values, addresses, and local codes necessary for other services inside the app. The user registration process that pops up at the initial running stores the ID and nickname of the Android device in the server. Even after reinstalling after deletion, it remembers the previous nickname using the Android device's ID. Figure 6 below shows the location setting execution screen.

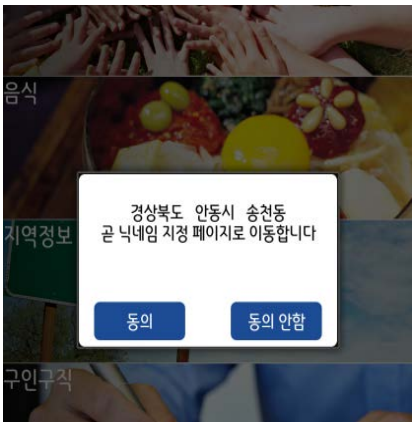


Fig. 6 Location Setting Execution Screen

When the user information is renewed, an alarm service is provided that enables the user to reset his/her location by agreeing to activation of the service. This alarm service can be set in the environment setting menu.

B. Tour API Module

This module provides information about local food, tourism, and cultural facilities to users of Tour API. When selecting a function, each menu completes different URLs to request of Tour API, so requests regarding food, tourism, and culture can be processed on a single module. Tour API is available in two types: XML and JSON. Because the JSON type has a comparative advantage in speed in the case of small-sized data processing, data processing was designed and implemented following JSON in this paper. The JSON object was objectified through the parsing process and stored in an array list to be used. Figure 7 below shows the screen where the information is being provided using Tour API based on the current location. The location information set by the user is used to provide a list of restaurants including information and images starting from the nearest restaurant. Clicking on "detailed view" delivers information about a particular restaurant and its accurate location.

Fig. 7 Local Information

V. Conclusion

LBS applications should be able to change adaptively according to the surrounding environment of the user's current location. However, previous LBS applications were inconvenient, as they were unable to provide information by designating a certain category or a certain distance by retrieving GPS coordinates, such as several meters away. The service application suggested in this paper was designed and implemented by supplementing these shortcomings and made it possible for users to choose selectively what information they want to receive by category so users can be provided with information more accurately and conveniently. To respond to technological developments and users' tastes, LBS applications should be able to show intellectually only information users want to see. With its active location setting and information provision abilities, we expect the LBS application suggested in this paper can contribute to the development and offering of diverse applications with decent usability in the future. Secure coding is essential for developing these applications. The developer should follow secure coding guidelines and examples of violations. Recent social network services (SNSs) also offer mobile services, rather than personal computer (PC) services, that can provide location-based services as a priority. In the future, these mobile services are likely to provide location-based services that use the user's present location as well as the user's scope of movement. Based on the results of this study, future studies are planned to evaluate various

ISBN: 978-1-61804-290-3

location-based services through the analysis of location information about the user's scope of movement and pattern analysis of other users' movements in similar locations.

REFERENCES

[1] http://blogs.wsj.com/korearealtime/2014/12/12/smartphone-usage-overtakes-pcs-in-south-korea/?mod=WSJBlog&mod=koreart
[2] http://play.google.com/
[3] http://api.visitkorea.or.kr/
[4] Crockford, Douglas. "The application/json media type for javascript object notation (json)." (2006).
[5] Hyeonpyo Hong, Hankyu Lim, Application of Smart Device Usage Control adopting Secure Coding for Hacking Protection, International Journal of Internet and Web Application, 2015

Iterative Form for Optimal FIR Filtering of Time-Variant Systems

Shunyi Zhao, Yuriy S. Shmaliy, Sanowar H. Khan, and Guoli Ji

***Abstract*—In this paper, the iterative forms of the optimal finite impulse response (OFIR) filter for time-variant systems is derived. With the resulting algorithm, the OFIR filter can be realized iteratively to avoid the inverse of matrices with large dimension. It shows that the iterative structure proposed is consistent with that developed previously to the unbiased FIR (UFIR) filter, which has a predictor/correcter format. Several special cases are also considered, and corresponding modifications of the results are obtained. Finally, an example is given to demonstrate the efficiency.**

***Keywords*—Iterative form, optimal filter, FIR filter, time-variant system.**

I. Introduction

As a counterpart to the Kalman filter (KF) having infinite impulse response (IIR), the finite impulse response (FIR) filter demonstrates several useful engineering properties such as the bound input/bound output (BIBO) stability [1], better robustness against temporary model uncertainties and round-off errors [2], and lower sensitivity to noise owing to embedded averaging [3]. The development of FIR filters, smoothers, and predictors have been provided during decades [4]–[11].

Unlike the recursive KF requiring a most recent measurement, the batch FIR filter utilizes N most recent measurements. If an averaging horizon N is long, all matrices and vectors acquire large dimensions and real-time applications of FIR filters become problematic in view of large computational burden and slow operation. An efficient solution implies computing FIR estimates iteratively using the Kalman procedure [12]. It makes the FIR filter about N times slower than KF and the former can operate almost as fast as the later using parallel computing. Following the approach, the Kalman-FIR filter was derived in [8] and optimal FIR (OFIR) filter with embedded unbiasedness addressed by Kwon, Kim and Han in [13] for prediction state-space models. For real-time state-space models [14], an iterative unbiased FIR (UFIR) algorithm was proposed by Shmaliy in [3], [12]. However, no iterative form still has been addressed for the batch OFIR filter [3], [15] in spite of the fact that optimal estimates are required for many applications.

In this paper, an iterative form of the OFIR filter is derived for discrete time-variant state space model with Gaussian white noise. Compared to the infinite impulse response (IIR) filters, the proposed method inherits advantages of FIR structures and is more robust against the temporary modeling uncertainties. The rest of the paper is organized as follows. In section II, we describe the systems and give the problem. The batch form of the OFIR filter is also derived in this section. In section III, we present the main result, where the iterative realization is proposed. A simulation example is conducted in Section IV, and conclusion are summarized in Section V.

We use the following notations: $\mathbb{R}^n$ denotes the n dimensional Euclidean space, $E\{\cdot\}$ denotes the statistical averaging of the stochastic process or vector, $\mathrm{diag}(\mathbf{e}_1 \cdots \mathbf{e}_m)$ represents a diagonal matrix with diagonal elements $\mathbf{e}_1, \cdots, \mathbf{e}_m$, $\mathrm{tr}(\mathbf{M})$ is the trace of $\mathbf{M}$, and $\mathbf{I}$ is the identity matrix of proper dimensions.

II. State-Space Model and Preliminaries

Motivated by the problems of state estimation and tracking which often arise in signal processing and wireless systems, we consider a linear discrete-time system represented in state-space with the time-variant model

$$\mathbf{x}_k = \mathbf{A}_k\mathbf{x}_{k-1} + \mathbf{B}_k\mathbf{w}_k\,, \tag{1}$$

$$\mathbf{y}_k = \mathbf{C}_k\mathbf{x}_k + \mathbf{D}_k\mathbf{v}_k\,, \tag{2}$$

where $\mathbf{x}_k \in \mathbb{R}^n$ is the state vector in Euclidean space, $\mathbf{y}_k \in \mathbb{R}^q$ is the measurement vector, $\mathbf{A}_k \in \mathbb{R}^{n\times n}$, $\mathbf{B}_k \in \mathbb{R}^{n\times u}$, $\mathbf{C}_k \in \mathbb{R}^{q\times n}$ and $\mathbf{D}_k \in \mathbb{R}^{q\times v}$ are time-variant matrices, which are assumed to be known. The process noise $\mathbf{w}_k \in \mathbb{R}^u$ and measurement noise $\mathbf{v}_k \in \mathbb{R}^v$ are zero mean Gaussian, $E\{\mathbf{w}_k\} = \mathbf{0}$ and $E\{\mathbf{v}_k\} = \mathbf{0}$, mutually uncorrelated and have known covariances $\mathbf{R} = E\{\mathbf{w}_k\mathbf{w}_k^T\}$, $\mathbf{Q} = E\{\mathbf{v}_k\mathbf{v}_k^T\}$.

Employing the discrete convolution, the FIR estimator can be defined as a linear combination of finite samples of measurements as

$$\hat{\mathbf{x}}_{k|k} = \mathbf{K}_k\mathbf{Y}_{k,l}, \tag{3}$$

where $l = k - N + 1$ is the starting point of the horizon, N is the horizon length, $\hat{\mathbf{x}}_{k|k}$ is the estimate, $\mathbf{Y}_{k,l}$ is a vector of measurements collected on a horizon $[l, k]$, and $\mathbf{K}_k$ is the filter gain determined by a given performance criterion.

Compared with the IIR structure, a distinct feature of FIR estimator is that N most recent measurements are employed at each time step, while only one most recent measurement is used in the IIR (Kalman) recursion. This leads to $\mathcal{O}(N)$ complexity. However, some good properties such as the BIBO stability and better robustness are achieved.

This investigation was supported by the Royal Academy of Engineering under the Newton Research Collaboration Programme NRCP/1415/140.

S. Zhao and G. Ji are with the Department of Automation, Xiamen University, Xiamen, Fujian, 361005, e-mail: shunyi.s.y@gmail.com.

Y. S. Shmaliy is with the Department of Electronics Engineering, Universidad de Guanajuato, Salamanca, 36885, Mexico e-mail: shmaliy@ugto.mx.

S. H. Khan is with the Department of Electronics Engineering, City University London, London, UK, e-mail: S.H.Khan@city.ac.uk.

To derive the OFIR filter on a horizon of N past measurements from l to k, we represent (1) and (2) in a batch form as

$$\mathbf{X}_{k,l} = \mathbf{A}_{k,l}\mathbf{x}_l + \mathbf{B}_{k,l}\mathbf{W}_{k,l}\,, \quad (4)$$

$$\mathbf{Y}_{k,l} = \mathbf{C}_{k,l}\mathbf{x}_l + \mathbf{H}_{k,l}\mathbf{W}_{k,l} + \mathbf{D}_{k,l}\mathbf{V}_{k,l}\,. \quad (5)$$

Here, $\mathbf{X}_{k,l} \in \mathbb{R}^{Nn}$, $\mathbf{Y}_{k,l} \in \mathbb{R}^{Nq}$, $\mathbf{W}_{k,l} \in \mathbb{R}^{Nu}$ and $\mathbf{V}_{k,l} \in \mathbb{R}^{Nv}$ are specified as, respectively,

$$\mathbf{X}_{k,l} = \left[\mathbf{x}_k^T \mathbf{x}_{k-1}^T \cdots \mathbf{x}_l^T\right]^T, \quad (6)$$

$$\mathbf{Y}_{k,l} = \left[\mathbf{y}_k^T \mathbf{y}_{k-1}^T \cdots \mathbf{y}_l^T\right]^T, \quad (7)$$

$$\mathbf{W}_{k,l} = \left[\mathbf{w}_k^T \mathbf{w}_{k-1}^T \cdots \mathbf{w}_l^T\right]^T, \quad (8)$$

$$\mathbf{V}_{k,l} = \left[\mathbf{v}_k^T \mathbf{v}_{k-1}^T \cdots \mathbf{v}_l^T\right]^T. \quad (9)$$

The extended model matrix $\mathbf{A}_{k,l} \in \mathbb{R}^{Nn\times n}$, process noise matrix $\mathbf{B}_{k,l} \in \mathbb{R}^{Nn\times Nu}$, observation matrix $\mathbf{C}_{k,l} \in \mathbb{R}^{Nq\times n}$, auxiliary process noise matrix $\mathbf{H}_{k,l} \in \mathbb{R}^{Nq\times Nu}$ and measurement noise matrix $\mathbf{D}_{k,l} \in \mathbb{R}^{Nq\times Nv}$ are all time-variant and dependent on the current time k and the horizon length N. Model (1) and (2) suggests that these matrices can be written as, respectively

$$\mathbf{A}_{k,l} = [\mathcal{A}_{k,l+1}^T, \mathcal{A}_{k-1,l+1}^T, \cdots, \mathcal{A}_{l+1,l+1}^T, \mathbf{I}]^T, \quad (10)$$

$$\mathbf{B}_{k,l} = \begin{bmatrix} \mathbf{B}_k & \mathcal{A}_{k,k}\mathbf{B}_{k-1} & \cdots & \mathcal{A}_{k,l+2}\mathbf{B}_{l+1} & \mathcal{A}_{k,l+1}\mathbf{B}_l \\ \mathbf{0} & \mathbf{B}_{k-1} & \cdots & \mathcal{A}_{k-1,l+2}\mathbf{B}_{l+1} & \mathcal{A}_{k-1,l+1}\mathbf{B}_l \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{B}_{l+1} & \mathcal{A}_{l+1,l+1}\mathbf{B}_l \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{B}_l \end{bmatrix}, \quad (11)$$

$$\mathbf{C}_{k,l} = \bar{\mathbf{C}}_{k,l}\mathbf{A}_{k,l}, \quad (12)$$

$$\mathbf{H}_{k,l} = \bar{\mathbf{C}}_{k,l}\mathbf{B}_{k,l}, \quad (13)$$

$$\mathbf{D}_{k,l} = \mathrm{diag}\left(\mathbf{D}_k\mathbf{D}_{k-1}\cdots\mathbf{D}_l\right), \quad (14)$$

with

$$\mathcal{A}_{\psi,\zeta} = \begin{cases} \mathbf{A}_\psi\mathbf{A}_{\psi-1}\cdots\mathbf{A}_\zeta, & \text{if } \psi > \zeta \\ \mathbf{A}_\psi, & \text{if } \psi = \zeta \end{cases}, \quad (15)$$

$$\bar{\mathbf{C}}_{k,l} = \mathrm{diag}\left(\mathbf{C}_k\mathbf{C}_{k-1}\cdots\mathbf{C}_l\right), \quad (16)$$

where $\psi \geqslant \zeta$. Note that the state equation specified by (4) and (5) at the initial point l is $\mathbf{x}_l = \mathbf{x}_l + \mathbf{B}_l\mathbf{w}_l$, suggesting that $\mathbf{w}_l$ is zero-valued. That is, the initial state $\mathbf{x}_l$ is required to be known or estimated optimally. In the following, we concentrate our attention on the fast iterative form for the OFIR filter developed in [15].

A. Batch Optimal FIR filter

Before discussing the iterative form for the OFIR filter, we consider and modify its batch form. In doing so, we substite (5) into (3) and write

$$\hat{\mathbf{x}}_{k|k} = \mathbf{K}_k\left(\mathbf{C}_{k,l}\mathbf{x}_l + \mathbf{H}_{k,l}\mathbf{W}_{k,l} + \mathbf{D}_{k,l}\mathbf{V}_{k,l}\right). \quad (17)$$

At this point, our objective is to achieve the optimal gain $\hat{\mathbf{K}}_k$ to minimize the covariance of estimation error in the MMSE sense. That is, the following cost criterion must be minimized

$$\hat{\mathbf{K}}_k = \arg\min_{\mathbf{K}_k} E\left\{\left(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k}\right)\left(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k}\right)^T\right\}. \quad (18)$$

In order to compute (18), the orthogonality principle can be employed. Specifically, the optimal filter gain $\hat{\mathbf{K}}_k$ should guarantee the estimation error $\mathbf{x}_k - \hat{\mathbf{x}}_{k|k}$ is uncorrelated with any of the measurements $\mathbf{Y}_{k,l}$, and also to any of the linear combination of these measurements. In this sense, an equivalent way to rewrite (18) is

$$E\left\{\left(\mathbf{x}_k - \hat{\mathbf{K}}_k\mathbf{Y}_{k,l}\right)\left(\mathbf{Y}_{k,l}\right)^T\right\} = \mathbf{0}\,. \quad (19)$$

The state model $\mathbf{x}_k$ required by (19) must be represented on the averaging horizon of N points. We provided this model as

$$\mathbf{x}_k = \mathcal{A}_{k,l+1}\mathbf{x}_l + \bar{\mathbf{B}}_{k,l}\mathbf{W}_{k,l}\,, \quad (20)$$

where $\bar{\mathbf{B}}_{k,l}$ is the first row vector of $\mathbf{B}_{k,l}$. Substituting (5) and (20) into (19), using the fact that the initial state $\mathbf{x}_l$, systems noise vector $\mathbf{W}_{k,l}$ and measurement noise $\mathbf{V}_{k,l}$ are pairwise independent, and taking the expectation and rearranging the terms, transform (19) to

$$\mathcal{A}_{k,l+1}\mathbf{\Theta}_{x,l}\mathbf{C}_{k,l}^T + \bar{\mathbf{Z}}_{w,k} = \hat{\mathbf{K}}_k\mathbf{Z}_{x+w+v,k}\,, \quad (21)$$

where auxiliary matrices are

$$\mathbf{\Theta}_l = E\left\{\mathbf{x}_l\mathbf{x}_l^T\right\}, \quad (22)$$

$$\mathbf{Z}_{x,k} = \mathbf{C}_{k,l}\mathbf{\Theta}_l\mathbf{C}_{k,l}^T\,, \quad (23)$$

$$\mathbf{Z}_{w,k} = \mathbf{H}_{k,l}E\left[\mathbf{W}_{k,l}\mathbf{W}_{k,l}^T\right]\mathbf{H}_{k,l}^T\,, \quad (24)$$

$$\mathbf{Z}_{v,k} = \mathbf{D}_{k,l}E\left[\mathbf{V}_{k,l}\mathbf{V}_{k,l}^T\right]\mathbf{D}_{k,l}^T\,, \quad (25)$$

$$\bar{\mathbf{Z}}_{w,k} = \bar{\mathbf{B}}_{k,l}E\left[\mathbf{W}_{k,l}\mathbf{W}_{k,l}^T\right]\mathbf{H}_{k,l}^T\,, \quad (26)$$

$$\mathbf{Z}_{x+w+v,k} = \mathbf{Z}_{x,k} + \mathbf{Z}_{w,k} + \mathbf{Z}_{v,k}\,. \quad (27)$$

Here, $\mathbf{Z}_{x,k}$, $\mathbf{Z}_{w,k}$ and $\mathbf{Z}_{v,k}$ denote the covariances of the initial state, system, and measurement noise respectively which are symmetric and invertible. We further multiply both sides of (21) from the right-hand side with term $\mathbf{Z}_{x+w+v,k}^{-1}$ and go to the optimal gain of

$$\hat{\mathbf{K}}_k = (\mathcal{A}_{k,l+1}\mathbf{\Theta}_l\mathbf{C}_{k,l}^T + \bar{\mathbf{Z}}_{w,k})\mathbf{Z}_{x+w+v,k}^{-1}\,. \quad (28)$$

Further, by multiplying $\mathbf{\Theta}_l$ on the right hand of (28) with the identity matrix $(\mathbf{C}_{k,l}^T\mathbf{C}_{k,l})^{-1}\mathbf{C}_{k,l}^T\mathbf{C}_{k,l}$, from the left-hand side, the optimal filter gain $\hat{\mathbf{K}}_k$ can be equivalently rewritten in a more compact form as

$$\hat{\mathbf{K}}_k = \bar{\mathbf{K}}_k\mathbf{Z}_{x,k}\mathbf{Z}_{x+w+v,k}^{-1} + \bar{\mathbf{Z}}_{w,k}\mathbf{Z}_{x+w+v,k}^{-1}\,, \quad (29)$$

where

$$\bar{\mathbf{K}}_k = \mathcal{A}_{k,l+1}\left(\mathbf{C}_{k,l}^T\mathbf{C}_{k,l}\right)^{-1}\mathbf{C}_{k,l}^T \quad (30)$$

is the unbiased filter gain [3]. One may now substitute (30) into (3), provide the averaging, and arrive at the conclusion that (29) guarantees the unbiasedness $E\left[\hat{\mathbf{x}}_{k|k}\right] = E\left[\mathbf{x}_k\right]$.

In order to compute (29), the covariance of the initial state $\mathbf{Z}_{x,k}$ must be specified. It has been shown in [3], [15] that

ISBN: 978-1-61804-290-3

$\mathbf{Z}_{x,k}$ can be found by solving the discrete algebraic Riccati equation (DARE),

$$\mathbf{Y}_{k,l}\mathbf{Y}_{k,l}^T\mathbf{Z}_{w+v,k}^{-1}\mathbf{Z}_{x,k}-\mathbf{Z}_{x,k}\mathbf{Z}_{w+v,k}^{-1}\mathbf{Z}_{x,k}-2\mathbf{Z}_{x,k}-\mathbf{Z}_{w+v,k}=\mathbf{0} \tag{31}$$

where

$$\mathbf{Z}_{w+v,k}=\mathbf{Z}_{w,k}+\mathbf{Z}_{v,k}\,. \tag{32}$$

Note that numerical solution of (32) is often unavailable at each time index k due to computational problems.

III. Iterative Form

In order to find an iterative form for (28), we use i as an iterative variable, and employ (10), (12) and (13) and decompose $\mathbf{C}_{i,l}$, $\mathbf{H}_{i,l}$, and $\mathbf{D}_{i,l}$ as, respectively,

$$\begin{aligned}
\mathbf{C}_{i,l} &= \left[(\mathbf{C}_i\mathcal{A}_{i,l+1})^T\ \mathbf{C}_{i-1,l}^T\right]^T, && (33)\\
\mathbf{H}_{i,l} &= \begin{bmatrix} \mathbf{C}_i\mathbf{B}_i & \mathbf{C}_i\mathbf{A}_i\bar{\mathbf{B}}_{i-1,l} \\ \mathbf{0} & \mathbf{H}_{i-1,l} \end{bmatrix}, && (34)\\
\mathbf{D}_{i,l} &= \begin{bmatrix} \mathbf{D}_i & \mathbf{0} \\ \mathbf{0} & \mathbf{D}_{i-1,l} \end{bmatrix}. && (35)
\end{aligned}$$

Using (33)-(35) allows representing (23)-(25) as

$$\begin{aligned}
\mathbf{Z}_{x,i} &= \begin{bmatrix} \mathbf{C}_i\mathcal{A}_{i,l+1}\mathbf{\Theta}_l\mathcal{A}_{i,l+1}^T\mathbf{C}_i^T & \mathbf{C}_i\mathcal{A}_{i,l+1}\mathbf{\Theta}_l\mathbf{C}_{i-1,l}^T \\ \mathbf{C}_{i-1,l}\mathbf{\Theta}_l\mathcal{A}_{i,l+1}^T\mathbf{C}_i^T & \mathbf{C}_{i-1,l}\mathbf{\Theta}_l\mathbf{C}_{i-1,l}^T \end{bmatrix} && (36)\\
\mathbf{Z}_{w,i} &= \begin{bmatrix} \mathbf{C}_i\bar{\mathbf{B}}_{i,l}\mathbf{R}_i\bar{\mathbf{B}}_{i,l}^T\mathbf{C}_i^T & \mathbf{C}_i\mathbf{A}_i\bar{\mathbf{Z}}_{w,i-1}^T \\ \bar{\mathbf{Z}}_{w,i-1}^T\mathbf{A}_i^T\mathbf{C}_i^T & \mathbf{Z}_{w,i-1} \end{bmatrix}, && (37)\\
\mathbf{Z}_{v,i} &= \begin{bmatrix} \mathbf{D}_i\mathbf{Q}\mathbf{D}_i^T & \mathbf{0} \\ \mathbf{0} & \mathbf{Z}_{v,i-1} \end{bmatrix}, && (38)
\end{aligned}$$

where

$$\bar{\mathbf{Z}}_{w,i-1}=\bar{\mathbf{B}}_{i-1,l}\mathbf{R}_{i-1}\mathbf{H}_{i-1,l}^T\,. \tag{39}$$

By defining $\mathbf{\Delta}_i$, $\mathbf{F}_i$ and $\mathbf{U}_i$ as, respectively,

$$\begin{aligned}
\mathbf{\Delta}_i &\triangleq \begin{bmatrix} \tilde{\mathbf{Q}}_i & \mathbf{0} \\ \mathbf{0} & \mathbf{Z}_{x+w+v,i-1} \end{bmatrix}, && (40)\\
\mathbf{U}_i &\triangleq \mathcal{A}_{i,l+1}\mathbf{\Theta}_l{\mathcal{A}_{i,l+1}}^T+\bar{\mathbf{B}}_{i,l}\mathbf{R}_i\bar{\mathbf{B}}_{i,l}^T, && (41)\\
\mathbf{F}_i &\triangleq \mathcal{A}_{i,l+1}\mathbf{\Theta}_l\mathbf{C}_{i,l}^T+\bar{\mathbf{B}}_{i,l}\mathbf{R}_i\mathbf{H}_{i,l}^T, && (42)\\
\tilde{\mathbf{Q}}_i &\triangleq \mathbf{D}_i\mathbf{Q}\mathbf{D}_i^T, && (43)
\end{aligned}$$

we provide

$$\mathbf{Z}_{x+w+v,i}=\mathbf{\Delta}_i+\mathbf{\Phi}_i\,, \tag{44}$$

where

$$\mathbf{\Phi}_i = \begin{bmatrix} \mathbf{C}_i\mathbf{U}_i\mathbf{C}_i^T & \mathbf{C}_i\mathbf{A}_i\mathbf{F}_{i-1} \\ \mathbf{F}_{i-1}^T\mathbf{A}_i^T\mathbf{C}_i^T & \mathbf{0} \end{bmatrix}. \tag{45}$$

Using the matrix inversion lemma [16], we represent the inverse of $\mathbf{Z}_{x+w+v,i}$ as

$$\mathbf{Z}_{x+w+v,i}^{-1}=\mathbf{\Delta}_i^{-1}-\mathbf{\Delta}_i^{-1}(\mathbf{I}+\mathbf{\Phi}_i\mathbf{\Delta}_i^{-1})^{-1}\mathbf{\Phi}_i\mathbf{\Delta}_i^{-1}\,. \tag{46}$$

We next decompose $\bar{\mathbf{B}}_{i,l}$ as $\bar{\mathbf{B}}_{i,l}=\left[\mathbf{B}_i\ \mathbf{A}_i\bar{\mathbf{B}}_{i-1,l}\right]$ and rewrite $\bar{\mathbf{Z}}_{w,i}$ equivalently as

$$\bar{\mathbf{Z}}_{w,i}=\left[\bar{\mathbf{B}}_{i,l}\mathbf{R}_i\bar{\mathbf{B}}_{i,l}^T\mathbf{C}_i^T\ \mathbf{A}_i\bar{\mathbf{Z}}_{w,i-1}\right]. \tag{47}$$

Similarly, we get

$$\begin{aligned}
\mathcal{A}_{i,l+1}\mathbf{\Theta}_l\mathbf{C}_{i,l}^T &= \mathcal{A}_{i,l+1}\mathbf{\Theta}_l\left[(\mathbf{C}_i\mathcal{A}_{i,l+1})^T\ \mathbf{C}_{i-1,l}^T\right] \\
&= \left[\mathcal{A}_{i,l+1}\mathbf{\Theta}_l{\mathcal{A}_{i,l+1}}^T\mathbf{C}_i^T\ \mathcal{A}_{i,l+1}\mathbf{\Theta}_l\mathbf{C}_{i-1,l}^T\right]. && (48)
\end{aligned}$$

Now, substituting (47) and (48) into (42) leads to

$$\mathbf{F}_i=\left[\mathbf{U}_i\mathbf{C}_i^T\ \mathbf{A}_i\mathbf{F}_{i-1}\right]. \tag{49}$$

and we notice that $\mathbf{U}_i$ can also be computed recursively as

$$\begin{aligned}
\mathbf{U}_i &= \mathbf{A}_i\left(\mathcal{A}_{i-1,l+1}\mathbf{\Theta}_l{\mathcal{A}_{i-1,l+1}}^T\right)\mathbf{A}_i^T \\
&\quad+\left[\mathbf{B}_i\mathbf{R}\ \ \mathbf{A}_i\bar{\mathbf{B}}_{i-1,l}\mathbf{R}_{i-1}\right] \\
&\quad\times\left[\mathbf{B}_i^T\ \bar{\mathbf{B}}_{i-1,l}^T\mathbf{A}_i^T\right]^T \\
&= \mathbf{A}_i\mathbf{U}_{i-1}\mathbf{A}_i^T+\mathbf{B}_i\mathbf{R}\mathbf{B}_i^T\,. && (50)
\end{aligned}$$

By using (46), (49), and (50) and taking into account that $\hat{\mathbf{K}}_{i-1}=\mathbf{F}_{i-1}\mathbf{Z}_{x+w+v,i-1}^{-1}$, we obtain

$$\begin{aligned}
\hat{\mathbf{K}}_i &= \left[\mathbf{U}_i\mathbf{C}_i^T\ \mathbf{A}_i\mathbf{F}_{i-1}\right] \\
&\quad\times\left(\mathbf{\Delta}_i^{-1}-\mathbf{\Delta}_i^{-1}(\mathbf{I}+\mathbf{\Phi}_i\mathbf{\Delta}_i^{-1})^{-1}\mathbf{\Phi}_i\mathbf{\Delta}_i^{-1}\right) \\
&= \left[\mathbf{U}_i\mathbf{C}_i^T\tilde{\mathbf{Q}}_i^{-1}\ \mathbf{A}_i\hat{\mathbf{K}}_{i-1}\right]-\left[\mathbf{U}_i\mathbf{C}_i^T\tilde{\mathbf{Q}}_i^{-1}\ \mathbf{A}_i\hat{\mathbf{K}}_{i-1}\right] \\
&\quad\times\left(\mathbf{I}+\mathbf{\Phi}_i\mathbf{\Delta}_i^{-1}\right)^{-1}\mathbf{\Phi}_i\mathbf{\Delta}_i^{-1}\,. && (51)
\end{aligned}$$

After some rearrangements, we arrive at

$$\hat{\mathbf{K}}_i=\left[\mathbf{U}_i\mathbf{C}_i^T\tilde{\mathbf{Q}}_i^{-1}\ \mathbf{A}_i\hat{\mathbf{K}}_{i-1}\right]\mathbf{S}_i^{-1}\,, \tag{52}$$

where

$$\mathbf{S}_i=\mathbf{I}+\mathbf{\Phi}_i\mathbf{\Delta}_l^{-1}=\begin{bmatrix} \mathbf{S}_{i11} & \mathbf{S}_{i12} \\ \mathbf{S}_{i21} & \mathbf{S}_{i22} \end{bmatrix}, \tag{53}$$

with

$$\begin{aligned}
\mathbf{S}_{i11} &= \mathbf{I}+\mathbf{C}_i\mathbf{U}_i\mathbf{C}_i^T\tilde{\mathbf{Q}}_i^{-1}\,, && (54)\\
\mathbf{S}_{i12} &= \mathbf{C}_i\mathbf{A}_i\hat{\mathbf{K}}_{i-1}\,, && (55)\\
\mathbf{S}_{i21} &= \mathbf{F}_{i-1}^T\mathbf{A}_i^T\mathbf{C}_i^T\tilde{\mathbf{Q}}_i^{-1}\,, && (56)\\
\mathbf{S}_{i22} &= \mathbf{I}\,. && (57)
\end{aligned}$$

Using the Schur complement of $\mathbf{S}_{i11}$ [17], the inverse of (53) can now be found as

$$\mathbf{S}_i^{-1} = \begin{bmatrix} \bar{\mathbf{S}}_{i11}^{-1} & -\bar{\mathbf{S}}_{i11}^{-1}\mathbf{S}_{i12} \\ -\mathbf{S}_{i21}\bar{\mathbf{S}}_{i11}^{-1} & \mathbf{I}+\mathbf{S}_{i21}\bar{\mathbf{S}}_{i11}^{-1}\mathbf{S}_{i12} \end{bmatrix}, \tag{58}$$

where

$$\begin{aligned}
\bar{\mathbf{S}}_{i11} &= \mathbf{S}_{i11}-\mathbf{S}_{i12}\mathbf{S}_{i22}^{-1}\mathbf{S}_{i21} \\
&= \mathbf{I}+\mathbf{C}_i\mathbf{N}_i\mathbf{C}_i^T\tilde{\mathbf{Q}}_i^{-1}\,, && (59)\\
\mathbf{N}_i &= \mathbf{U}_i-\mathbf{A}_i\hat{\mathbf{K}}_{i-1}\mathbf{F}_{i-1}^T\mathbf{A}_i^T\,. && (60)
\end{aligned}$$

ISBN: 978-1-61804-290-3

Then substituting (59) into (52) and using (50), (51), and (53) give us

$$\hat{\mathbf{K}}_i = \left[\mathbf{G}_i \quad \mathbf{A}_i\hat{\mathbf{K}}_{i-1} - \mathbf{G}_i\mathbf{C}_i\mathbf{A}_i\hat{\mathbf{K}}_{i-1}\right], \tag{61}$$

where

$$\mathbf{G}_i = \mathbf{N}_i\mathbf{C}_i^T\left(\tilde{\mathbf{Q}}_i + \mathbf{C}_i\mathbf{N}_i\mathbf{C}_i^T\right)^{-1}. \tag{62}$$

At this point, the batch OFIR filtering estimate (3) $\hat{\mathbf{x}}_i$ can be computed iteratively using the Kalman recursion as

$$\begin{aligned}\hat{\mathbf{x}}_i &= \left[\mathbf{G}_i \quad \mathbf{A}_i\hat{\mathbf{K}}_{i-1} - \mathbf{G}_i\mathbf{C}_i\mathbf{A}_i\hat{\mathbf{K}}_{i-1}\right]\begin{bmatrix}\mathbf{y}_i\\ \mathbf{Y}_{i-1,l}\end{bmatrix}\\ &= \mathbf{A}_i\hat{\mathbf{x}}_{i-1} + \mathbf{G}_i\left(\mathbf{y}_i - \mathbf{C}_i\mathbf{A}_i\hat{\mathbf{x}}_{i-1}\right).\end{aligned} \tag{63}$$

To arrive at the final iterative form of $\hat{\mathbf{x}}_i$, some extra transformations are required. Namely, by setting $i = i+1$, we provide

$$\mathbf{N}_{i+1} = \mathbf{U}_{i+1} - \mathbf{A}_i\hat{\mathbf{K}}_i\mathbf{F}_i^T\mathbf{A}_i^T. \tag{64}$$

Using the recursions of $\mathbf{U}_i$, $\hat{\mathbf{K}}_i$ and $\mathbf{F}_i$, we obtain

$$\mathbf{N}_{i+1} = \mathbf{A}_i\mathbf{N}_i\mathbf{A}_i^T + \mathbf{B}_i\mathbf{R}\mathbf{B}_i^T - \mathbf{A}_i\mathbf{G}_i\mathbf{C}_i\mathbf{N}_i\mathbf{A}_i^T, \tag{65}$$

in which we used a property $\mathbf{U}_i = \mathbf{U}_i^T$. Substituting $\mathbf{G}_i$ with (62) and setting $i = i-1$, we also find

$$\begin{aligned}\mathbf{N}_i &= \mathbf{A}_i\mathbf{N}_{i-1}\mathbf{A}_i^T + \mathbf{B}_i\mathbf{R}\mathbf{B}_i^T - \mathbf{A}_i\mathbf{N}_{i-1}\mathbf{C}_i^T\\ &\quad\times\left(\tilde{\mathbf{Q}}_i + \mathbf{C}_i\mathbf{N}_{i-1}\mathbf{C}_i^T\right)^{-1}\mathbf{C}_i\mathbf{N}_{i-1}\mathbf{A}_i^T.\end{aligned} \tag{66}$$

The iterative OFIR filtering algorithm can finally be summarized in the following theorem.

Theorem 1: Given the discrete time-variant state space model (1) and (2) with zero mean and mutually independent noise vectors $\mathbf{w}_k$ and $\mathbf{v}_k$ having Gaussian distributions and known covariances, the iterative form of the OFIR estimator can be stated by

$$\mathbf{N}_l = \mathbf{\Theta}_l + \mathbf{B}_l\mathbf{R}\mathbf{B}_l^T, \tag{67}$$

$$\hat{\mathbf{x}}_l = \mathbf{N}_l\mathbf{C}_l^T\mathbf{Z}_{x+w+v,l}^{-1}\mathbf{y}_l, \tag{68}$$

$$\begin{aligned}\hat{\mathbf{x}}_i &= \mathbf{A}_i\hat{\mathbf{x}}_{i-1} + \mathbf{N}_i\mathbf{C}_i^T\left(\tilde{\mathbf{Q}}_i + \mathbf{C}_i\mathbf{N}_i\mathbf{C}_i^T\right)^{-1}\\ &\quad\times\left(\mathbf{y}_i - \mathbf{C}_i\mathbf{A}_i\hat{\mathbf{x}}_{i-1}\right),\end{aligned} \tag{69}$$

where i ranges from $l+1$ to k, $\mathbf{N}_i$ is computed recursively by (66), and the initial mean square state $\mathbf{\Theta}_l$ can be obtained by solving (31).

Proof: The proof has been provided by (33)-(66). ■

It is seen clearly that (69) has the Kalman-like form, where the term $\mathbf{A}_i\hat{\mathbf{x}}_{i-1}$ predicts the estimate from $i-1$ to i, the term $\mathbf{G}_i\left(\mathbf{y}_i - \mathbf{C}_i\mathbf{A}_i\hat{\mathbf{x}}_{i-1}\right)$ corrects the predicted value using the residual, and $\mathbf{G}_i$ plays the role of the Kalman gain.

TABLE I. SIMULATION RESULTS

Performance	BF	IF	KF
Average RMSE of first state	0.894	0.894	0.725
Average RMSE of second state	0.076	0.076	0.071

IV. EXAMPLE OF APPLICATIONS

As an example of applications, we employ the two-state polynomial model, (1) and (2), specified with $\mathbf{B} = \mathbf{I}$, $\mathbf{D} = \mathbf{I}$, $\mathbf{C} = [\,1\ 0\,]$, and

$$\mathbf{A} = \begin{bmatrix}1 & \tau\\ 0 & 1\end{bmatrix}$$

where τ is a constant in unit of time. We let $\tau = 0.1$ s and suppose that the system and measurement noise variances are $\sigma_{w1}^2 = 1\times10^{-4}$, $\sigma_{w2}^2 = 1\times10^{-4}/\mathrm{s}^2$, and $\sigma_v^2 = 1\times10^2$. The batch form (denoted as BF) and iterative form (denoted as IF) of OFIR filter were simulated over 2000 subsequent points. KF was employed as a benchmark. Typical estimation errors are given in Table I. As can be seen, both the batch form (BF) and iterative form (IF) of the OFIR filter inherently have the same accuracy. Although an example shown in Table I indicates a bit more accuracy in the KF estimate, in average, both the OFIR and KF filters produce equal errors.

Another experiment was conducted to compare the computation time required by the OFIR and KF algorithms. Toward this end, we set the measurement noise variance as $\sigma_v^2 = 1\times10^2$ and range the noise-to-noise ratio $\alpha = \sigma_{w1}^2/\sigma_v^2$ from 10^{-3} to 10^2. For each α, we define the "optimal" horizon N_{opt} for FIR filters by the value of N which provides the OFIR estimate closely related to the KF estimate. We then evaluate the computation time for each filter in the same computer environment.

The results are shown in Fig. 1. As can be seen in Fig. 1a, the BF is computationally low efficient, but the IF reduces the computation cost significantly. On the other hand, the computation time of the IF is about N_{opt} times larger than in the KF, because IF requires N_{opt} recursions to produce iteratively an optimal estimate.

To increase the computation rate, we further conducted the same experiment using parallel commutating. The computation time consumed by the filters is reflected in Fig. 1b. As expected, the computation time of FIR filters was reduced by the factor of N_{opt}. It is also seen that the parallel computation has made the IF algorithm operation as fast as the KF.

V. CONCLUSIONS

In this paper, we have derived a Kalman-like iterative algorithm for OFIR filtering of real time-variant discrete state-space models with white Gaussian noise. The algorithm proposed has the predictor/corrector structure which is inherent to the KF. Unlike the batch OFIR algorithm, the iterative OFIR algorithm does not require the inversion of matrices having large dimensions. It also saves essentially the computational resources without affecting the estimation accuracy.

ISBN: 978-1-61804-290-3

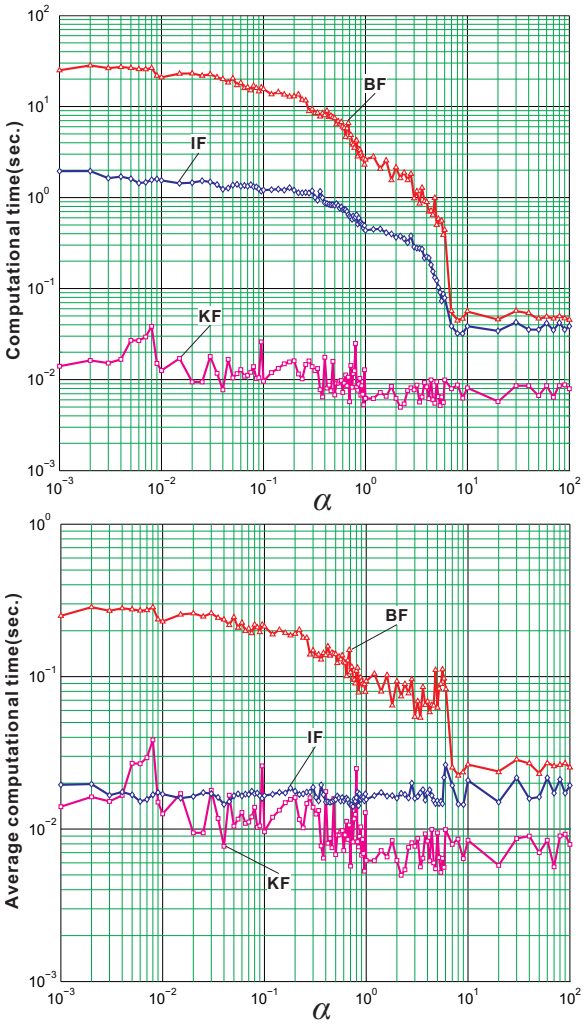


Fig. 1. Computation time required by the batch OFIR filter (BF), iterative OFIR filter (IF), and KF: (a) direct computation and (b) parallel computation.

REFERENCES

[1] A. H. Jazwinski, *Stochastic Processes and Filtering Theory*, New York: Academic, 1970.

[2] W. H. Kwon and S. Han, *Receding Horizon Control: Model Predictive Control for State Models*, London, U.K.: Springer, 2005.

[3] Y. S. Shmaliy, "Linear optimal FIR estimation of discrete time-invariant state-space models," *IEEE Trans. Signal Process.*, vol. 58, no. 6, pp. 3086-2010, Jun. 2010.

[4] A. H. Jazwinski, "Limited memory optimal filtering," *IEEE Trans. Autom. Control*, vol. 13, no.10, pp. 558-563, Oct. 1968.

[5] C. T. Mullis and R. A. Roberts, "Finite-memory problems and agorithms," *IEEE Trans. Inf. Theory*, vol. IT-20, no. 4, pp. 440-445, Jul. 1974.

[6] O. K. Kwon, W. H. Kwon, and K. S. Lee, "FIR filters and recursive forms for discrete-time state-space models," *Automatica*, vol. 25, no. 5, pp. 715-728, Sep. 1989.

[7] L. Danyang and L. Xuanhuang, "Optimal state estimation without the requirement of a prior statistics informantion of the initial state," *IEEE Trans. Autom. Control*, vol. 39, no. 10, pp. 2087-2091, Oct. 1994.

[8] W. H. Kwon, P. S. Kim, and P. Park, "A receding horizon Kalman FIR filter for discrete time-invariant systems," *IEEE Trans. Autom. Control*, vol. 99, no. 9, pp. 1787-1791, Sep. 1999.

[9] L. Mirkin, "On the H1 fixed-lag smoothing: how to exploit the information preview," *Automatica*, vol. 39, no. 8, pp. 1495-1504, Aug. 2003.

[10] B. K. Kwon, S. Han, O. K. Kim, and W. H. Kwon, "Minimum variance FIR smoothers for discrete-time state space models," *IEEE Trans. Signal Process. Lett.*, vol. 14, no. 8, pp. 557-560, Aug. 2007.

[11] Y. S. Shmaliy, "Optimal gains of FIR estimations for a class of discrete-time state-space models," *IEEE Trans. Signal Process. Lett.*, vol. 15, pp. 517-520, 2008.

[12] Y. S. Shmaliy, "An iterative Kalman-like algorithm ignoring noise and initial conditions," *IEEE Trans. Signal Process.*, vol. 59, no. 6, pp. 2465-2473, Jun. 2011.

[13] W. H. Kwon, P. S. Kim, and S. Han, "A receding horizon unbiased FIR filter for discrete-time state space models," *Automatica*, vol. 38, no. 3, pp. 545-551, Mar. 2002.

[14] J. Salmi, A. Richter, and V. Koivunen, Detection and tracking of MIMO propagation path parameters using state-space approach, *IEEE Trans. Signal Process.*, vol. 57, no. 4, pp. 1538C1550, Apr. 2009.

[15] Y. S. Shmaliy and O. I. Manzano, "Time-variant linear optimal finite impulse response estimator for discrete state-space models," *Int. J. Adapt. Contrl Signal Process.*, vol. 26, no. 2, pp. 95-104, Sep. 2012.

[16] G. H. Golub and G. F. van Loan, *Matrix Computation*, 3rd ed. Baltimore, MD: Johns Hopkins Univ. Press, 1996.

[17] S. Zhao, Y. S. Shmaliy, and F. Liu, "Fast computation of discrete optimal FIR estimates in white Gaussian noise," *IEEE Signal Processing Letters*, 2015, 22(6): 718-722.

ISBN: 978-1-61804-290-3

Performance analysis of synchronization in chaotic DSSS-CDMA system under Jamming attack

A. Tayebi, S.M. Berber, and A. Swain

***Abstract*— This paper investigates the performance of synchronization of the direct sequence spread spectrum (DSSS-CDMA) system under jamming. A result of initial investigation shows that the synchronization in DSSS-CDMA is vulnerable to the jamming attacks. This vulnerability has a tragic effect on the system's error rate. A mathematical expression for of the probability of detection and probability of false alarm has been derived and being used to study the performance of synchronization. . Further, we have computed the amount of jamming power which is required to collapse the synchronization. Performance analysis using Monte Carlo simulation in MATLAB conforms the theoretical results.**

Keywords*— Physical layer security; jamming attacks; synchronization; wireless sensor networks

I. INTRODUCTION

WRELESS sensor networks (WSN) are valuable to the different security attacks [1]. Because WSNs are playing an important role in various applications such as environment monitoring and target detection, it is important to evaluate their performance against security attacks. Amongst various types of security attacks, jamming is one of the simplest and the most dangerous attack [2]. A jammer can target different sections of a communication link [3]. In this study we focus on synchronization in the chaotic DSSS-CDMA system and evaluate its performance against jamming.

The communication system considered in this study is DSSS-CDMA, which uses chaotic sequences. In [4] the chaotic DSSS-CDMA performance is analyzed in the case of Gaussian noise for *N* single-users. The synchronization for this system is mathematically modeled and simulated in [5]. Also, the synchronization performance with the presence of channel noise and fading is analyzed in [6-8]. In addition, in [9], the authors propose an algorithm to attack the synchronization packets. In their investigation, the synchronization packets and data packets are assumed to be separated. Since the synchronization packets are fixed and periodic, it is possible to detect and attack them. However, in DSSS-CDMA system the synchronization bits are sent with the data bits at the same time. The robustness of synchronization in OFDM to different security attacks is investigated in [10, 11]. In addition, the synchronization of the MIMO-OFDM system is analyzed in [12].

This paper is organized as follows: In section-II, we describe the chaotic DSSS-CDMA system and its block-schematic. In section-III, the importance of the synchronization in DSSS-CDMA system is explained followed by block-schematic of acquisition section of synchronization. In the next section, we develop the mathematical model for the probability of detection and probability of false alarm. In section V, we analyze the mathematical model and show the simulation result for verifying the theoretical results obtained through mathematical expression.

II. SYSTEM DESCRIPTION

A. DSSS-CDMA system

The system block schematic is shown in the fig. 1. This system is introduced in [13] which is based on the chaotic communication. DSSS-CDMA system uses sequences to present a bit. These sequences have to be orthogonal so that multiple users can share the same channel. Each member of the sequences is called a chip. The chips can be generated based on different functions [13]. In this study, we used Chebyshev maps, which has following function

$$X_{K+1} = 2X_K^2 - 1 \tag{1}$$

These sequences have the mean value equal to 05. In order to make comparison with the conventional binary sequences , we multiply them by $\sqrt{2}$. In addition, their probability density function is equal to

$$f_{C_i}(c_i) = \frac{1}{\pi\sqrt{2-c_i^2}}, \text{ for } -\sqrt{2} \le c_i \le \sqrt{2} \tag{2}$$

Moreover, the mean value will be equal one and also,

$$E\{c_i^4\} = \int_{-\sqrt{2}}^{\sqrt{2}} c_i^4 \frac{1}{\pi\sqrt{2-c_i^2}} dc_i = \frac{3}{2} \cdot \tag{3}$$

In fig. 1, the bit b_j^1 is the jth bit of the message which is sent by the user one. Based on the nature of DSSS-CDMA system, b_j^1 is spread by sequence of chips (c_i^1). It is then passed through a modulation and interleaver block. The effect of modulation and interleaver are studied in [14].

In the channel, noise and jammer are introduced to the signal and also signal is get affected by the delay. On the receiver side, the signal is demodulated and passed through deinterleaver. Next, it is multiplied by the chip sequences

generated locally by the sequence synchronization block. Then, it enters to the correlator. Finally, the decision making circuit constructs $b_j^{1\prime}$.

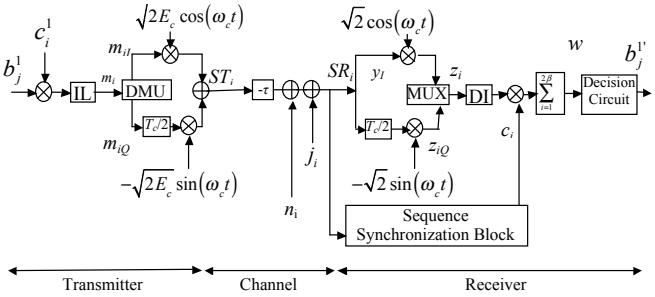


Fig. 1. Direct sequence spread spectrum system block schematic.

B. Jammer

In the present study, we assume that the jammer uses wideband Gaussian noise due to lack of information about the system's operating frequency. The modulation used in our system is a narrow band signal. The overlap of the narrowband signal and the wideband jammer is a narrowband jammer, which can describe as

$$j(k) = j_I(k)\sqrt{2E_j}\cdot\cos(\omega_c t) - j_Q(k)\sqrt{2E_j}\cdot\sqrt{2E_c}\cos(\omega_c t)\cdot \tag{4}$$

III. SYNCHRONIZATION

A. The importance of synchronization in DSSS-CDMA

Without the synchronization block, there would be a delay between received signal and the locally created chips.

In order to find the importance of the synchronization in the DSSS-CDMA system, we show the effect of the delay on probability of error rate. The output of the correlator in the case of delay is given by

$$W_{unsyn} = \sqrt{E_c}\sum_{i=1}^{2\beta} c_i^1 \cdot c_{(i-\tau)}^1 + \sqrt{E_N}\sum_{i=1}^{2\beta} n_i \cdot c_{(i-\tau)}^1 + \sqrt{E_J}\sum_{i=1}^{2\beta} j_i \cdot c_{(i-\tau)}^1 \cdot \tag{5}$$
$$= A + B + C$$

Where E_c is the energy of each chip, and E_N and E_J are energies of the noise jammer that are equal to N_0 and N_j respectively. Also, τ represents a delay in the system and 2β is the number of chips per bit.

The probability of error is equal to:

$$P_e(w) = \frac{1}{2}erfc\left(\frac{E[w]}{\sqrt{2\cdot \mathrm{var}[w]}}\right). \tag{6}$$

Its mean value is equal to

$$E^2\left[W_{unsyn}\right] = E[A+B+C] = E[A]+E[B]+E[C]. \tag{7}$$

So,

$$E[A] = E\left[\sqrt{E_c}\sum_{i=1}^{2\beta} c_{(i-\tau)}^1 c_i^1\right] = 0. \tag{8}$$

The autocorrelation of the chaotic sequences has impulse behavior. Thus it is maximum when τ is equal to zero, and for the rest of τ it is near to zero. More detail of the chaotic sequence properties are studied in [15].

Also,

$$E[B] = E[C] = 0. \tag{9}$$

Therefore,

$$E\left[W_{unsyn}\right] = 0. \tag{10}$$

by replacing (10) in (6), the probability of error becomes

$$p_{e(unsyn)} = \frac{1}{2}erfc(0) = 0.5. \tag{11}$$

Thus, if there is a delay between received signal and the locally generated sequences on the receiver side, the probability of error experiences the dramatic effect. In the following section, the synchronization process in DSSS-CDMA system is explained.

B. Schematic of Synchronization Block

In DSSS-CDMA system, the synchronization happens in two phases: acquisition and tracking [5].

In the acquisition phase, the delay will be removed with the resolution of a chip duration (*Tc*). Then, during the tracking phase, a fine tuning is done to eliminate the delay within a chirp duration.

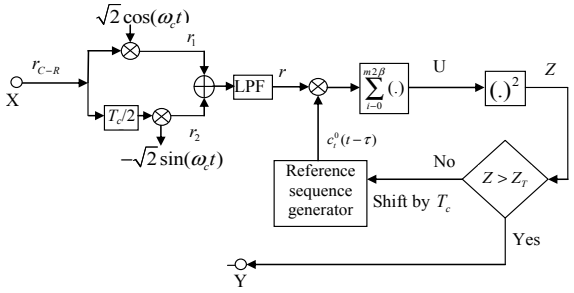


Fig .2. Acquisition section of synchronization block schematic.

In this paper, we assume that the delay is a multiple of *Tc*. Therefore, only acquisition phase is investigated here. The performance of the tracking part against jamming is the subject of our future studies.

As mentioned earlier, the synchronization is crucial for DSSS-CDMA system. In order to perform the synchronization, a synchronization signal is sent with the message signal. The synchronization signal always has a value equal to +1. Similar to the message, each

synchronization bit is spread with a sequence.

The acquisition section's block diagram is illustrated in the Fig. 2. In the acquisition section, the received signal is demodulated. Then, it is multiplied by the locally generated signal and passed through the correlator. The outcome of the correlator is given by

$$U = \sqrt{E_c}\sum_{i=1}^{2\beta} c_i^0 . c_{(i-\tau)}^0 + \sqrt{E_N}\sum_{i=1}^{2\beta} n_i \cdot c_{(i-\tau)}^0 + \sqrt{E_J}\sum_{i=1}^{2\beta} j_i \cdot c_{(i-\tau)}^0 . \quad (12)$$

where, c_i^0 presents the chip sequence of the synchronization signal. Please note that, the synchronization block also receives the message signal. However, due to orthogonality of message chips and synchronization chirp, the effect of the message signal in the synchronization becomes negligible.

In the next step, the correlator outcome is passed through the square law device. Similar to the autocorrelation function, the square law device has its maximum when τ is equal to zero. In other words, the synchronization signal and locally generated chips are aligned. Therefore, a threshold value (Z_T) is set (Z_T). If the square law device outcome becomes greater than Z_T, the synchronization is successful. Else, the locally generated chips are shifted for a chip and the process is repeated again utill the outcome of the square law device becomes greater than Z_T.

IV. MATHEMATICAL MODEL

In this section, the performance of DSSS-CDMA synchronization is modeled mathematically. Our performance analysis is based on the two key factor of the synchronization block: probability of detection and probability of false alarm.

Probability of detection is the probability in which the synchronization block detects the delay when the two signals are aligned. On the other hand, the probability of false alarm is the probability in which the synchronization detects the delay but two signals are not aligned. In this case, the outcome of the synchronization is an error.

A. The probability of the false alarm

A false alarm can happen when the square law device outcome become greater than the threshold value (Z_T). In order to find that, we calculate the probability density function of the square law device in the case of misalignment.

The correlator outcome, in the case of misalignment, can be expressed as

$$\begin{aligned} U_{unsyn} &= \sqrt{E_c}\sum_{i=1}^{2\beta} c_i^0 . c_{(i-\tau)}^0 + \sqrt{E_N}\sum_{i=1}^{2\beta} n_i \cdot c_{(i-\tau)}^0 + \sqrt{E_J}\sum_{i=1}^{2\beta} j_i \cdot c_{(i-\tau)}^0 \\ &= A + B + C. \end{aligned} \quad (13)$$

The mean value of (13) is given by

$$E\left[U_{unsyn}\right] = E[A+B+C] = 0, \quad (14)$$

Also, the variance of U_{syn} is obtained from

$$\begin{aligned} \sigma_{U_{unsyn}}^2 &= E[{U_{unsyn}}^2] - E^2\left[U_{unsyn}\right] = E[(A+B+C)^2] - E^2\left[U_{unsyn}\right] \\ &= E[A^2] + E[B^2] + E[C^2]. \end{aligned} \quad (15)$$

So,

$$\begin{aligned} E[A^2] &= E\left[\left(\sqrt{E_c}\sum_{i=0}^{2\beta}(c_{i-\tau}^0)(c_i^0)\right)^2\right] = E_c 2\beta E\left[\left(c_i^0\right)^2\right] E\left[\left(c_{i-\tau}^0\right)^2\right] \\ &\quad + E_c 2\beta(2\beta-1) E\left[\left(c_i^0\right)\right] E\left[\left(c_{i-\tau}^0\right)\right] E\left[\left(c_i^0\right)\right] E\left[\left(c_{i-\tau}^0\right)\right] \\ &= E_c 2\beta, \end{aligned} \quad (16)$$

$$E[B^2] = E\left[\left(\sqrt{E_N}\sum_{i=1}^{2\beta} n_i \cdot c_{(i-\tau)}^0\right)^2\right] = 2\beta N_0/2 \quad (17)$$

and

$$E[C^2] = E\left[\left(\sqrt{E_J}\sum_{i=1}^{2\beta} j_i \cdot c_{(i-\tau)}^0\right)^2\right] = 2\beta N_j/2 . \quad (18)$$

Thus,

$$\begin{aligned} \sigma_{U_{unsyn}}^2 &= E[A^2] + E[B^2] + E[C^2] \\ &= E_c 2\beta + 2\beta N_0/2 + 2\beta N_j/2 . \end{aligned} \quad (19)$$

So, U_{unsyn} can be presented as

$$U_{unsyn} \approx G\left(0, E_c 2\beta + 2\beta N_0/2_0 + 2\beta\varphi N_j/2\right). \quad (20)$$

The square law device outcome is equal to

$$Z = U^2 \quad (21)$$

Therefore, the PDF of the "*Z*" can be written by the chi-square distribution [16].

$$P_Z = \frac{1}{\sigma_{U_{unsyn}}\sqrt{2\pi Z}} \exp\left(-\frac{1}{2}(Z/\sigma_{U_{unsyn}}^2)\right) \quad (22)$$

As mentioned earlier, the false alarm happens when the system detects the delay by mistake. Therefore, the outcome of the square law device, in case of misalignment, should be greater than Z_T. Thus, probability of false alarm can be expressed as

$$P_F = \int_{z_T}^{\infty} \frac{1}{\sigma_{U_{unsyn}}\sqrt{2\pi Z}} \exp\left(-\frac{1}{2}(Z/\sigma_{U_{unsyn}}^2)\right) dZ \quad (23)$$

By replacing $\sqrt{Z} = x$, we have

$$P_F = \int_{\sqrt{z_T}}^{\infty} \frac{\sqrt{2}}{\sigma_{U_{unsyn}}\sqrt{\pi}} \exp\left(-\frac{1}{2}(x^2/\sigma^2_{U_{unsyn}})\right)dx$$
$$= \frac{\sqrt{2}}{\sigma_{U_{unsyn}}\sqrt{\pi}} \times \frac{\sqrt{2}\sqrt{\pi}\sigma_{U_{unsyn}}}{2} erf\left(\frac{x}{\sqrt{2\sigma^2_{U_{unsyn}}}}\right)\Bigg]_{\sqrt{z_T}}^{\infty} \quad (24)$$
$$= 1 - erf\left(\sqrt{\frac{z_T}{2\sigma^2_{U_{unsyn}}}}\right)$$

B. The probability of the false alarm

As mentioned before, detection happens when the outcome of the square law device, in case of the alignment, become greater than Z_T. Similar to the calculation of the probability of false alarm, , first, the distribution parameters of the correlator, in case of the aligned signals, is calculated. Then, the distribution of the square law device outcome is calculated. Finally, based on the probability density function of the square law device the probability of detection can be found.

Thus, the outcome of the correlator, in case of alignment, is given by

$$U_{syn} = \sqrt{E_c}\sum_{i=0}^{2\beta}(c_i^0)^2 + \sqrt{E_N}\sum_{i=1}^{2\beta} n_i \cdot c_i^0 + \sqrt{E_J}\sum_{i=1}^{2\beta} j_i \cdot c_i^0 \quad (25)$$
$$= D + F + G.$$

U_{syn} has the mean value equal to:

$$E\left[U_{syn}\right] = E\left[D + F + G\right] = E[D] + E[F] + E[G]. \quad (26)$$

Therefore,

$$E[D] = E\left[\sqrt{E_c}\sum_{i=0}^{2\beta}(c_i^0)^2\right] = \sqrt{E_c}\,2\beta, \quad (27)$$

and

$$E[G] = E[F] = 0 \quad (28)$$

Thus

$$E\left[U_{syn}\right] = \sqrt{E_c}\,m2\beta. \quad (29)$$

The variance of the U_{syn} is equal to

$$\sigma^2_{U_{syn}} = E[Z^2_{U_{syn}}] - E^2\left[U_{syn}\right] = E[(D+F+G)^2] - E^2\left[U_{syn}\right] \quad (30)$$
$$= E[D^2] + E[F^2] + E[G^2] - E^2\left[U_{syn}\right],$$

$$E[D^2] = E\left[\left(\sqrt{E_c}\sum_{i=0}^{2\beta}(c_i^0)^2\right)^2\right]$$
$$= E_c 2\beta E\left[\left(c_i^0\right)^4\right]$$
$$+ E_c 2\beta(2\beta-1)E\left[\left(c_i^0\right)^2\right]E\left[\left(c_i^0\right)^2\right] \quad (31)$$
$$= E_c 4\beta^2 + E_c\beta,$$

$$E[F^2] = E\left[\left(\sqrt{E_N}\sum_{i=1}^{2\beta} n_i \cdot c_i^0\right)^2\right] = 2\beta N_0/2 \quad (32)$$

and

$$E[G^2] = E\left[\left(\sqrt{E_J}\sum_{i=1}^{2\beta} j_i \cdot c_i^0\right)^2\right] = 2\beta N_j/2. \quad (33)$$

Thus,

$$\sigma^2_{U_{syn}} = E[D^2] + E[F^2] + E[G^2] - E^2\left[U_{syn}\right],$$
$$= E_c 4\beta^2 + E_c\beta + 2\beta N_0/2 + 2\beta N_j/2 - E_c 4\beta^2 \quad (34)$$
$$= E_c\beta + 2\beta N_0/2 + 2\beta N_j/2$$

Therefore "U" can be described as

$$U_{syn} \approx G\left(\sqrt{E_c}\,2\beta\ ,\ E_c\beta + 2\beta N_0/2 + 2\beta N_j/2\right) \quad (35)$$

Same as before, $Z=U^2$. Therefore, PDF of "Z" can be express as a chi-square distribution [16]

$$p_Z = \left(\frac{1}{\sigma_{U_{syn}}\sqrt{2\pi Z}} \exp\left(-\left(Z+\lambda/\sigma^2_{U_{syn}}{}^2\right)/2\right)\right), \quad (36)$$
$$\left(\times\cosh\left(\sqrt{Z/\sigma^4_{U_{syn}}\lambda}\right)\right)$$

where

$$\lambda = E\left[U_{syn}\right] = E_c 4\beta^2. \quad (37)$$

The probability of detection can be express as:

$$P(D)=\int_{z_T}^{\infty}\left(\frac{1}{\sigma_{U_{syn}}\sqrt{2\pi Z}}\exp\left(-\frac{\left(Z/\sigma^2_{U_{syn}}+\lambda/\sigma^2_{U_{syn}}\right)}{2}\right)\times\cosh\left(\sqrt{Z/\sigma^4_{U_{syn}}\lambda}\right)dZ\right) \quad (38)$$

Same as before, by using $\sqrt{Z}=x$, we have

$$p_D=\frac{\sqrt{2}}{\sigma 2\sqrt{\pi}}\int_{z_T}^{\infty}\exp\left(-\frac{\left(x^2+\lambda\right)}{2\sigma^2_{U_{syn}}}\right)\left(\exp\left(\sqrt{\lambda/\sigma^4_{U_{syn}}}x\right)+\exp\left(-\sqrt{\lambda/\sigma^4_{U_{syn}}}x\right)\right)dx \quad (39)$$

$$\begin{aligned}p_D&=\frac{\sqrt{2}}{\sigma_{U_{syn}}2\sqrt{\pi}}\int_{z_T}^{\infty}\exp\left(-\frac{\left(x^2+\lambda\right)}{2\sigma^2_{U_{syn}}}\right)\exp\left(\sqrt{\lambda/\sigma^4_{U_{syn}}}x\right)dx\\&+\int_{z_T}^{\infty}\exp\left(-\frac{\left(x^2+\lambda\right)}{2\sigma^2_{U_{syn}}}\right)\exp\left(-\sqrt{\lambda/\sigma^4_{U_{syn}}}x\right)dx\\&=K+L\end{aligned} \quad (40)$$

$$\begin{aligned}K&=\frac{\sqrt{2}}{\sigma_{U_{syn}}2\sqrt{\pi}}\int_{z_T}^{\infty}\exp\left(-\frac{\left(x^2+\lambda\right)}{2\sigma^2_{U_{syn}}}\right)\exp\left(\sqrt{\lambda/\sigma^4_{U_{syn}}}x\right)dx\\&=\frac{\sqrt{2}}{\sigma_{U_{syn}}2\sqrt{\pi}}\int_{z_T}^{\infty}\exp\left(-\frac{x^2-2\sqrt{\lambda}x+\lambda}{2\sigma^2_{U_{syn}}}\right)dx\\&=\frac{\sqrt{2}}{\sigma_{U_{syn}}2\sqrt{\pi}}\int_{z_T}^{\infty}\exp-\left(x/\sqrt{2}\sigma_{U_{syn}}-\sqrt{\lambda}/\sqrt{2}\sigma_{U_{syn}}\right)^2dx\\&=\frac{\sqrt{2}}{\sigma_{U_{syn}}2\sqrt{\pi}}\frac{\sqrt{\pi}\sqrt{2}\sigma_{U_{syn}}}{2}erf\left(x/\sqrt{2}\sigma_{U_{syn}}-\sqrt{\lambda}/\sqrt{2}\sigma_{U_{syn}}\right)\Big]_{Z_T}^{\infty}\\&\frac{1}{2}erf\left(\sqrt{Z/2\sigma^2_{U_{syn}}}-\sqrt{\lambda/2\sigma^2_{U_{syn}}}\right)\Big]_{Z_T}^{\infty}\\&=\frac{1}{2}\left(1-erf\left(\sqrt{Z_T/2\sigma^2_{U_{syn}}}-\sqrt{\lambda/2\sigma^2_{U_{syn}}}\right)\right)\end{aligned} \quad (41)$$

and

$$L=\frac{1}{2}\left(1-erf\left(\sqrt{Z_T/2\sigma^2_{U_{syn}}}+\sqrt{\lambda/2\sigma^2_{U_{syn}}}\right)\right). \quad (42)$$

Therefore,

$$\begin{aligned}p_D&=K+L\\&=1-\frac{1}{2}\left(erf\left(\sqrt{Z_T/2\sigma^2_{U_{syn}}}-\sqrt{\lambda/2\sigma^2_{U_{syn}}}\right)+erf\left(\sqrt{Z_T/2\sigma^2_{U_{syn}}}+\sqrt{\lambda/2\sigma^2_{U_{syn}}}\right)\right)\end{aligned} \quad (43)$$

V. Result and Discussion

In previous section, we derived the mathematical expression of the probability detection of false alarm under jamming attack. Here, we analyze our mathematical expressions in different scenarios. In all of these analytic scenarios, we set number of chips per bit (2β) equal to 300. Also, we normalize the power of each bit (E_b=1).

Fig. 3 shows the ROC plot for jammer with different power. To focus on the effects of the jammers, in this scenario, we assume that there is no environment noise i.e SNR=inf. As can be seen in Fig. 3, even a jammer, which can cause SJR=5 dB, can have a dramatic effect on the system's performance. In this scenario, we set Z_T to 200.

In addition, to evaluate our mathematical expressions we run Monte Carlo simulation on Matlab. The results are shown in the green straight lines, and the theoretical outcomes are shown by blue dashed lines. As can be seen from the figure, the simulation results match with the theory. This simulation is run for 10^4 times.

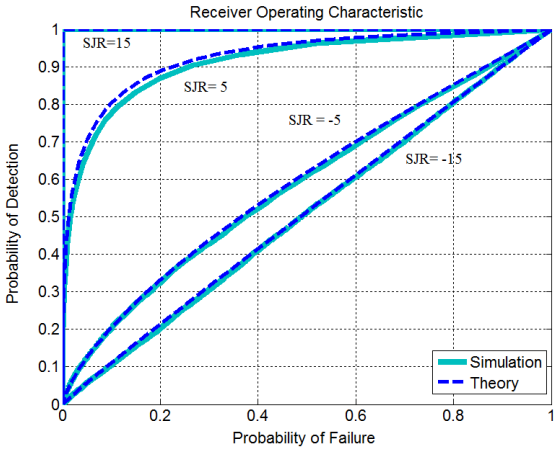


Fig. 3. ROC curves in the case of jammer with deferent power based on the theory and simulation.

(43) and (24) show that the probability of detection and false alarm are sensitive to the choice of Z_T. Therefore, to optimize the system performance, it is necessary to find Z_T in which, the probability detection become maximum and the false alarm become minimum. Fig. 4 demonstrates the probability of detection and false alarm based on Z_T in the different SJR. The straight blue line and dotted blue line are probabilities of detection and false alarm in case of no existing jammer (SJR=inf). In this case, Z_T can be chosen smaller than 200. As the jammer power increases, the probability of detection decreases and the probability of false alarm increases. However, the probability of false alarm seems to get

more affected than probability of detection. For SJR=-10 dB, the probability of false alarm is getting close to the probability of detection.

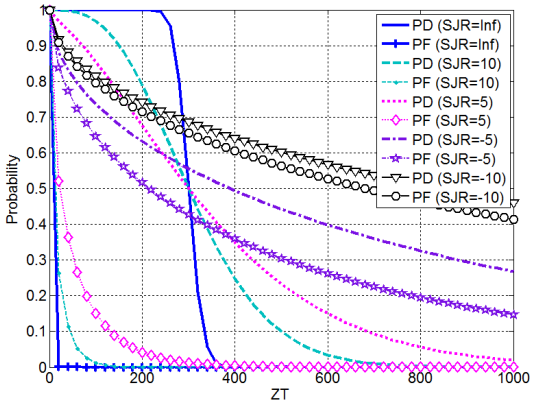


Fig. 4. Probability of detection and false alarm based on Z_T for deferent jammer powers.

Fig. 5 presents the probability of detection and false alarm based on the jammer power (SJR) when Z_T is equal to 200, 300 and 400. As shown in the figure, for Z_T=200, the probability of detection is better compared to the others. However, it has the worse probability of false alarm. Consequently, for Z_T=400 has the best probability of false alarm and worse probability of detection. Also, the probability of false alarm rises after SJR=10 dB. After this point, the communication system starts to experience a noticeable effect on its error rate.

As can be seen in Fig. 5, in Area I, when the jammer power increases, the probability of detection is decreased, and the probability of false alarm is increased. By decreasing Z_T, we can increase the chance of detection; however, the probability of false alarm is increased as well which is not desirable.

On the other hand, both probability of detection and false alarm is increasing by jammer power in the area II. In fact, in that area, the probability of false alarm and detection become equal due to the high level of the jamming signal.

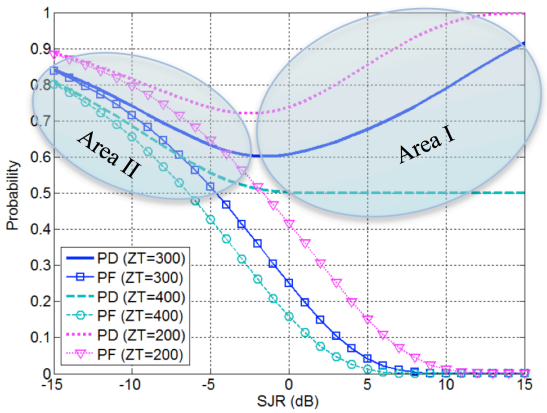


Fig. 5. Probability of detection and false alarm based on SJR.

VI. Conclusion

In this paper, we investigate the effects of jammer on DSSS-CDMA system using chaotic sequences. We specifically study the effect jamming attack on accusation section of the synchronization block. We show that the synchronization in DSSS-CDMA system is quite vulnerable to the jamming attacks. In order to consider the effect of these jammers on synchronization block, mathematical expressions of the probability of detection and false alarm are developed in closed form. Also, we run a simulation based on Monte Carlo method to evaluate our mathematical expressions.

Our result also shows how the vulnerability of synchronization block to jamming signals can cause tragic effects on the system probability of error. Our investigation reveals that even low power jammer (with SJR=10 dB in our scenario) can make the whole system collapse.

References

[1] A. Tayebi, S. Berber, and A. Swain, "Department of Electrical and Computer Engineering, University of Auckland, Auckland, New Zealand," in *Sensing Technology (ICST), 2013 Seventh International Conference on*, 2013, pp. 97-102.

[2] D. Fudenberg and J. Tirole, "A" signal-jamming" theory of predation," *The RAND Journal of Economics,* pp. 366-376, 1986.

[3] W. Xu, K. Ma, W. Trappe, and Y. Zhang, "Jamming sensor networks: attack and defense strategies," *Network, IEEE,* vol. 20, pp. 41-47, 2006.

[4] G. S. Sandhu and S. M. Berber, "Investigation on operations of a secure communication system based on the chaotic section shift keying scheme," in *Information Technology and Applications, 2005. ICITA 2005. Third International Conference on*, 2005, pp. 584-587 vol.2.

[5] S. M. Berber and B. Jovic, "Sequence synchronization in a wideband CDMA system," 2007.

[6] R. Vali, S. M. Berber, and S. K. Nguang, "Effect of Rayleigh fading on non-coherent sequence synchronization for multi-user chaos based DS-CDMA," *Signal Processing,* vol. 90, pp. 1924-1939, 2010.

[7] B. Jovic, C. Unsworth, G. S. Sandhu, and S. M. Berber, "A robust sequence synchronization unit for multi-user DS-CDMA chaos-based communication systems," *Signal Processing,* vol. 87, pp. 1692-1708, 2007.

[8] R. Vali, S. M. Berber, and N. Sing Kiong, "Analysis of Chaos-Based Code Tracking Using Chaotic Correlation Statistics," *Circuits and Systems I: Regular Papers, IEEE Transactions on,* vol. 59, pp. 796-805, 2012.

[9] C. Yuan, H. Fei, Y. Jian, C. Xiang, and G. Yuantao, "A smart tracking-based jamming scheme for signals with periodic synchronization sequences," in *Wireless Communications and Signal Processing (WCSP), 2011 International Conference on*, 2011, pp. 1-5.

[10] M. J. La Pan, T. C. Clancy, and R. W. McGwier, "An Assessment of OFDM Carrier Frequency Offset Synchronization Security for 4G Systems," in *Military Communications Conference (MILCOM), 2014 IEEE*, 2014, pp. 473-478.

[11] M. J. La Pan, T. C. Clancy, and R. W. McGwier, "Section warping and differential scrambling attacks against OFDM frequency synchronization," in *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*, 2013, pp. 2886-2890.

[12] C. Shahriar, S. Sodagari, and T. C. Clancy, "Performance of pilot jamming on MIMO channels with imperfect synchronization," in *Communications (ICC), 2012 IEEE International Conference on*, 2012, pp. 898-902.

[13] S. Berber and S. Feng, "Theoretical Modeling and Simulation of a Chaos-Based Physical Layer for WSNs."

ISBN: 978-1-61804-290-3

[14] S. M. Berber, "Fading mitigation in an interleaved noise-based DS-CDMA system for secure communications," in *Proceedings of the Fifth IASTED International Conference on Signal Processing, Pattern Recognition and Applications*, 2008, pp. 260-265.
[15] G. S. Sandhu and S. M. Berber, "Investigation on orthogonal signals for secure transmission in multiuser communication," 2007.
[16] H. O. Lancaster and E. Seneta, *Chi-Square Distribution*: Wiley Online Library, 1969.

Arash Tayebi received the B.S degree in electrical engineering from Shiraz University, Shiraz, Iran in 2009, and Master degree in telecommunication engineering from University of Melbourne, Melbourne, Australia in 2011. He is currently working toward the PhD degree in electrical engineering at University of Auckland, Auckland, New Zealand. His research interests include chaos-based communication systems, physical layer security, and CDMA and OFDM communication. He is student member of IEEE.

Stevan Mirko Berber was born in Stanisic, Serbia, former Yugoslavia. He completed his undergraduate studies in electrical engineering in Zagreb, master studies in Belgrade, and PhD studies in Auckland, New Zealand. Currently Stevan is with the Department of Electrical and Computer Engineering at Auckland University, New Zealand. He was appointed Visiting Professor at the University of Novi Sad in 2004 and Visiting Scholar at the University of Sydney in 2008. His teaching interests are in communication systems, information and coding theory, discrete stochastic signal processing and wireless sensor and computer networks. His research interests are in the fields of digital communication systems and signal processing with the emphasis on applications in CDMA systems and wireless computer, communication and sensor networks. He is the author of more than 80 refereed journal and conference papers, 8 books and three book chapters. Dr Berber is a referee for papers in leading journals and conferences in his research area. He has been leading or working on a large number of research and industry projects. Dr. Berber is a senior member of IEEE, a member of New Zealand Scientists, and an accredited NAATI translator for English language.

Akshya Swain received B.Sc. Engineering degree in Electrical Engineering and M.Sc. Engineering degree in Electronic Systems and Communication from Sambalpur University, India, in 1985 and 1988, respectively. From 1994 to 1996 he was Commonwealth Scholar in the United Kingdom and obtained Ph.D. degree from the Department of Automatic Control and Systems Engineering at the University of Sheffield in 1997. From 1986 to 2002, he worked as Lecturer, Assistant Professor and Professor of Electrical Engineering at the National Institute of Technology, Rourkela, India. During 1988/1989 he was Assistant Director in the Ministry of Energy for the Indian government. He joined the Department of Electrical and Computer Engineering at The University of Auckland in September 2002. His research interests include nonlinear system identification & control, biomedical signal processing, sensor networks and control applications to power system and inductive power transfer systems. Currently he acts as a member of the Editorial Board of the International Journal of Automation and Control and International Journal of Sensors, Wireless Communications and Control.

Agent simulator-based control architecture for rapid development of multi-robot systems

Ismael Fabricio Chaile and Lluís Ribas-Xirgo

***Abstract*— Development of high complexity supervisory control and data acquisition (SCADA) software for multi-robot systems in manufacturing plants or logistic facilities is a long, difficult process. In order to relieve part of this overburden, software is organized in different application aspects and abstraction entities. Particularly, it can be divided into application-specific and internal logistics and transportation aspects. Furthermore, each aspect can be resolved by sets of agents. In this paper, we focus on the internal transportation aspect and propose a control architecture that allows rapid development of SCADA software by re-using system simulation in the final products. By using this construction model, we have built a prototype of a transport system for an automated laboratory as an example for future developments. Results show that it is possible to concurrently run the actual system and its simulation, which is used to update system model state and to execute the higher abstraction level parts of the controller. By doing so, the time to prototype and to final deployment can be drastically reduced.**

***Keywords*— Agent-based controllers, mobile robots for internal transportation, multi-robot controller architecture, robot systems.**

I. INTRODUCTION

DEVELOPMENT of SCADA applications is a highly complex process, particularly for fault-tolerant flexible manufacturing systems, where internal transportation plays an important role. In this paper, we shall focus on this aspect [1] and, particularly, on the data representation and control software. Noteworthy, monitoring and control are key parts of any SCADA and transportation problem is relatively transversal because mobile robots are becoming commonplace in many modern industrial facilities and share common problems.

Mobile robots require minimal infrastructure to do the transportation and enable the system to be flexible and robust, as paths can be dynamically altered on the events (machinery temporarily out of order, obstacles, bottlenecks, et cetera) along the way.

SCADA systems are used to monitor mobile robot teams' states and control their operations to some extension, depending on their degree of autonomy. For instance, they can control internal transportation in manufacturing plants, warehouses or train systems by taking orders from application-specific components (i.e. interfaces with e.g. production planners, management systems or transport authorities) and transforming them into requests for vehicles in accordance to system state, which includes individual vehicle information.

However, the autonomous control of teams of such robots is very complex because of the number of possible situations and of the interaction with other peer-robots and elements of the system. Therefore, human intervention is often required to resolve conflicts due to unplanned robot systems' states.

Automated laboratories are particular instances of manufacturing plants that produce sample analyses' reports. Sample distribution is focused on throughput and controlled by laboratory information management systems (LIMSs), a kind of SCADA systems. As for the transportation, the task of a LIMS is to plan how samples are picked at collection spots and delivered to several analysis machines before being carried to some drop zone. While planning can be statically determined when all elements of the plant and the samples are known beforehand, plan execution must face a number of events that require re-planning: priority samples added on-the-fly, test repetition because of erroneous data, machines temporarily out-of-order, traffic incidences, et cetera. All of these situations create problems quite common in other multi-robot environments, and solutions can be shared.

The goal of this work is to relieve part of the complexity of SCADA development through re-use, abstraction and hierarchy, which are well supported by agents and agent technology [2].

In fact, there exist a number of agent-based solutions to develop SCADA for distributed embedded systems, including multi-robot systems (Section 2). The methodology begins by building an agent-based model (ABM) of the system and, after validation, move on to implementation by progressively transforming the multi-agent model into software for the actual machines (computers, robots, local controllers, et cetera). In our work, however, the idea is to take profit of as much as possible of the ABM specification to minimize the

This work was supported in part by a PhD Scholarship grant from the Universitat Autònoma de Barcelona (UAB).

I. F. Chaile was with the Department of Microelectronics and Electronic Systems, Universitat Autònoma de Barcelona, Barcelona, 08193, Spain. He is now with K-LAGAN company (e-mail: ChaileIF@gmail.com).

Ll. Ribas-Xirgo is with the Department of Microelectronics and Electronic Systems, Universitat Autònoma de Barcelona, Barcelona, 08193, Spain. (phone: +34-935811078; fax: +34-935813033; e-mail: Lluis.Ribas@uab.cat).

amount of code to be further transformed to reach an implementable version, i.e. to be embedded into specific elements.

The proposed ABM (Section 3) organizes agents into two classes: the ones that act as interfaces with the rest of the systems and the ones that have to do with the transportation aspect. The software architecture [3] of this last class of agents consists of a set of three layers of which only the lower one has to be embedded into the mobile robots while the other two can be kept in the original form, in the ABM.

The higher levels of the transportation agents are run in the model simulation, together with the lowest level, which acts as a physical system state estimator. The top level is the one responsible for agent's decisions and interaction, and uses the low level to estimate vehicle status. If this low level is replaced by real robots, then, the simulation of the ABM controls the real system. Our solution, though, does not replace the lowest layer but introduce an intermediate level to synchronize the simulation of the lowest layer with the reality (Section 4). By doing so, ABM simulation can be used as human-machine interface (HMI) and controllers can estimate next states in advance, if required.

The result design environment (Section 5) is quite simple, as it takes only a software able to develop and simulate ABMs, and a methodology which is supported by software templates and a synchronizer.

The methodology has been applied to develop the ABM for the transportation of samples in an automated laboratory prototype [4]. Results show that it is possible to develop effective multi-robot controllers that are partially run on their ABM simulators (Section 6).

II. Agent-based Models and Controllers

Analyses of social behavior of individuals and how it is affected by changes on individual behavior [5] are easy to perform with ABM simulations. Conversely, any kind of group behavior can be validated against some requirements, particularly that of multi-robot systems. Following this line of thought, ABMs can be taken as system's models and used to control them by generating the commands to the individuals so that they behave as required by the corresponding applications.

A. ABMs as System Simulators

In transport logistics, as in other domains, agent-based modeling is used mainly to support decision taking [6] but not to automate processes, i.e. not as system controllers. In fact, agents can be used to distribute the problem into its participants, which collaborate to solve their local problems.

For instance, agents can model the elements involved in supply chain [7] and, hence, help to manage them. Or they can be used to model a carpooling application [8], where passengers can share cars that move autonomously in a network with independent traffic lights and local conflict solving at intersections.

Both examples above can be extended to the transportation field, as supply chain planning imply sequencing transport orders within a given set of time-varying constraints, and carpooling can be an option when carriers can take several loads, as they should be grouped so that each group may follow the same minimal route.

B. ABMs as System Controllers

Most applications require their controllers dealing with dynamically changing demands, and those with transportation systems are not an exception. Autonomic systems [9] relieve part of the application design effort.

In autonomic systems, components tell others what they want and not how to attain the corresponding goals. Following this principle, the transportation systems can be divided into two parts: the one of the carriers (or AGVs) and the one for the application, which tells the first one what is needed but not how it must be fulfilled. In other words, transport order tasks are allocated in a distributed fashion [10].

Furthermore, provided that ABMs can be simulated and that these simulations can be run concurrently with physical agents, the difference between expected and sensed behavior can be minimized via an additional controlling level. This controller level on top of the others can be system-wide [9] or local to each agent. The first approach depends on a single module or a coordinating agent for analyzing the dynamics of the system and tuning other agents' operations by sending appropriate messages thus achieving better cost-effective performances. The last one does the same locally thus not guaranteeing global optimality but minimizing inter-agent communication and maximizing agent independency, which can, in the end, result in obtaining even better yields from systems.

However, the main problem to use ABM simulation as a controller is that ABM must run in real time with the physical requirements of the system and its application.

C. ABMs for Traffic and Transport Control

Systems of agents have already been used or proposed to manage traffic in urban areas [11]–[14], in warehouses [15]–[17], and in many other applications. Particularly, they have been used for controlling AGV-based systems [18]–[22].

The general idea is to have a traffic system that can be self-regulated from individual choices and that requires as little assistance as possible from agents at a higher level of hierarchy. In other words, the transport orders from applications are handled by transportation agents in an autonomous manner, with minimal information from other agents, including those who may act as planners and routers.

Our approach offers a complete development environment for agent-based systems, as in [23], and uses an ABM of the transport system that accepts inputs from the rest of the system and outputs control data for the physical transportation units as well as other data to the system. Differently from their proposal and other works alike, our approach uses a single ABM tool to simplify the development framework and minimize the development costs.

ISBN: 978-1-61804-290-3

As previously introduced, the idea is that the ABM simulation is run concurrently with the on-board vehicle controllers to relieve code transformation to the part that has to be embedded. The next section shall be devoted to detail the transport agent architecture that makes this strategy possible and to explain the resulting controller architecture.

III. ABM of a Transport System Controller

The diagram of such he diagram of a multi-agent system (MAS) for a controller like this is shown in Fig. 1. Agents are divided into two classes, namely the application-specific and the transportation one. The former are agents ($\{A_i\}$) that link the latter ones ($\{B_i\}$) with the rest of the elements in the application. To name a few, $\{A_i\}$ include the human-machine interface (HMI), the ones to interact with remote terminal units (RTUs) in manufacturing plants or the interface with the enterprise resource planner (ERP). Particularly, there must be an application-specific agent able to supply transport orders to transportation agents.

Transportation agents $\{B_i\}$ correspond to automated-guided vehicles (AGVs). Therefore, agent-based controllers of AGVs are organized such that each transportation agent can control an AGV and, at the same time, interact with the rest of the agents in the system.

Besides $\{A_j\}$ and $\{B_i\}$, our proposed ABM relies on other resources to be run, namely agent communication language (ACL) services, intra-agent communication services, and a physical plant simulator (P).

Additionally, there can be an agent to help controlling the traffic in cases where distributed, *heterarchical* decisions are not enough. Note that, most of the time, AGVs can move around with only local information and, eventually, they have to be solving conflicts with others, thus creating temporary hierarchies among them. In some scenarios, particularly in high-density traffic networks, traffic coordinator agents would minimize inter-agent communication to solve conflicts and the number of conflicts (for examples on algorithms these agents should include, see [24] and [25].)

A. Transportation Agent Architecture

Transportation agents are those that control vehicle sensors and actuators, as well as communicate with other agents. While the former operations require be done in real time, the latter can be performed under less stringent requirements, particularly with lengthier time periods than the first ones. Taking profit of this fact, they are split into two layers: the lowest one is in charge of controlling the vehicle in accordance with the requests from the topmost one. The top layer is the one capable of communicating with other agents and, therefore, of taking into consideration system state when taking decisions on accepting and completing transport orders.

The resulting architecture for a transportation agent or *taxi* is shown in Fig. 2, where an additional, intermediate layer has been added. This interface layer includes all services to communicate the top layer with the bottom one in a safe and secure way. Basically, it includes procedures to send requests to the lowest level layer and to get answers from it.

With this organization, the high level layer or L1 is detached from the low level layer or L0. Consequently, implementation of L1 and L0 are independent, provided that they share the communication language. Note that even though L1 and L0 can use the same language and services as if they were different agents they are not.

Therefore, any taxi B_i is divided into two levels, L1 or T_i, and L0, which is either virtual (V_i) or real (R_i). In our approach, though, both L0s can co-exist and the interface layer synchronizing L1 and L0s when they do.

The main advantage of this option is that simulation and control is done concurrently, with simulation helping to maintain a symbolic view of the system for all agents and to foresee results of individual choices.

The virtual representation of the system includes the state of the plant as well as the state of the L0s of taxis. In fact, all $\{V_i\}$ interact with a plant environment simulator P and, as a result, all $\{T_i\}$ have access to system-wide information without communicating with other agents. For instance, they know the symbolic position of any other taxi to determine

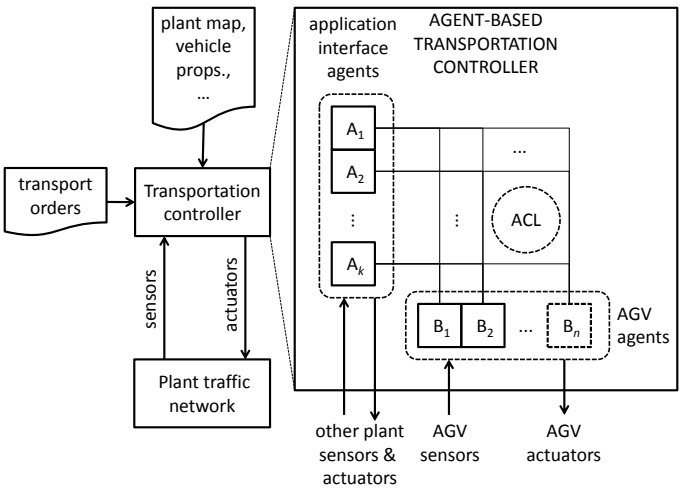


Fig. 1. Agent-based model of a multi-robot controller

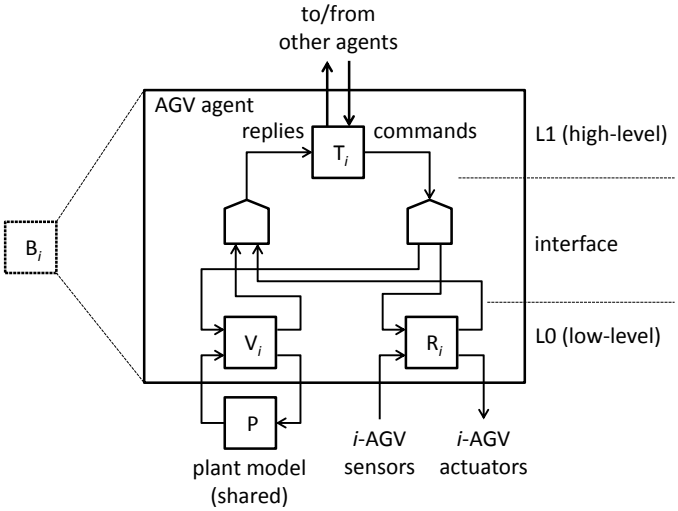


Fig. 2. Software architecture of transportation agents

collision-free routes or to solve conflicts at crossings. The *symbolic position* is the position in the traffic network as represented in the simulated plant environment, which has to be accurate enough for the application, even though it is not close enough to obtain a realistic representation on a screen.

Each R_i is "controlled" by commands from T_i, which stands for the topmost controller level of taxi B_i. Commands depend on replies from R_i, but also on differences between V_i and R_i replies, which are monitored and controlled by the synchronizing interface, on messages from other $T_{j\neq i}$ and A_k, and on global information stored in P.

The former vehicle controlling scheme is the core of the architecture of system's controller, which is explained next.

B. Transportation System Controller Model

Fig. 3 illustrates the result control loop with our ABM. Note that the topology of the plant and the number of taxis are among the variables that configure the model that is used for controlling the real plant.

The model is run under inputs that come from external agents and physical elements and generates outputs for the latter ones. This control loop might be too slow for many applications unless physical elements have embedded some controllers and relation with the ABM is done at a higher level of abstraction. However, even with this solution, ABM has to be executed fast enough to interact at real time with the physical elements. This requires agents to be efficient in taking decisions, which usually goes against reflexive, elaborated behaviors and to have simple communication protocols that enable negotiations to occur within a few messages.

The higher level modules of the taxis ($\{T_i\}$) get orders from agents that represent other modules of the application ($\{A_j\}$) and try to fulfill them.

The variability of incoming orders increases the complexity of a central planner and/or a traffic coordinator thus making it difficult to attain any gain in cost or throughput. Consequently, the taxis operate autonomously, without extra coordinating agents. Although this work mode has less guarantee of optimality, it has advantages with respect to flexibility and robustness.

To fulfill the transport orders, taxis must negotiate with application agents $\{A_j\}$ and other workmates which jobs they take and, when in transit, how can they be done in the more efficient way.

In taking the decisions, taxis have knowledge of their own state and the state of their lower-level counterparts ($\{V_i\}$). Results of deliberations are transformed into requests to the $\{V_i\}$ and also to the real robots $\{R_i\}$. The last set of requests is, in fact, the output of the ABM controller. And the inputs include the replies to these requests from robots, hence closing the loop between the controller and the controlled system.

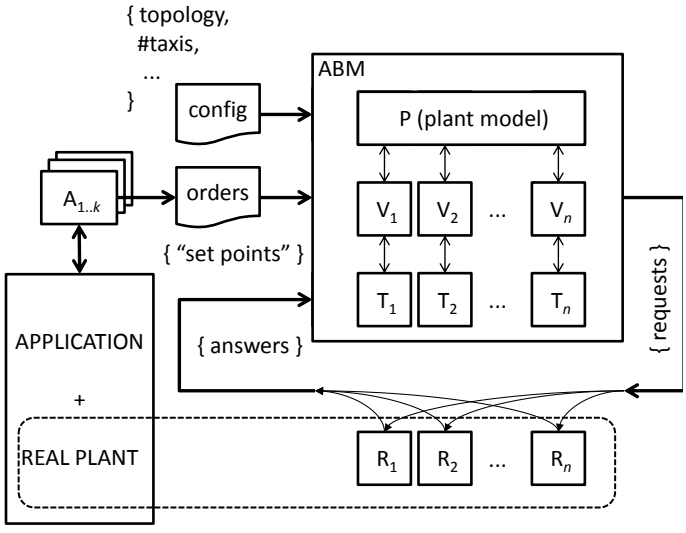


Fig. 3. Controller scheme

IV. Synchronization of Simulation with Reality

As previously stated, ABM simulation can be used to monitor, supervise and control transport systems of applications. For the first part, it is important that simulated reality do not differ from actual reality, i.e. that the expected system state matches timely with the measured one.

At the L1, taxis have a topological map of the traffic network that consists of a graph with annotated nodes and arcs. Annotations include time to perform actions at nodes and average speed for arcs. In fact, these data come from $\{V_i\}$ simulation so that expected behavior at L1 matches simulated reality. The problem arises when simulation has to run concurrently with actual reality.

L1 parts of taxis $\{T_i\}$ record delays between requests and acknowledgements from corresponding L0. In case there is exclusively a R_i, or a V_i, responses from L0 are automatically sent to L1. Otherwise, the interface layer routines will first synchronize $\{V_i\}$ with corresponding $\{R_i\}$.

To synchronize simulation with reality for the *i*-th taxi means to force taxi simulation to keep up with messages from reality occurring ahead of what it is expected, or to suspend it while waiting for them. In other words, a synchronized simulation requires that messages from V_i and R_i occur at the same time (or at the same control cycle.)

Synchronization mechanism deals with *events*, i.e. time-tagged messages. There are two classes of events: the ones caused by requests from T_i and the ones that are generated directly from L0. The former have to be synchronized in hard-real time while the latter allow some mismatch between reality and simulation, thus being soft real-time.

Hard real-time synchronization events (HSEs) include requests emitted by $\{T_i\}$ and the corresponding replies from $\{(V, R)_i\}$, which are expected to occur at the same time.

Soft real-time synchronization events (SSEs) include informative messages from $\{(V, R)_i\}$, which refer to the occurrence of conditions that are autonomously managed by the L0. For instance, detecting an obstacle or running low in battery are situations that L0s handle locally and that do not require immediate attention by corresponding L1s. However, it is important that L0 counterpart run synchronized to keep

ISBN: 978-1-61804-290-3

virtual representation more accurate to reality and, in case of mixed-reality operation, to have reality working together with simulated-only taxis.

A. Synchronization Mechanism

The synchronization mechanism is internal to each taxi and the rest of the section focuses on a single one. Events are denoted *h* or *s*, depending on them being HSE or SSE. We shall use subscripts to indicate the source and a star in superscript when they correspond to synchronizing error events. Subscripts can be *T*, *S*, *R* or *V* to mean that events come from L1, *synchronizer*, real AGV or simulated AGV, respectively. For example, h_T corresponds to a HSE from L1, h_S^*, to a HSE from the synchronizer with reporting some error, and s_R and s_V, to SSE from real and simulated AGVs, respectively.

Note that, for every pair of events from L0 the synchronizer must produce an equivalent, outgoing event, i.e. for any pair of h_V and h_R in response to an h_S caused by some h_T there should be an h_S to L1. If something fails, then appropriate synchronization error events, h_S^*, are sent to L1 and, eventually, to L0.

A typical, error-free communication protocol (Fig. 4.*a*) starts by an h_T request, which is sent to the synchronizer, which, at its turn, sends the request (h_S) to L0. After that, the synchronizer waits for h_V and h_R to happen. At this state, when they both occur at the same control cycle and corresponding messages are equal, the synchronizer sends the acknowledgement (h_S) to L1, with the message contents from L0. If messages are not the same, the synchronizer emits an h_S* error event to L1 and L0.

If L0 events do not happen simultaneously (i.e. at the same *instant*), the synchronizer either waits for h_R or causes the simulation to catch up to reality, i.e. causes to h_V happen.

The previous situations (see Fig. 4.*b*) activate either the *event discovery method* (EDM) or the *immediately synchronization method* (ISM) of the synchronizer.

Obstacle detections and other SSEs happen at L0 and help synchronizing virtual representation and reality, though they admit some mismatch. In this case, the synchronizer tries to pair each s_R with the corresponding s_V with a similar strategy than with h_V and h_R, however, it allows that there is a tolerance in the time s_R and s_V occur. Again, when messages are equal s_S is sent to L1 and s_S* with error notification to L1 and to silent L0 part when they are not. The sending of s_S* to V_i if no s_V has occurred allows V_i to perceive unexpected reality events and the other way round, i.e. sending s_S* to R_i if no s_R has occurred enables R_i perceiving simulation-only stimuli.

To keep the synchronizer independent from L1 and L0, events' data frames include some parameters to help it solve problems when events do not match in time or message contents. These parameters are explained below.

T_{out} is a timeout for h_R with respect to an h_S caused by h_T. If h_R does not occur within this period of time, an error event h_S* is sent to both L1 and L0. In this case, h_S* has a message that contains the *timeout message* from the initial h_T. In other words, is a timeout for the EDM.

I_{max} is the maximum number of allowed ABM runs to cause the simulation to fire a h_V corresponding to a previous h_R. It is a limit for the ISM. When ISM fails, i.e. the synchronizer does not receive the h_V before I_{max}, a h_S* with a predefined message is sent to both levels.

SSEs are allowed to occur in different time instants within a period shorter than T_{tol}, thus T_{tol} is a tolerance time threshold before starting an EDM for s_R.

T_{outs} is a timeout for SSEs from reality, just as T_{out} for h_R.

Note that, in a similar way than with HSEs, the synchronizer starts an ISM to generate a s_V for any unmatched s_R, and that these mismatches cause s_S* that can be used to appropriately update the virtual representation of the system.

V. Development Methodology and Platform

The process that leads to multi-robot system controllers begins with building ABMs. In our case, designers must follow our model architecture, which implies programming interface agents $\{A_i\}$, taxis $\{B_i\}$ and a plant simulator *P* according to model templates.

Initial ABMs can be used for functional validation and for taxi characterization through simulation.

In this context, a *characterization* is a process to determine model parameters or characteristics. For instance, to compute the average speed of a vehicle to travel along a path segment and tag with it the corresponding arc in the graph of the topological map of the associated T_i.

Note that first characterizations will draw data from *P*, so that they will be mostly derived from data-sheets of system components and should be taken as initial guesses for model parameters.

Final ABMs have access to real data through $\{R_i\}$ and characterizations can be more close to reality.

The framework also includes mechanisms to measure worst-case execution times (WCETs) of models, and to monitor whether the control loop is closed fast enough with respect to events coming from reality.

Again, at the first stages of the development, WCET computation enables estimating computation power requirements to run the ABM simulator and controller.

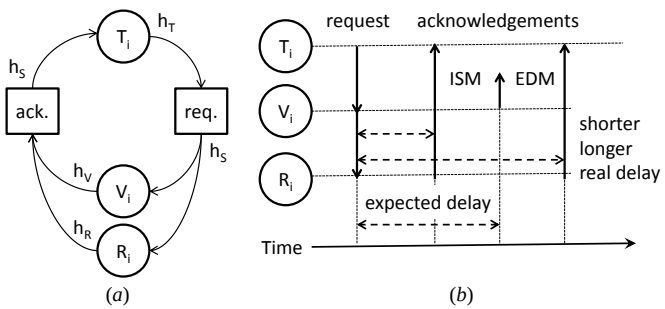


Fig. 4. (*a*) Message passing in successful synchronizations, and (*b*) chronogram with sync. mechanisms when expected delay is larger or shorter than real delay

A. ABM Development

Before developing the model for the controller, designers have to develop a plant simulator, which must include the workplace simulation for the taxis and robot sensor and actuators for every class of taxis in the plant. Obviously, sensors take data from the environment and actuators change it. With mobile robots, actuators can change their position and orientation in the workplace.

On top of robot sensor and actuators, L0 must be programmed so that communication with L1 is done by using the communication services of the synchronizer.

It is quite recommendable to have a secondary L1 that implements some HMI to enable direct control on that level.

Main L1 can be developed afterwards, with low-level tested.

If not all taxis are of the same class, the previous procedure should be repeated for every class.

Once finished, L1 can be verified, first by using secondary interface agents $\{A'_j\}$ and, finally using the real ones $\{A_j\}$.

Templates of all agents are provided so designers can reuse most of the code in them and adapt or prepare the rest of it to the target transportation system.

The last stage of this development phase is to embed L0 code into corresponding AGV and program the interface R_i in the model. R_i is in charge of communicating with vehicle on-board controllers and send/receive messages to/from T_i via synchronizer services.

At this point, it is possible to connect simulation with reality so actual ABM characterization and behavioral tests can be done.

B. Characterization of Taxis

Model accuracy depends on good characterization of the actual plant. Static data such as traffic network and nominal characteristics of vehicles such as average speed and energy consumption can be used for functional validation of the system and as a set of initial values for the model. However, in order to control a real plant, parameters should be as accurate as possible so they have to be estimated from a series of test runs.

Our model includes a mechanism for parameter identification and updating that can be used for both plant characterization and continuous model adaptation.

Plant characteristics are of two types: the ones that define its traffic network and the ones that define the functional and non-functional behavior of the taxis. We assume the traffic network be constant and defined by a topological graph that is known to all taxis of the system.

Each taxi tags the topological graph with data related to the cost it takes to itself to get to a node or to perform some action at a node.

In a simple version, the cost data consists of the time to go from a node to another and the time devoted at each node to decide which outgoing arc to take.

For instance, the characterization of an arc for a given taxi consists of measuring how long it takes to travel from the origin to the destination. And the characterization of the time required to perform an action is done by measuring the time to complete it after being requested to. Such measures are done indirectly from messages between $\{T_i\}$ and the corresponding taxi L0s.

For every order request from a T_i to a R_i, the delay time that takes to T_i to get a reply from R_i is recorded. This delay is compared to the previous one in the same node or arc of the map graph and updated accordingly so that further decisions of T_i and the reactive behavior of V_i are more accurate to the reality. Note that the characterization is made at every communication so taxis may end up by having very different "views" of the traffic network and behaving in a very different manner.

Other characteristics can be measured by the robots and transmitted with the acknowledgement messages but, in the first version of the proposed model, these are not taken into account.

C. Real-Time Monitoring

All delays are compared to the WCET of the body of the main control loop so to guarantee that no inputs from the plant will be lost or taken into account out of time. Therefore the control loop has a cycle period only compatible with robots whose embedded controllers can understand quite complex instructions, with execution times larger than the WCET of the model.

This is the usual case in transport systems with lower-level parts of taxis executing actions such as "go to the next landmark", "take the next turning to the right" or "dock at the machine pier".

For every request-ack. pair between T_i and $\{R_i, V_i\}$, if the actual delay is longer, the view of the corresponding agent remains stand still until the time gap is covered. On the other side, if the real delay is shorter than the expected one, the view is updated for the missed, un-simulated time.

To prevent ABM from missing input data or sending outdated orders and, subsequently, from having a misrepresentation of the symbolic information about the system (P), our model controls that all measured delays go well above its WCET.

In case delays are closer to WCET, there are alternatives to preserve coherence between simulation and reality such as including time-stamps into the messages or minimizing the WCET by appropriately modifying the scheduling of agent execution [26].

VI. Study Case: Transport System Controller of an Automated Laboratory

Laboratories of clinical analyses have progressively been transformed into complex "manufacturing" facilities, able to produce thousands of analyses per hour from blood and other body fluids' samples. In these facilities, samples are dropped into tubes that are placed in racks which are delivered to different analyzing machines by using a conveyor system [27].

Unfortunately, some tests done by analyzers have to be repeated, not all racks have to stop at the same analyzers and there can be several analyzers which can do the same job, though with different workload capacities.

As a result, the complexity of managing this kind of laboratories is quite high, even though they use relatively simple transport infrastructures. In these systems, small AGVs can successfully replace conveyors [28]: They add more degrees of freedom to the system but relieve plant manager from operating with lots of data and make it possible to gain flexibility and robustness [29].

A. Plant

To include most of the characteristics of actual plants of automated laboratories, the case study includes four different analyzers: one ion-counting unit, one coagulometry analyzer and two biochemical units, as most of the samples require measuring biochemical factors.

The layout of the plant (Fig. 5) is quite similar to that of a conveyor system where conveyors are replaced by autonomous AGVs, thus not requiring much infrastructure. In this case, to simplify vehicle operations, robots move around by following a line with marks, which are used by AGVs to self-locate within the plant map. In fact, they are used to indicate a programming spot, a bifurcation or a junction. The type of the mark is determined by AGVs in accordance with their location in the plant.

The programming spots at the loading dock (bottom left at plant prototype and on screen) and at the beginning of the return lane (second to topmost and rightmost cross) are places where the LIMS (laboratory information management system) tell taxis which kind of tests should be done on the samples (transport orders) and which tests have been done successfully (transport order changes), respectively.

There is a re-circulating lane (middle line) that can be used by AGVs that carry samples that wait for acknowledgement of their tests or to repeat them, in case the tests go wrong.

At the beginning of the returning lane (topmost rightmost mark), AGVs have their tube racks unloaded, and, at the waiting queue, they have their batteries re-charged (if needed), and follow their pace to the programming spot.

B. Application-Specific Agents

As already indicated, the overall planning is done by the LIMS, which link samples and tests and, consequently, samples to sets of analyzers. These data are used by a LIMS interface agent to create transport orders for taxis.

The simplicity of the traffic network, with only one collection and one ending spots, relieves the problem of transport order generation and sequencing, which can be solved through auctions that follow a greedy approach. There is, however, a better option by implementing evolutionary learning processes on top of the auctions [30].

Apart from the LIMS agent, there are agents that represent the analyzers in the laboratory. They are responsible for interfacing with AGVs so that analyzers can take samples and perform tests.

C. Transportation Agents

Each taxi features an AGV that is aware of its own position, recognizes the environment and communicates with others to coordinate their movements. AGVs use information about the plant to determine to which analyzer they should go to satisfy the requirements of their loads the fastest they can. Currently,

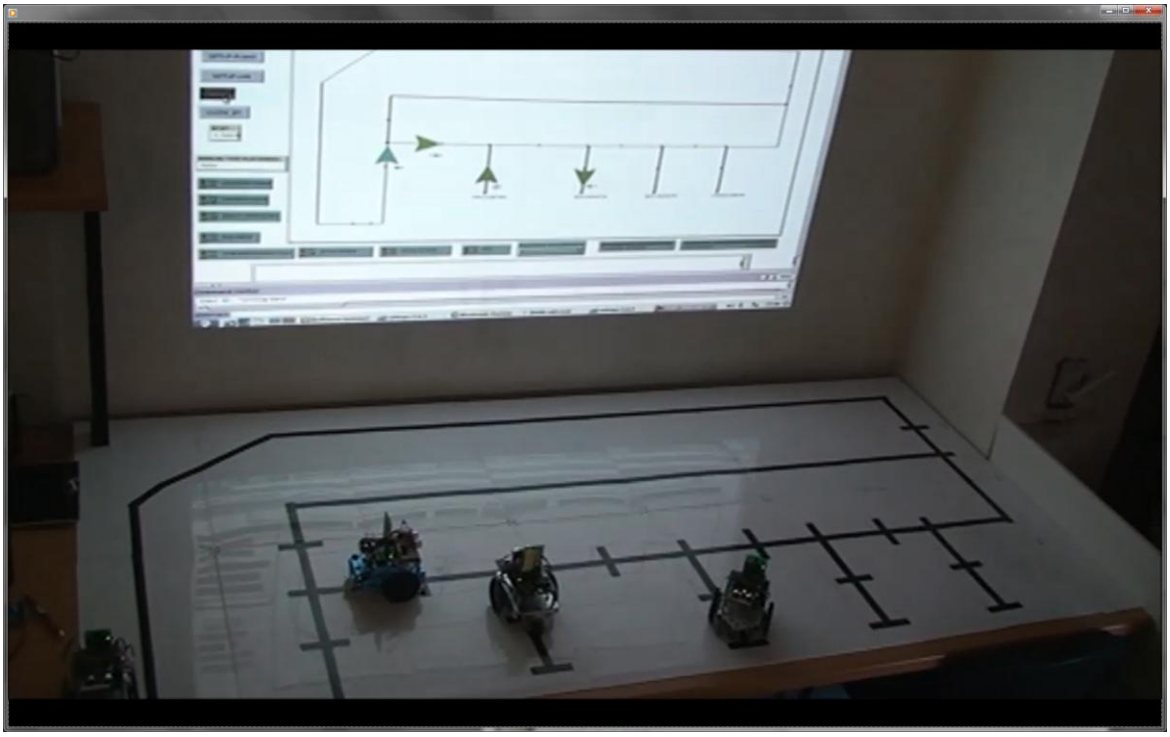


Fig. 5. Plant prototype at the bottom and controller screen on top

ISBN: 978-1-61804-290-3

in our model, AGVs randomly choose from compatible goals, i.e. they can go to either biochemical analyzer on a random basis, as the focus of this work is about validating the proposed ABM-based controller.

When an AGV arrives at its destination, it docks at the port of the corresponding analyzer so that it can begin with its work. In case it is busy, the taxi puts itself on hold in a parking area (short wait) or goes on to a compatible destination or to the re-circulation lane (long wait).

D. Experimental Results

The framework has been developed with Netlogo and so the agents of the system. Real AGVs have been implemented with Parallax Boebots, which have been programmed to behave as modeled within Netlogo. The result prototype (see Fig. 5) consists of Boebots with embedded L0 software, a printed circuit on a surface, and a computer running the Netlogo program for the ABM [4].

To test the maximum load of the circuit, a series of simulations of the case study with 20 AGVs performing random transport orders have been done to estimate the WCET. The average WCET was 16ms, and the estimated communication time (we have real data only for up to 4 robots) for 20 taxis increased this value to 36ms. As a consequence, the ABM controller can handle real time at frequencies of 14 cycles per second.

This frequency implies that simulated ABM can control 20 real robots $\{R_i\}$ with an spatial resolution under the cm, which is acceptable for the laboratory previously presented, even if working at 25% more than the peak throughput of the top current analyzers (8000 tests/hour). Note that marks and objects are more than one cm away from each other.

After system functionality and estimates were validated, a prototype with three AGVs was used to verify the characterization, WCET control, and deployment stages.

1) Pre-Runtime Characterization

The characteristics of each robot with respect to the plant traffic network have been obtained by averaging the travel time at each segment for 30 runs. The full circuit is 11 m long and took an average time of 112.654, 125.456 and 123.169 s per robot to be completed.

Note that, though the framework enables a continuous update of these data, it is important to perform an off-line characterization per robot so to make simulation agree with reality and to help $\{T_i\}$ making better decisions on robot actions.

2) WCET Control

The characterization stage can also be used to validate the estimates on WCET. In this case, the ABM-based controller was run on an inexpensive laptop (64-bit MS Windows 7 OS on a 1.65GHz AMD E-450 CPU, 4 GB of RAM machine) that communicated with robots via serial protocol over Bluetooth.

During characterization, the WCET was 431 ms, which is unsatisfactory for real-time control of multiple robots, as they move at speeds up to 10 cm/s. In order to get rid of this problem, it is possible to use a better machine with a tailored OS to improve execution times and avoid application-unrelated processes interfere with controller processes. However, taking into account 1) that the synchronizers would make real robots wait, and 2) that this WCET and other peaks over 50 ms happen only at 1.3% of all control cycles, it is not a big issue to run the ABM controller on that machine.

3) Synchronization Quality

The more accurate is the characterization of the physical elements in a plant the better the control and, subsequently, the efficiency of the system. A measure of this accuracy can be obtained from the synchronizer: the less the number of EDM and ISM runs the more exact goes the simulation and so works the controller.

The quality of the synchronization during the pre-runtime characterization is measured in terms on percentages of synchronizers' states with respect to the total number of HSE states they go through. As expected, simulation tends to be optimistic and goes ahead reality about 33% of the time, as time delays at segments are initially set to zero. However, most of the time (66%), simulation and reality go together and a mere 1% of time is devoted to run ISM to make simulation keep up with reality.

VII. Conclusion

Rapid design and deployment of systems of AGV drastically reduces costs while agent-based technologies provide resulting controllers with flexibility, adaptability and scalability. Similarly to other approaches, we have proposed a methodology based on ABM. Differently from them, the organization of ABM enables taking profit of it to keep as much as possible of the controller code in the original ABM simulator version.

We have presented an ABM for transportation system controllers that classifies agents into two classes, the application-specific ones and the transportation ones, which, at their turn, are divided into three tiers: L1, synchronizer and L0. The last one is the one to be concurrently run in simulator and on real AGVs, thus requiring to be synchronized.

We have shown that simulators of such type of ABMs can be used as a HMI for the corresponding controllers and, especially, to control AGVs.

We have proposed a development methodology on top of the ABM architecture where only a part of the controllers' code has to be actually embedded into mobile robots.

A key factor in the proposed framework is to have an API that makes it easy to build transportation agents in accordance with the above-mentioned architecture and that makes the synchronization mechanisms as transparent to designers as possible.

Experimental results show that systems like the automated laboratory used as a case study can be developed in less than two man-months and give a clue on how the proposed strategy can contribute to minimize the time-to-prototype and the time-to-market, taking into account that the development platform can be the same that the deployment one.

In the near future we expect complete the development platform with a library of typical transportation problem-solving methods (e.g. path planners) for L1.

REFERENCES

[1] S. Schreiber, and A. Fay. "Requirements for the benchmarking of decentralized manufacturing control systems." In *Proc. of Emerging Technologies and Factory Automation* (ETFA), 2011.

[2] M. Maggio, H. Hoffmann, A.V. Papadopoulos, J. Panerati, M.D. Santambrogio, A. Agarwal, and A. Leva. "Comparison of decision-making strategies for self-optimization in autonomic computing systems." In *ACM Trans. Auton. Adapt. Syst.* 7, 4, Article 36 (Dec. 2012).

[3] Ll. Ribas Xirgo, I.F. Chaile. "An agent-based model of autonomous automated-guided vehicles for internal transportation in automated laboratories." In *5th Int'l. Conf. on Agents and Artificial Intelligence* (ICAART), Barcelona (Spain), 15–18 Feb. 2013.

[4] Ll. Ribas-Xirgo, I.F. Chaile. "Multi-agent-based controller architecture for AGV systems". In*18th IEEE Int'l. Conf. on Emerging Technologies and Factory Automation* (EFTA), Cagliari (Italy), 10–13 Sept. 2013.

[5] A. Kashif, X.H. Binh Le., J. Dugdale, and S. Ploix. "Agent-based framework to simulate inhabitants' behaviour in domestic settings for energy management." In *Proc. of Int'l. Conf. on Agents and Artificial Intelligence* (ICAART). 2011.

[6] P. Davidsson, L. Henesey, L. Ramstedt, J. Törnquist, and F. Wernstedt. "An analysis of agent-based approaches to transport logistics." In *Transportation Research Part C* 1. 2005. 255–271.

[7] L.A. Santa-Eulalia, G. Halladjian, S. D'Amours, and J.-M. Frayret. "Integrated methodological frameworks for modeling agent-based advanced supply chain planning systems: A systematic literature review." In *J. Ind. Eng. & Management*, JIEM, 2011 – 4(4):624-668.

[8] M. Armendáriz, J.C. Burguillo, A. Peleteiro, G. Arnould, and D. Khadraoui. "Carpooling: A multiagent simulation in Netlogo." In *Proc. of the 25th European Conf. on Modelling and Simulation* (ECMS), Kraków (Poland), 2011.

[9] T. De Wolf and T. Holvoet. "Towards autonomic computing: Agent-based modelling, dynamical systems analysis, and decentralised Control." In *Proc. of the First Int'l. Workshop on Autonomic Computing Principles and Architectures*. 2003.

[10] K. Zhang, E.G. Jr. Collins, and D. Shi. "Centralized and distributed task allocation in multi-robot teams via a stochastic clustering auction." In *ACM Trans. Autonom. Adapt. Syst*. 7, 2, Art. 21 (July 2012).

[11] J.Z. Hernández, S. Ossowski, and A. García-Serrano. "On multiagent coordination architectures: A traffic management case study. In *System Sciences, 2001. 34th Annual Hawaii Int'l. Conf. on*, Jan. 2001, 3–6.

[12] O. Banias, R.-E. Precup, and D. Curiac. "Multiagent architecture applied in decentralized real-time urban road traffic control." In *5th Int'l. Symp. on Applied Computational Intelligence & Informatics.* 28–29 May, 2009. DOI=10.1109/SACI.2009.5136255

[13] A. Guerrero-Ibáñez, J. Contreras-Castillo, R. Buenrostro, A. Martí, and A. Muñoz. "A policy-based multi-agent management approach for intelligent traffic-light control." In *Intelligent Vehicles Symposium*, 2010, 694–699.

[14] M. Behrisch, L. Bieker, J. Erdmann, and D. Krajzewicz. "SUMO – Simulation of Urban MObility: An Overview". In *The Third Int'l. Conf. on Advances in System Simulation* (SIMUL 2011), 63–68.

[15] R.-S. Chen, K.-Y. Lu, and C.C. Chang. "Intelligent warehousing management systems using multi-agent." In *Int. J. Comput. Appl. Technol.* 16, 4 (July 2003), 194–201. DOI=10.1504/IJCAT.2003.000325

[16] M. Cossentino, C. Lodato, S. Lopes, and P. Ribino. "Multi agent simulation for decision making in warehouse management." In *Computer Science and Information Systems (FedCSIS), 2011 Federated Conference on*, 18–21 Sept. 2011, 611–618.

[17] H. L. Liang, J. Verriet, R. Hamberg, and B. van Wijngaarden. "Graphical configuration of agent-based warehouse management and control systems." In *Advances on Practical Applications of Agents and Multi-Agent Systems, Advances in Intelligent and Soft Computing*, Vol. 155, 2012. 265–268.

[18] P. Farahvash and T.O. Boucher. "A multi-agent architecture for control of AGV systems." In *Robotics and Computer-Integrated Manufacturing*, Volume 20, Issue 6 (Dec. 2004), 473–483.

[19] A. Wallace. "Multi-agent negotiation strategies utilizing heuristics for the flow of AGVs." In *International Journal of Production Research*, Vol. 45, Issue 2, 2007.

[20] S.C. Srivastava, A.K. Choudhary, S. Kumar, and M.K. Tiwari. "Development of an intelligent agent-based AGV controller for a flexible manufacturing system." In *The Int'l. J. of Advanced Manufacturing Technology*, Vol. 36, Issue 7-8 (March 2008), 780–797.

[21] M.H.F. bin Md Fauadi, H. Lin, and T. Murata. "Dynamic task assignment of autonomous AGV system based on multi agent architecture." In *Progress in Informatics and Computing (PIC), IEEE International Conference on*, Vol.2 (10-12 Dec. 2010), 1151–1156.

[22] R. Erol, C. Sahin, A. Baykasoglu, and V. Kaplanoglu. "A multi-agent based approach to dynamic scheduling of machines and automated guided vehicles in manufacturing systems." In *Appl. Soft Comput*. 12, 6 (June 2012), 1720–1732. DOI=10.1016/j.asoc.2012.02.001

[23] A. Fernández-Caballero and J.M. Gascueña. "Developing multi-agent systems through integrating Prometheus, INGENIAS and ICARO-T." In *Int'l. Conf. on Agents and Artificial Intelligence* (ICAART). 2009.

[24] T. Standley and R. Korf. "Complete algorithms for cooperative pathfinding problems". In *Proc. of the 22nd Int'l. Joint Conf. on Artificial Intelligence*. Barcelona, Catalonia, Spain, 16-22 July 2011. 668–673.

[25] J. Yu and S.M. LaValle. "Planning optimal paths for multiple robots on graphs." In *eprint arXiv*:1204.3830, 04/2012.

[26] P. Mathieu and Y. Secq. "Environment updating and agent scheduling policies in agent-based simulators." In *Proc. of Int'l. Conf. on Agents and Artificial Intelligence* (ICAART), 2012.170-175.

[27] Ll. Ribas-Xirgo, A. Miró-Vicente, I.F. Chaile, A.J. Velasco-González. "Multi-agent model of a sample transport system for modular in-vitro diagnostics laboratories." In *17th IEEE Int'l. Conf. on Emerging Technologies and Factory Automation* (EFTA), Kraków (Poland), 17–21 Sept. 2012.

[28] Ll. Ribas-Xirgo, J.M. Moreno-Villafranca, I.F. Chaile. "On using automated guided vehicles instead of conveyors." In1*8th IEEE Int'l. Conf. on Emerging Technologies and Factory Automation* (EFTA), Cagliari (Italy), 10–13 Sept. 2013.

[29] J. Himoff, G. Rzevski, M. Hinton, and P. Skobelev. "MAGENTA technology: Multi-agent logistics i-scheduler for road transportation." In *Proc. of AAMAS*, Hokkaido, Japan, 2006.

[30] J. Wojtusiak, T. Warden, and O. Herzog. "Agent-based pickup and delivery planning: The learnable evolution model approach." In *Proc. of Int'l. Conf. on Complex, Intelligent and Software Intensive Systems* (CISIS), 2011. 1-8.

Two pronged Strategy for Energy Optimization in WSNs by using In-network Compression and Synthesis of Multiple Queries at Base-Station

Vandana Jindal, A.K.Verma, Seema Bawa

***Abstract*—** Wireless Sensor Networks (WSNs) find applications in environmental monitoring, healthcare monitoring, military surveillance, traffic monitoring etc. Immense data may be collected with the help of densely deployed sensors through the process in three steps - Data Acquisition, Data Processing and Data Communication. Desired information is extracted from this data by multiple queries. Sensor nodes of these networks are constrained in resources like energy and bandwidth. Due to high energy consumption in data communication major energy reduction works have focused on reducing this component only. In-network Data aggregation reduces quantity of data in communication for a simple query. Optimization of multiple queries reduces number of queries. In this paper query optimization at base station and in-network data compression techniques for achieving maximum benefit have been discussed. Both these, result into energy saving, thereby extending the life of node.

***Keywords*—**Wireless Sensor Network, Multiple Query Optimization, MEMS, Epoch, SunSPOT, Solarium.

I. INTRODUCTION

Wireless Sensor Networks (WSNs) are made up of a large number of tiny sensor nodes spread randomly over a large geographical area. Miniaturization of sensor nodes with advances in Micro Electrical Mechanical Systems (MEMS) [1] technology has reduced the production costs of these nodes appreciably. Small size and affordable costs of nodes are the desirable features to enable their use in diverse applications like industry, science, transportation, civil infrastructure and security etc. To keep the size and cost at affordable levels these nodes are manufactured with limited resources like processing power, bandwidth and electrical energy. Power is consumed in every activity of these nodes i.e., sensing, computation and communication thus making it the fastest depleting resource of these nodes. As the sensors are deployed randomly in many inaccessible locations it is not possible to replenish the power. Once power of a node is depleted, it makes the node useless which affects the whole network badly. Therefore, primary concern is to use this resource judiciously. Cost in terms of energy consumption for sensing and computation is far less than energy consumed in communication. Therefore, besides optimizing sensing and computation activities, main thrust is on reduction in radio communication messages either through data compression in the network itself or through reduction in redundant queries at the base station or both.

In this paper, In-network compression and energy efficient multiple query optimization schemes are considered to reduce the quantity of data and the number of monitoring queries that are running in the sensor network. When a new query is submitted at the base station, the scheme checks whether any reduction is possible if the query is rewritten by synthesizing it with one of the already running queries through specifically designed algorithm. If so, the algorithm rewrites the newly injected query using the running queries. The rewritten query is then evaluated at the base station by making use of the results of currently running queries without being injected into the sensor network. As a result, the number of queries injected into the sensor network is reduced, resulting in lower energy consumption. Query rewriting is done at the base station therefore it does not affect the network, In-network aggregation/ compression is used by sensor nodes to reduce the quantum of data communicated to the base station for a particular query. Compression techniques, as the name suggests squeezes the data within the network and pass on just a small amount of the data to the sink.

The following section gives a brief summary of the works accomplished in related area. Sec III gives us a brief overview of energy as a constrained resource in WSN. Sec IV depicts the two pronged strategy for energy optimization used - in network compression and query optimization at the base station. Sec V describes the simulation setup for carrying out the proposed scheme and expresses the simulation results followed by the conclusion in Sec VI.

II. RELATED WORK

Cougar [2] and TinyDB [3] are the two most widely used database management systems to extract information from the database generated by sensor nodes. However, in these systems the focus has largely been on optimizing and executing a single long running query only. Unlike query processing in traditional databases, query processing in WSNs is different because of its own semantics, constraints and objectives etc. Same WSN is used for varying applications by multiple users therefore multiple queries pertaining to individual application are encountered in WSN. Handling of these multiple queries in an optimized fashion is area of research. Few studies and solutions for management of

multiple queries with a sensor proxy to control sampling rates have been proposed through Fjords architecture [4] proposed by Madden et al. and another proposal of SwissQM project [5] at ETHZurich. However, these proposals were successful in eliminating data redundancy to some extent and provided an approximate answer only (not the exact value). Algorithms too were proposed to optimize multiple region based queries. These were divided into partial aggregation sharing approach [6] and equivalence class approach [7, 8]. The approaches were not suitable for majority of the queries.

Energy efficient protocols have been devised. Low Energy Adaptive Clustering Hierarchy (LEACH) protocol proposed by Wendi B. Heizelman et al. [9] is one of the widely used protocol. Many of its application specific variants [10-18] have been devised by the researchers.

Work conducted on WSNs using data compression has also been the focus for quite some time now. Pradhan et al. [19] came up with the idea of distributed compression where source and channel coding were used. As a result the data transfer between the nodes due to compression was reduced. A distributed match source channel communication architecture and reconstruction method from noisy projections was proposed by Rabat et al. [20]. Similar gossip communication approach was proposed by Wagner et al. [21]. Although Wagner et al. [22, 23] came up with the architecture for distributed wavelet analysis eliminating the hypothesis about the grid regularity but failed to depict the method in selecting an optimal path for the compression and spatial correlation. Works on audio and video compression in WSNs have been found in [24, 25]. Efforts in improvising routing, aggregation, indexing and storage, energy balancing using compression could be seen in [21]. Sadlen and Martonosi came up with a new version of lossless LZW algorithm [26] which could compress 528 bytes data block. This S-LZW algorithm showed a significant amount of energy saving locally and globally, which employed buffering of data prior to its transmission. Comparisons of various compression codes updates for reconfiguring of nodes [27] was carried out by Tsiftes et al. presenting an algorithm with preprocessing and coding they experimentally showed 67% of energy saving using GZIP. Packet compression method (based on its frequency), was done by Ju and Cui [28] while they stated that packets with randomly changing fields were transferred uncompressed. Difference coding on Length variable coding were used for sensor readings. It was stated that 50% compression was permissible after removal of redundant data. The present work is an endeavor to obtain desired results in cost reduction through multi query optimization at base station and In-network compression. In this work the solution is sought by looking into base-station optimization in which data redundancy is removed followed by filtering of the result, ultimately moving towards decrease in the communication cost in terms of energy consumption by reducing the number of queries. Sensor nodes employ compression techniques thus a two pronged strategy to filter out redundant data has been devised. Taking into consideration the characteristics of WSNs and the constraints involved with respect to energy, our main objective is to minimize communication costs. Our work here presents the achievable energy savings when the sensor readings are compressed at the originating node.

III. MOTIVATION

Economic viability and long term reliability of WSN has always been the area of interest for research community. Every endeavor is to find ways and means which are frugal on energy consumption so as to increase the lifetime of the WSN. Energy is a major constraint in WSN. Therefore once energy is exhausted, node is dead which affects the network badly. The energy consumption involved for execution of a single instruction is 1nJ. However data transmission costs are manifold higher than data processing costs.

1Kb data transmission = processing three million instructions.

Therefore to achieve maximum benefit in terms of energy consumption, it is essential that the number and content of transmissions is reduced without loss of other essentials such as data security, data fidelity and acceptable latency.

IV. PROPOSED TWO PRONGED STRATEGY FOR ENERGY OPTIMIZATION

Data is collected with the help of densely deployed sensors through steps of data acquisition, processing and communication. Initially when the Query set is empty, the query is injected into the WSN for data acquisition. The acquired data has to be communicated to the base station. The level of data compression achieved within the network, has a positive impact over the network's energy consumption. Following depicts the two pronged strategy employed for optimizing the energy usage in a network.

A. In-network compression

There are many successive bytes within the input data stream, so we have selected compression algorithms which deliberately exploit these data structures. The other goal was the demand to achieve good compression gains while compiling to sizes within the resources available on current mote/ node platforms. Compression algorithms compared on Simulator were: Huffman and LZW. Besides the In-network compression results of which has been evaluated on simulator and is presented in the next section, base-station optimization as described below is also employed.

B. Base station Optimization

We are emphasizing on optimization of multiple queries at base station as this will help in reduction of energy consumption significantly. The queries are considered to be long running ones. In case of long running queries, mostly result dissemination messages are left in the network so a metric in terms of number of result dissemination messages in a unit time is considered to calculate or arrive at the cost involved in answering a query. Gain in cost reduction can be evaluated by finding the difference between the energy cost of newly injected query and synthesized query.

If we assume that each sensor node then has equal chance of being queried, selectivity of a predicate from the data is sel(p),

in case of multi hop query with sensor at depth d, cost of data acquisition query (Wq_i) shall be:

$$Wq_i = \frac{sel(p).d}{S_i}$$

where S_i is sampling rate of the query.

Sel(p) depends on the distribution of attributes. Selectivity is computed over the whole data range. If a system has A attributes for querying and each attribute has a range which falls between [min_i, max_i] (i =1, 2,…A) the selection criteria of the predicate p having the range (upper $value_i$, lower $value_i$) may be written as:

$$Sel(p) = \sum_{i=1}^{A} \frac{\text{upper value}_i - \text{lower value}_i}{max_i - min_i}$$

If an existing query set Q_1 is synthesized into new query set Q_2 the cost difference is given as:

Cost difference = $W_{Q1} - W_{Q2}$

Where W_{Q1} is cost of all the queries in Q_1 and W_{Q2} is the cost of queries in new set Q_2. Value must be positive for synthesis to be beneficial. While synthesizing queries into new one it is to be ensured that new query is super set of queries being merged. Semantic correctness is also essential in case of data aggregation queries. Positive difference is denoted by a metric called *Gain*. Gain metric quantifies the saved cost in query rewriting. If we merge two queries q_1 and q_2 into one synthetic query q', it should be such that all the data requested by q_1 and q_2 must be requested by q'.

$Gain_{12}=sel(p_1)/s_1+ sel(p_2)/s_2 - sel(p_1 U\ p_2)/GCD(s_1,s_2)$

We shall write q_1 and q_2 into q' if and only if $Gain_{12} > 0$.

$Gain_{12}$>0 only if

GCD (s_1,s_2)==s_1 *or GCD*(s_1,s_2)==s_2 Theorem[33]

('s' refers to the sample period or the 'epoch')

In case Gain < 0, new queries are not integrated and go directly into the existing synthetic query list. If Gain > 0, queries are synthesized, new synthetic query is then checked for positive Gain with other queries of the set. The iterative algorithm is so designed that any achievable positive Gain is exploited fully. Due to generation and storage of data continuously by nodes in a sensor network a WSN may be considered similar to a distributed database [22, 29]. Assumption that the data is distributed in a database makes the usage of data more comfortable as modification of data becomes easy. Queries are injected to extract information from this database. Queries may be categorized as one shot queries and continuous queries. Queries reporting the current data only once are termed as one shot queries where as continuous queries are those where the sensors produce and report the data periodically. Multiple queries being studied may be of any of these types.

C. Synthesis of Multiple Queries

The section here presents multi-query optimization algorithms. The base-station is the interface between the network and the user. User sends queries and obtains the result at the base station. Base station is not resource constrained as the nodes. Therefore, base-station is used to filter out the redundant load of multiple queries into the network. Multi-query optimization algorithm rewrites a set of similar queries into a new set of queries before injecting them into the network, so that redundant data requests can be eliminated as much as possible. Correctness of semantics of queries is to be ensured while rewriting new query. All this is achieved with the help of different algorithm designed to obtain maximum Gain.

Treatment of a new query:

If a new query q_n arrives at a base station where results of a synthetic query set Q_s are already being calculated or obtained from the network, the algorithm will evaluate the benefit of rewriting the new query with the existing synthetic queries and find the most beneficial(in terms of cost) one. A new query is generated by merging the new query and existing query where the Gain is maximum. If there is no such query then new query is directly added to the set q_s.

Iterative evaluation of cost reduction by integration of new synthetic query in the q_s:

When a synthetic query is generated on arrival of new query as explained above the previously existing query set Q_s gets modified. It is evaluated through successive iterations through algorithm whether new synthetic query can be synthesized with any of the existing queries in Q_s to achieve further Gain. If so the pair is rewritten and the new one is again checked. This iterative process is continued till no further Gain is achievable.

When duration of a new injected query gets finished it is to be removed from the query set being evaluated. Query list is updated again to the previously optimized status prevailing before injection of the query.

In this way Multiple Query Optimization (MQO) design should be scalable-every new query is treated individually for positive Gain, energy efficient and adaptable-no redundant query should be there in the system.

Query Rewriting is employed for answering the newly injected queries by reusing the results of the already existing queries. This results into optimization of resource usage as duplicate data requests can be removed.

The notations used in the algorithm are:

Q = {q_1, q_2 ,…, q_n} the set consisting of already running queries;

a_n = The attributes like temperature, pressure; etc;

q'_{new} = a rewritten query of q_{new} defined over q_1,q_2,…, q_n;

EP = Epoch is the time interval of taking the readings; sampling rate;

SC(q) = Selection Criteria of q;

A(q) = Set of attributes listed;

S(q) = Set of attributes that are in selection criteria SC;

(sc= $p_1 \wedge\ p_2$ …..) e.g.sc= (light > 100) $\wedge$ (temp<30)

Decomposition of query q_{new} into d_{a1}, d_{a2}, ….., $d_{an;}$

$d_{ai} = \begin{cases} \prod_{nodeid,\ ai}(\sigma_{sc}(q_{new},a_i)(sensors)), \text{ if } a_i \in A(q_{new}) \\ \prod_{nodeid}(\sigma_{sc}(q_{new},a_i)(sensors)), \text{ if } a_i \in S(q_{new}) - A(q_{new}) \end{cases}$

For $1 \leq i \leq n$.

e.g. $d_{temp} = \prod_{nodeid,\ temp}(\sigma_{temp}(q_{temp<30})(sensors))$

$d_{light} \prod_{nodeid}(\sigma_{light<100})(sensors))$

Algorithm for Rewriting the Query

1. Let $A(q_{new}) \cup S(q_{new}) = \{a_i\}$;
where i = 1,2,3…..n;
/* composing the Prospect Query set*/
2. Q' = Q;
3. for j=1 to m
    for $q_j \in Q'_j$ do
        if (($EP(q_{new})$ % EP(q)!=0)^(SC(q)^ $SC(q_{new})$ == false))
            remove q from Q';
        end
    end
end
4. for i=0 to n, $A(q_{new}) \cup S(q_{new})$
    do
    Q' a_i = {q | (q ϵ Q') ^ (a_i ϵ P(q))};
end
5. for each Q'a_i ($1 \leq i \leq n$)
    do compute test $C(q_{new}, a_i) \rightarrow V_{q \in Q'\ ai}$ SC(q)
        if test == true
            hold;
        else
            return;
end
6. for i=1 to n in $A(q_{new})$
    do $d_{ai} = U_{q \in Q'\ ai} \prod_{nodeid,\ a_i\sigma_{sc}(q_{new},a_i)(q))}$;
end
7. for i =1 to n
    $S(q_{new}) - A(q_{new})$
    $d_{ai} = U_{q \in Q'\ ai} \prod_{nodeid,\ (\sigma_{sc}(q_{new},a_i)(q))}$;
end
8. Return $q'_{new} = d_{a1} \bowtie_{a2} \bowtie \ldots \bowtie d_{an}$;

V Simulation Setup

The goal of this simulation study is the evaluation of the achievable energy gains when data compression is applied prior to packet transmission. We examine a multi-hop scenario, where a node periodically delivers the data to the sink (base-station). The data collected through high density distributed WSNs are immense. In applications like temperature monitoring, the data collected at the free nodes needs to be transferred to the base-station periodically so that the data available is up-to-date. When readings are taken at regular time intervals, they are not expected to change significantly, but need to be transferred to the base-station at regular intervals.

SunSPOT mote used in the study is a WSN mote developed by Sun Microsystems. The device is built upon the IEEE 802.15.4 standard. Unlike other mote systems, The SPOT is built on Squawk VM (Virtual Machine) [30]. A SPOT is about the size of a 3×5 card with 32-bit ARM9 CPU, 1 MB RAM and 8 MB of Flash memory, a 2.4 GHz radio and a USB interface. The network platform of SPOT has built-in sensors along with the capability of interfacing with external devices. Two kinds of SPOTs i.e. free-range SPOT and base-station SPOT are present. The anatomy of free range SPOT has a battery processor board, a sensor board and a sunroof.

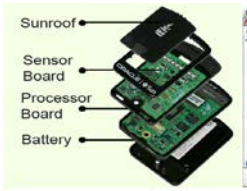


Figure 1: SPOT Anatomy

Figure 2: Solarium Application

The sensors present are capable of measuring acceleration, temperature and light intensity. The base-station SPOT does not have a sensor board. It acts as an interface between the base station application running on the host (PC with Windows platform) and those running on the targets. The host application is a J2SE program [31] and target application is a Squawk Java Program. Two of the platforms used here are the SPOT Manager tool and Net Beans 7.0 Integrated Development Environment. The SPOT Manger tools are – the SPOT Manager and Solarium [32]. Each SunSPOT has an IEEE network number. Solarium has an emulator for running applications on a virtual SPOT.

The compressed data is then transmitted to the user's end via the base station thus reducing the quantum of data transmission. The reduction in the size of the data leads to saving energy of the nodes in the WSN.

VI Test Cases, Test Results and Test Analysis

A. *Test Case I*

Test Objectives

The objective of the test case I is to verify that the functionality of one of the approaches i.e. in-network compression as proposed in the two pronged approach works in the direction of energy saving. The test is executed on the test bed "Solarium". It has to be verified if the result obtained using the synthetic data fulfils our proposed approach. The compression algorithm used in Test Case-I is Huffman Compression.

Test Results

The final graph plotted between the number of queries executed against their respective sizes (in bits) i) without being subjected to compression and ii) with Huffman Compression is as under:

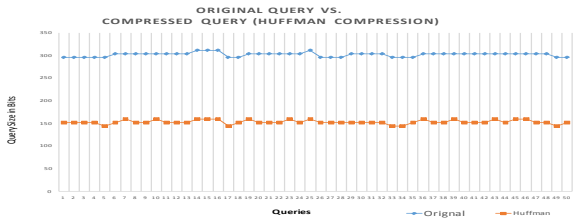


Figure 3(a): Graphs showing results of Original Query vs. Compressed Query performed with Huffman compression

Test Result Analysis

In case of in-network optimization the technique used is data compression. Use of data compression results into reduction of the number of bits that are to be transmitted thus reducing the number of radio transmissions. Lesser the radio transmission, lesser will be the energy used. The graphical results depict that the acquired data gets compressed with the help of Huffman. The compression factor accomplished is almost 50.73%.

B. *Test Case II*

Test Objectives

The objective of the test case II is to verify that the functionality of one of the approaches i.e. in-network compression as proposed in the two pronged approach works in the direction of energy saving. The test is executed on the test bed "Solarium". It has to be verified if the result obtained using the synthetic data fulfils our proposed approach. The compression algorithm used in Test Case-II is LZW Compression.

Test Results

The final graph plotted between the number of queries executed against their respective sizes (in bits) i) without being subjected to compression and ii) with LZW Compression is as under:

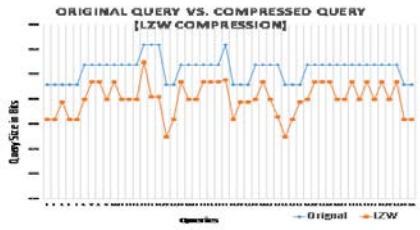


Figure 3(b): Graphs showing results of Original Query vs. Compressed Query performed with LZW compression

Test Result Analysis

In case of in-network optimization the technique used is data compression. Use of data compression results into reduction of the number of bits that are to be transmitted thus reducing the number of radio transmissions. Lesser the radio transmission, lesser will be the energy used. The graphical results depict that the acquired data gets compressed with the help of LZW algorithm. The compression factor accomplished is almost 40.42%.

C. *Test Case III*

Test Objectives

The objective of the test case III is to verify that the functionality of one of the approaches i.e. base-station optimization proposed in the two pronged approach works in the direction of energy saving. The test is executed on the test bed "Solarium". It has to be verified if the result obtained using the synthetic data fulfils our proposed approach.

Test Results

The graph in Figure 4 is obtained by employing three techniques i.e. i) firing independent queries, ii) query merging and iii) query rewriting along with merging of the already fired queries at the base-station. The values along the y-axis show the readings that were sensed and transmitted from the nodes to the base-station.

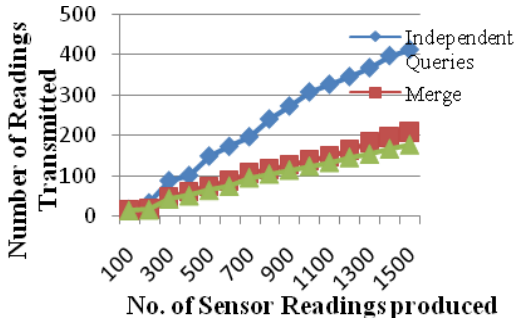


Figure 4: Graph showing results obtained from Number of Sensor readings transmitted and received

Test Result Analysis

It is found that merging of different queries proves to be beneficial. Secondly, when merging of queries is followed by rewriting of a resultant query by adopting merging, the technique proves to be beneficial. Both the methods resulted in reducing the number of transmissions. Out of the two techniques tested, the latter resulted into reducing greater number of radio transmissions proving more beneficial in energy conservation.

VII. CONCLUSION AND FUTURE SCOPE

In this paper we have proposed an energy efficient compression technique in-network to reduce traffic in WSN facilitating a better life time of the network. At the base-station query re-writing method is used for writing a new query if possible from the already injected queries. Both methods resulting into reduced energy consumption of the nodes. The work is implemented using the SPOT wireless platform. After the data has been sensed by the sensors of a free-range SPOT that is installed the data compression algorithm then comes into play. This compressed data in then transmitted to a base station SPOT which in turn sends them via USB to a PC. Even though various compression schemes are still under development, experimental results indicate that their compression rate and power reduction manners are quite

ISBN: 978-1-61804-290-3

impressive. They are one possible mode to diminish resource constrain of wireless sensor nodes. On testing, the query optimization and processing algorithm described in this paper results in cutting the cost in terms of energy and bandwidth usage, both scarce resources in a WSN in sensing, computation and communication. Communication is the most energy intensive operation.

In the future as technology progresses the application areas of WSNs will become broader than the existing scenario. Their availability will become more to the masses than right now. There still exist many obstacles to be overcome for practical use of sensor networks

REFERENCES

[1] Gardner,J.W., Varadan,V.K., and Awadelkarim,O.O.: *Microsensors, MEMS and Smart Devices*. New York: Wiley, 2001.

[2] Demers,A., Gehrke,J., Rajaraman,R., Trigoni,N., and Yao,Y.: The cougar project: A work-in-progress report. *SIGMOD Record*, 32(4),2003.

[3] Madden, S., Franklin, M.J., Hellerstein, J.M., and Hong, W.: TinyDB: An acquisitional query processing system for sensor networks. *ACM TODS*, 30(1), November 2005.

[4] Madden, S. and Franklin, M.J.: Fjording the stream: An architecture for queries over streaming sensor data. In *Proc. of ICDE*,2002.

[5] Muller, R. and Alonso,G.: E±cient sharing of sensor networks. In *2006 IEEE International Conference on Mobile Adhoc and Sensor Systems (MASS)*, pages 109-118, 2006.

[6] Emekci,F., Yu,H., Agrawal,D., and Abbadi,A.E.: Energy-conscious data aggregation over large-scale sensor networks. In *Technical Reportof UCSB*,2003.

[7] Trigoni, N., Yao, Y., Gehrke, J., Rajarama, R., and Demers, A.: Multi-query optimization for sensor networks. In *Proc. of DCOSS*,2005.

[8] Trigoni, N., Guitton, A., and Skordylis, A.: Routing and processing multiple aggregate queries in sensor networks. In *SenSys*,2006.

[9] Wendi B. Heinzelman, Anathan P. Chandraskan, and Hari Blakrisshnan, "Energy-Efficient Communication Protocol for Wireless Microsensor Networks", Proceedings of the 33rd Hawaii IEEE International Conference on System Sciences, pp. 1-10, 2000.

[10] Fuad Bajaber and Irfan Awan, "Dynamic/ Static Clustering Protocol for Wireless Sensor Network", 2nd IEEE UKSIM European Symposium on Computer Modeling and Simulation, pp. 524-529, 2008.

[11] Wei Wang, Bingwen Wang, Zhuo Liu, Lejiang Guo, and Wei Xiong, "A cluster-based and tree-based power efficient data collection and aggregation protocol for wireless sensor networks", Information Technology Journal, vol.10, no. 3, pp. 557-564, 2011.

[12] Erfan. Arbab, Vahe. Aghazarian, Alireza. Hedayati, and Nima. Ghazanfari Motlagh, "A LEACH-Based Clustering Algorithm for Optimizing Energy Consumption in Wireless Sensor Networks", 2nd International Conference on Computer Science and Information Technology (ICCSIT2012), pp.147-150, 2012.

[13] Razieh Sheikhpour, Sam Jabbehdari, and Ahmad khademzadeh, "A Cluster-Chain based Routing Protocol for Balancing Energy Consumption in Wireless Sensor Networks", International Journal of Multimedia and Ubiquitous Engineering vol. 7, no. 2, pp. 1-16, 2012.

[14] H.Srikanth.Kamath, "Energy Efficient Routing Protocol for Wireless Sensor Networks", International Journal of Advanced Computer Research, vol. 3, no. 2, issue 10, pp. 95-100, 2013.

[15] Messai Mohamed-Lamine, **"**New Clustering Scheme for Wireless Sensor Networks" 8th IEEE International Workshop on Systems, Signal Processing and their Applications (WoSSPA), pp. 487-491, 2013.

[16] Zahra Beiranvand, Ahmad Patooghy, and Mahdi Fazeli, "I-LEACH: An efficient routing algorithm to improve performance & to reduce energy consumption in Wireless Sensor Networks", 5th IEEE Conference on Information and Knowledge Technology (IKT), pp. 13-18, 2013.

[17] Surender Kumar, M.Prateek, N.J.Ahuja, and Bharat Bhushan, "DE-LEACH: Distance and Energy Aware LEACH", International Journal of Computer Applications, vol. 88, no.9, pp.36-42, 2014.

[18] Xiaowen Ma and Xiang Yu, "Improvement on LEACH Protocol of Wireless Sensor Network", Proceedings of the 2nd International Symposium on Computer, Communication, Control and Automation (ISCCCA-13), pp.338-341, 2013.

[19] Pradhan S., Kusuma J., and Ramchandran K., "Distributed Compression in a Dense Microsensor Network," *IEEE Signal Processing Magazine*, vol.19, no. 2, pp. 51-60, 2002.

[20] Rabbat M., Haupt J., Singh A., and Nowak D., "Decentralized Compression and Pre-distribution via Randomized Gossiping," *in Proceedings of International Conference on Information Processing in Sensor Networks*, USA, pp. 51-59, 2006.

[21] Wagner R., Choi H., Baraniuk R., and Delouille V., "Distributed Wavelet Transform for Irregular Sensor Network Grids," *in Proceedings of IEEE Workshop on Statistical Signal Processing*, USA, pp. 1196-1201, 2005.

[22] Wagner S., Baraniuk G., Du S., Johnson B., and Cohen A, "An Architecture for Distributed Wavelet Analysis and Processing in Sensor Networks," *in Proceedings of International Conference on Information Processing in Sensor Networks*, USA, pp. 243-250, 2006.

[23] Roy O. and Vetterli M., "Distributed Compression in Acoustic Sensor Networks Using Oversampled A/D Conversion," *in Proceedings of IEEE International Conference on Acoustic, Speech and Signal Processing*, vol. 4, France, pp.165-168, 2006.

[24] Gehrig N. and Dragotti L., "Distributed Compression in Camera Sensor Networks," *in Proceedings of International Conference on Image Processing*, UK, pp.82-85, 2001.

[25] Gehrig N. and Dragotti L., "Distributed Sampling and Compression of Scenes with Finite Rate of Innovation in Camera Sensor Networks," *In Proceedings of Data Compression Conference, pp*.83-92, 2006.

[26] Sadler C. M. and Martonosi M., "Data Compression Algorithms for Energy-Constrained Devices in Delay Tolerant Networks," in *Proceedings of the 4th International Conference on Embedded Networked Sensor Systems (SenSys)*, 2006.

[27] Tsiftes N., Dunkels A., and Voigt T., "Efficient Sensor Network Reprogramming through Compression of Executable Modules," in *Proceedings of the 5th Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks (SECON)*, 2008.

[28] Ju H. and Cui L., "EasiPC: A Packet Compression Mechanism for Embedded WSN," *In Proceedings of the 11th IEEE International Conference on Embedded and Real-Time Computing Systems and Applications (RTCSA)*, 2005.

[29]Bajwa W., Haupt J., Sayeed A., and Nowak R,, "Compressive Wireless Sensing," *In Proceedings of Information Processing in Sensor Networks*, USA, pp.134-142, 2006.

[30] Simon, D., Cifuentes, C., Cleal, D., Daniels, J., White, D. Java™ on the Bare Metal of Wireless Sensor Devices— The Squawk Virtual Machine. In VEE'06 June 14–16, Ottawa, Ontario, Canada.ACM Press.

[31]J2SE program. Available at: http://www.oracle.com/technetwork/java/javase/1-5-0-139765.html

[32] Goldman, Ron. Sun Oracle. 16.9.2010. Available at: www.sunspotworld.com/docs/AppNotes/AccelerometerAppNote.pdf

[33] Xiang, S., Lim, H.B., Tan, K.L.:"Impact of Multi-query Optimization in Sensor Networks", *Proceedings of the 3rd International Workshop on Data Management for Sensor Networks (DMSN'06)*, Seoul, South.

BIOGRAPHY:

Vandana Jindal is currently working as an Assistant Professor in the department of Computer Science at D.A.V College, Bathinda. She holds degrees of B.Tech, MCA, M.Phil. Since January 2009, she has been with the Thapar University, Patiala in Punjab as a Ph.D. student. Her research interests include database management systems and wireless sensor networks. She is a member of IEEE and IEI.

A. K. Verma is currently working as Associate Professor in the department of Computer Science and Engineering at Thapar University, Patiala in Punjab (INDIA). He received his B.S. and M.S. in 1991 and 2001respectively, majoring in Computer Science and Engineering. His research interests include wireless networks, routing algorithms and securing ad hoc networks.

Seema Bawa holds M.Tech (Computer Science) degree from IIT Kharagpur and Ph.D. from Thapar Institute of Engineering & Technology, Patiala. She is currently Professor in the department of Computer Science and Engineering at Thapar University, Patiala in Punjab (INDIA).. Her areas of interest include Parallel and distributed computing, Grid computing, VLSI Testing and network management. Prof. Bawa is member of IEEE, ACM, Computer society of India and VLSI Society of India.

Extraction of urban land features from TM Landsat image using the land features index and Tasseled cap transformation

R. Bouhennache, T. Bouden , A. A. Taleb and A.Chaddad

Abstract— In this paper we propose a method to map the urban areas. The method uses an arithmetic calculation processed from the land features indexes and Tasseled cap transformation TC of multispectral Thematic Mapper Landsat TM image. For this purpose the derived indexes image from the original image such SAVI the soil adjusted vegetation index, UI the urban Index, and EBBI the enhanced built up and bareness index were staked to form a new image and the bands were uncorrelated, also the Spectral Angle Mapper (SAM) and Spectral Information Divergence (SID) supervised classification approaches were first applied on the new image TM data using the reference spectra of the spectral library and subsequently the four urban, vegetation, water and soil land cover categories were extracted with their accuracy assessment. The urban features were represented using a logic calculation applied to the brightness, UI-SAVI, NDBI-greenness and EBBI- brighteness datasets. The study applied to Blida and mentioned that the urban features can be mapped with an accuracy ranging from 92 % to 95%.

***Keywords*—**EBBI, SAVI, Tasseled Cap Transformation, UI.

I. Introduction

The fast urbanization and urban expansion have significant impact on conditions of urban ecosystems. Updated information on the status and trends of urban ecosystems is needed to develop strategies for reasonable development and to improve the livelihood of cities. The necessity to monitor urban land-cover/land-use changes is highly desirable by the policy decision makers and regional planners. Remote sensing materials in the form of satellite images are usually converted into useful information such as land cover maps using software and hardware processing and allow the opportunity to accurate and actualize the specific mapping, besides the useful indexes of studying LC/LU can more helping interpretation of the evolution and the management of land features such as NDVI which has been widely used for mapping vegetation and NDBI for interpreting urban. [1] were used a technique of combining the NDBI and the NDVI with a specific processes calculation to extract the built-up areas and were retched an accuracy of 92.6%, [2] developed an approach to detect urban LC/LU changes by quantifying sub-pixel percent imperviousness using Landsat and high-resolution imagery and they have mentioned The satisfactory of method based on a comparison using independent reference data and that information on sub-pixel imperviousness allows the data user to quantify urban LC/LU based on their own threshold.[3] transformed a raw correlated image to a new uncorrelated dataset using the features land index NDBI, Soil Adjusted Vegetation index SAVI and Normalized Difference Water Index NDWI and Through a supervised classification, a principal components analysis, and a logic calculation on the new image, the urban built-up lands were finally extracted with an overall accuracy ranging from 91.5% to 98.5 %, many studies have combined the changes like [4] witch attempted to employ a quantitative approach in exploring the relationship LC/LU index to detect urban between temperature and several indices, including NDVI, NDWI and NDBI, they have proposed a new index, the Normalized Difference Bareness Index NDBaI to extract bare land from the satellite images, others studies were compiled and analyzed using the famous post classification comparison PCC like in [5,6]. In our previously study [7], we have followed the expansion of urban tissue using the method of difference soil adjusted vegetation index DSAVI, difference normalized difference built up index DNDBI and also the post classification of multispectral and multi-temporal L5 and L7 Landsat satellite using a MLC and we have mentioned the fast urban growth with a rate of 0.5% annually. In this study we attempt to extract the urban features using the proposed urban land features indexes combined with the Tasseled cap transformation, using only NDBI needed a complex processing and a various process calculation besides the not satisfactory of accuracy, however if the maximum likelihood supervised classification method cluster the image pixels into classes corresponding to the defined training classes [8], The SAM and SID supervised classifications are based on match pixels spectrum to the identified of reference spectra. SAM compares the angle between the reference spectrum vector and each pixel vector in n-D space where smaller angles represent closer matches to the reference spectrum [9]. SID uses a divergence measure to match pixels to reference spectra, the pixels will be cluster in the class if the divergence will be smaller [11]. This study used the SAM and SID classification methods because the LC/LU features extracted have known identified reference spectra. Besides The urban features were represented using a logic calculation

ISBN: 978-1-61804-290-3

applied to the brightness, UI-SAVI and EBBI- brightness and NDBI-greenness datasets.

II. MATERIALS AND METHODS

A. Study area and data used

Situated at the north of Africa, Blida is a town near of Algiers the capital of Algeria as shown in Fig 1. Acquired at the 19 may 2010 the TM Landsat satellite image was georefferenced to UTM projection WGS 84 datum zone 31, the radiometric characteristics of TM Landsat is shown in tableI .

Fig.1 Location of Blida the study area

B. Data preprocessing

The principal visualization RGB is the composite 742 and it's available to distinguishing between vegetation (green), urban (magenta) and water (blue). To minimizing the effects of sensor a radiometric correction is involved using Lmin/Lmax calibration of radiance image as in Table II

Table I Acquisition of scene and radiometric characteristics sensors

Scene size =170×183 Km^2		Swatch = 185 KM	
Sensors	**Bands**	**spectral Resolution (µm)**	**spatial Resolution (m)**
TM5	Band1: blue Band2 : red Band3 : green Band4 :near IR Band5 : mid IR Band 6 : Thermal Band7 : mid IR	0,45 – 0,52 0,52 – 0,60 0,63 – 0,69 0,76 – 0,90 1,55 – 1,75 10.4 – 12.5 2,08 – 2,35	30 30 30 30 30 120 30

Table II the coefficients Lmin/Lmax used for calibration

Bands	**Lmin**	**Lmax**
Band1	-1.52	193.00
Band2	-2.84	365.00
Band3	-1.17	264.00
Band4	-1.51	221.00
Band5	-0.37	30.20
Band7	-0.15	16.50

C. Methodology

a. Derived indexes image

The UI index is used instead the normalized difference built up index NDBI because the urban and soil features are more distinguishable in UI rather than NDBI and more correct results were obtained when using band 7 instead band 5 like in [10] as in Fig 2, the SAVI is more sensitive to the vegetation and the EBBI is a remote sensing index that applies wavelengths ranged from 0.83 µm to 1.65 µm, and 11.45 µm (NIR, SWIR, and TIR, respectively) to Landsat ETM+ images. These wavelengths were selected based on the contrast reflection range and absorption in built-up and bare land areas.

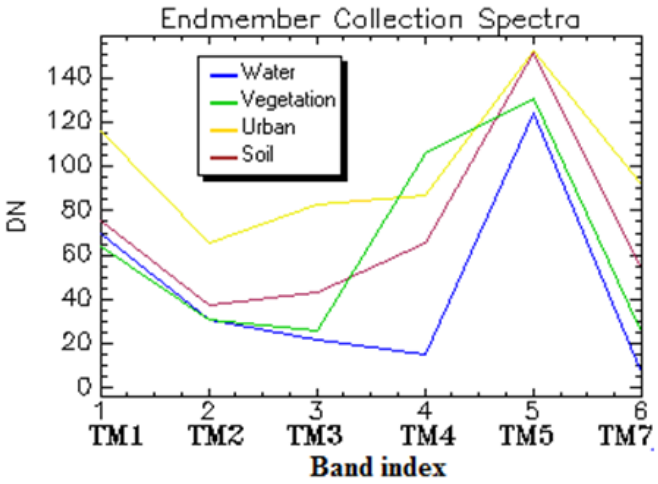


Fig. 2 Spectral response of the four land features in the six TM bands

The expression of land features indexes is

$$UI = \frac{TM7 - TM4}{TM7 + TM4} \quad (1)$$

$$NDBI = \frac{TM5 - TM4}{TM5 + TM4} \quad (2)$$

$$EBBI = \frac{TM5 - TM4}{10\sqrt{TM5 + TM6}} \quad (3)$$

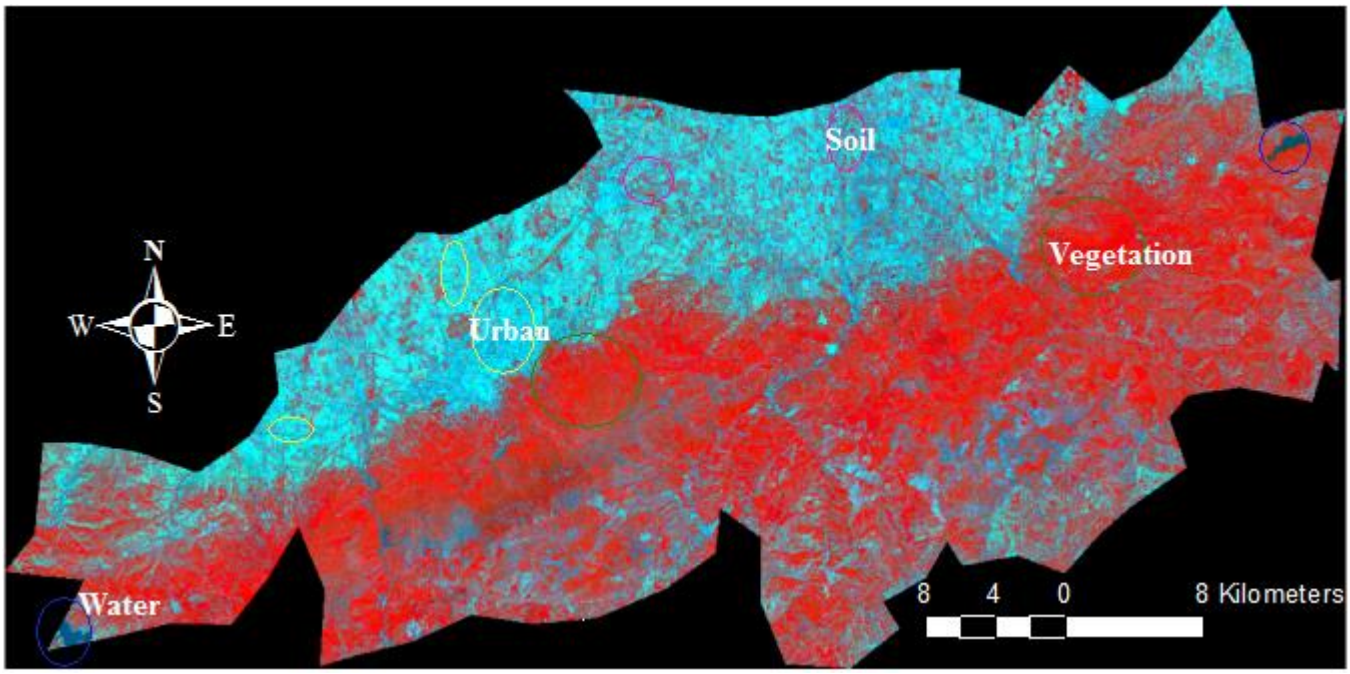


Fig 3 the false color RGB combination of derived indexes, SAVI is displayed as Blue, NDBAI as Green and UI as Red.

$$SAVI = \frac{(TM4 - TM3) \times 1.5}{TM4 + TM3 + 0.5} \quad (4)$$

$$NDBAI = \frac{TM5 - TM6}{TM5 + TM6} \quad (5)$$

b. Tasseled Cap transformation

The Tasseled Cap transformation is a practical vegetative index and spectral enhancement which transforms six TM bands [1-5] and 7 to basically three outputs bands named respectively brightness, greenness, and soil/vegetation wetness or third component [19], however to represent urban features the datasets outputs aren't used singly, the NDBI – greenness and EBBI-brightness were used to highlight the urban features.

c. Classification and accuracy assessment

To produce a single thematic map coming from all bands and reflecting the real ground land cover, theclustering pixels to classes is required and a spectral angle mapper SAM, spectral information divergence classification SID [12] were performed and subsequently the four urban, vegetation, water and soil land cover categories were extracted with their accuracy assessment and Kappa coefficient as in Table III and Fig 4. Accuracies assessment were carried out using ground truth data acquired from topographic map.

d. Extraction of urban features

The urban land cover was extracted from the binary

Table III Overall Accuracy, Kappa Coefficient of SAM and SID classification and the statistics of DNs of the four covers in the six spectral TM bands.

SAM Overall Accuracy = 92.55% (2986767/3226965) Kappa Coefficient = 0.89			SID Overall Accuracy =86.12% (2778695/3226965) Kappa Coefficient = 0.80	
		Min Max Mean Std		Min Max Mean Std
Water	Band 1	54 107 74.01 6.44		57 255 76.84 19.87
	Band 2	21 63 34.10 5.04		21 255 36.65 20.02
	Band 3	16 68 24.83 5.83		16 255 28.15 22.56
	Band 4	10 59 18.87 7.39		10 255 21.98 22.85
	Band 5	5 59 16.77 7.53		5 255 20.47 24.79
	Band 7	3 38 9.73 4.13		3 255 12.85 22.39
Urban	Band 1	53 255 101.89 15.89		54 255 99.58 14.76
	Band 2	22 255 54.91 11.11		21 248 53.97 9.85
	Band 3	17 255 66.87 15.70		17 255 67.43 13.33
	Band 4	24 255 75.62 13.02		20 255 76.2211.60
	Band 5	30 255 121.0125.95		26 255 130.48 22.42
	Band 7	13 255 75.17 19.47		10 255 80.17 16.32
Vegeta -tion	Band 1	52 98 64.74 4.03		52 115 66.10 5.02
	Band 2	21 52 31.15 3.17		21 65 32.08 3.75
	Band 3	17 60 27.67 4.51		17 80 29.43 5.71
	Band 4	39 172 89.94 14.41		38 172 87.77 14.23
	Band 5	28 136 71.65 11.77		29 148 74.75 13.15
	Band 7	10 74 28.00 6.18		11 121 30.56 7.75
Soil	Band 1	53 195 77.87 8.94		53 190 80.17 8.55
	Band 2	22 92 39.90 6.23		21 148 41.27 5.88
	Band 3	18 135 44.10 10.62		17 148 46.31 9.30
	Band 4	35 147 78.22 10.31		25 135 77.2910.28
	Band 5	34 255 102.13 21.68		27 255 105.39 19.11
	Band 7	14 206 52.62 15.43		10 195 55.5012.78

ISBN: 978-1-61804-290-3

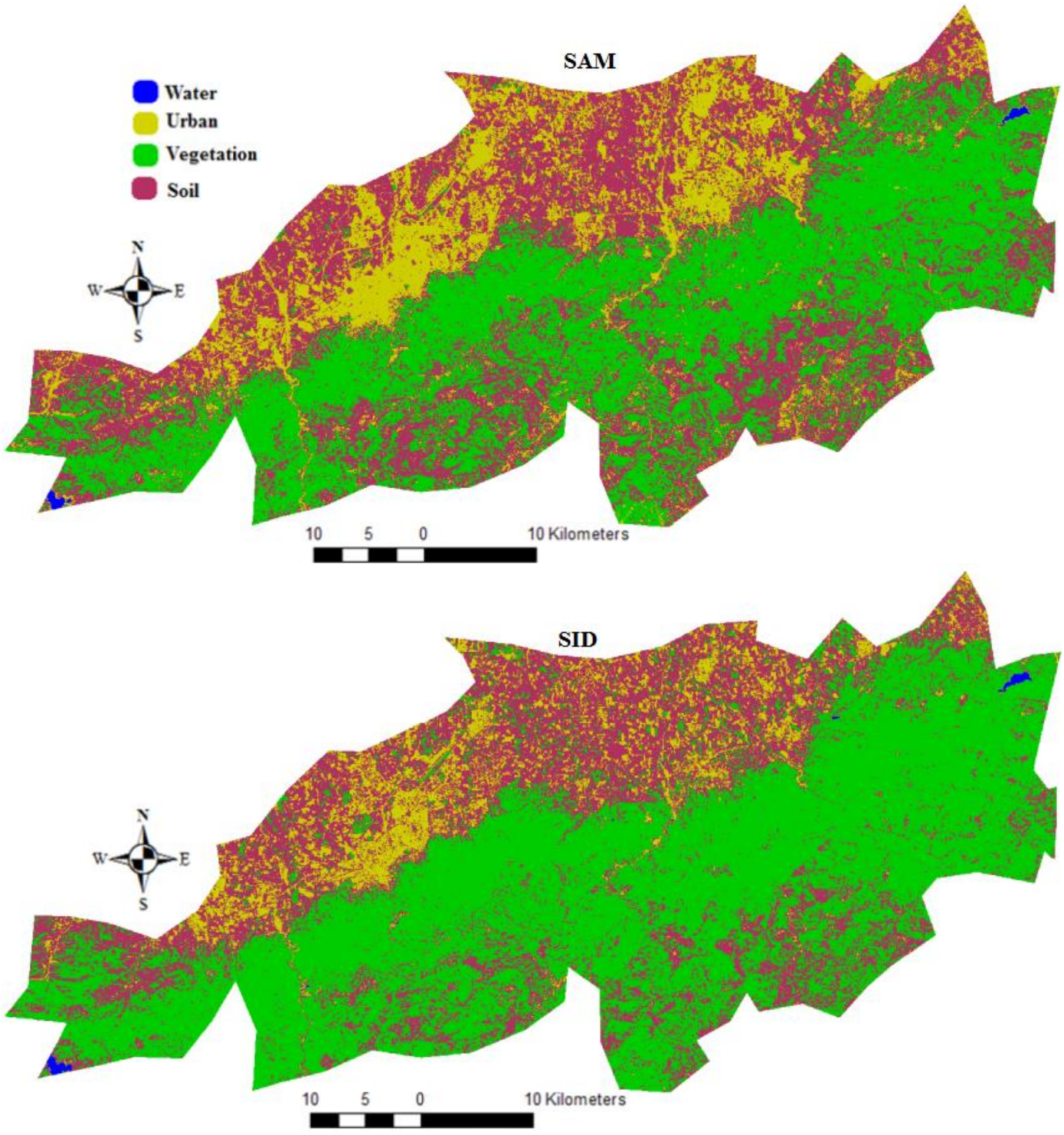


Fig.4 The output classes results from the SAM and SID classification

SAM and SID classification and from the new datasets UI-SAVI, BRIGHTNESS, NDBI-GREENESS and EBBI-BRIGHTNESS using the recoded manipulation method as in Fig.6 and the results are shown in Fig 7 and theirs Overall Accuracy are shown in Fig 8.

III. Results and Discussion

Fig.2 shows the spectral response of the considering LC/LU in the six bands of original image for what the process of land features indexes was chosen, indeed the urban have the high response for band 5, band7 and the low response for band 4, band 3 but the response spectral illustrates also the difficulty to separate the urban from the soil because both have the high reflectance in the same short wavelength, Fig.3 shows the displayed new image for false color SAVI NDBAI UI band combinations, the vegetation appears in red, urban areas are cyan blue, soils in light browns and water is blue and it's obvious that the land features are more distinguishable in the new image rather in the original because the bands in the new image are uncorrelated as in Fig.5. Table III shows the good overall accuracy and Kappa coefficient for both classification but the SAM classification was more accurate than SID and the statistic explains the more difficult to separate the urban from the soil because the values of brightness were very near. The Fig 7 and Fig 8 show that the methods of EBBI-BRIGHTNESS and BRIGHTNESS can extract the urban features with a high Overall Accuracy of 94.77% and 94.65% respectively; also the SAM, SID and UI-SAVI binary images represent the urban land cover with a good Overall Accuracy contrarily with the technique of NDBI-GREENESS which present the worse mapping of the urban with an Overall Accuracy of 92.03%.

IV. Conclusion

In this paper a methods for urban features extraction from TM Landsat based on the urban land features indexes and Tasseled Cap transformation as well as theirs overall accuracy were discussed, the methods show that the process of urban land features can be extracted using several process of spectral classification and logic calculation, for spectral classification the spectral angle mapper and spectral information divergence are very useful and can be mapped the urban land features with an overall accuracy of 94.48% and 94%, for the logic calculation the methods of UI-SAVI, EBBI-BRIGHTNESS and BRIGHTNESS can be represent the urban land use with a high overall accuracy of 93.77%, 94.77% and 94.65%.

References

[1] Y. ZHA , J. GAO, ''Use of normalized difference built-up index in automatically mapping urban areas from TM imagery'', International Journal of Remote Sensing, 2003, Vol. 24, No. 3, pp. 583–594

[2] Y. Limin, X. George, ''Urban Land-Cover Change Detection through Sub-Pixel Imperviousness Mapping Using Remotely Sensed Data'', Photogrammetric Engineering & Remote Sensing, September 2003, Vol. 69, No. 9, pp. 1003–1010.

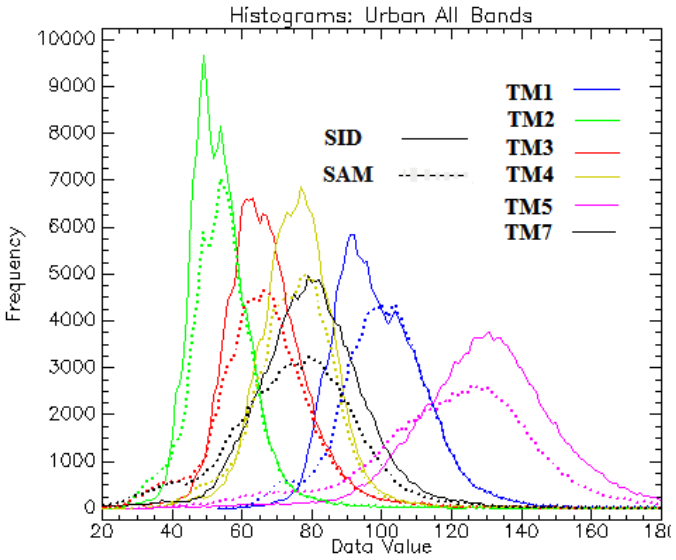


Fig.5 Histograms of urban land cover in the six bands for SAM and SID classification showing that's the bands are mush correlated.

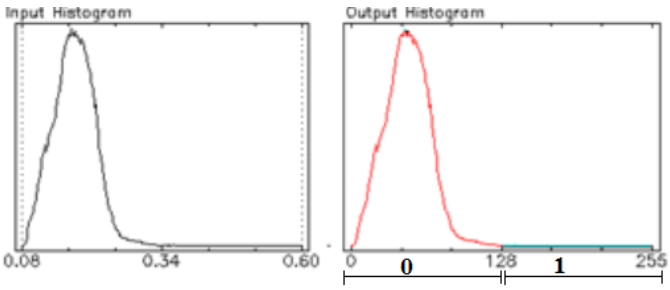


Fig.6 The linear transformation used for recoded images.

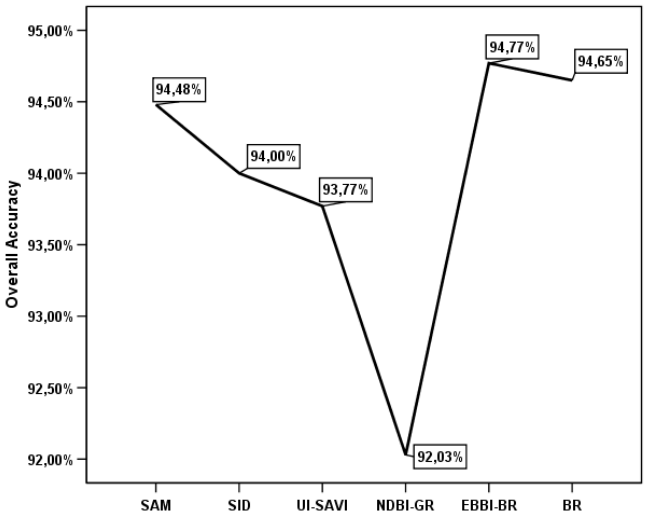


Fig.8 Overall Accuracy of the binary coding images showing that NDBI-GREENESS is the worse extraction

[3] X .Hanqiu, ''Extraction of Urban Built-up Land Features from Landsat Imagery Using a Thematic oriented Index Combination Technique'', Photogrammetric Engineering & Remote Sensing, September 2007, Vol. 73, No. 12, pp. 1381–1391.

[4] L. C. Xiao, M. Z. Hong, X. L. Ping, Y. Zhi, ''Remote sensing image-based analysis of the relationship between urban heat island and land use/cover changes'', Remote Sensing of Environment,2006, vol. 104,pp. 133–146

[5] M. Kenneth, M. Gunter, ''Monitoring Land-Use Change in Nakuru (Kenya) Using Multi-Sensor Satellite Data'', Advances in Remote Sensing, 2012, vol.1, pp. 74-84

[6] Y .Fei, E . Kali, C. Brian, E. Marvin, ''Land cover classification and change analysis of the Twin Cities (Minnesota) Metropolitan Area by multitemporal Landsat remote sensing'', Remote Sensing of Environment, 2005, vol.98 , pp. 317 – 328

[7] R .Bouhennache, T. Bouden, ''Change Detection In Urban Land Cover Using Landsat Images Satellites, A Case Study In Algiers Town'', IEEExplore, 2014, vol.14, pp. 4799-562

[8] J.A. Richards, ''Remote Sensing Digital Image Analysis''. Springer-Verlag, Berlin, 1999

[9] F.A. Kruse, A.B. Lefkoff, J.B. Boardman, K.B. Heidebrecht, A.T. Shapiro, P.J. Barloon, A.F. Goetz, ''The Spectral Image Processing System (SIPS) – Interactive Visualization and Analysis of Imaging Spectrometer Data''. Remote Sensing of the Environment 44, 145–163, 1993.

[10] P. S. Pratibha , M. H. Priya, S. D. Duhita, ''Fusion Classification of Multispectral and Panchromatic Image using Improved Decision Tree Algorithm'', IEEExplore, 978-1-4799-3140-8/14/$31.00 ©2014 IEEE

[11] H . Du, H. Chang, F.M. Ren, J.O. D'Amico, J. Jensen, ''New hyperspectral discrimination measure for spectral characterization. Optical Engineering 43, 1777–1786, 2004.

[12] B.Sunil , P. Shanka , M. Ramnarayan ''Per-pixel and object-oriented classification methods for mapping urban features using Ikonos satellite data'', Applied Geography 30 (2010) 650–665

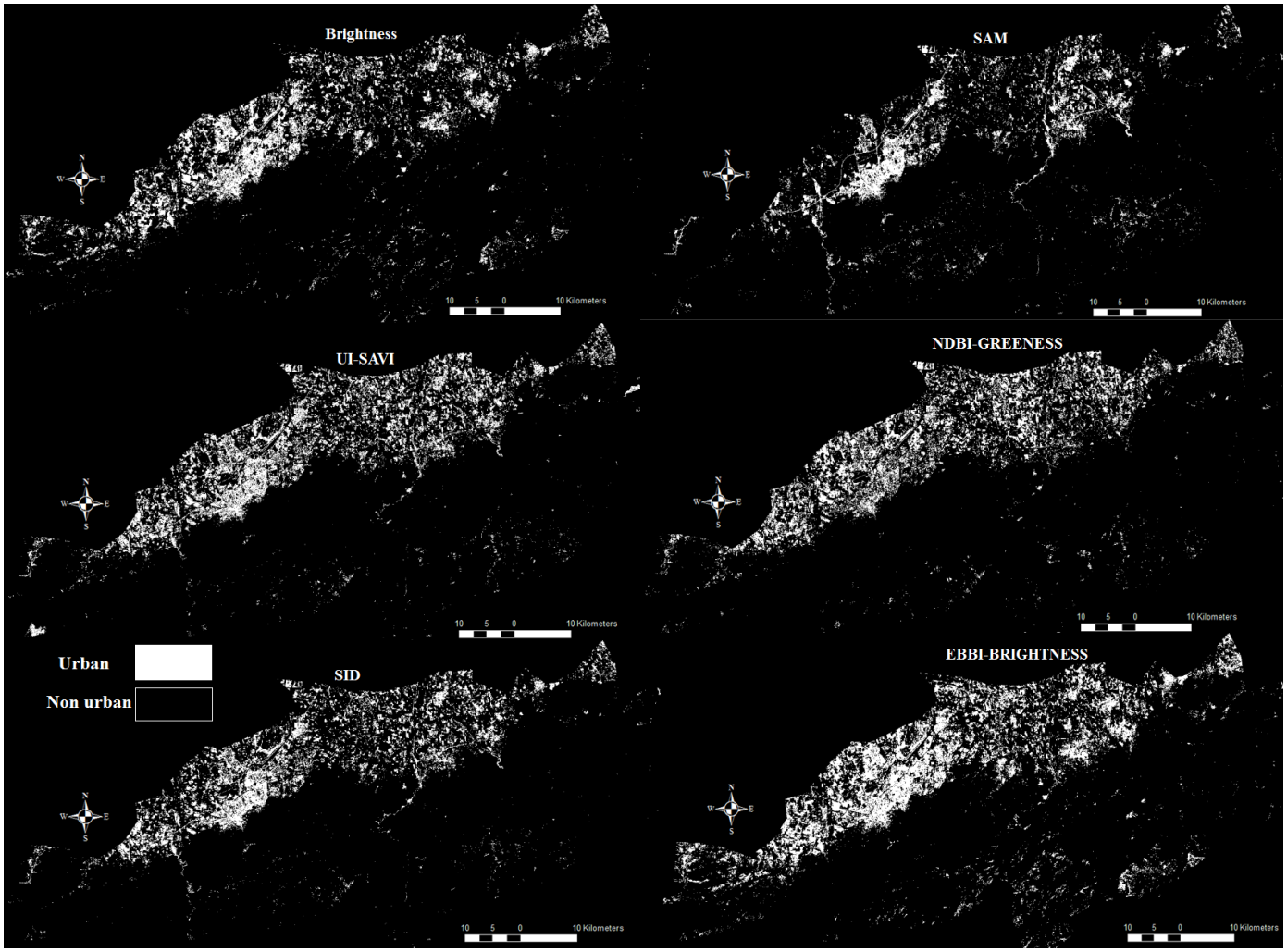


Fig.7 The output urban land features extraction from binary coding images.

On Riccati-Genetic Algorithms Approach for Non-convex Problem Resolution. Case of Uncertain Linear System Quadratic Stabilization

K. DCHICH, A. ZAAFOURI and A. CHAARI

***Abstract*—**This paper deals with algorithms leading to the design of a state-feedback control law minimizing a cost function J. For quadratic stabilization of uncertain linear systems, with norm bounded uncertainties, the comparison between the Algebraic Riccati Equation (ARE) algorithm, using convex optimization approach and the Algebraic Riccati Equation- Genetic Algorithms (ARE-GA), designed for non-convex optimization, shows that the second algorithm gives better performances than the first one. The case of an uncertain second order system is considered in order to illustrate the efficiency of the ARE-GA approach.

***Keywords*—** Quadratic Stabilization, Uncertain System, Riccati Equation, Convex, Non-convex, Genetic Algorithm.

I. Introduction

IN 1993, Rockafellar did affirm that "[..] the great watershed in optimization is not between linearity and nonlinearity, but convexity and non-convexity" [1]. The main objective of this work is to apply this sentence in the context of the convex optimization and non-convex optimization field [2], [3].

The optimization provides a rich algorithmic framework for all areas of applied sciences. There are two branches of deterministic optimization: convex programming and non-convex programming. A convex optimization problem is defined by the minimization of a convex function (objective) within convex constraints. When the double convexity, in the objective and the constraints, is not satisfied the problem falls into non-convex optimization field.

In this paper is presented a new approach to solve the problem of quadratic stabilizability of uncertain linear systems. The proposed idea is to split the problem in two parts: a convex part, involving a large number of decision variables that requires solving an Algebraic Riccati Equation (ARE) [4]-[7], and a non-convex part, involving a small number of decision variables, which solution is estimated using a Genetic Algorithm (GA) [8]-[10].

Genetic algorithms [8] are an optimization technique developed in 1975 by J. Holland that have been inspired by Charles Darwin's theory of biological populations evolution. Genetic algorithms are based on the principle of the survival of the fittest, which means that the best suited structures and the most closer to the desired result, using genetic operators such as selection, crossover and mutation. The fitness of a particular individual is measured using a fitness function, which evaluates how close the individual is to the objective [9]-[12].

When using a quadratic Lyapunov function, it is possible to synthesize a control law which takes into account the variations of uncertainty that guarantee the quadratic stability of a closed loop system [13]-[15].

The organisation of the paper is as follows. The quadratic stabilizability problem of an uncertain linear system is given in the second section. Then, in the third section, the application of the convex optimization approach using Riccati solvers (ARE) is presented.

The ARE solvers and the genetic algorithms (GA), introduced in section 4, are used to solve the problem splitted into a convex and a non convex subproblem and to optimize the P and ε parameters of the Riccati matrix Equation.

In section 5, the approach is applied to synthesize an adequate control law for an uncertain second order system. Finally, conclusions are given in section 6.

II. Quadratic Stabilization of Uncertain Systems. Problem Statement

Consider the following linear system, described in state space by

$$\begin{cases} \dot{x}(t) = Ax(t) + Bu(t) \\ y(t) = Cx(t) \end{cases} \quad (1)$$

K. Dchich is with the University of Tunis, Unit C3S, Higher School of Sciences and Techniques of Tunis (ESSTT), 5 Av. Taha Hussein, BP 56, 1008 Tunis, Tunisia (E-mail: Khira.Dchich@fsb.rnu.tn).

A. Zaafouri is with the University of Tunis, Unit C3S, Higher School of Sciences and Techniques of Tunis (ESSTT), 5 Av. Taha Hussein, BP 56, 1008 Tunis, Tunisia (E-mail: abderrahmen.zaafouri@isetr.rnu.tn).

A. Chaari is with the University of Tunis, Unit C3S, Higher School of Sciences and Techniques of Tunis (ESSTT), 5 Av. Taha Hussein, BP 56, 1008 Tunis, Tunisia (E-mail: assil.chaari@esstt.rnu.tn).

$x(t) \in \Re^n$ is the state vector, $u(t) \in \Re^m$ the control vector, $A \in \Re^{n \times n}$ and $B \in \Re^{n \times m}$ constant matrices, such that (*A*, *B*) is a controllable pair and *B* has full column rank.
The state feedback control law has the following general form

$$u(t) = Kx(t) \tag{2}$$

the closed-loop system can be described by

$$\dot{x}(t) = (A + BK)x(t) \tag{3}$$

The design of this control law, minimizing the following cost function *J* [16]

$$J = \int_0^\infty \left(x^T Q x + u^T R u \right) dt \tag{4}$$

where $Q \geq 0$ and $R > 0$ are, respectively, the state and the input weighting matrices of the criterion *J*, needs the resolution of an Algebraic Ricatti Equation (ARE) in the form

$$A^T P + PA - PBR^{-1}B^T P + Q = 0 \tag{5}$$

It comes the solution P of equation (5) and the control gain *K* such that

$$K = R^{-1}B^T P \tag{6}$$

Suitable choices of the *Q* and *R* introduced matrices lead to the computation of the gain *K*. Therefore, the eigenvalues and eigenvectors, of the closed-loop characteristic matrix, can be chosen and the system performances provided.

In the case where the uncertainty is in norm bounded type, the system (1) can be described by [5], [17]

$$\dot{x}(t) = (A + \Delta A)x(t) + Bu(t) \tag{7}$$

where ΔA represents the uncertainty of the studied system such that

$$\Delta A = DFE \tag{8}$$

D and *E* are constant matrices of appropriate sizes and *F* a matrix satisfying the following matrix inequality

$$F^T F \leq I \tag{9}$$

The stability of the resulting closed-loop uncertain system (7) is then established by the use of a quadratic Lyapunov function.

Definition

The uncertain linear system (7) is said to be quadratically stabilizable, if there exists a linear feedback control law (2), a positive semi-definite symmetric matrix $P \in \Re^{n \times n}$ and a constant parameter $\alpha > 0$ such that the following condition holds: Given any admissible uncertainly *F*, the Lyapunov function $V(x) = x^T Px$ is such that [17]

$$\dot{V} = x^T \left[A^T P + PA \right] x + 2x^T PDFEx \leq -\alpha \|x\|^2 \tag{10}$$

The synthesis of such control law algorithm, needing the resolution of an algebraic Riccati equation, is proposed in [17] where is introduced a necessary and sufficient condition for quadratic stabilizability by linear state feedback.

Theorem

The uncertain linear system (7) is said to be quadratically stabilizable, if there exists a linear feedback control law (2), a constant ε such that, for any positive-definite symmetric matrix *R*, the Riccati equation

$$A^T P + PA - PBR^{-1}B^T P + \varepsilon PDD^T P + \varepsilon^{-1}E^T E + Q = 0 \tag{11}$$

or the inequalities

$$A^T P + PA - PBR^{-1}B^T P + \varepsilon PDD^T P + \varepsilon^{-1}E^T E < 0 \tag{12}$$

have a positive semi-definite symmetric matrix solution $P \in \Re^{n \times n}$, expressed by (6).

For fixed parameter ε, the resolution of equation (11) corresponds to a convex problem which becomes non-convex, for free ε.

Problem formulation

Find the parameters *P* and ε, solutions of equation (11) minimizing the criterion (4). *A*, *B*, *Q* and *R* are respectively real matrices of dimensions *n* x *n*, *n* x *m*, *n* x *n* and *m* x *m*, *Q* a symmetric positive semi-definite matrix and *R* a symmetric positive definite matrix.

III. Convex optimization problem resolution algorithm

Many problems of uncertain systems can be solved through convex optimization tools. In this case, the computation time to find a solution is reasonable; the result corresponds to a global minimum of the criterion [18], [31].
Let a function $f : E \subset \mathbb{R}^n \to \mathbb{R}$ such that *E* is a convex set, figure 1.

f is convex iff

$$f\left(\lambda x_1 + (1-\lambda)x_2\right) \leq \lambda f(x_1) + (1-\lambda) f(x_2) \tag{13}$$

$\forall \lambda \in [0,1] \subset \mathbb{R}, \forall (x_1, x_2) \in E^2$ and λ a constant parameter.

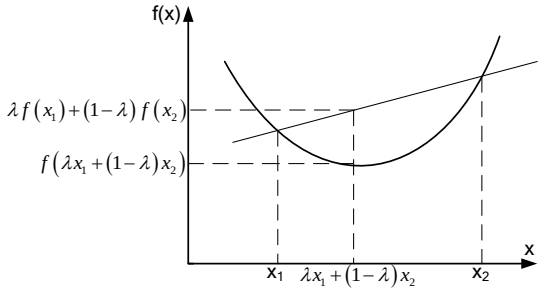


Fig. 1 Convex function

Proposed algorithm

The stabilization approach by the control law of the process (7), defined by (2), required for the resolution of the Riccati equation (11), follows the following steps

(i) choose the weighting matrices *Q* and *R* such that *R* = *Q* = *I* and take ,

(ii) solve the equation (11); if the solution , the system is quadratically stabilizable; then, calculate *K*. Otherwise, go to (iii),

(iii) replace by ; if , stop; the system is not quadratically stabilizable. Otherwise, return to (ii).

This algorithm is used, in section 5, for the second order uncertain system study.

IV. Non-convex optimization problem resolution algorithm

Consider the problem of finding a positive semi-definite solutions to the Riccati equation (11) which appears often in a wide range of control applications.

In the case of an n^{th}-order system, the parameters to determine are $\varepsilon > 0$ and $P>0$, $P = \{p_{ij}\}$, such that $p_{ij} = p_{ji}\ \forall i, j$.

The optimization problem to solve becomes the following one.

minimize ε

subject to: $A^T P + PA - PBR^{-1}B^T P + \varepsilon PDD^T P + \varepsilon^{-1}E^T E < 0$

$P > 0$

For any fixed value $\varepsilon = \varepsilon_0$, the problem is convex and can be solved via Riccati solvers and the solution *P*, which depends on $N = 1/2(n + 1)n$ decision parameters p_{ij}, $\forall i, j$, determined.

Then, for non fixed , the problem is non-convex and cannot be solved using standard solvers. This observation motivates the idea to split the original problem into a small non-convex part solved by Genetic Algorithm (GA) and a large convex part solved with a Riccati solver, as shown in Figure 1. Then, to do this, we propose [7]

- ✓ to let a fast and efficient Riccati solver takes care of the large convex part of the problem: for a given ε find the unique solution *P* (if it exists),
- ✓ and to let GA, which may be unreliable for a large number of decision parameters, deal with the smaller non-convex part and search over ε (which usually depends on the controller and other parameters)

Thus, GA is used to construct the constant ε and a Riccati solver applied to calculate *P* (if it exists). The full chromosomes are constructed by adjoining the decision parameters in and *P*.

If a standard GA is used alone to solve the original problem, the GA chromosomes must code both ε and P, and if P is large, the chromosome, consequently, will be too long for an efficient and reliable solution.

Reducing the dimension of the solution space for the GA which doesn't only accelerate the evolution process, but also increases the chances of converging to the global solution of the problem.

The overall algorithm is given in Figure 2. Once the full chromosomes are constructed and the fitness evaluated.

It is applied, in the next section, to a second order uncertain system and its efficiency compared to ARE approach.

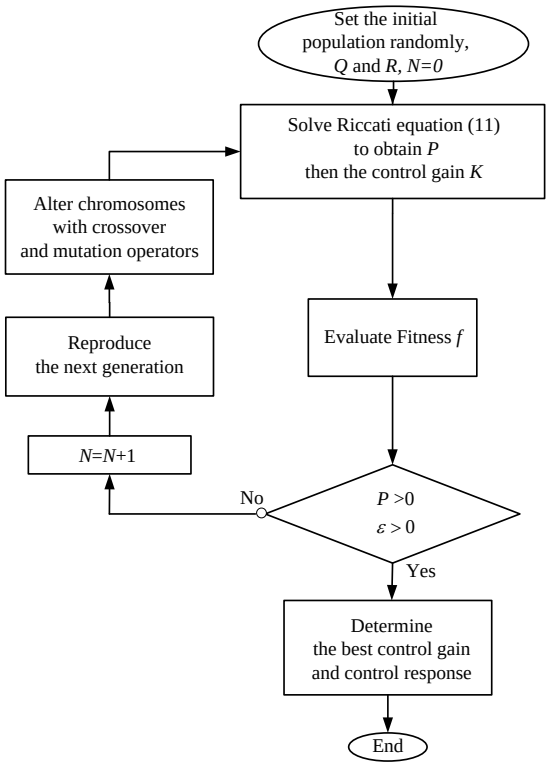


Fig. 2 The structure of the proposed ARE-GA algorithm

V. Second order uncertain system case

For stabilization longitudinal movement of hunting plane F4E study, in two operating points, let consider the reduced model described by (1) such that [24]

$$A = \begin{pmatrix} -1.82 & 17.76 \\ 0.17 & -0.75 \end{pmatrix},\ B = \begin{pmatrix} -91.24 \\ 0 \end{pmatrix},$$

$$E = \begin{pmatrix} 0.2 & 2 \\ 0.4 & 2 \end{pmatrix} \text{ and } D = \begin{pmatrix} 1.6 & 0 \\ 0 & -0.5 \end{pmatrix}$$

To show the performances of the proposed ARE-GA formalism in terms of convergence, let run, several times, the implemented algorithm of optimization in order to obtain data on and P parameters.

Then, the best solutions of ARE-GA introduced in Figure 2, obtained by the use of Matlab 7.10, are

$\varepsilon = 0.038$

$$P = \begin{bmatrix} 1.1075 & 4.8679 \\ 4.8679 & 33.5459 \end{bmatrix}$$

$$K = \begin{bmatrix} 4.8679 & 33.5459 \end{bmatrix}$$

and the best solutions found by the use of ARE are

$\varepsilon = 0.062$ for ε in the interval $[0.001,1]$

$$P = \begin{bmatrix} 0.2658 & 1.0030 \\ 1.0030 & 6.9121 \end{bmatrix}$$

$$K = \begin{bmatrix} 1.0030 & 6.9121 \end{bmatrix}$$

For these gains, the stability is satisfied for the studied system [5],[20].

By running 20 times the proposed ARE-GA algorithm, regardless of the initial population, we note that its convergence is always located in the same area of the search space. This means that the ARE-GA reaches each time the most interesting region of the search space. The results, obtained by simulation using the optimal settings of figures 3, 4, 5 and 6, show the effectiveness of the proposed ARE-GA algorithm compared to the solution obtained by the Algebraic Riccati Equation (ARE).

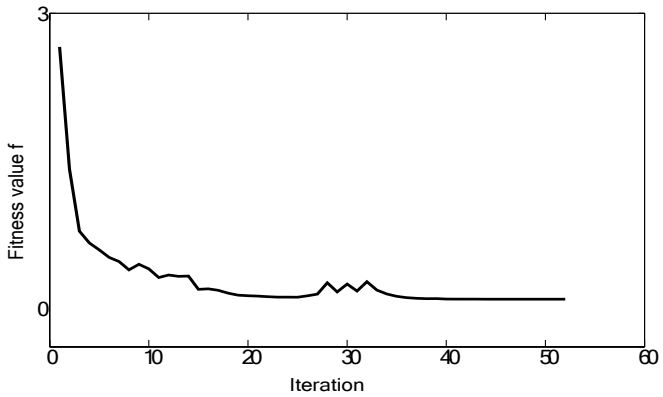


Fig. 3 Convergence result of the algorithm ARE-GA for non-convex problem

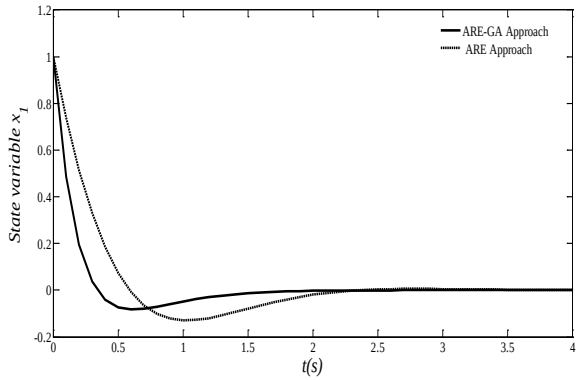


Fig. 4 State variable x_1 evolutions of the uncertain system obtained by a non-convex optimization (ARE-GA) and by a convex optimization (ARE) approaches

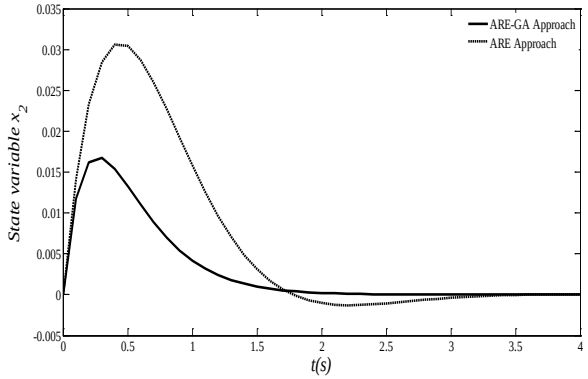


Fig. 5 State variable x_2 evolutions of the uncertain system obtained by a non-convex optimization (ARE-GA) and by a convex optimization (ARE) approaches

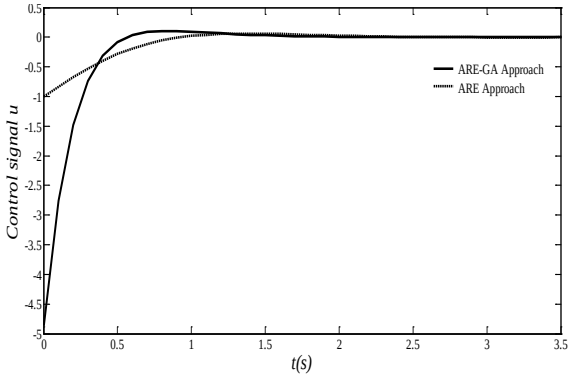


Fig. 6 Control signal u evolutions of the uncertain system obtained by a non-convex optimization (ARE-GA) and by a convex optimization (ARE) approaches

VI. Conclusion

In this paper, the proposed approach, based in the use of the ARE technique associated to GA, for solving a non-convex optimization problem, is formulated and tested, with success, and for the quadratic stabilizability study of uncertain linear systems. The comparison of these results with those obtained

by the use of ARE technique shows that the best performances, in terms of speed convergence, quality solutions and simplicity implementation and robustness, are obtained when the proposed ARE-GA is applied.

APPENDIX

Proof

Sufficiency

The system (7) is quadratically stabilizable by the control law $u(t) = Kx(t)$ if and only if it exists $P = P^T > 0$ such that [5], [20]

$$\left(A + DFE - BK\right)^T P + P\left(A + DFE - BK\right) < 0 \tag{14}$$

with

$$x^T \left(A^T P + PA - 2PBK + 2PDFE\right) x < 0 \tag{15}$$

$$\forall x \in \Re^n, x \neq 0$$

We have

$$\left(\sqrt{\varepsilon} D^T P - \frac{1}{\sqrt{\varepsilon}} FE\right)^T \left(\sqrt{\varepsilon} D^T P - \frac{1}{\sqrt{\varepsilon}} FE\right) \geq 0 \tag{16}$$

for ε >0 and then

$$\begin{aligned} &\varepsilon PDD^T P + \varepsilon^{-1} E^T E \geq \varepsilon PDD^T P + \varepsilon^{-1} E^T F^T FE \\ &\geq PDFE + E^T F^T D^T P \end{aligned} \tag{17}$$

For the solution $K = R^{-1} B^T P$, the inequalities (12) lead to(15).

Necessary condition

If the system is quadratically stabilizable then $\exists P_i > 0$ such that

$$x^T \left(A^T P_i + P_i A - 2P_i BK + 2P_i DFE\right) x < 0 \tag{18}$$

$$\forall x \in \Re^n, x \neq 0$$

This implies the following inequalities

$$x^T \left(A^T P_i + P_i A + 2P_i DFE\right) x < 0 \tag{19}$$

which are equivalent to

$$x^T \left(A^T P_i + P_i A - P_i BR_i^{-1} B^T P_i + 2P_i DFE\right) x < 0 \tag{20}$$

$$\forall x \in \mathbb{R}^n, x \neq 0$$

according to the theorem of Finsler [17], for matrix $R_i > 0$. Thus

$$\begin{aligned} &x^T \left(A^T P_i + P_i A - P_i BR_i^{-1} B^T P_i\right) x < -2 \max \\ &\left\{x^T P_i DFEx : F^T F \leq I\right\} \leq 0 \text{ for } x \in \mathbb{R}^n \end{aligned} \tag{21}$$

From (21), it comes

$$\begin{aligned} &\left[x^T \left(A^T P_i + P_i A - P_i BR_i^{-1} B^T P_i\right) x\right]^2 > \\ &4\left(\max\left\{x^T P_i DFEx : F^T F \leq I\right\}\right)^2 \end{aligned} \tag{22}$$

or [17]

$$\left(\max\left\{x^T P_i DFEx : F^T F \leq I\right\}\right)^2 = x^T P_i DD^T P_i xx^T E^T Ex \tag{23}$$

and therefore

$$\left[x^T \left(A^T P_i + P_i A - P_i BR_i^{-1} B^T P_i\right) x\right]^2 > 4x^T P_i DD^T P_i xx^T E^T Ex \tag{24}$$

If (16) is satisfied, then it exists real constant $\varepsilon_i > 0$ such as [17]

$$A^T P_i + P_i A - P_i BR_i^{-1} B^T P_i + \varepsilon_i P_i DD^T P_i + \frac{1}{\varepsilon_i} E^T E < 0 \tag{25}$$

For any matrix $R > 0$, it exists $\varepsilon^* > 0$, such as the condition

$$\frac{1}{\varepsilon^*} P_i BR^{-1} B^T P_i \geq P_i BR_i^{-1} B^T P_i \tag{26}$$

is satisfied for $\varepsilon \in \left(0,\ \varepsilon^*\right]$. By substituing the left-hand side of inequalities (25) in (26), it comes

$$A^T P_i + P_i A - \frac{1}{\varepsilon^*} P_i BR_i^{-1} B^T P_i + \varepsilon_i P_i DD^T P_i + \frac{1}{\varepsilon_i} E^T E < 0 \tag{27}$$

Dividing these inequalities by ε^*, for $P = P_i / \varepsilon^*$ and $\varepsilon = \varepsilon_i \varepsilon^*$, we obtain the desired result.

Remark

Based on the fact that if it exists a solution of equation (5) for $R_i > 0$ and given ε_i, there exist also for $\forall\ R > 0$ and a range of values of $\varepsilon \in \left(0,\ \varepsilon^*\right]$; it is proposed in [17] an iterative search strategy for computing $P > 0$, based on decreasing values of ε, i. e. an iterative algorithm to solve the Riccati equation

$$A^T P + PA - PBR^{-1} B^T P + \varepsilon PDD^T P + \varepsilon^{-1} E^T E + Q = 0 \tag{28}$$

The solution of this equation is independent of the choice of the positive definite matrix Q.

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for their helpful comments and suggestions to improve the original manuscript.

REFERENCES

[1] R. T. Rockafellar, "Lagrange multipliers and optimality", *SIAM Review* vol. 35, no. 2, pp. 183-238, 1993.

[2] E. Pérez, C. Arino, F. X. Blasco and M. A. Martinez, "Maximal closed loop admissible set for linear systems with non-convex polyhedral constraints", *Journal of Process Control, Science Direct*, vol. 21, no. 4, pp. 529–537, April 2011.

[3] H. Gao and T. Chen, "A Poly-quadratic Approach to Quantized Feedback Systems", *IEEE Conference on Decision and Control, Manchester, CA*, December 2006.

[4] A. Haddi and N. El Alami, "Stabilisation d'un microsatellite par approche quadratique", *4th International Conference on Sciences of Electronic, Technologies of Information and Télécommunications, SETIT*, March 2007.

[5] B. Barmish, "Necessary and Sufficient Conditions for Quadratic Stabilization of an Uncertain System", *Journal Optimization Theory Appl.*, vol. 46, no. 7, pp. 399-408, 1985.

[6] P. Khargonekar, I. R. Petersen and K. Zhou, "Robust stabilization of uncertain linear systems: Quadratic stability and H_∞ control theory", *IEEE Trans. on Autom. Contr.*, vol. 35 no. 3, pp. 356-361, 1990.

[7] A. Farag and H. Werner, "A Riccati genetic algorithms approach to fixed-structure controller synthesis", *American Control Conference*, Boston, Massachusetts, June, 2004.

[8] A. J. Rojas, "Closed-form solution for a class of continuous-time algebraic Riccati equations", *48th IEEE Conference on Decision and Control*, Shanghai, December 2009.

[9] S. Carrière, S. Caux and M. Fade, "Synthèse LQ pour le contrôle en vitesse d'un actionneur synchrone autopiloté accouplé directement à une charge mécanique incertaine", *Revue e-STA*, vol. 6, no. 1, 2009.

[10] D. Cheng, C. Martin and J. Xiang, "An Algorithm for Common Quadratic Lyapunov Function", *The 3rd World Congress on Intelligent Control and Automation*, Hefei, P. R, 2000.

[11] K. Somyot and P. Manukid, "Genetic-Algorithm-Based Fixed-Structure Robust H_∞ Loop Shaping Control of a Pneumatic Servosystem". *Journal of Robotics and Mechatronics*, vol. 16, no. 4, pp. 362–373, 2004.

[12] P. Borne and M. Benrejeb, "*Des algorithmes d'optimisation. La nature, source d'inspiration pour l'ingénieur*". L'essor des "métaheuristiques pour l'optimisation difficile" basées sur le comportement de la nature. L'ingénieur, (260), pp. 12–15, 2010.

[13] B-S. Chen and Y-M. Cheng, "A Structured-Specified H∞ Optimal Control Design for Practical Applications: A Genetic Approach", *IEEE Transactions on Control Systems Technology*, vol. 6, no. 6, pp. 707–718, 1998.

[14] D. E. Goldberg, "*Algorithmes génétiques : exploration, optimisation et apprentissage automatique*". Addison–Wisley, France, 1994.

[15] J. H. Holland, "*Adaptation in Natural and Artificial Systems*", University of Michigan Press, Ann. Arbor, 1975.

[16] M. Jungers, "Jeux différentiels LQ de Stackelberg avec une pondération temporelle commune", *Revue e-STA*, vol. 3, no. 4, 2006

[17] I. R. Petersen, "A Stabilization algorithm for a class of uncertain linear systems", *Systems and Control Letters*, vol. 8, pp. 351-357, 1987.

[18] O. Yeniay, "Penalty Function Methods for Constrained Optimization with Genetic Algorithms", Mathematical and Computational Applications, vol. 10, no. 1, pp. 45–56, 2005.

[19] I. Tseveendorj and D. Fortin, "Global Optimization and Multik-napsack: a percolation algorithm", *European Journal of Operational Research*, vol. 154, Issue 1 , pp. 46-56, April 2004.

[20] K. Zhou and P. Khargonekar, "Robust stabilization of linear system norm bounded time-varying uncertainty", *Systems and Control Letters*, vol. 10, pp. 17-20, 1988.

[21] *G. Garcia and J. Bernussou, "H_2* guaranteed cost design by output feedback *", SIAM J. of Control and Optimization, 1995*.

[22] W. Colminares, "*Sur la robustesse des systèmes linéaires incertains: Approche quadratique, retour de sortie*", Thèse Docteur-Ingénieur, Université Paul Sabatier de Toulouse, 1996.

[23] A. Zaafouri, A. Kochbati, J. Bernussou and M. Ksouri, "Robust Control of Uncertain Linear Systems: Quadratic Approach", *IEEE CESA'96 IMACS Multiconference, Symposium on Control, Optimization and Supervision*, Lille, vol. 2, pp. 1135-1140, July 1996.

[24] P. Luis Dias Peres, "*Sur la robustesse des systèmes linéaires: Approche par programmation linéaire*", PhD, Université de Paul Sabatier de Toulouse, 1989.

[25] D. Cheng, L. Guo and J. Huang, "On Quadratic Lyapunov Functions", *IEEE Transactions on Autom. Contr*., vol. 48, no. 5, May 2003.

[26] W. J. Mao and J. Chu, "Quadratic Stability and Stabilization of Dynamic Interval Systems", *IEEE Transactions on Autom. Contr.*, vol. 48, no. 6, June 2003.

[27] T. Hu, "Nonlinear Control Design for Linear Differential Inclusions via Convex Hull Quadratic Lyapunov Functions", *The 2006 American Control Conference*, Minneapolis, Minnesota, June 2006

[28] T. Hu, A. R. Teel and L. Zaccarian, "Non-quadratic Lyapunov functions for performance analysis of saturated systems", *IEEE Conference on Decision and Control, The European Control Conference 2005*, Séville, December 2005

[29] I. Tseveendorj, "*Conditions d'optimalité en optimisation globale: contributions à l'optimisation combinatoire*", PhD, Université de Versailles, St. Quentin en Yvelines, 2007

[30] B. Barmish, "Stabilization of Uncertain Systems via Linear Control", *IEEE Trans. on Autom. Control*, vol. AC 28, no. 8, pp. 848-850, August 1983

[31] S. El Hani and N. El Alami, "A parametric robust control of an electric power generator", *6th International Workshop on Electronics, Control, Measurement and Signals ECMS'03*, Liberec, 2003.

Khira Dchich was born in Gabés (Tunisia) in 1981. She obtained the Master degree in electrical engineering (Automatic and Informatics industrial) from the High School of Sciences and Techniques of Tunis (ESSTT) in 2005. She received the Master's Degree in automatic at the same university in 2008. Research electrical engineering to C3S (ESSTT) . She is currently Assistant Professor contractual at the Faculty of Science of Bizerte. Her research interests are synchronous machine with permanent magnets, extended Kalman filter, robust control, convex optimization and non convex, quadratic stability.

Abderrahmen Zaafouri was born in 1967 in Sidi Bouzid, Tunisia. He received a B.Sc. in electrical engineering from the High Normal School of Technical Education of Tunis (ENSET) in 1993. In 1995, he obtained the Master's degree in automatic control at the same university, and the Ph.D degree from the National Engineering School of Tunis in 2000, he joined the High School of Sciences and Techniques of Tunis (ESSTT) at Tunis University as an assistant professor. Now, he is a member of the Research Unit on Control, Monitoring and Safety of Systems (C3S) at ESSTT. His main research interests are approaches to robust control of uncertain systems: performances study.

Abdelkader Châari was born in Sfax, Tunisia, on November 25, 1957. He received his DEA degree in Automatics from the Ecole Normale l'Enseignement (ENSET), Tunis, Tunisia, in 1982, and his Ph.D. degree in Electrical Engineering with a focus on Self-adjustment Application for DC Motors from the Habilitation University, Tunis, Tunisia, in 2008. Since 1982, he has been with the School of Sciences and Techniques at Tunis, (Ecole Supérieure des Sciences et Techniques de Tunis (E.S.S.T.T.)) Tunis, Tunisia, as an Assistant Professor. He is presently responsible of a research unit on control, monitoring, and system safety (C3S: Commande, Surveillance fonctionnement des Systèmes) at E.S.S.T.T. His current research interests include the identification and control of nonlinear systems, robust estimation and robust filtering. Systems and Superieure de Technique de Tunis Identification and Control et Sûreté de fonctionnement des systèmes) at ESSTT. His current research interests include the identification and control of nonlinear systems, robust estimation and robust filtering.

Characteristics Analysis of Reflection and Transmission According to Building Materials in the Millimeter Wave Band

Byeong-Gon Choi, Won-Ho Jeong, Kyung-Seok Kim

***Abstract*—**Millimeter wave is the propagation of short wavelength and wide bandwidth. Since the advantages of the millimeter wave band include miniaturization, weight reduction of components and a lot of information transfer, it has become an alternative frequency band for next-generation mobile communications. Also, it is strongly affected by climate and geographic features. So, characteristics analysis of reflection and transmission is required to use the millimeter wave band effectively. In this paper, we measure and analyze reflection and transmission characteristics according to various building materials from 13GHz to 28GHz in the millimeter wave band. For reflection measurement, we measured received power reflected from materials with changing incidence and reflection angles. For transmission measurement, we measured received power penetrating materials with changing reception angle. From these measurement results, we analyzed reflection and transmission characteristics according to building material and change of angle from 13GHz to 28GHz in the millimeter wave band.

***Keywords*—**Reflection, Transmission, Millimeter wave, Building materials.

I. Introduction

THE millimeter wave band has a frequency range of 30 to 300GHz and a 1 to 10mm wavelength. Because of the difficulty in controlling it and the disadvantages of transmission loss, the utilization of millimeter waves has been low. However, since the advantages of the millimeter wave band include miniaturization, weight reduction in components, and a lot of information transfer, it has become an alternative frequency band for next-generation mobile communications. Also, the use of smart devices such as the smartphone and tablet PCs has exploded, so high-speed, broadband and high-definition communications has been required. As a result, the need for research into millimeter waves has been increasing [1], [2]. In particular, one essential research area on millimeter waves is characteristics analysis of reflection and transmission because it is strongly affected by climate and geographic features [1], [3].

Byeong-Gon Choi and Won-Ho Jeong are with the Smart Radio Communication System Lab, Department of Electrical and Electronic Engineering, Chungbuk National University, Cheongju, Chungbuk, Rep. of Korea. (e-mail: byung717@naver.com, whjeong@cbnu.ac.kr).

Kyung-Seok Kim is with the Smart Radio Communication System Lab., Department of Electrical and Electronic Engineering, Chungbuk National University, Cheongju, Chungbuk, Rep. of Korea. (corresponding author, phone: +82-10-8802-5823; e-mail: kseokkim@cbnu.ac.kr).

In this paper, we measure and analyze reflection and transmission characteristics according to various building materials from 13GHz to 28GHz in the millimeter wave band by using a signal generator, a spectrum analyzer and directional antennas. Also, we measure and analyze reflection and transmission characteristics according to change of angle by changing incidence and reflection angles in each reflection measurement and reception angle in transmission measurement. In each measurement, we measure received power and loss of reflection and transmission with comparison to a reference material. Then, we derive reflectance and transmittance according to the material and change of angle. From these results, we confirm reflection and transmission characteristics according to material and change of angle from 13GHz to 28GHz in the millimeter wave band.

This paper is organized as follows. Section II discusses reflection and transmission systems and the measurement procedure. Reflection measurement results are shown in Section III, and transmission measurement results are shown in Section IV. From these results, characteristics of reflection and transmission are analyzed in Section V. Finally, Section VI concludes this paper.

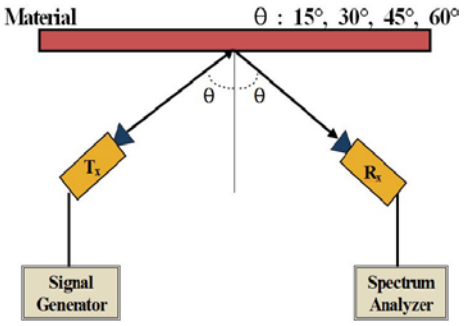
II. Reflection and transmission measurement Systems

As seen in Fig. 1, we compose reflection and transmission measurement systems by using a signal generator, a spectrum analyzer and directional antennas [4]-[8]. Also, we dispose a propagation absorber around the measurement system to prevent unwanted reflections. The materials used in this measurement are representative building materials, and the distance between material and antenna is 30cm. We set the material sizes as shown in Table I to prevent diffraction.

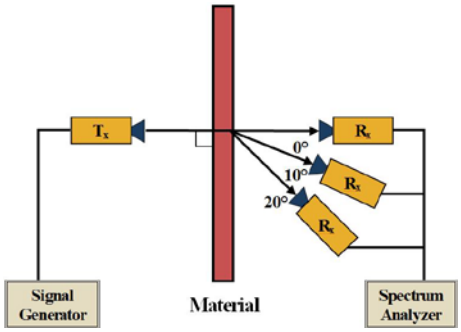
TABLE Ⅰ. The types of material

Material	Width (mm)	Height (mm)
Glass	600	610
Marble	600	400
Concrete	400	400
Particleboard	1300	600
Tile	400	248
Plasterboard	900	840
Wood	625	385

ISBN: 978-1-61804-290-3



(a)Reflection



(b)Transmission

Fig 1. Reflection and transmission measurement systems

To analyze reflection and transmission characteristics according to change of angle, we change incidence and reflection angles for each reflection measurement. Also, we fix the transmission angle at vertical to the materials, and change reception angle in each transmission measurement.

In reflection measurement, if incidence and reflection angles are below 5°, propagation reflected from materials is prevented by propagation transmitted from the transmission antenna. Also, if incidence and reflection angles are greater than 75°, propagation is transmitted to the reception antenna directly, not reflected from the materials [5]. So, we restrict incidence and reflection angles to between 15° and 60° with 15° resolution. In transmission measurement, if the reception angle is greater than 30°, propagation is not received at the reception antenna, as seen in Fig. 2 regardless of frequency band and types of material.

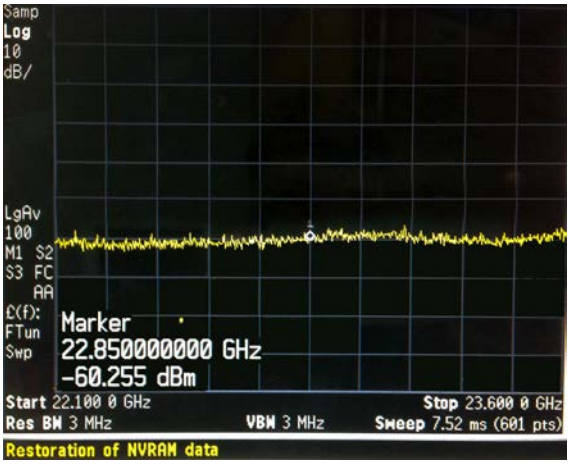


Fig 2. Transmission measurement result at a 30° reception angle

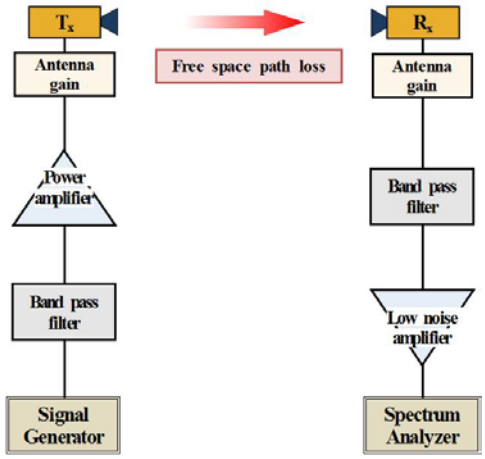


Fig 3. Measurement system

TABLE Ⅱ. Transmission power according to frequency

Frequency (GHz)	Free space path loss (dB)	System gain (dB)	Transmission power (dBm)
13.5	50.62	64.48	-50
22.85	50.62	54.37	-30
25.75	50.62	49.14	-30
28	50.62	47.46	-30

The waveform shown in Fig. 2 is the same when it is measured at the receiver without the transmitter. This is just the waveform of the noise level. So, we restrict reception angle to between 0° and 20° with 10° resolution.

In this paper, we measured reflection and transmission characteristics at 13.5GHz, 22.85GHz, 25.75GHz and 28GHz. In the measurement system in Fig. 3, because system gains are different according to frequency band, we set transmit power in each frequency band as shown in Table II before measuring reflection and transmission characteristics.

III. Reflection Measurement Results

In the reflection measurement results, the average received power readings at 13.5GHz, 22.85GHz, 25.75GHz and 28GHz are shown in Table III. We set a metal plate, where the reflection coefficient is assumed to be 1, as a reference to compare reflection characteristics according to the material [5], [6], [7], [9]. From this result, we calculated reflection loss by comparing received power for test materials and the reference material, showing them in Fig. 4.

In the results of received power in Table III and reflection loss in Fig. 4, there are no clear characteristics according to change of angle. We determined that this result is due the characteristics of the directional antenna used for measurement. In the reflection characteristics according to material, glass on average has the smallest reflection loss at -4.59dBm, and wood on average has the biggest reflection loss at -18.57dBm. In the

ISBN: 978-1-61804-290-3

TABLE Ⅲ. Received power from reflection measurement (dBm)

Material \ Degree	15°	30°	45°	60°
Reference	-39.95	-39.47	-39.33	-40.81
Glass	-44.22	-42.75	-42.82	-48.12
Concrete	-45.02	-45.56	-45.43	-47.34
Marble	-46.86	-47.80	-49.22	-48.67
Tile	-48.47	-47.50	-48.89	-49.50
Particleboard	-50.65	-51.07	-50.91	-51.46
Plasterboard	-51.50	-52.00	-50.62	-52.95
Wood	-56.89	-60.00	-58.81	-58.12

TABLE Ⅳ. Received power in transmission measurement (dBm)

Material \ Degree	0°	10°	20°
Reference	-35.47	-44.60	-51.67
Glass	-36.46	-46.01	-52.62
Tile	-38.08	-46.88	-54.05
Plasterboard	-39.71	-47.14	-55.09
Particleboard	-40.51	-49.13	-54.03
Marble	-41.47	-49.59	-54.72
Wood	-48.81	-53.01	-58.83
Concrete	-59.71	-59.66	-59.75

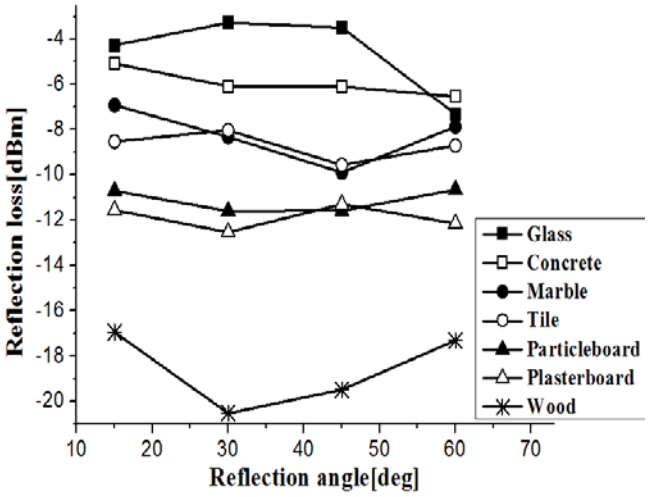


Fig 4. Reflection loss according to material

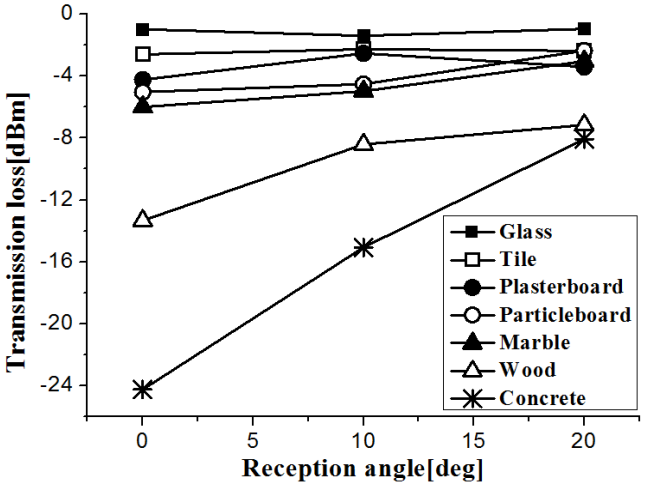


Fig 5. Transmission loss according to material

reflection measurement results, material that has a smooth surface has a smaller reflection loss.

IV. Transmission Measurement Results

In the transmission measurement results, the average received power readings at 13.5GHz, 22.85GHz, 25.75GHz and 28GHz are shown in Table IV. We set line-of-sight (LOS), where the transmission coefficient is assumed to be 1, as a base reference to compare transmission characteristics according to material [6]-[8]. From the results, we calculated transmission loss by comparing received power of the test materials and the reference material, as shown in Fig. 5.

In the results for received power in Table IV, the biggest average received power was measured when reception angle is 0°. When we changed the reception angle, average additional attenuations of -6.98dBm at 10° and -12.57dBm at 20° were generated.

In the results of transmission loss in Fig. 5, the biggest average transmission loss was measured when the reception angle is 0°. Also, it measured from -0.99dBm to -24.24dBm, according to the material. When we changed the reception angle, transmission loss measured from -1.4dBm to -15.06dBm at 0° and from -0.96dBm to -8.08dBm at 10°. It was because that received power was more sharply reduced according to the change of reception angle in LOS condition. In the transmission characteristics according to material, glass has the smallest transmission loss, and concrete has the biggest transmission loss.

V. Characteristics Analysis of Reflection and Transmission

From the reflection and transmission measurement results, we derived reflectance and transmittance to analyze reflection and transmission characteristics.

To derive reflectance, we used (1). $P_{reference}$ means received power reflected from the metal plate, $P_{material}$ means received power reflected from the material [5].

$$R = P_{material} / P_{reference(metal\ plate)} \quad (1)$$

Reflectance according to material and change of incidence

ISBN: 978-1-61804-290-3

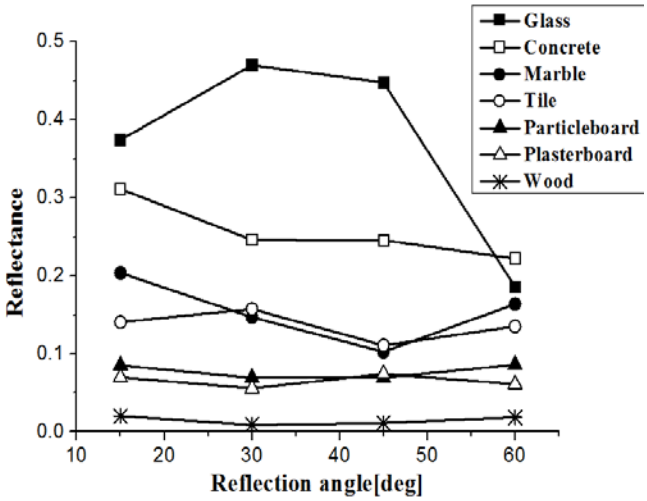


Fig 6. Reflectance according to material

and reflection angles is shown Fig. 6. In reflection measurement, when incidence and reflection angles are 30°, glass has the smallest reflection loss at -3.27dBm, and the biggest reflectance at 0.4699. On the other hand, when incidence and reflection angles are 30°, wood has the biggest reflection loss at -20.53dBm and the smallest reflectance at 0.0089. At the smallest reflectance, little propagation is received at the reception antenna. Almost all the reflectance from the materials measured between 0 and 0.5. So, there is a lot of loss when a millimeter wave is reflected from materials.

To derive transmittance, we used (2). $P_{reference}$ means received power in LOS, $P_{material}$ means received power penetrating the material [5].

$$T = P_{material} / P_{reference(LOS)} \quad (2)$$

Transmittance according to material and change of reception angle is shown Fig. 7. In transmission measurement, when the reception angle is 20°, glass has the smallest transmission loss at

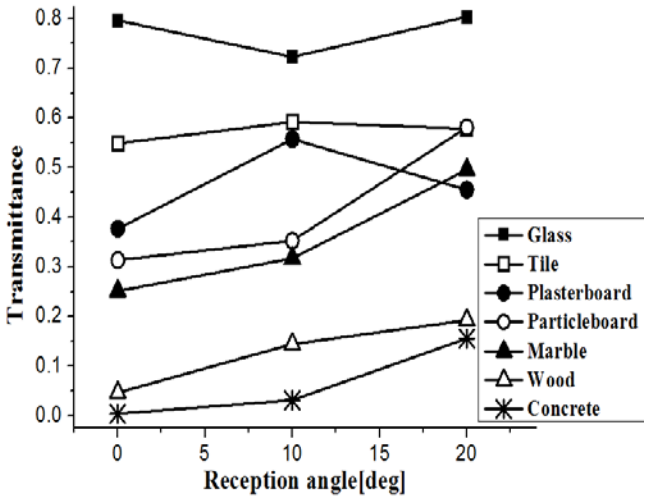


Fig 7. Transmittance according to material

TABLE Ⅴ. The maximum and minimum values of reflectance and transmittance according to material

Material	Reflectance		Transmittance	
	Max	min	Max	min
Glass	0.4699	0.1858	0.8035	0.7228
Tile	0.1574	0.1107	0.5916	0.5483
Plasterboard	0.0743	0.0058	0.5572	0.3767
Particleboard	0.0861	0.0692	0.5808	0.3133
Marble	0.2037	0.1026	0.4955	0.2512
Wood	0.0202	0.0089	0.1923	0.0463
Concrete	0.3112	0.2223	0.1556	0.0038

-0.96dBm, and the biggest transmittance at 0.8035. At the biggest transmittance, most of the propagation is received at the reception antenna. On the other hand, when reception angle is 0°, concrete has the biggest transmission loss at -24.24dBm, and the smallest transmittance at 0.0038. At the smallest transmittance, little propagation is received at the reception antenna. Almost all transmittance for the materials measured between 0 and 0.8. So, there is a lot of loss when a millimeter wave penetrates materials. But it is relatively small compared to the loss of reflection. Also, transmittance showed big differences according to material.

Table V shows the maximum and minimum values of reflectance and transmittance according to material. In reflection measurement, when a millimeter wave was reflected from the each material, more than half of propagation was lost due to attenuation. Especially, when a millimeter wave was reflected from some materials such as tile, plasterboard, particleboard and wood, most of propagation was lost due to attenuation. So, very small amount of propagation was received at the reception antenna. Also, in transmission measurement, when a millimeter wave penetrated some materials such as marble, wood and concrete, more than half of propagation was lost due to attenuation. Especially, when a millimeter wave penetrated wood and concrete, most of propagation was lost due to attenuation. So, very small amount of propagation was received at the reception antenna. In both reflection and transmission results, wood made much attenuation, and glass made little attenuation. Also, a millimeter wave was strongly affected by reflection than transmission relatively.

VI. Conclusion

In this paper, we measured and analyzed reflection and transmission characteristics according to materials and change of angle from 13GHz to 28GHz in the millimeter wave band. In each case, we measured received power and loss of reflection and transmission with comparisons to a reference material. And

ISBN: 978-1-61804-290-3

then, we derived reflectance and transmittance.

In the reflection measurement results, there are no clear characteristics according to change of angle. We determined that this is due to the characteristics of the directional antenna used for measurement. As for reflection characteristics according to material, glass has the smallest reflection loss and the biggest reflectance. On the other hand, wood has the biggest reflection loss and the smallest reflectance. Generally, material that has a smooth surface has a smaller reflection loss.

In the transmission measurement results, the biggest transmission loss was measured when the reception angle is 0°, and the smallest transmission loss was measured when the reception angle is 20°. As for transmission characteristics according to material, glass has the smallest transmission loss and the biggest transmittance. On the other hand, concrete has the biggest transmission loss and the smallest transmittance.

In both reflection and transmission results, there is a lot of loss when a millimeter wave reflects from and penetrates materials. In particular, when a millimeter wave reflects from the material, a lot of loss is generated.

From these results, we confirmed reflection and transmission characteristics according to material and change of angle from 13GHz to 28GHz in the millimeter wave band. As for future research, the research about characteristic analysis in millimeter wave band regarding of indoor and outdoor condition is need because millimeter wave is strongly affected by climate features. Moreover, additional research into different environments for measurements, such as antenna type, the size of the material and the frequency bands, will help to utilize the millimeter wave band and build communications systems effectively.

REFERENCES

[1] Dooyoung Youn, “Concepts and trends in millimeter-wave”, KIDSI Information and Communications Policy, August 2006.

[2] Seung Eun Hong et al., “A standardization trend of millimeter-wave wireless transmission technology for base on the 60GHz frequency band”, TTA Journal, July 2010

[3] Christian Jansen, Radoslaw Piesiewicz, Daniel Mittleman, Thomas Kurner and Martin Koch, “The Impact of Reflections From Stratified Building Materials on the Wave Propagation in Future Indoor Terahertz Communication Systems”, IEEE Transaction on Antennas and Propagation, May 2008

[4] Fricke, A et al., “Reflection and transmission properties of plastic materials at THz frequencies”, 2013 38th International Conference on Infrared, Millimeter, and Terahertz Waves(IRMMW-THz), 2013

[5] Javad Ahmadi-Shokouh, Sima Noghanian and Hengameh Keshavarz, “Reflection Coefficient Measurement for North American House Flooring at 57-64GHz”, IEEE Antennas and Wireless Propagation Letters, October 2011

[6] R.Piesiewicz, R.Geise, J.Schoebel and T.Kurner, “Reflection and Transmission Properties of Building Materials in W-band”, The Second European Conference on Antennas and Propagation, 2007

[7] Inigo Cuinas and Manuel Garcia Sanchez, “Measuring, modeling, and characterizing of indoor radio channel at 5.8GHz”, IEEE Transactions on Vehicular Technology, March 2001

[8] Hang Zhao et al., “28GHz millimeter wave cellular communication measurements for reflection and penetration loss in and around buildings in New York city”, 2013 IEEE International Conference on Communications(ICC), 2013

[9] Inigo Cuinas, David Martinez, Manuel Garcia Sanchez and Ana Vazquez Alejos, “Modelling and Measuring Reflection Due to Flat Dielectric Surfaces at 5.8GHz”, IEEE Transactions on Antennas and Propagation, April 2007

Extended Filtering for Self-Localization over RFID Tag Grid Excess Channels – II

Moises Granados-Cruz, Yuriy S. Shmaliy, and Sanowar H. Khan

***Abstract*—In the first part of this paper, we have modified the extended Kalman filter (EKF) algorithm and developed a new extended unbiased finite impulse response (EFIR) filtering algorithm for mobile robot self-localization over radio frequency identification (RFID) tag grid excess channels. In the second part, we provide simulations and show that redundant information captured from the tags allows increasing both the localization accuracy and system stability. The common factor here is that the number of tags required to increase accuracy is limited in the target nonlinear medium, by about six in our case. It is also shown that target state observation over the RFID tag excess channels allows mitigating effect of the imprecisely defined noise statistics on the EKF performance and preventing divergence in EKF.**

Keywords*—*RFID tag information grid, extended unbiased FIR filter, extended Kalman filter, self-localization.

I. Introduction

MOBILE robot self-localization over the radio frequency identification (RFID) tag information grids require highest accuracy. In such grids, information delivery is typically combined with sensing and wealth of information captured critically depends on accuracy of the information receptor localization.

To increase the localization accuracy, various modifications are exploited of the Kalman filter (KF) [1]–[5], particle filter (PF) [6]–[8], and unscented KF (UKF) [9]. Algorithms utilizing the extended Kalman filter (EKF) require white noise approximation as well as known noise statistics, initial conditions, and initial error statistics in order for the EKF to be suboptimal. Otherwise, accuracy provided by EKF may be low [10] and unacceptable for information grids. Another flaw is that EKF can be unstable and demonstrate divergence under the uncertainties [11] and large nonlinearities with intensive noise [12]. The problem of not exactly known noise statistics also arises in UKF, although this filter demonstrates better performance than EKF for highly nonlinear systems. The PF is free of many disadvantages peculiar to EKF. However, PF based on the Monte Carlo approach often requires large data and time and cannot always be used in real-time localization.

It is also known that the Gauss's least squares (LS) often give accuracy that is superior to the best available EKF [13]. Thus, methods of averaging implemented in LS and finite impulse response (FIR) filters may be more preferable. So, there is still room for discussion of the best estimator for RFID tag information grids. The FIR filter has been under the development for decades [11], [14]–[20]. It has been shown that this filter is more robust than KF under the unbounded disturbances [18]. The FIR filter is also lesser sensitive to noise [19] and produces smaller round-off errors [14] owing to averaging. Of practical importance is that complex optimal FIR (OFIR) structures [14], [15] do not demonstrate essential advantages against simple unbiased FIR (UFIR) ones [19] which ignore the noise statistics [11], [15]. The effect is due to averaging leveling the difference between OFIR and UFIR on large averaging horizons. The latter has made the UFIR filter a strong rival to the Kalman filter. Recently, the UFIR algorithm was developed in [20] to the extended UFIR (EFIR) algorithm following the same strategy as for the Kalman filter. First applications of the EFIR filter to localization problems [21], [22] have already shown some promising results. It was revealed [21] that the EFIR filter initiated by EKF can be much more successful in accuracy and stability in the triangulation-based localization. It was also noticed [22] that the EFIR filter has much stronger protection against divergency and instability than EKF in the RFID tag grid-based localization.

In the first part of this paper [23], we have discussed the extended filtering algorithms for self-localization over RFID tag grid access channels. To learn effect of redundant information captured from excess tags on the localization accuracy, below we consider a vehicle travelling on an indoor floorspace in different RFID tag environments. Each tag has a circular detection area with the detection range r. Both the short range tags ($r < 15$ cm) [1] and long range tags ($r < 15$ m) [2] can be used. In our case, we suppose that a vehicle has a reader which is able to measure distances to $k_n \geqslant 2$ tags at once employing the maximum RSSI given by the Friis equation $P_r = P_t \frac{\eta}{D_i^2}$ [1], [24], in which P_r is the received power, P_t is the transmitted power, D_i is a reader-to-(ith)tag distance, and η is a coefficient dependent on the transmitter and received antenna gains, system loss factor, and wavelength. A FOG installed on a vehicle directly measures a pose angle Φ_n.

All noise sources are supposed to be additive, stationary, zero mean, white Gaussian, and uncorrelated. Accordingly, we introduce the estimation error variances σ_x^2, σ_y^2, and σ_Φ^2 and specify the noise covariance matrix with the main diagonal $\mathtt{diag}\,\mathbf{Q} = [\,\sigma_x^2\ \sigma_y^2\ \sigma_\Phi^2\,]$ and all other components zeros. For the noise variances $\sigma_{\rm L}^2$ and $\sigma_{\rm R}^2$ in the inputs $d_{{\rm L}n}$ and $d_{{\rm R}n}$, we specify the relevant noise covariance matrix with the main diagonal $\mathtt{diag}\,\mathbf{L} = [\,\sigma_{\rm L}^2\ \sigma_{\rm R}^2\,]$ and all other

This investigation was supported by the Royal Academy of Engineering under the Newton Research Collaboration Programme NRCP/1415/140.

Moises Granados-Cruz and Yuriy S. Shmaliy are with the Department of Electronics Engineering, Universidad de Guanajuato, Mexico (e-mail: shmaliy@ugto.mx).

S. H. Khan is with the Department of Electronics Engineering, City University London, London, UK, e-mail: S.H.Khan@city.ac.uk.

components zeros. Finally, for the measurement noise variances σ_{vi}^2, $i \in [1, k_n]$, we specify the covariance $\mathbf{R}_n$ as $\mathtt{diag}\,\mathbf{R}_n = [\,\sigma_{v1}^2\ \sigma_{v2}^2\ \ldots\ \sigma_{vk_n}^2\ \sigma_{\phi}^2]$ with all other components zeros.

II. Self-Localization Using RFID Tags Nested on Indoor Boundaries

We first consider a scheme in which a vehicle is localized on an indoor floorspace with 12 tags mounted with equal intervals of 2 m on the indoor space boundaries (Fig. 1). We assume that a reader has range of $r = 6$ m and is thus able to detect simultaneously several tags at the signal-to-noise ratio (SNR) level of 10 dB. Two tags, Tag1 $(0, 0)$ and Tag4 $(0, 6\,\mathrm{m})$, are used to form "linear" measurements $\mathbf{y}_n$ as will be shown below. The state and input noise standard deviations are set as $\sigma_x = \sigma_y = \sigma_{\mathrm{L}} = \sigma_{\mathrm{R}} = 1$ mm and $\sigma_{\Phi} = 0.5°$. Note that, in [1], σ_x, σ_y, and σ_{Φ} are supposed to be zeroth. Following [8], we allow noise in the measured distances to have $\sigma_{v1} = \sigma_{v2} = 5$ cm and let $\sigma_{\phi} = 2°$. We also suppose that test measurements are available and thus the test trajectory $\mathbf{x}_n$ is known. Simulations are conducted at 5000 points.

In the nonlinear problem considered in [23], linear measurements of x_n and y_n are unavailable. Because a vector $\mathbf{y}_n$ combining linear measurements is required by the EFIR algorithm, we exploit distances D_1 and D_2 measured to tags T1 and T4 which have only one nonzero coordinate $\mu_4 = 6$ m and solve the inverse problem to define "linear" measurements $\tilde{x}_n$ and $\tilde{y}_n$ of x_n and y_n as

$$\tilde{x}_n = \sqrt{A_{1n} - \frac{1}{4\mu_4^2}(\mu_4^2 - A_{2n} + A_{1n})^2}\,, \qquad (1)$$

$$\tilde{y}_n = \frac{1}{2\mu_4^2}(\mu_4^2 - A_{2n} + A_{1n})\,, \qquad (2)$$

where $A_{1n} = D_{1n}^2 - c_1^2$, $A_{2n} = D_{2n}^2 - c_2^2$, and $c_1 = c_2 = 1$ m. Here, c_1 and c_2 correspond to tags $T1$ and $T4$ respectively. We then unite $\tilde{x}_n$, $\tilde{y}_n$, and $\tilde{\Phi}_n$ in a vector $\mathbf{y}_n = [\,\tilde{x}_n\ \tilde{y}_n\ \tilde{\Phi}_n\,]^T$.

Provided accurate values of $\mathbf{x}_n$ via test measurements, a straightforward way to find N_{opt} is to minimize the MSE by N as [25]

$$N_{\mathrm{opt}} = \arg\min_{N} [\,\mathrm{tr}\,\mathbf{P}_{12}(N)\,]\,, \qquad (3)$$

where $\mathrm{tr}\,\mathbf{P}_{12}(N)$ means the trace of $\mathbf{P}$ defined by (21) in [23] with the third state removed, because the third state is defined in angular units while the first and second states are in meters. Using (3), we find $N_{\mathrm{opt}} = 89$. An example of measured vehicle location for a spiral trajectory is given in Fig. 1. Here, we also show the EKF and EFIR estimates of location for exactly known noise covariances $\mathbf{R}_n$, $\mathbf{Q}$, and $\mathbf{L}$, initial state $\hat{\mathbf{x}}_0 = \mathbf{y}_0$, initial error $\mathbf{P}_0 = \mathbf{0}$, and $N_{\mathrm{opt}} = 89$. Because such conditions are "ideal", the estimates sketched in Fig. 1 are the best available by the EFIR filter and EKF. As can be seen, measurements are too rough here for straightforward localization. In turn, the estimates are quite accurate and consistent with each other. A more precise look at the localization errors is thus required. Below we consider several special cases.

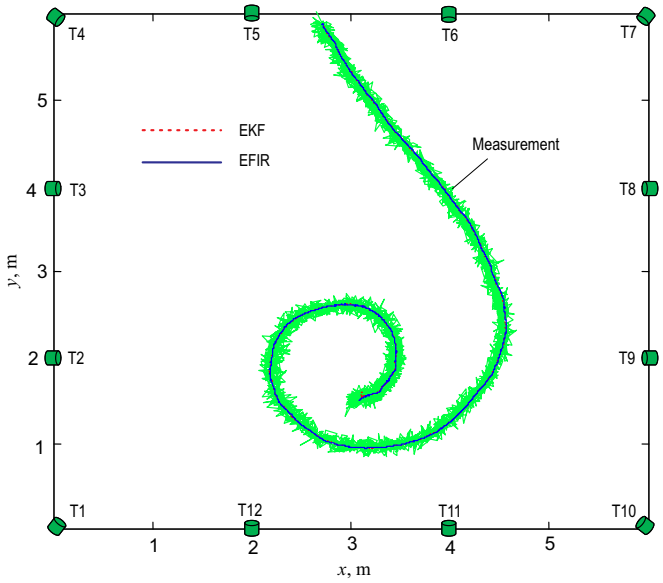


Fig. 1. EKF and EFIR filtering estimates of location of a vehicle travelling spirally on an indoor floorspace. The indoor space is equipped with 12 nested RFID tags. Measurement is obtained from Tag1 and Tag4. Localization is provided for exactly known noise statistics and $N_{\mathrm{opt}} = 89$.

TABLE I. Sets (k_n) of Observed SAW Tags (t) Corresponding to Fig. 2

k_n	t									
2	1	4	–	–	–	–	–	–	–	–
3	1	4	7	–	–	–	–	–	–	–
4	1	4	7	10	–	–	–	–	–	–
5	1	4	7	10	2	–	–	–	–	–
6	1	4	7	10	2	5	–	–	–	–
7	1	4	7	10	2	5	8	–	–	–
8	1	4	7	10	2	5	8	11	–	–
9	1	4	7	10	2	5	8	11	3	–
10	1	4	7	10	2	5	8	11	3	6

A. Fully Known Noise Statistics

To investigate effect of excess tags interacting with a target, we increase the reader sensitivity and add more detected tags as shown in Table I. For each of these cases, we repeat measurements and estimates 10 times and compute the localization errors by the trace of $\mathbf{P}_{12}(N_{\mathrm{opt}})$. The results are sketched in Fig. 2a and Fig. 2b for EKF and EFIR filter respectively along with the average errors. As can be seen, the localization errors reach peak-values in both filters by $k_n = 3$, undergo reduction with $4 \leqslant k_n \leqslant 7$, and then become constant in average by further increase in k_n. What else flows from this simulation and previously published outcomes is the following:

- Extra tags interacting with a target create an environment for error reduction. In linear systems with white Gaussian noise, the output noise variance is reduced by the factor of k_n. In nonlinear systems such as that formalized in Fig. 1 in [23], the error reduction function is rather complex and the effect can be lesser pronounced. It implies a finite value for k_n which is about six in Fig. 1.

ISBN: 978-1-61804-290-3

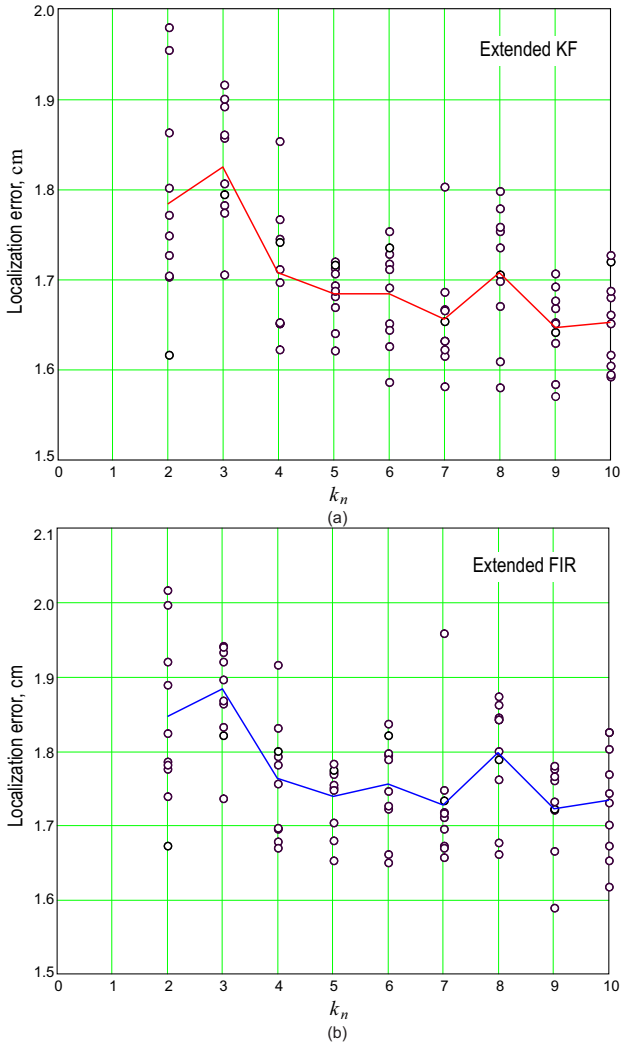


Fig. 2. Instantaneous (circles) and averaged (bold) localization errors $\sqrt{\mathrm{tr}\,\mathbf{P}_{12}(N_{\mathrm{opt}})}$ as functions of the number k_n of the observed tags in Fig. 2. Localization is provided for exactly known noise statistics and $N_{\mathrm{opt}} = 89$: (a) EKF and (b) EFIR filter.

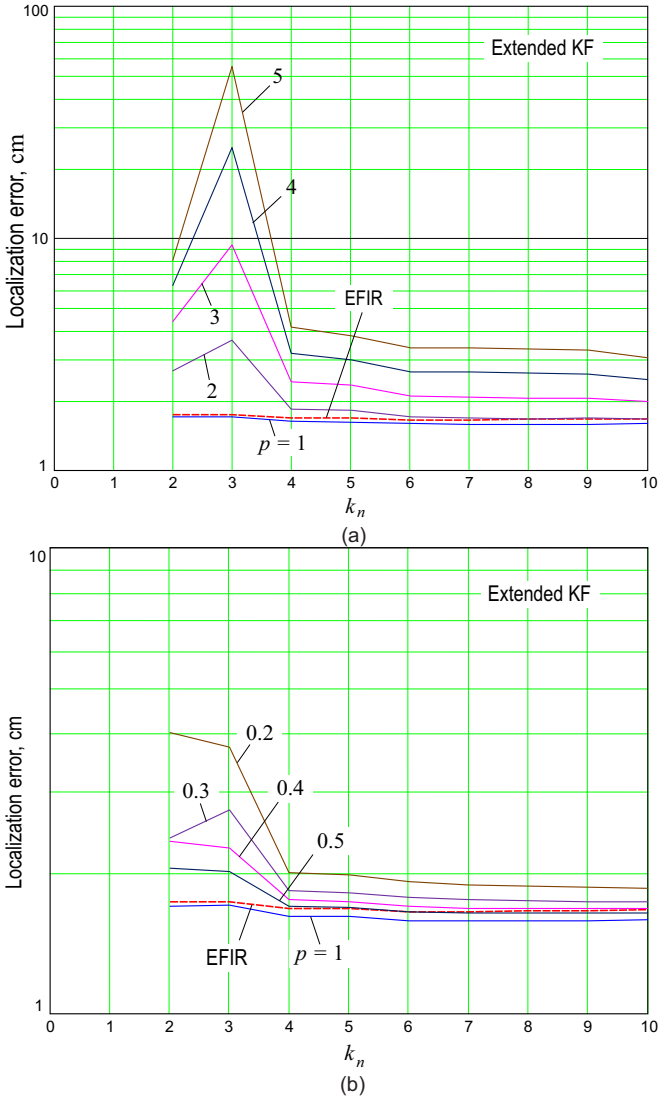


Fig. 3. Localization errors $\sqrt{\mathrm{tr}\,\mathbf{P}_{12}(N_{\mathrm{opt}})}$ in the EKF as functions of k_n and p corresponding to Fig. 2. EFIR estimates provided with $N_{\mathrm{opt}} = 89$ are dashed. Localization is obtained for $\hat{\mathbf{x}}_0 = \mathbf{y}_0$: (a) $p = 1, 2, 3, 4, 5$ and (b) $p = 0.2, 0.3, 0.4, 0.5$.

- Because noise in the extended model is not actually white, extended filtering techniques may not be very successful in accuracy due to the bias error.
- In white Gaussian medium generated for PF, hybrid structures such as PF/EKF and PF/EFIR may be more accurate owing to excess tags. An evidence for this statement can be found in [8] where the authors pointed out that the PF/EKF exploited in the short-range ($r = 50$ cm) tag environment has reduced the localization error from 1.2 cm to 0.52 cm by increasing k_n from 2 to 5.

B. Not Fully Known Noise Statistics

The noise statistics are typically not well-known to the engineer [10]. To evaluate a maximum effect of imprecisely defined $\mathbf{Q}$, $\mathbf{R}_n$, and $\mathbf{L}$ on the output accuracy, we introduce a correction coefficient p as $p^2\mathbf{R}_n$, $\mathbf{Q}/p^2$, and $\mathbf{L}/p^2$ [22] and compute the trace of $\mathbf{P}_{12}(N_{\mathrm{opt}})$ as function of p. The localization errors are shown in Fig. 3. As expected, the EKF is more accurate here than EFIR filter when $p = 1$ (ideal case). But the error difference between two filters is very small: about several mm. Some other useful observations can also be made:

- It is only when $0.5 < p < 2$ that errors in EKF are smaller than in EFIR filter. Moreover, EKF estimates do not seem to be acceptable with $p > 4$.
- Similarly to Fig. 2, error reduction is provided here by increasing k_n for any reasonable p.
- Error reduction is more noticeable when $p > 1$ (Fig. 3a) and lesser appreciable if $p < 1$ (Fig. 3b). For example, with $p = 4$ (Fig. 3a) the localization error of 24 cm by $k_n = 3$ reduces to 2.6 cm by $k_n = 6$.

ISBN: 978-1-61804-290-3

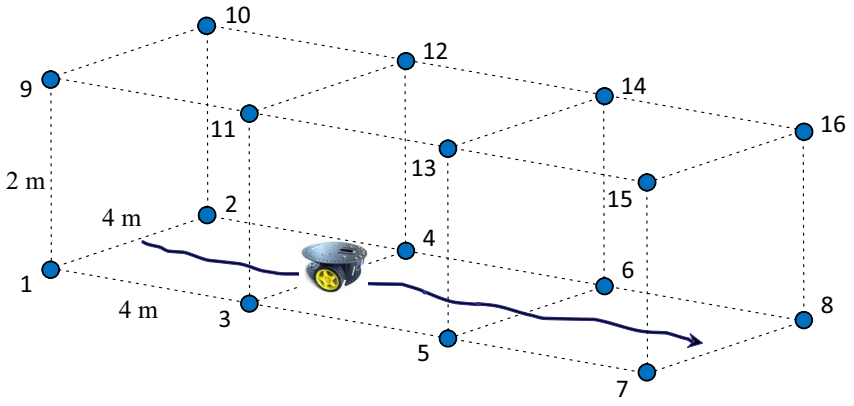


Fig. 4. Schematic diagram of a vehicle platform travelling on an indoor passway in the RFID tag environment with 8 tags mounted on a ceiling and 8 tags mounted on a floor.

III. Robot Travelling on a Passway

A most common organization of the RFID tag-based environment is a square grid with tags nested on a floor or/and ceiling [2]. This environment is simple, since the tag coordinates can easily be predicted based on its geometrical location. We construct such an environment in an indoor passway with the tags nested as in Fig. 4. We suppose that a reader is able to detect 8 tags at once at each floorspace point. It is also implied that not each tag is available at each floorspace point by some reasons (out of service, isolated by furniture, low power, etc.). However, at least two tags are always available.

The following noise statistics are allowed: $\sigma_x = \sigma_y = 1$ mm, $\sigma_L = \sigma_R = 0.1$ mm, and $\sigma_\Phi = 0.5^\circ$. Because the tags mounted on a ceiling are most far distanced from a vehicle platform, we set different measurement noise statistics as $\sigma_{v1} = \cdots = \sigma_{v8} = 5$ mm and $\sigma_{v9} = \cdots = \sigma_{v16} = 10$ mm. The noise standard deviation in measured Φ_n is set to be $\sigma_\phi = 2^\circ$. Using (3), we then find $N_{\text{opt}} = 12$.

In such an environment, simple inverse solution to the measurement equation is unavailable to form a "linear" measurement vector $\mathbf{y}_n$. We therefore run the EKF (Table I in [23]) with roughly set covariances ($p = 5$) and initial values (allowing errors of 10%) and use its output $\hat{\mathbf{x}}_n^{\text{EKF}}$ as $\mathbf{y}_n$ in the EFIR filter (Table II in [23]) on a horizon of first N_{opt} points. Thereby, we exploit the EFIR/Kalman algorithm.

Taking into account that not all 8 tags can be detected at each floorspace point, we voluntary change the observed tag-sets each 2 m along a passway referring to the nearest tags as in Table II. The instantaneous localization errors are sketched in Fig. 5 for the case of exactly known noise statistics. Even a quick look at this figure suggests that the EKF and EFIR estimates are almost identical so that the EFIR/Kalman filter ignoring the noise statistics is virtually as successful in accuracy as EKF. Minimal localization errors in Fig. 5a correspond to 5 detected tags (second interval) and maximal errors to 3 detected tags (third interval). Although this deduction is supported by Fig. 2 and Fig. 3, we admit that the effect may vary in other nonlinear models that requires further investigations.

TABLE II. RFID Tags Detected in 6 intervals (in m) Along the Passway Shown in Fig. 5. Tags 12, 14 – 16 were not Available.

m	Tag											
	1	2	3	4	5	6	7	8	9	10	11	13
0–2	x	x	–	–	–	–	–	–	x	x	–	–
2–4	x	x	x	–	–	–	–	–	x	x	–	–
4–6	–	–	x	x	–	–	–	–	–	–	x	–
6–8	–	–	x	–	x	x	–	–	–	–	–	x
8–10	–	–	–	–	x	x	–	–	–	–	–	x
10–12	–	–	–	–	–	–	x	x	–	–	–	–

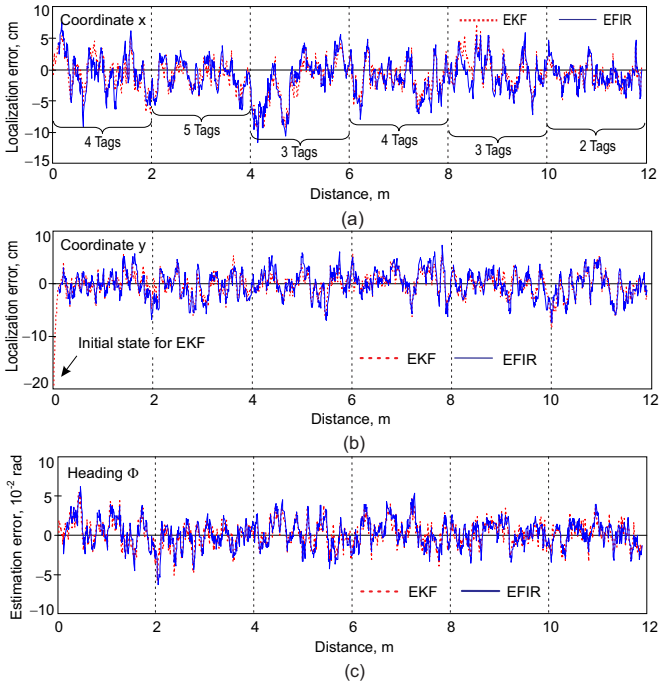


Fig. 5. Typical localization errors in 6 intervals specified in Table **??** for exactly known noise covariances and $N_{\text{opt}} = 12$: (a) coordinate x, (b) coordinate y, and (c) heading Φ.

ISBN: 978-1-61804-290-3

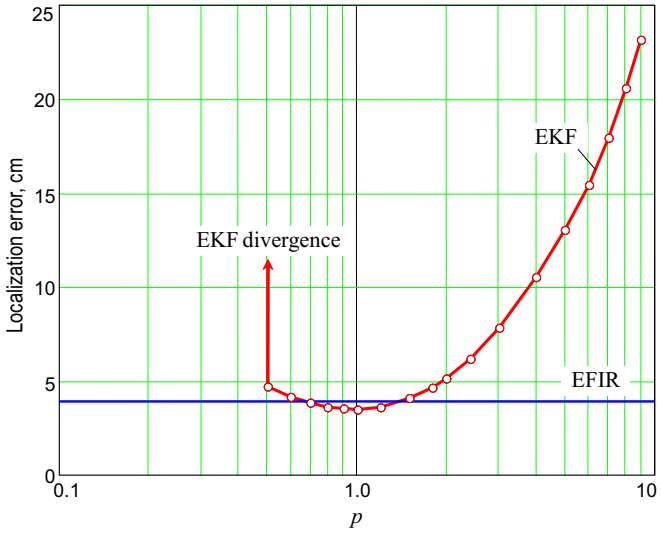


Fig. 6. Typical localization errors by EKF and EFIR filter as functions of p. The EFIR filter is p-invariant. The EKF diverges when $p < 0.5$.

TABLE III. AVERAGE LOCALIZATION ERROR STATISTICS (EXCLUDED $N_{\text{opt}} = 12$ INITIAL POINTS OF TRANSIENTS) CORRESPONDING TO FIG. 5 FOR EFIR FILTER AND EKF.

	x, cm		y, cm		Φ, 10^{-4} rad	
	bias	σ	bias	σ	bias	σ
EFIR	-1.094	3.094	-0.235	1.954	25.381	176.3
EKF ($p = 3$)	1.372	5.426	-2.05	4.981	154.0	303.0
EKF ($p = 1$)	-0.866	2.702	-0.131	1.749	19.608	164.0
EKF ($p = 0.3$)	-0.702	6.188	0.055	3.488	8.386	316.2

A situation changes dramatically if to define the noise covariances imprecisely owing to typically insufficient knowledge about noise. As follows from Fig. 6, the EKF retains here a slight advantage in accuracy of several mm, but only within a narrow region of $0.7 < p < 1.3$. Beyond this region, errors in the EKF grow rapidly and the EFIR filter becomes more advantageous. As an evidence, Table III gives typical average errors for the EFIR filter ($N_{\text{opt}} = 12$) and EKF assuming $p = 3$, $p = 1$, and $p = 0.3$ with 10% of inaccuracy in the appointment of the initial state.

What was unexpected that the EKF will diverge with $p < 0.5$ (Fig. 6). The known causes of divergence in EKF are larger nonlinearities and intensive noise [12]. Figure 6 definitely points out at another *sourse of divergence – errors in the noise covariances*. To learn it in more detail, we repeat the estimates for $p < 0.5$ with different noise realizations and select three specific appearances shown in Fig. 7. One can recognize here local instabilities (Fig. 7a) which may not be treated as divergence. Figure 7b demonstrates a brightly pronounced single divergence between 4 m and 6 m. Note that, within this interval, the reader detects 3 tags (Table II) and the localization error is maximal (Fig. 2). Multiple divergence is illustrated in Fig. 7c and we notice that EKF diverges here every time when a target interacts with a small number of the tags. A conclusion one may arrive at is the following: *an increase in the number of tags interacting with a target*

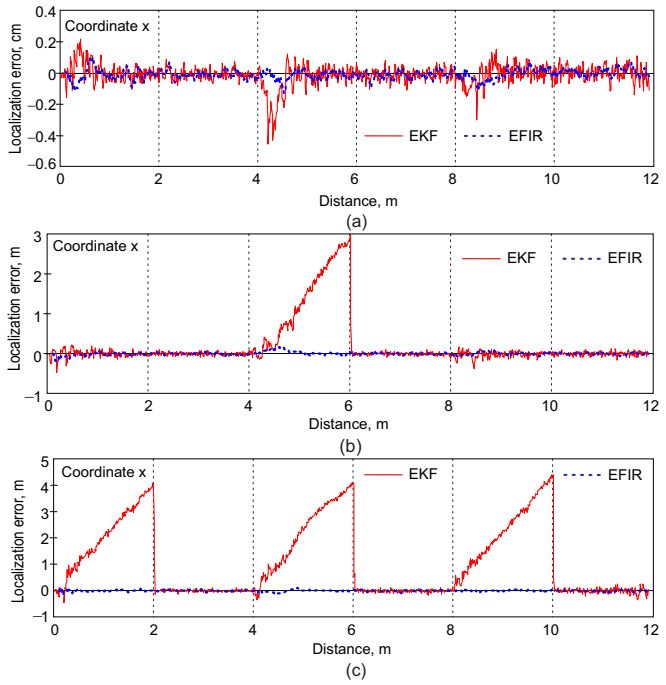


Fig. 7. Failures of EKF due to imprecisely defined noise covariances with $p < 0.5$: (a) local instabilities, (b) single divergence, and (c) multiple divergence.

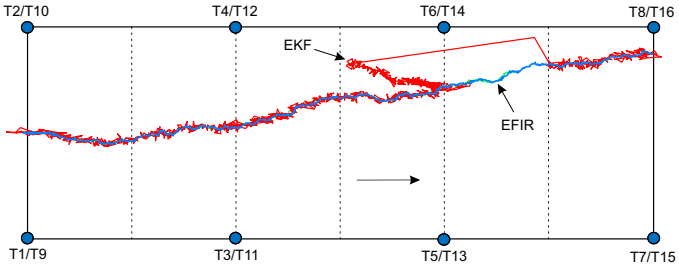


Fig. 8. Typical appearance of the EKF divergence between 8 m and 10 m with 3 tags (T5, T6, and T13) in a view. The EKF diverges due to imprecisely defined noise covariances with $p < 0.5$.

prevents the EKF divergence caused by imprecisely defined noise statistics.

Just to illustrate the EKF divergence in the x–y plane, in Fig. 8 we give an example of a single effect along the passway. Note that the EFIR/Kalman algorithm (Table II in [23]) is p-invariant and thus protected against such kind of failures. However, it may fail during the first N_{opt} points if the initiating EKF diverges in this interval as shown in [22].

IV. CONCLUDING REMARKS

In general, the EKF, EFIR, and EFIR/Kalman algorithms become more successful in accuracy by increasing the number of detected tags. However, this number (six in our case) is limited in the target nonlinear model. Besides, if the noise statistics are specified imprecisely with $p > 1$, then an increase in accuracy in EKF is most noticeable. It was also revealed that the EKF becomes addicted to divergence if $p < 1$. Hereby,

ISBN: 978-1-61804-290-3

we state that *deviations from the actual noise covariances in nonlinear state-space models may cause divergence in EKF*. Note that the EFIR filter is p-invariant and thus protected against such sort of failures. But, the EFIR filter requires about N times more computation time to complete iterations.

We finally notice that the EFIR/Kalman algorithm developed in this paper should be tested by uncertainties and non-Gaussian noise often peculiar to applications and compared to the PF and UKF. Divergence in EKF caused by errors in the noise covariances also needs further investigations.

REFERENCES

[1] S. S. Saab and Z. S. Nakad, "A standalone RFID indoor positioning system using passive tags," *IEEE Trans. Ind. Electron*., vol. 58, no. 5, pp. 1961–1970, May 2011.

[2] E. DiGiampaolo and F. Martinelli, "A passive UHF-RFID system for the localization of an indoor autonomous vehicle," *IEEE Trans. Ind. Electron.*, vol. 59, no. 10, pp, 3961–3970, Oct. 2012.

[3] V. Savic, A. Athalye, M. Bolic, and P. M. Djuric, "Particle filtering for indoor RFID tag tracking," in Proc. *IEEE Statist. Signal Process. Workshop (SSP)*, 2011, pp. 193–196.

[4] M. Boccadoro, F. Martinelli, and S. Pagnotelli, "Constrained and quantized Kalman filtering for an RFID robot localization problem," *Auton. Robots*, vol. 29, no. 3-4, pp. 235–251, Nov. 2010.

[5] J. Pomarico-Franquiz, M. Granados-Cruz, and Y. S. Shmaliy, "Self-localization over RFID tag grids excess channels using extended filtering techniques," *IEEE J. of Selected Topics in Signal Process.*, vol. 9, no. 2, pp. 229–238, Mar. 2015

[6] S. Park and H. Lee, "Self-recognition of vehicle position using UHF passive RFID tags," *IEEE Trans. Ind. Electron*., vol. 60, no. 1, pp. 226–234, Jan. 2013.

[7] A. Howard, "Multi-robot simultaneous localization and mapping using particle filters," *Int. J. of Robotics Research*, vol. 25, no. 12, pp. 1243–1256, Dec. 2006.

[8] E. DiGiampaolo and F. Martinelli, "Mobile robot localization using the phase of passive UHF-RFID signals," *IEEE Trans. Ind. Electron.*, vol. 61, no. 1, pp. 365–376, Jan. 2014.

[9] F. Martinelli, "Robot localization: comparable performance of EKF and UKF in some interesting indoor settings," in Proc. *16th Mediterranean Conf. on Contr. Autom.*, Ajaccio, France, June 25-27, 2008, pp. 499–504.

[10] B. Gibbs, *Advanced Kalman Filtering, Least-Squares and Modeling*, New York: Wiley, 2011.

[11] Y. S. Shmaliy, "An iterative Kalman-like algorithm ignoring noise and initial conditions," *IEEE Trans. Signal Process.*, vol. 59, no. 6, pp. 2465–2473, Jun. 2011.

[12] R. J. Fitzgerald, "Divergence of the Kalman filter," *IEEE Trans. Autom. Control*, vol. AC-16, no. 6, pp. 736–747, Dec. 1971.

[13] F. Daum, "Nonlinear filters: beyond the Kalman filter," *IEEE Aerosp. Electron. Syst. Mag.*, vol. 20, no. 8, pp. 57–69, Aug. 2005.

[14] W. H. Kwon and S. Han, *Receding Horizon Control: Model Predictive Control for State Models*. London: Springer, 2005.

[15] Y. S. Shmaliy, "Linear optimal FIR estimation of discrete time-invariant state-space models," *IEEE Trans. Signal Process.*, vol. 58, pp. 3086–3096, Jun. 2010.

[16] A. M. Bruckstein and T. Kailath, "Recursive limited memory filtering and scattering theory," *IEEE Trans. Inf. Theory,* vol. IT-31, no. 3, pp. 440–443, May 1985.

[17] W. H. Kwon, Y. S. Suh, Y. I. Lee, and O. K. Kwon, "Equivalence of finite memory filters," *IEEE Trans. Aerospace Electron. Syst.*, vol. 30, no. 8., pp. 968–972, Jul. 1994.

[18] A. H. Jazwinski, *Stochastic Processes and Filtering Theory*, New York: Academic, 1970.

[19] Y. S. Shmaliy, "Unbiased FIR filtering of discrete time polynomial state space models," *IEEE Trans. Signal Process.*, vol. 57, no. 4, pp. 1241–1249, Apr. 2009.

[20] Y. S. Shmaliy, "Suboptimal FIR filtering of nonlinear models in additive white Gaussian noise," *IEEE Trans. Signal Process.*, vol. 60, no. 10, pp. 5519–5527, Oct. 2012.

[21] J. Pomarico-Franquiz, S. Khan, and Y.S. Shmaliy, "Combined extended FIR/Kalman filtering for indoor robot localization via triangulation," *Measurement*, vol. 50, pp. 236–242, Apr. 2014.

[22] J. Pomarico-Franquiz and Y.S. Shmaliy, "Accurate self-localization in RFID tag information grids using FIR filtering," *IEEE Trans. Ind. Inform.*, vol. 10, no. 2, pp. 1317–1326, May 2014.

[23] M. Granados-Cruz and Y. S. Shmaliy, "Extended filtering for self-localization over RFID tag grids excess channels – II," in Proc. *The 2015 Int. Conf. on Systems, Control, Signal Process. Informatics*, Barcelona, Spain, April 7-9, 2015.

[24] J. Zhou and J. Shi, "RFID localization algorithms and applications – a review," *J. Intell. Manuf.*, vol. 20, no. 6, pp. 695–707, Dec. 2009.

[25] F. Ramirez-Echeverria, A. Sarr, and Y. S. Shmaliy, "Optimal memory for discrete-time FIR filters in state space," *IEEE Trans. Signal Process.*, vol. 62, no. 3, pp. 557–561, Feb. 2014.

Optimal control of multi-missile system based on analytical method

Xing Liu, Yongji Wang, Shuai Dong, Lei Liu

***Abstract*—The minimum-time synchronous control problem of multi-missile system is considered in this paper. By designing the optimal controller like a bang-bang structure which has a single switch dependent on a time parameter, the states inputs can be expressed mathematically and analytically. With the specific expressions of the states inputs, the original time optimal problem is transformed into a standard nonlinear programming problem. After optimizing all the parameters via optimization tools, the minimum-time problem is solved. Compared with the numerical methods which should discrete the states and control inputs, the proposed strategy perform less computational burden and more precision. Simulation results demonstrate the efficiency and validity of the proposed method.**

***Keywords*—muti-vehicle, analytical method, synchronization, minimum-time optimal control, centralized control**

I. INTRODUCTION

OVER the past few years, the research of cooperative control of multiple vehicles has attracted more and more attentions and efforts[1-3]. Some mature and technical methods have been employed in the formation of spacecrafts[4-5], multiple unmanned aerial vehicles (UAVs)[6-8] and missiles[9-10]. Compared with one single vehicle, muti-vehicle can achieve more complex and diverse tasks within common communication network.

Since the increasing military demands for the attack and defense systems, a group of well-organized and low-cost missiles is confronting more advantages than a single one. And the cooperative fighting manner will definitely be the main pattern in the future battlefield. Therefore, it is significant to study on the research of cooperative control of multi-missile system (MMS).

The cooperative attack of MMS has two common manners: cooperative game and saturation attack. The concept of cooperative game is proposed in recent years and the study on it has become a very popular area for research. Shima[11] investigated the optimal cooperative pursuit and evasion strategies against a homing missile based on the optimal control theory. Rubinsky[12] researched the three-player conflict by differential game. And Andrey[13] proposed the cooperative differential games strategies for active aircraft protection from a homing missile. It is noted that the strategy of cooperative game requires the MMS should only take the manner of centralized control. However, either the centralized control[14] or decentralized control[15-19] can achieve saturation attack, which is the common and useful way of cooperative engagement. An MMS can destroy and pierce the defense system by attacking the target simultaneously. Jeon[14] proposed a homing guidance law called cooperative proportional navigation (CPN) for cooperative attack of multiple missiles based on centralized control. Zhao[16] proposed a cooperative guidance scheme where coordination algorithms and local guidance laws are combined together. Zhang[19] considered the cooperative interception of a moving target by multiple vehicles with tolerance of actuator or network failures.

This paper studies the MMS cooperative control that aims at a simultaneous attack on a static target and mainly considered the minimum-time synchronous control problem (MTSCP) of an MMS. The purpose is to obtain the minimum time when all the states of missiles are synchronous by optimal control inputs, which can reduce the time of cooperative attack. Since the decentralized control is not optimal due to lack of global information, the centralized control is applied here to solve the optimal control problem. Furthermore, the constraint of overload is considered owing to the restriction of the vehicle's structure. Then the boundary of control input defined in this paper is fixed. With the designed controller like a bang-bang structure which has a single switch that is dependent on a time parameter, the analytical expressions of state inputs are obtained. According to the expressions, we can get a region in the state space. And it is worth noting that the time of synchronization is shortest when the boundary of each missile's state region has a common point with the others. Based on the above principle, the original time optimal problem can be transformed into a standard nonlinear programming problem. Finally, by optimizing all the parameters existing in the analytical expressions, the minimum time and optimal control inputs of all the missions are obtained.

The proposed method can be applied to achieve the saturation attack and sequence attack at the same time. The sequence attack demands that the missiles attack the target with

This work was supported in part by the National Nature Science Foundation of China (No. 61203081 and 61174079), Doctoral Fund of Ministry of Education of China (No. 20120142120091), Fundamental Research Funds for the Central Universities of HUST (NO. 2013054), and Precision Manufacturing Technology and Equipment for Metal Parts (No. 2012DFG70640).

Corresponding author, Xing Liu is with Key Laboratory of Ministry of Education for Image Processing and Intelligent Control,School of Automation,Huazhong University of Science and Technology, Wuhan, China (e-mail: gliuxxin@gmail.com).

YongJi Wang (Invited-Dimitrova) is a professor at the School of Automation, Huazhong University of Science and Technology, Wuhan, China. (e-mail: wangyjch@mail.hust.edu.cn).

Shuai Dong is with the School of Automation, Huazhong University of Science and Technology, Wuhan, China. (e-mail: hustacds@hust.edu.cn).

a fixed time sequence. Therefore, the key to the optimal control strategy is to guarantee the synchronization of all the missies on a fixed time sequence. The mathematic descriptions of the MTSCP about both saturation attack and sequence attack are presenting in this paper.

The organization of this paper is as follows. Sec. II introduces the dynamic equation of missile agent and then presents the description of the MTSCP. Sec. III proposes a strategy that transforming the time optimal problem to a nonlinear programming problem. In Sec. IV an example is given to demonstrate the effectiveness of the proposed approach. Finally, Sec. V summarizes some conclusions of this article.

II. Problem Statement

The cooperative guidance of n similar missiles which are denoted as 1, 2…. n is considered in the article. Assume that the missile agent can only change the direction of the velocity and the speeds of all the missiles are a constant v. And then decompose the trajectory of missiles into motions on a longitudinal plane and a lateral plane, and design independent controllers. When the missile is close to a static target, it almost remains stagnant upon the lateral plane. Without loss of generality, we assume all missiles are located and move in a same longitudinal plane, and ignore the collision that occurs among the group of missiles [20]. The relative motion of missile i is shown in Fig.1, and the XOY coordinates are established, which can obtain the corresponding equation:

$$\begin{aligned} &\begin{bmatrix}\dot{x}\\ \dot{y}\end{bmatrix}=\begin{bmatrix}v_x\\ v_y\end{bmatrix},\ \begin{bmatrix}\dot{v}_x\\ \dot{v}_y\end{bmatrix}=\begin{bmatrix}a_x\\ a_y\end{bmatrix}\\ &v=\sqrt{v_x^2+v_y^2},\ R=\sqrt{x^2+y^2}\\ &\begin{bmatrix}a_x\\ a_y\end{bmatrix}=a\cdot\begin{bmatrix}\dfrac{v_x}{\sqrt{v_x^2+v_y^2}}\\ -\dfrac{v_y}{\sqrt{v_x^2+v_y^2}}\end{bmatrix}\end{aligned} \tag{1}$$

where $[x,y]^T$ is the position vector, and $[v_x,v_y]^T$ is the velocity vector, $[a_x,a_y]^T$ is the acceleration vector, R is the distance scalar between the missile and the target and a is the acceleration scalar. Define λ and θ as the anticlockwise angles from the OX axis to the vector $[x,y]^T$ and $[v_x,v_y]^T$ respectively, and there exists

$$\lambda=\angle(x,y),\theta=\angle(v_x,v_y) \tag{2}$$

Because the guidance law a is difficult to design based on the model (1) and the sensor assembled on the missile can only measure the states of R, λ and θ, the model below is adopted to design the controller

$$\begin{aligned}&\dot{R}=-v\cos(\lambda-\theta)\\ &\dot{\lambda}=\frac{v\sin(\lambda-\theta)}{R}\\ &\dot{\theta}=-\frac{a}{v}\end{aligned} \tag{3}$$

Due to the restriction of each vehicle's structure the value of vehicle's overload has a boundary: $|a|\le n_x$. Then according to (3) there is

$$-\frac{n_x}{v}+\frac{v\cdot\sin(\lambda-\theta)}{R}\le\frac{a}{v}+\frac{v\cdot\sin(\lambda-\theta)}{R}\le\frac{n_x}{v}+\frac{v\cdot\sin(\lambda-\theta)}{R} \tag{4}$$

Since the value of $\dfrac{v\cdot\sin(\lambda-\theta)}{R}$ is tiny that can be ignored compared with $\dfrac{n_x}{v}$, there exists

$$-\frac{n_x}{v}\le\frac{a}{v}+\frac{v\cdot\sin(\lambda-\theta)}{R}\le\frac{n_x}{v} \tag{5}$$

Assume $r=R/v$, $\varsigma=\lambda-\theta$ and, $u=\sin\varsigma/r+a/v$, so the equation (3) can be rewritten as

$$\begin{aligned}&\dot{r}=-\cos\varsigma\\ &\dot{\varsigma}=u\\ &|u|\le\bar{u}\end{aligned} \tag{6}$$

where $\bar{u}=n_x/v$. n_x, v and $\bar{u}$ are all constants.

Consider the cooperative guidance of n similar missiles and the missile i has the following dynamic equation

$$\begin{cases}\dot{r}_i=-\cos\varsigma_i\\ \dot{\varsigma}_i=u_i\\ -\bar{u}_i\le u_i\le\bar{u}_i\end{cases} \tag{7}$$

The initial states of missile i are $r_i(0)=r_{i0}$ and $\varsigma_i(0)=\varsigma_{i0}$. It is obvious that there exist $\varsigma_i\in(0,\pi/2)$ and $r_i\ge 0$ according to the definitions of the states. Furthermore, we assume all the control inputs have the same boundary $\bar{u}$, so there has $\bar{u}_i=\bar{u}$.

This paper aims to design a centralized control for a simultaneous attack on a static target, so the description of the MTSCP is

$$\begin{aligned}&\min_{u_i\in[-\bar{u},\bar{u}]}t_f\\ &s.t.\quad (1)\ r_1(t_f)=r_2(t_f)=\cdots=r_n(t_f)\\ &\qquad (2)\ \varsigma_1(t_f)=\varsigma_2(t_f)=\cdots=\varsigma_n(t_f)\end{aligned} \tag{8}$$

where t_f is the time of synchronization, which is the objective of the MTSCP.

By designing the optimal controller u_i^*, the MTSCP can be transformed and solved analytically.

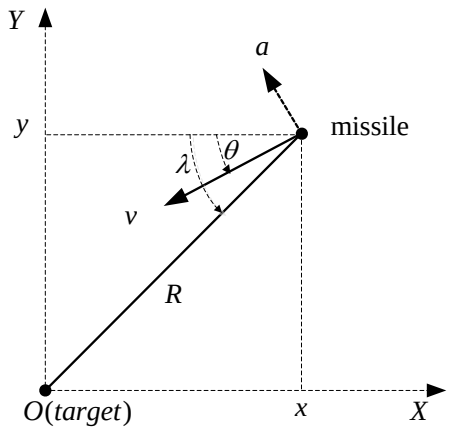


Fig.1 relative motion between the missile and the target

III. Optimal Controller Generated

According to (7), the states at the time t are expressed as followed

$$\begin{aligned} \varsigma_i(t) &= \varsigma_{i0} + \int_0^t u_i(\tau)d\tau \\ r_i(t) &= r_{i0} - \int_0^t \cos\varsigma_i(\tau)d\tau \end{aligned} \tag{9}$$

Assume the controller u_i of missile i is designed as

$$u_i(t) = \begin{cases} s\cdot\bar{u} & 0 \le t < t_c \\ -s\cdot\bar{u} & t_c \le t < T \end{cases} \tag{10}$$

Where $t_c \in [0,T]$ and $s \in \{-1,1\}$ .It is evident that the designed controller is similar to the bang-bang structure and the switch of it is the t_c .Substituting (10) into (9), one can have

$$\varsigma_i(t) = \begin{cases} \varsigma_{i0} + s\cdot\bar{u}\cdot t & 0 \le t < t_c \\ \varsigma_{i0} + 2s\cdot\bar{u}\cdot t_c - s\cdot\bar{u}\cdot t & t_c \le t < T \end{cases} \tag{11}$$

And the expression of r_i is

$$r_i(t) = \begin{cases} r_{i0} - \dfrac{1}{s\cdot\bar{u}}\sin(\varsigma_{i0} + s\cdot\bar{u}\cdot t) & 0 \le t < t_c \\ r_{i0} - \dfrac{1}{s\cdot\bar{u}}[2\sin(\varsigma_{i0} + s\cdot\bar{u}\cdot t_c) - \sin(\varsigma_{i0}) & t_c \le t < T \\ \qquad -\sin(\varsigma_{i0} + 2s\cdot\bar{u}\cdot t_c - s\cdot\bar{u}\cdot t)] & \end{cases} \tag{12}$$

It is obvious that the state $\varsigma_i(t)$ changes linearly with the time t . When the t_c varies from 0 to T , each value of t_c is corresponding to a final state $\varsigma_i(T)$, where $\varsigma_i(T) = \varsigma_i(t)\big|_{t=T}$. So at the time T , all the values of $\varsigma_i(T)$ form the range of $\varsigma_i(T)$ .Because of the condition $\varsigma_i \in (0,\pi/2)$ and (9), it is obviously noting that r_i is monotonically increasing function of ς_i .Therefore, the range of $r_i(T)$ can be obtained. Finally, the boundary of $(\varsigma(T), r(T))$ in the state space can be depicted by the analytical expressions (11) and (12).

For a single missile, assume the initial states are $\varsigma_0 = \pi/3$ and $r_0 = 30$ .The value of T is 10 and $\bar{u}$ is 0.05. Fig.2 shows the boundary of $(\varsigma(T), r(T))$ .And the three-dimensional trajectory showed in Fig.3 presents the boundary of $(\varsigma(t), r(t))$ at different time t .

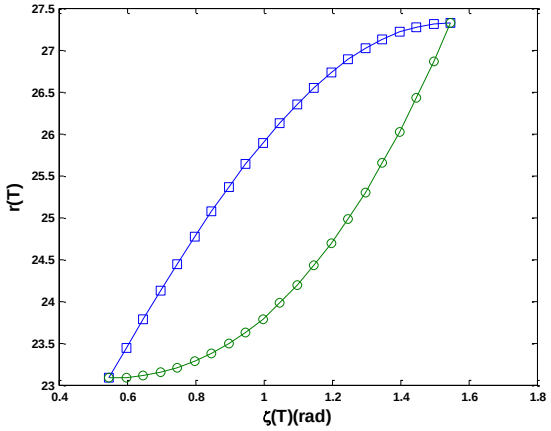


Fig.2 the boundary of $(\varsigma(T), r(T))$ when $T = 10$s

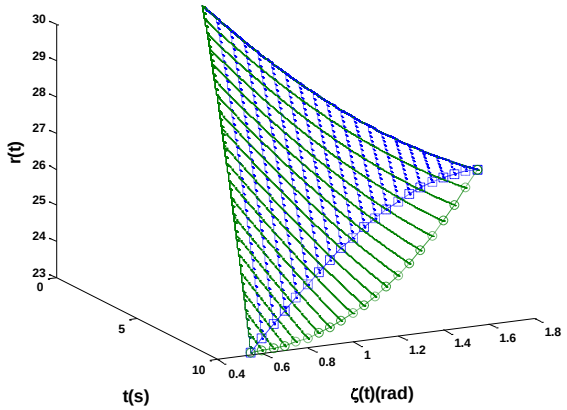


Fig.3 the boundary of $(\varsigma(t), r(t))$ at different time t

From Fig. 2 and Fig. 3, the conclusion is easily obtained that the time of synchronization t_f is shortest when boundary in the state space of each missile as showed in Fig.3 has a common point with the others, which means that the states of all the missiles are equal on the boundaries. Therefore, the expressions (11) and (12) can be applied to solve the optimal control problem (8) by transforming it into a nonlinear programming one. Then some optimization tools in MTLAB are employed here to optimize the parameters t_f , t_{ci} and s_i .Finally the optimal t_f^* and controller u_i^* are obtained.

In this work, the proposed method can achieve saturation attack and also sequence attack. Each problem is descripted as follows. The saturation attack means all the missiles move from the same time and achieve synchronization at the same moment. So the MTSCP statement of saturation attack is

$$\begin{aligned} &\min_{t_{ci}, s_i, t_f} t_f \\ s.t.\quad &1)\ r_1(t_f) = r_2(t_f) = \cdots = r_n(t_f) \\ &2)\ \varsigma_1(t_f) = \varsigma_2(t_f) = \cdots = \varsigma_n(t_f) \end{aligned} \tag{13}$$

The sequence attack requires all the missiles reach a

consensus on a fixed time sequence. Donate Δt_i as the time interval of missile i, which is a constant. So the MTSCP statement of sequence attack is

$$\min_{t_{ci},s_i,t_f} t_f$$

$$s.t.\quad 1)\ r_1(t_f)=r_2(t_f+\Delta t_1)=\cdots=r_n(t_f+\Delta t_1+\cdots+\Delta t_{n-1}) \quad (14)$$

$$2)\ \varsigma_1(t_f)=\varsigma_2(t_f+\Delta t_1)=\cdots=\varsigma_n(t_f+\Delta t_1+\cdots+\Delta t_{n-1})$$

Fig.4 clearly demonstrates the MTSCP of sequence attack. For the convenience of solving the MTSCP, the above statement can be transformed into another one that the missiles begin to conduct the cooperative guidance from the different time but achieve synchronization at the same time, which illustrates in Fig.5. Therefore, the MTSCP can be easily solved by analytical expressions (11) and (12).

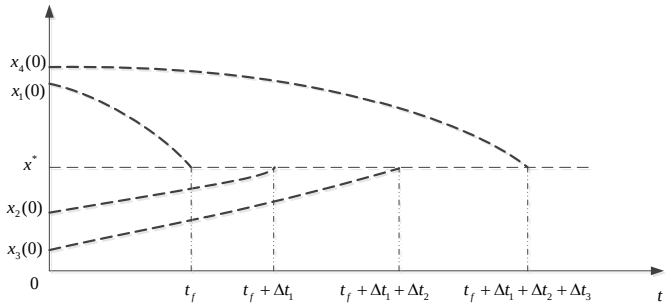


Fig.4 the MTSCP of sequence attack from the same time

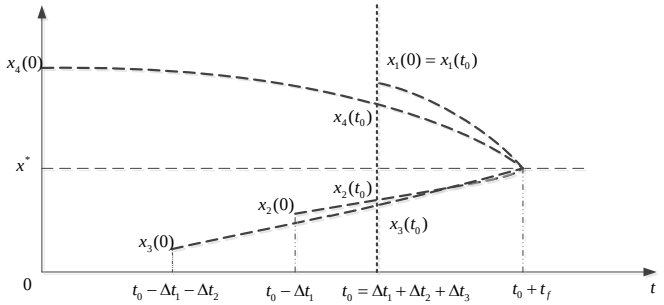


Fig.5 the MTSCP of sequential attack from the different time

IV. Nonlinear Simulation

In this section, an MTSCP of sequence attack is provided for simulation to demonstrate the effectiveness of the theoretical results in this article.

The cooperative attack of three missiles is considered in this example. The values of time intervals are $\Delta t_1=1$, $\Delta t_2=5$ and the controller boundary is $|u_i|\le 0.05$. Assuming that the initial states are $r_{10}=18$, $\varsigma_{10}=\pi/5.5$, $r_{20}=21$, $\varsigma_{20}=\pi/3$, $r_{30}=20$, $\varsigma_{30}=\pi/4$.

According to the above conditions and the proposed method, the results of simulation are obtained. And the optimal time to achieve synchronization is $t_f^*=6.9942s$. The values of other parameters are $t_{c3}^*=7.3502s$, $s_3^*=1$, $t_{c2}^*=7.7620s$, $s_2^*=-1$, $t_{c1}^*=6.4922s$ and $s_1^*=1$. Therefore, the specific expressions of all the three controllers are obtained.

Fig.8 shows the trajectories of the three optimal controllers $u_1^*(t)$, $u_2^*(t)$ and $u_3^*(t)$. It is clearly to notice that the optimal control inputs are similar to the bang-bang structure and the switch is dependent on t_{ci}. Trajectories of the state inputs $r_i(t)$ and $\varsigma_i(t)$ of all the three missiles are shown in Fig.6 and Fig.7 respectively. In Fig.9 the curves represent the boundaries of the state inputs at the time $t=t_f^*$.

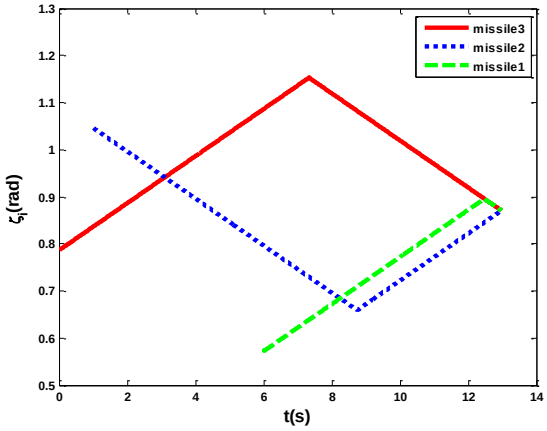


Fig.6 the trajectories of $\varsigma_i(t)$

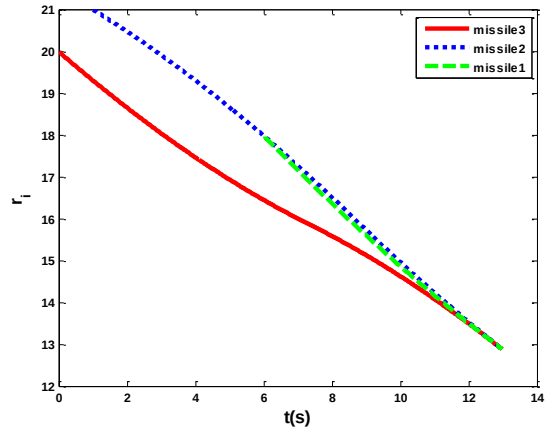


Fig.7 the trajectories of $r_i(t)$

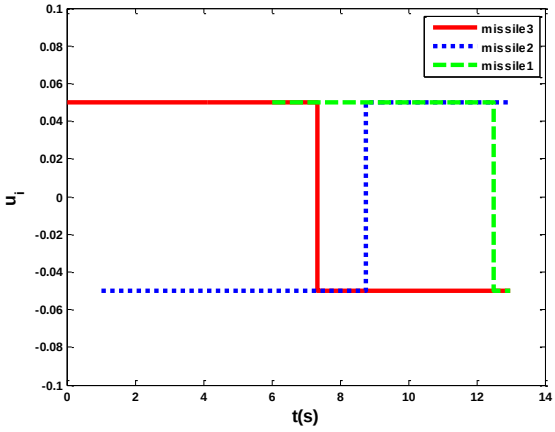


Fig.8 the trajectories of $u_i^*(t)$

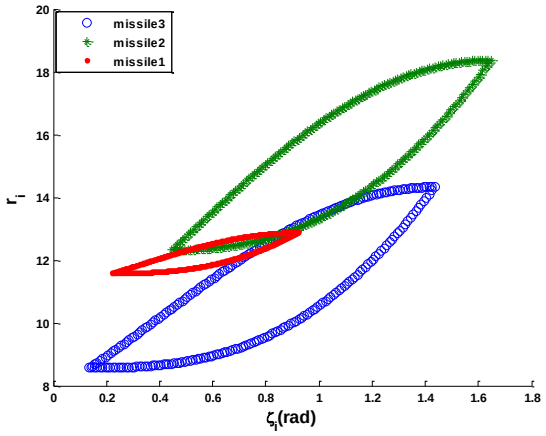


Fig.9 the boundaries of the three missiles when $t = t_f^*$

The trajectories of $r_i(t)$ and $\varsigma_i(t)$ in Fig.6 and Fig.7 illustrate that the states of the three missiles are synchronous for the first time at the time t_f^* , which proves the rationality of the results. Fig.9 shows that the state boundaries of three missiles have a common point exactly at the time t_f^* and it demonstrates this analytical method is significant and efficient.

V. Conclusion

The cooperative guidance problem with overload constraints based on centralized control is studied in this paper. The proposed strategy is applied to make the MMS achieve synchronization within minimum time. By designing optimal controller like a bang-bang structure which has a single switch dependent on a time parameter, the analytical expressions of state inputs are obtained. So the original time optimal problem can be transformed into a standard nonlinear programming one. Finally, the MTSCP is solved by optimizing all the parameters. This analytical method is capable to complete the saturation attack and sequential attack. Moreover, the results of simulation demonstrate the less computational burden and more precision compared with numerical solutions. In future study several factors such as the different boundary of each missile's overload and impact angle constraint will be considered for practical implementations. Furthermore, the proposed method will be expanded and completed theoretically to achieve the optimal cooperative guidance.

References

[1] Ratnoo, A. and T. Shima, "Formation-flying guidance for cooperative radar deception", *Journal of Guidance Control And Dynamics*, Vol.35, No.6, 2012, pp. 1730-1739.

[2] Campa, G., et al., "Design and flight-testing of non-linear formation control laws", *Control Engineering Practice*, Vol.15, No.9, 2007, pp. 1077-1092.

[3] Ousingsawat, J. and M.E. Campbell, "Optimal cooperative reconnaissance using multiple vehicles", *Journal of Guidance Control And Dynamics*, Vol.30, No.1, 2007, pp. 122-132.

[4] Bando, M. and A. Ichikawa, "Active formation flying along an elliptic orbit",*Journal of Guidance Control And Dynamics*,Vol.36, No.1,2013, pp. 324-332.

[5] Massari, M., F. Bernelli-Zazzera and S. Canavesi, "Nonlinear control of formation flying with state constraints",*Journal of Guidance Control And Dynamics*, Vol.35, No.6, 2012, pp. 1919-1925

[6] Ollero, A. and I. Maza, "Multiple Heterogeneous Unmanned Aerial Vehicles", Berlin, Springer-Verlag, 2007

[7] Dydek, Z.T., A.M. Annaswamy and E. Lavretsky, "Adaptive configuration control of multiple UAVs", *Control Engineering Practice*, Vol.21, No.8, 2013, pp. 1043-1052.

[8] Moon, S., E. Oh and D.H. Shim, "An Integral Framework of Task Assignment and Path Planning for Multiple Unmanned Aerial Vehicles in Dynamic Environments", *Journal of Intelligent & Robotic Systems*, Vol.7, No.1, 2013, pp. 303-313.

[9] Wei, C., et al., "Optimal formation keeping control in missile cooperative engagement", *Aircraft Engineering And Aerospace Technology*, Vol.84, No.6, 2012, pp. 376-389.

[10] Cui, N., et al., "Study on missile formation reconfiguration optimized trajectory generation and control", *Journal of Applled Machanics*, Vol.77, No.5, 2010, pp. 051501.

[11] Shima, T., "Optimal cooperative pursuit and evasion strategies against a homing missile",*Journal of Guidance Control And Dynamics*, Vol. 34, No.2, 2011, pp. 414-425.

[12] Rubinsky, S. and S. Gutman, "Three-player pursuit and evasion conflict", *Journal of Guidance Control And Dynamics*, Vol. 37, No.1,2014, pp. 98-110.

[13] Perelman, A., T. Shima and I. Rusnak, "Cooperative differential games strategies for active aircraft protection from a homing missile", *Journal of Guidance Control And Dynamics*, Vol.34, No.3, 2011, pp. 761-773.

[14] Jeon, I., J. Lee and M. Tahk, "Homing guidance law for cooperative attack of multiple missiles", *Journal of Guidance Control And Dynamics*, Vol.33, No.1, 2010, pp. 275-280.

[15] Weitz, L.A., J.E. Hurtado and A.J. Sinclair, "Decentralized cooperative-control design for multivehicle formations", *Journal of Guidance Control And Dynamics*, Vol.31, No.4, 2008, pp. 970-979.

[16] Zhao, S. and R. Zhou, "Cooperative Guidance for Multi-missile Salvo Attack", *Chinese Journal of Aeronautics*, Vol.21, No.6, 2008, pp. 533-539.

[17] Mclain, T.W. and R.W. Beard, "Coordination variables, coordination functions, and cooperative-timing missions", *Journal of Guidance Control And Dynamics*, Vol.28, No.1, 2005, pp. 150-161.

[18] Zheng, Y., Y. Zhu and L. Wang, "Consensus of heterogeneous multi-agent systems", *IET Control Theory and Applications*, Vol.16, No.5, 2011, pp. 1881-1888.

[19] Zhang, P., et al., "Fault tolerance of cooperative interception using multiple flight vehicles", *Journal of the Franklin Institute*, Vol. 350, No.9, 2013, pp. 2373-2395.

[20] Y. Wang, S. Dong, L. OU, and L. Liu, "Cooperative control of multi-missile systems", *IET Control Theory & Application*, to be published.

Biography:

Yongji Wang he received his PhD in Power Plant Engineering from Huazhong University of Science and Technology, China in 1990. He is currently a Professor in School of Automation at Huazhong University of Science and Technology. His research interests include neural network, system identification and control, flight vehicle control.

Relay Node Placement for Lost Connectivity Restoration in Partitioned Wireless Sensor Networks

Virender Ranga, Mayank Dave, Anil Kumar Verma

***Abstract*—** Due to low cost devices used in wireless sensor networks (WSNs), their applications in harsh surroundings, i.e. combat field reconnaissance, border protection, space exploration, etc. have become very common in the recent years. Due to unhealthy environments in which the network has to be operated sometimes result in large scale damage of the backbone nodes that causes the network to split into multiple disjoint segments. Placement of healthy relay nodes (RNs) as recovery nodes is the only way to connect the partitioned network, but the higher cost of RNs then becomes an addressable issue of their placement. With this motivation, we have suggested a new solution based on spiral format of Fermat points towards the centre of deployment called Restore Relay Lost Connectivity using CenTroiD (RRLC-CTD) for RNs' placements. The simulation results confirm the effectiveness of our proposed approach.

***Keywords*—** Connectivity Restoration, Nodes Failure, , Relay Node Placements, Spider Web-1C.

.

I. Introduction

The role of wireless sensor networks (WSNs) has become really useful in the real life in the recent years. The applications such as combat field reconnaissance, border protection, space exploration, etc. operate in the harshest environments, where sensor nodes reduce the danger of the human life [1, 2]. Since a sensor node is typically constrained in its energy, computational and communication resources, a large set of sensors is involved to ensure area coverage and increase the fidelity of the collected data. Due to small form factor and limited on board energy supply, a sensor is very susceptible to the failure. Due to hostile environments in which the network operates result in large scale damage of the nodes that causes network partitioning and converts into disjoint segments as shown in Fig. 1. For e.g., some sensors may be buried under snow or sand after the storm or in the field of battle, a component of the deployment area may be assaulted by the explosives and, thus a set of sensor nodes in the neighborhood would be ruined. Thus, repairing of large scale partitioned WSN is the latest hot research topic in the recent years. Deploying the RNs in the disconnected network is the solitary path to tie the large scale damaged network. RN is a more up to node with significantly more energy reserve and longer communication range than sensor nodes. Although RNs can, in principle, be equipped with sensor circuitry; mainly perform data aggregation and forwarding. Unlike sensor nodes, a RN may be mobile and has some navigation capabilities. The RNs is favored in the retrieval process, because these are easily to accurately place relative to the sensor nodes, and their communication range is even larger, which facilitates and expedites the connectivity restoration among the disjoint segments effectively and efficiently. Intuitively, RNs are more expensive and thus, minimum number of RNs should to be used for the recovery of the partitioned network. The minimum number of RNs can be found out using Steiner Minimum Tree (SMT), but it is shown to be NP-hard problem [3]. Therefore, some good heuristics are to be required to find the minimum number of deployed RNs in the partitioned network.

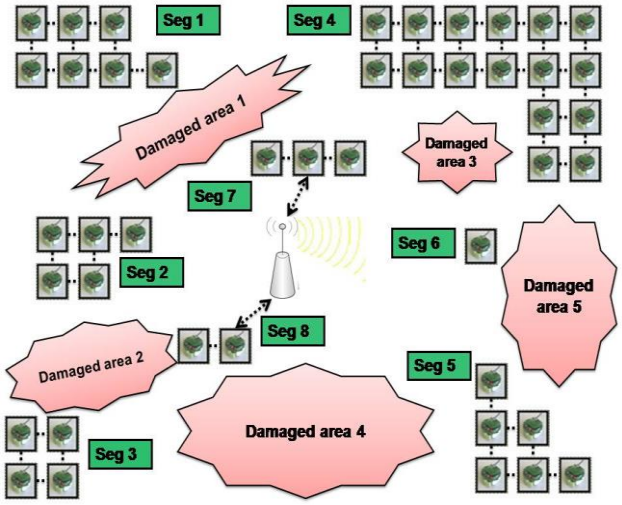


Fig. 1 Articulation of segmented area due to large scale failure of nodes

Two major RNs placement heuristics are published in the recent years. In the first approach, the authors proposed bio-inspired heuristic and use a spider web like RNs placement technique known as Spider Web-1C approach [3] with the

Virender Ranga is with the National Institute of Technology, Kurukshetra, Haryana, India (corresponding author to provide phone: 01744-233545; fax: 01744-233805; e-mail: virender.ranga@nitkkr.ac.in).

Mayank Dave is Professor in the Department of Computer Engineering, National Institute of Technology, Kurukshetra, Haryana, India (e-mail: m.dave@ieee.org).

Anil Kumar Verma is Associate Professor in Computer Science & Engineering Department, Thapar University, Patiala, (e-mail: akverma@thapar.edu).

segments situated at the perimeter of the network. The idea is to form the stronger connectivity, achieve better sensor coverage and enables balanced distribution of traffic load on the employed relays as well. One of the primary advantages of the Spider Web-1C approach that it plugs into the segmented network very efficiently and effectively, but the major issue is the big number of deployed RNs for the retrieval.

The second approach is based on SMT called ORC-SMT [4]. The idea is the use of SMT by considering three outer segments that are formed after applying the convex hull algorithm recursively in the cyclic fashion. The points thus obtained are then applied recursively to find more Steiner Minimum Point (SMP) for the RNs placements. The multiple points that come in the radio range of the node then become a single point for RN placement. In this manner, the process repeats itself till all the outer segments for which the run was made are not less than three. Then, RNs are placed on these points by applying Minimum Spanning Tree (MST) algorithm such as Kruskal or Prim algorithm. The main advantage of ORC-SMT that it connects various segments quickly and efficiently with a small number of RNs placements. In this research paper, we consider the similar type of problem, i.e. a large scale node failure with large number of partitions in the network. The distinction of our work as compared with previous proposed works is that in our proposed approach a large scale node's failure issue is addressed and solved by using of spiral format of Fermat points, which is not as yet proposed in other resolutions. However, in [13], the authors have considered Fermat point in the data propagation to reduce data transmission distance among the nodes to enhance network lifetime.

The remainder of the paper is organized as follows: In Section II, related work is described. Section III gives the problem statement of our proposed approach. Section IV explains the various approaches for comparison and section V shows performance evaluation of our proposed approach through simulations and compares with traditional approaches to prove its effectiveness. In section VI, the article is concluded with future scope.

II. Related Works

Many advances have been proposed till last one year to endure a large scale node failure in WSNs. The authors of [1, 2] have given the comprehensive survey of the network partitioning recovery approaches based on different standards. All approaches are classified into two broad categories: a) Centralized approaches, and b) distributed or semi-distributed approaches. The classification is further divided into three different categories, i.e. proactive, reactive and hybrid approaches. For proactive schemes, many approaches have been pursued to tolerate node's failure in the works. A similar method is applied for reactive and hybrid approaches. In all proposed approaches, controlled node mobility has been used to restore the partitioned network. For example, in [5], a robot called Packbot has been used to serve as a mobile RN. The use of robot enables the recovery of partitioned network, or break links. An algorithm is applied to determine the trajectory of moving robot in the network. A similar type of work is presented by Wang et al. [6]. The authors have used mobile RNs within 2-hop of the sink in the network to restore the partitioned network. Unlike [5], the idea is that RNs do not need to travel the long distance in the network. The use of Packbots and similar types of devices is inefficient due to unexpected delays in data delivery even multiple such devices are used in the network. The reason is the slow motion of devices to cover every individual best point in the network.

Wang et al. [7] exploited node controlled mobility in order to cover the coverage holes which are not covered by sensor nodes during their initial deployment. The idea behind this work is to identify some spare nodes from different parts of the network that can be relocated to coverage-hole places. Since moving a node for long distance can drain significant node power, a cascaded movement is proposed if the sufficient number of sensor nodes is available on the way.

Recently, some centralized/ semi-centralized approaches have been proposed to handle large scale failures in the network by using cascaded control mobility of the nodes. In [8], the recovery problem is formulated as an Integer Linear Program (ILP). The objective of the ILP based optimization model is to form a connected topology while minimizing the individual travelling distance of the nodes.

The author of [9] strives to restore connectivity by a Multi-Integer Linear Program (MILP) based on transportation network flow model. The idea is to restore the connectivity with minimum travelling distance of nodes with the assumption that every node should be able to go to all destinations i.e. reach to all other nodes when network connectivity restoration is to be required. Due to centralized in nature, this approach does not scale well.

Another approach is proposed by Sentruk et al. [10] to improve the scalability by reducing the number of candidate locations. A RN placement algorithm is used to find the set of locations which can guarantee the connectivity if RNs are to be deployed to these locations.

Vemulapalli et al. [11] described another distributed approach based on nodes' knowledge of full path to sink. The pre-failure route information is used to determine the location of the failed nodes. The location of nodes is obtained when paths are established. Thus, upon partitioning, nodes can attempt to re-establish the path towards sink node by moving to the next hop location. However, many nodes do the same in the partition, recovery cost can be high. To limit the recovery cost, the recovery process elects only one node as a leader node based on its distance from the failed node or sink. When the leader node moves towards sink, a cascaded movement of nodes within the partition is also required in order to sustain intra-segment connectivity.

Another approach based on game theory is proposed by Senturk et al. [12] by assuming the complete knowledge of the location of partitions, number of partitions, and failed nodes. Each partition is used as a player in the game. The payoff function is based on nodes' degree and elects a partition

representative (P_a) . The elected P_a opts to maximize the payoff of its partition which motivates the partitions to move each other. Due to the centralized nature of this approach, each representative node must know the payoff function of the other partitions and eventually network reaches to Nash equilibrium when all partitions are connected with each other.

III. SYSTEM MODEL AND PROBLEM STATEMENT

We assume a WSN in which a large number of sensor nodes are deployed throughout an area of interest and sink node is located in the middle of deployment. Without losing the generality, this assumption ensures that there is a balanced traffic load in the network. Due to the harsh environment of the application like in a battlefield, where sensor nodes could be destroyed by enemy explosives, thus causing a large scale node's failure which leads to multiple disjoint partitions in the network. For e.g., Fig. 1 shows the partitioned WSN with 8 segments having sink node is in the middle of the network. Thus, RNs are used to connect this disjoint network.

Our problem can be defined as follows: "*N sensor nodes that know their location using some localization algorithm are randomly deployed in an area of interest. Let us assume that j disconnected sub-networks are formed as a result of failure of a large scale nodes in the network. Each sub-network* G_i *has* n_i *sensor nodes where* $0 < n_i \ll N$ *. Our goal is to implement an algorithm that will ensure the lost connectivity among the disconnected sub-networks* G_i *by using minimum number of RNs placements and thus, create a new connected network* ".

IV. COMPARISON WITH SIMILAR SOLUTIONS

The following approaches are used to compare with our proposed solution, i.e. RRLC-CTD:

A. *Basic Deployment (BD):* It is a very basic approach in which the isolated segments apply the Graham Scan i.e. convex hull algorithm in the partitioned network. The outer segments then deploy the RNs along the borders of the convex hull in the circular fashion. Similarly, the inner segments deploy the RNs towards the nearest RN as shown in Fig. 2. It is assumed that all nodes know the complete topology of the network. The main issue of this proposed approach is the number of required RNs for the repairing of the disjoint network.

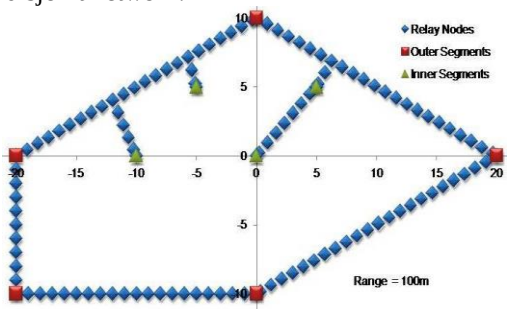


Fig. 2 RNs placements using BD approach

B. *Centre Deployment with Convex Hull (CDCH)*: In this approach, first we calculate the representative nodes like ORC-SMT. Then, all representative nodes apply the convex hull algorithm to find the outer segments. The obtained outer segments calculate the centre of mass (CoM). Further, each representative node of the outer segment deploys the RNs towards CoM. Moreover, the inner representative nodes deploy the RNs towards the nearest deployed nodes as shown in Fig. 3. The main advantage of CDCH is that it requires a small number of RNs placements as compared with the BD approach.

C. *Spider Web-1C heuristic:* The key idea behind Spider Web-1C deployment strategy is to place the RNs inwardly of the damaged area to yield better network connectivity and coverage. To balance the inter-segment path length in terms of the number of hops, RNs are placed toward the estimated CoM of the segments. Basically, from each partition to the CoM, Spider Web-1C has gradually deployed nodes until all the partitions are connected efficiently. In this way, it not only increases the total coverage of the network, but also reduces the possible number of cut vertices in the network as well. Before placing of RNs, Spider Web-1C first needs to identify the outer segments in the area of interest. To do this, it randomly picks the representative nodes from each partition and runs a convex hull algorithm. The convex hull algorithm returns a subset of representative nodes that sit on the corners of a convex polygon. After finding the convex polygon, it determines the CoM of the polygon. RNs are then deployed along the line between a segment and the CoM. Obviously, the relays around the CoM will be in the communication range of each other, and the segments then become connected. Fig. 4 shows the pictorial representation how to connect the disjoint network by using Spider Web-1C heuristic approach.

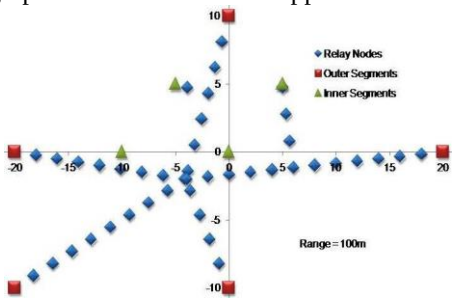


Fig. 3 Illustration of repairing of network using CDCH approach

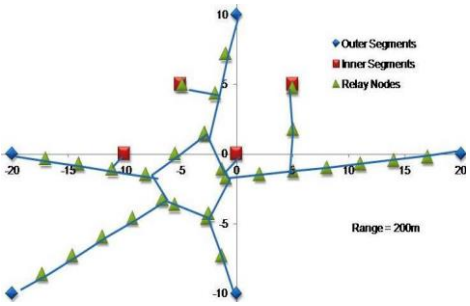


Fig. 4 RNs deployment in Spider Web-1C approach

ISBN: 978-1-61804-290-3

D. *Optimal Relay Node Placement Algorithm using Steiner Minimum Tree (SMT) on Convex Hull (ORC-SMT):* ORC-SMT pursues greedy heuristic that has two main phases: (1) identify the Steiner points (SPs) at which RNs would be placed with the objective of minimizing the number of deployed RNs to connect the segments, and (2) deploy additional RNs in order to form a fully connected inter SPs topology considering the communication range (r) of a RN. The first phase has further two main steps that are repeated until all the necessary SPs are calculated. In the first step, ORC finds the convex hull to identify the boundary segments. Then, the SPs that connect every three neighboring boundary segments are identified. These SPs, is called first tier SPs. For the unengaged segments, the convex-hull is again computed to identify boundary terminals (i.e. segments or first tier SPs) that are used in the second round and then the second tier SPs are found. The third tier SPs will be identified based on the second tier and so on. In other words, the two steps are repeated recursively for m rounds until the number of points considered for computing a convex hull is less than three or they form a complete graph in terms of communication range of a RN. ORC-SMT then switches to the second phase in which the identified SPs and segments are stitched together. Basically, every segment Seg_i identifies the closest SP and RNs get placed on the line from Seg_i to such SP. The same procedure applies for the first tier SP to connect them to the second tier and so on. As mentioned above, in the first phase, ORC operates in rounds. In the first round $(m = 0)$, ORC identifies a set of segments in the damaged area, which forms the smallest polygon that contains the other segments. Considering the segments as terminals, the convex hull of all segments is used to identify the boundary segments. The authors assume that there exist at least three non-collinear segments such that the convex hull ch_0 found in the first round forms a closed polygon as seen in Fig. 5. To find a convex hull the authors use the Graham Scan algorithm. The main advantage of this approach is that it requires small number of RNs to connect the large damaged area.

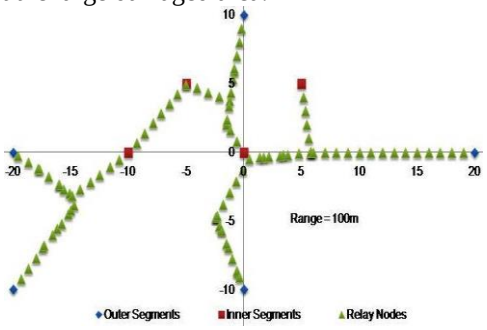


Fig. 5 Network partitioning recovery using ORC-SMT approach

E. *Restore Relay Lost Connectivity using Centroid (RRLC-CTD):* This approach, unlike ORC-SMT, considers three segment groups as a triangle and finds the centroid (CoD) of triangle instead of calculating SMT that behaves like a Fermat point (F_p) of a triangle for angle less than 120 degree. The Fermat point is a point within a triangle at which the sum of the distances between a point and the three vertices of the triangle is minimized [13]. Our proposed approach exploits this mathematical property of F_p to place the RNs. Fig. 6 shows an example how to calculate F_p. The point F_p denotes the Fermat point of Δxyz. It can be defined as follows. First three equilateral triangles i.e. $\Delta x'yz$, $\Delta y'xz$ and $\Delta z'xy$ are drawn on each side of Δxyz. These equilateral triangles are connected with three extended straight lines i.e. xx', yy' and zz'. The common point of intersection of three straight lines is a Fermat point F_p. Three angles $\angle x\,F_p y$, $\angle x\,F_p z$ and $\angle y F_p z$ will be equal to 120 degrees such that the sum of the distances between F_p and vertices x, y, z is minimized. Our proposed approach RRLC-CTD adopts the algorithm proposed by Ssu et al. [13] to perform vector calculations which quickly converges to an approximate value of the F_p for the placement of RNs. We also check to see whether F_p exist inside the triangle or not. If it is the case, we use Weiszfeld algorithm proposed in [13] to identify the F_p locations. Otherwise, CoD of the triangle will be chosen as F_p for the convergence. In case multiple segments exist in the network as we have taken in our scenario, then the idea is to place the RNs on the chaining path of the multiple consecutive F_p. Initially three random segments are chosen and calculate the first F_p . Consequently, the given segments are sorted in clockwise direction from the first segment. To understand this, let us consider the scenario as shown in Fig. 7 where first F_p is computed by taking three segments at a time. Indeed, we get a chain of connected tree by combining the calculated F_{p_s}. Thus, in the nutshell; the key idea is to deploy the RNs towards the CoD of the triangle instead of finding the SPs. Moreover, tons of algorithms are available in the literature to find the centroid of the triangle (in case triangle is not an equilateral triangle) like Napoleon point, Spieker Center and Nine-point Center etc. The Spieker Center is an easiest and simplest method to find the centroid of a triangle [14]. Furthermore, the main advantage of our proposed RRLC-CTD is that it requires a small number of RNs to connect disjoint network and can work for any number of disjoint segments. Fig. 8 shows the connected network using RRLC-CTD. The key idea is to deploy the RNs towards the CoD of the triangle instead of finding the SPs separately.

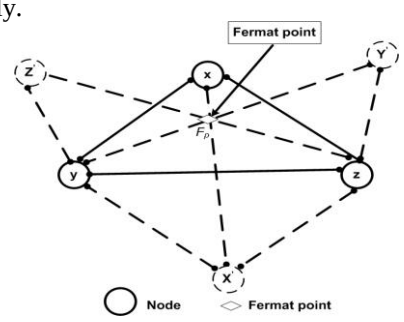


Fig. 6 Example of calculation of F_p of three segments

V. Simulations and Performance Evaluation

The purpose of simulation experiments acts as a proof of concept for the designed protocol. Using simulations, it can be determined whether the designed protocol adheres to the design criteria and requirements. This section evaluates the performance of our proposed approach RRLC-CTD through simulation. The goal of simulation is also to observe that the proposed approach outperforms over other approaches like ORC-SMT, CDCH and Spider Web-1C. Our proposed approach is implemented and validated in C++ environment. Table 1 shows the simulation parameters used in the simulation.

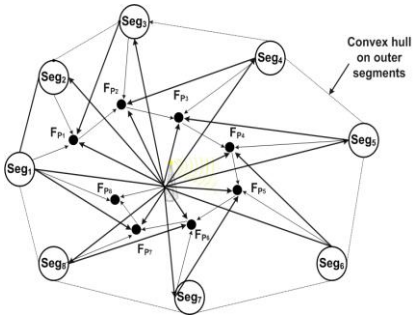


Fig. 7 Example of calculation of chain of F_{p_z}

Table 1 Simulation parameters

Parameter	Value
Simulation Area	2000m X2000m
Nodes	100-500
Radio Model	Path loss model
MAC Layer	IEEE 802.15.4
Communication Range (*R*)	50m-200m
Node Initial Energy(E_i)	100 joules
Total Number of Partitions	6-13
Channel Frequency	2.4GHz
Packet Size	512 bytes
Antenna Model	Omni-directional
Mobility Model	On demand mobility
Failure Model	Random
Data Transmission Rate	15 packets/sec
Simulation Time	1000s

The following three parameters are considered in our experiments for simulation:

a) Number of segments (N_s).
b) Communication range (r) of an RN.
c) Number of placed RNs.

In the simulation experiments, we have taken different topologies with the number of outer segments varies from 4 to 8 and inner segments varies from 2 to 5 i.e. total disjoint segments varies from 6 to 13 that are randomly located in an area of interest (i.e. 2000m × 2000m). While studying the impact of r on the performance, it is varied between 50m to 200m. The results of the individual experiments are averaged over 30 trials of different topologies. We have observed that with 95% confidence level, the simulation results stay within 5%-10% of the sample mean. We consider the number of placed RNs for evaluating the performance of our proposed approach and compare with the existing approaches.

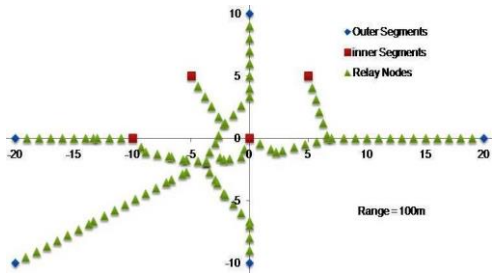


Fig. 8 Partitioned recovery using RRLC-CTD

A. Number of placed RNs. This metric report the total number of required RNs to restore the lost connectivity in the network. As aforementioned, RNs are usually more expensive than sensor nodes. Thus, this metric reflects the total cost of repairing of the partitioned network. Figs. 9a, 9b show the number of required RNs while varying node radio range in the configuration. Therefore, it is clear from the simulation graphs that ORC-SMT performs better than our proposed approach RRLC-CTD only when the number of partitions is less than 5, however our proposed approach performs well for any number of partitions. Moreover, BD approach shows a large number of RNs placements for the repairing of partitioned network due to deployment of nodes along the border of the convex hull in the circular fashion. Furthermore, Spider Web-1C shows similar result like our proposed approach RRLC-CTD as the node radio range increases due to making of large size web structure for the placement of RNs towards CoM. The ORC-SMT fails when the number of outer segments is more than 5 as shown in Fig. 10 and Fig. 11. The reason is that in the random topologies, when number of outer segments becomes larger than 5, more than one of the angles of the Steiner triangle of SMT comes out to be greater than 120 degree (obviously some are less than 120 degree), therefore, the calculated SPs comes out to be on the segment itself. This questions the convergence ability of the ORC-SMT algorithm towards the centre for which the authors have claimed. The situation becomes more intensive as the number of segments grows and algorithm fails in the simulations as we observed in our experiments. In a nutshell, we can say that ORC-SMT behaves best when it serves with a small number of segments (i.e. less than 5), as we have verified in the simulation. Furthermore, The Spider Web-1C heuristics run almost parallel to the CDCH when the node radio range is smaller. This is because the web formed by Spider Web-1C would be much closer to the CoM as explained earlier, and lesser number of nodes is required for the repairing of the lost connectivity. The RRLC-CTD shows good results as compared with ORC-SMT, Spider Web-1C, CDCH and BD as the number of outer segments increases. The reason is the deployment of small number of

RNs towards CoD as explained earlier. Figs. 10-11 also confirm the effectiveness of our proposed approach.

VI. Conclusion and Future Work

In this paper, we have proposed a novel soultion based on spiral format of Fermat points called Restore Relay Lost Connectivity using CenTroiD (RRLC-CTD) for RNs' placements in large scale nodes failure WSN. The main strength of our proposed approach is the use of small number of RNs placements and works for any number of disjoint segments as compared with the existing approaches. The simulation results confirm the goodness of our proposed approach over previously proposed approaches. In the future, our study can focus on simulation of our proposed approach RRLC-CTD to evaluate the actual network performance parameters like throughput, end-to-end delay, packets loss, delivery ratio, etc. with recovery process.

References

[1] M. Younis, Izzet F. Sentruk, Kemal Akkaya, Sookyoung Lee and Fatih Senel, "Topology Management techniques for Tolerating node failures in WSNs: A Survey", *Computer Networks (Elsevier)*, pp. 1-30, 2013.

[2] Virender Ranga, Mayank Dave and A.K. Verma, "Network Partitioning Recovery Mechanisms used in WSANs: A Survey", *Wireless Personal Communications (Springer)*, vol. 72, no. 2, pp. 857-917, September 2013.

[3] Fatih Senel, Mohamed F. Younis and Kemal Akkaya, "Bio-Inspired Relay Node Placement Heuristics for Repairing Damaged Wireless Sensor Networks", *IEEE Transactions* on *Vehicular Technology*, vol. 60, no. 4, pp. 1835-1848, May 2011.

[4] Sookyoung Lee and Mohamed Younis, "Optimized relay node placement for connecting disjoint wireless sensor networks", *Computer Networks (Elsevier)*, vol. 56, no. 12, pp. 2788-2804, August 2012.

[5] Kansal, A. Somasundara, DJM Srivastava and D. Estrin, "Intelligent fluid infrastructurefor embedded networks", *in Proceeding of the 2nd International Conference on mobile systems, applications and services (MobiSys'04)*, Boston, MA, 2004, pp. 1-14.

[6] W. Wang, V. Srinivasan and K. Chu, "Using mobile relays to prolong the lifetime of WSNs", *in Proceeding of the 11th annual International Conference on mobile computing and networking (Mobicom'05)*, Cologne, Gemany, 2005, pp. 270-283.

[7] Wang, G. Cao, TL. Porta and W. Zhang, "Sensor relocation in mobile sensor networks", *in Proceeding of the 24th International annual joint Conference of IEEE Computer and Communication societies (INFOCOM'05)*, Miami, FL, 2005, pp. 2302-2312.

[8] Alfadhly, U. Baroudi, M. Younis, "Optimal node repositioning for tolerating node failure in wireless sensor actor network*", in Proceeding of the 25th Biennial Symposium on Communication(QBSC'10)*, Kingston, Canada, 2010, pp.67-71.

[9] M. Sir, I. Senturk, F. Sisikoglu and K. Akkaya, "An optimization-based approach for connecting partitioned mobile sensor/actuator networks", *in Proceeding of 3rd International workshop on wireless sensor, actuator and robot networks (WiSARN), in conjunction with IEEE INFOCOM'11*, Shanghai, China, 2011, pp.525-530.

[10] I.F. Senturk, K. Akkaya and F. Senel, "An effective and scalable connectivity restoration heuristic for mobile sensor/actor networks", *in Proceeding of the IEEE Global Communications Conference (GLOBECOM'12)*, Anaheim, CA, 2012, pp. 518-523.

[11] S. Vemulapalli and K. Akkaya, "Mobility-based self route recovery from multiple node failures in mobile sensor networks", *in Proceeding of IEEE International Workshop on Wireless Local Networks (WLN'10)*, Denver, CO, October 2010, pp. 699-706.

[12] I. F. Senturk, K. Akkaya and S. Yilmaz, "A game-theoretic approach to connectivity restoration in wireless sensor and actor networks", *in Proceeding of IEEE International Conference on Communications*, Ottawa, Canada, June 2012, pp. 7110-7114.

[13] K. F. Ssu, C. H. Yang, C. H. Chou and A. K. Yang, "Improving routing distance for geographic multicast with Fermat points in MANETs", *Computer Networks*, vol. 53, no. 15, pp. 2663-2673, 2009.

[14] Katherine E. Strauss, "Investigating Centers of Triangles: The Fermat Point", M.S. Thesis, Miami University U.S.A, pp. 1-17, May 2011.

(a) (b)

Fig. 9 Number of relay nodes (a) vs. node radio range when outer segments are 4 and inner segments varies from 2-8, (b) vs. node radio range when outer segments are 5 and inner segments varies from 2-8

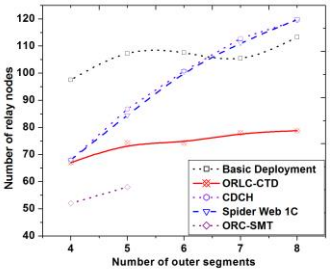


Fig. 10 Number of relay nodes vs. number of outer segments with node radio range 100m

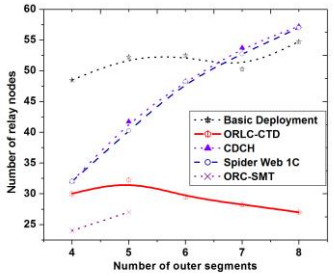


Fig. 11 Number of relay nodes vs. number of outer segments with node radio range 200m

High-Speed Architecture for Direct Computation of DCT

Higinio Mora-Mora*, María Teresa Signes-Pont, Jorge Azorín-López, Lázaro Corral Sánchez

***Abstract*.—** The Discrete Cosine Transform (DCT) is one of the transforms most widely used in digital image processing due to its energy compaction capability and therefore its effectiveness in the digital image compression process. In this work, we have proposed an architecture for high-performance computing of Discrete Cosine Transforms. The architecture applies techniques in developing arithmetic operators to create a compact structure that computes the cosine transform from its direct formulation. Previously, we reviewed the main methods and implementations for calculating this operator to discover ways to improve them. The proposed architecture has been implemented and simulated in reconfigurable cards for digital signal processing to evaluate their performance in terms of area, delay and power consumption. In addition, performance has been calculated with a technology-independent implementation model to compare it with other techniques in a homogeneous way. The results show that the proposed method is competitive in speed with the best-known methods.

Keywords.— Discrete Cosine Transforms, Circuits and Systems, Distributed Arithmetic. Signal Processing.

I. INTRODUCTION

The intensive arithmetic processing requirements of compression methods based on DCT justify the development of specialized hardware architectures that can meet time constraints while complying with the limitations of being embedded in new consumer devices such as media players, mobile phones or tablets. In this very competitive industry, any progress in device development and construction to improve processing capability will be welcome if it increases performance and quality of service to users.

In this work, we describe a calculation method that applies techniques of arithmetic operator development to the whole DCT calculation to take advantage of advances in this area and improve the temporal performance of the transform. These techniques look for organization in the calculation processes to compute them in an integrated way. The methodology starts with the basic description of the DCT calculation because it has the simplest formulation of the transform. From there, it compares that basic description with other arithmetic calculation methods based on distributed arithmetic (DA) or on a flow-graph algorithm (FGA).

We focus on 2-D 8x8 size samples as more easily handled input by the DCT function when it applies the image compression algorithm [1]. Thus, the implementation of the proposed method is optimized for functions with that input size.

* H. Mora-Mora, MT. Signes-Pont, J. Azorín-López and L. Corral-Sánchez are with the Department of Computer Technology and Computation, University of Alicante, Spain, 03690, San Vicente del Raspeig, Alicante, Spain. e-mail: ({hmora, teresa, jazorin, lcs21}@dtic.ua.es).

The remainder of this paper is organized as follows: the 2-D DCT expressions and relevant properties are briefly described in Section II. In section III, we review the main design techniques and hardware implementations of the algorithm. Section IV exposes the proposed architecture and section V evaluates its performance in area, delay and power consumption. Section VI provides a comparison with other techniques, and conclusions are drawn in Section VII.

II. DESCRIPTION OF DCT ALGORITHM

The 2-D DCT of an 8x8 block sample {f(x, y), x = 0, 1, ..., 7, and y = 0, 1, ..., 7} is generally expressed as follows:

$$F(u,v) = \frac{C(u)C(v)}{4}\sum_{x=0}^{7}\sum_{y=0}^{7} f(x,y) \cos\left[\frac{(2x+1)k\pi}{16}\right]\cos\left[\frac{(2y+1)k\pi}{16}\right] \quad (1)$$

where

$$(u), C(v) = \begin{cases} \frac{1}{\sqrt{2}}, & if\ u,v = 0 \\ 1, & otherwise \end{cases} \quad (2)$$

The efficient calculation of that transform is made using its separability property. That is, the 2-D DCT is organized in two consecutive 1-D DCT calculations. The transform is first calculated in row order and then in column order.

The 1-D DCT of eight samples {f(x), x = 0, 1, ..., 7} can be expressed as:

$$F(u) = \frac{C(u)}{2}\sum_{x=0}^{7} f(x)\cos\left[\frac{(2x+1)u\pi}{16}\right] \quad (3)$$

where C(u) is defined with the same (2) equation.

The scaling factor (1/2) can be eliminated without loss of generality to facilitate calculation. So, in a deployed way, the above formula can be written as:

$$F(u) = C_{u0}f(0) + C_{u1}f(1) + ... + C_{u7}f(7) \quad (4)$$

where: C0j, ∀ j ∈ {0..7}, is $C0j = \cos\left[\frac{\pi}{4}\right]$

and Cuj, with u ∈ {1..7} and j ∈ {0..7}, is $Cuj = \cos\left[\frac{(2j+1)u\pi}{16}\right]$

Therefore, the constants set is as follows:

$$Cuj \in C / C \equiv \{\cos\left[\frac{\pi}{16}\right], \cos\left[\frac{\pi}{8}\right], \cos\left[\frac{3\pi}{16}\right], \cos\left[\frac{\pi}{4}\right], , \cos\left[\frac{5\pi}{16}\right], \cos\left[\frac{3\pi}{8}\right], \cos\left[\frac{7\pi}{16}\right]\} \quad (5)$$

The above expression (3) can be written in vector notation as follows:

$$F = \frac{1}{2} T(N)_8 f \quad (6)$$

where $F = [F(0), ..., F(7)]^T$, $f = [f(0), ..., f(7)]^T$ and $T(N)_8$ is an 8 x 8 matrix whose (u, j)th component is:

$$T(u,j)_8 = C(u)\ cos\left[\frac{(2j+1)u\pi}{16}\right] \quad (7)$$

This expression facilitates computation by calculating inner products methods. Moreover, expression (6) can be divided, in turn, into an even and an odd part to achieve a simpler matrix subexpression.

III. Related Work

The direct computation of 2-D DCT requires 4096 multiplications and 4032 additions, while the separated calculation into row-column and row-wise 1-D DCT computations needs only 1024 multiplications and 896 additions. However, that computational cost is still very high to meet the real time bandwidth constraints of modern applications. Consequently, numerous proposals exist of methods and designs focused on reducing this computational cost. Moreover, the proliferation in recent years of mobile devices has created a need to embed these architectures in small machines but with a strong processing capability in this area.

A. Calculation Methods

In general, methods to implement the DCT algorithm can be classified by calculation procedures used in processing.

First, an important group of methods is based on the vectorial expression of the DCT calculation described in (6) and (7). As mentioned in the previous section, equation (6) can be decomposed recursively to obtain a final expression based on the calculation of inner products in which the vector formed by the coefficients is fixed [9], [10]. Therefore, regular computation patterns are obtained; this facilitates their implementation for signal processing applications [11].

Other methods take advantage of symmetries of the calculations derived from equation (3) to reduce the number of necessary operations. Thus, numerous designs have been proposed to minimize the necessary multiplications. Some of the most popular proposals are Chen [4], Wang [5], Lee [6], Vetterli [7] and Loeffer [8] methods. Loeffer [8] performs the DCT calculation with only 11 multiplication operations. The above methods are characterized by calculation flow-graph results and butterfly configurations.

B. Hardware Implementations

Hardware resource consumption is a critical aspect in the image compression process, whether in applications or consumer electronics devices with timing constraints. It is so critical that there are numerous hardware architectures that perform the VLSI implementation of the methods described in the previous section, which take into consideration performance issues.

Distributed Arithmetic (DA) Implementations from the vector formulation of the DCT is an efficient technique to process these operations when one of the vectors is fixed due to the regularity of its implementation. Recently, extensive research for this method of calculation of the DCT has been developed. There are two broad approaches to implementing a DA computation unit: ROM-based and adder-based. The ROM-based or conventional DA approach accelerates multiplication operations, pre-calculating them for each input value. Then, the results are produced by *shift and add* procedures. This method requires a ROM or look-up tables (LUTs) in which to store these precalculated products [11], [20], [21]. The second type, adder-based, does not requires LUTs and can exploit the distribution of binary value patterns of constant operands [24], [25], [26]. Nevertheless, these methods needs more addition operations and are slower than previous ROM-based approaches.

In consideration of the flow-graph algorithm (FGA) implementations for DCT calculation, some of these architectures try to minimize the impact of involved FP multiplications, replacing them by additions and shifts [12]. Other proposals perform the calculations without multiplications by approximating these operations with power of two denominators fractions. So, each multiplication is replaced by a series of integer additions and a subsequent shift [14], [15]. Although this technique avoids multiplications, slightly lower quality of the compressed image is obtained. In the same way, other techniques replace expensive multiplications by means of the CORDIC method, and perform the butterfly configuration calculations by rotating angles [16], [4], [17]. In these methods, the necessary final compensation of the CORDIC method is included in the quantization stage of the image compression. There are studies that describe other variations of the above methods to minimize the operations involved in calculating the algorithm. In [13], a method that improves the performance offered by the CORDIC variants from Loeffler by placing multiplications in the last stage of the flow-graph is presented.

Finally, there are variations of the method to adapt to specific applications, for example, Prime-length DCT [27] or shape adaptive DCT [28] and other custom VLSI-implementations that attempt to exploit various aspects of the transform to construct specific designs, such as [18], [29], [30], [31].

IV. Proposed Architecture

In this paper, we aim at presenting a more efficient hardware architecture for DCT using arithmetic computation techniques to processing the addition and multiplication schemes involved.

We use direct computation of DCT formulated in (4) because it is the simplest expression of the transform. According to this formula, the calculation of each F(u) comprises eight multiplications and seven additions. The operands of each of these multiplications are an f(x) input value and a constant (Cuj) defined in (5).

The defining characteristics of the proposed architecture are two: how multiplication and addition operations required for

each F(u) are integrated, and how the constants defined in (5) are including in the processing of each product.

With regard to the first issue related to the sequence of multiplication and addition operations, the standard method of calculating each multiplication operation consists of the well-known following stages [32]: partial product generation, partial product reduction and final addition. The proposed idea is simple: compute the eight multiplications and seven additions of each F(u) in a combined way, so that the partial product reductions are integrated with the additions required in each calculation. Thus, it is possible to build with them an addend reduction tree to make a VLSI implementation of the DCT in a more efficient way. In general terms, the proposed architecture implements a design for each F(u) composed of the three multiplication stages: partial product generation, partial product reduction and final addition.

There are several solutions for determining the partial products of the multiplication using constants. In this research, two shifts are analysed: *register-based partial product generation* and *direct partial product generation.*

A. Register-based partial product generation

This design does not need a ROM to store the results of the multiplications by constants, but only a collection of records with some multiples of the constants. Thus, the register outputs can be sent out to the inputs of all operators involved simultaneously through a shared bus without extra delays in memory access [33].

The multiples of the constants stored shall be defined by the size of operators and the number of partial products generated. If k is considered as operator's length, values that will need to be stored to implement this design will consist of multiplying the seven signed constants (5) by the odd numbers between 0 and 2^k-1. That is, it will be necessary to precalculate $7 \cdot 2^k/2$ values and keep them in records to be used in processing.

For example, for k = 4, the following extended constants set C' for each element of the set C defined in (5) will precalculate:

$$C' \equiv \{\forall c_i \in C, R_{Ci} = \{c_i, 3c_i, 5c_i, 7c_i, 9c_i, 11c_i, 13c_i, 15c_i\}\} \quad (8)$$

As an implementation decision, all multiples of each constant can be placed in a single register of $m \cdot 2^k/2$ length, where m is constant precision. For example, for k = 4 and m = 12, it would take only 7 x 96-bit registers to store all of the data.

With the previous register configuration, processing of each multiplication is performed by fragmenting the multiplicands (f(x)) in k-bit lengths and determining the partial products, hardwiring the extended constants set C'. Thus, n/k partial products of each multiplication are generated, where n is the length of the input f(x). The following figure depicts this approach (Fig. 1).

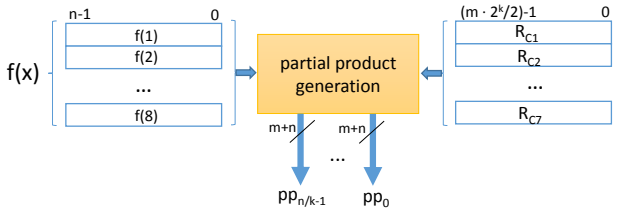


Fig. 1. Register-based partial product generation Scheme

Addressing by hardwiring these registers is done by multiplexer elements with k control inputs. For example, the next figure shows in detail the partial product generation process with k = 4 bits length (Fig. 2).

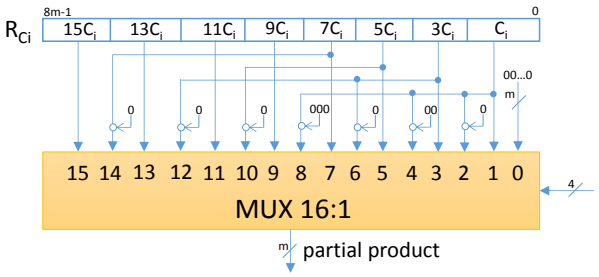


Fig. 2: Multiplexor Partial Product Selection.

Two's complement format is used to encode the sign of the partial products generated. In this case, the addition of a single one in the least significant part to perform the complement is compensated by also extending the sign bits to the right. The calculation error of each product due to this compensation method will be in the least significant bits of the result. Given the precision required in the DCT function for a length of n and m, the above error will remain within the discarded bits.

B. Direct partial product generation

Another design alternative of this module is intended to avoid the use of multiplexers and, consequently, save the area cost involved. The new proposed method uses neither single direct generation due to the large number of partial products generated, nor Booth's method because of the additional delay that incorporates [34]. Therefore, the proposal is to make a direct generation of only non-zero digits. For this, the distribution of bits of each fixed point is analysed, and then a custom circuit is created for each case. Table I shows the constants Ci (5) with their decimal and binary values.

TABLE I
CALCULATION CONSTANTS OF DCT

Ci	Decimal value	Binary value
cos[π/16]	0.9807852	0.111110110001
cos[π/8]	0.9238795	0.111011001000
cos[3π/16]	0.8314696	0.110101001101
cos[π/4]	0.7071067	0.101101010000
cos[5π/16]	0.5555702	0.100011100011
cos[3π/8]	0.3828125	0.011000100000
cos[7π/16]	0.1953125	0.001100100000

TABLE II
CSD CODED CALCULATION CONSTANTS OF DCT

Ci	Binary value	CSD binary value	Partial Products
cos[π/16]	0.111110110001	$1.000\bar{1}0110001$	5
cos[π/8]	0.111011001000	$1.00\bar{1}011001000$	5
cos[3π/16]	0.110101001101	$0.11010101001\bar{0}$	6
cos[π/4]	0.101101010000	0.101101010000	5
cos[5π/16]	0.100011100011	$0.10010\bar{0}100100$	4
cos[3π/8]	0.011000100000	0.011000100000	3
cos[7π/16]	0.001100100000	0.001100100000	3

To improve the generation of partial products, we use signed-digit format to represent those constants and so minimize the

number of nonzero digits while keeping the error within the 12 bits length. The signed-digit format used is Canonic Signed Digit (CSD) because the number of nonzero digits is minimal [39]. Table II shows the signed-digit representation and the resulting partial products in each case.

Generator circuits directly produce direct or inverse partial products according to the sign of each digit. The following figure depicts the schematic of this calculation method of partial product generators.

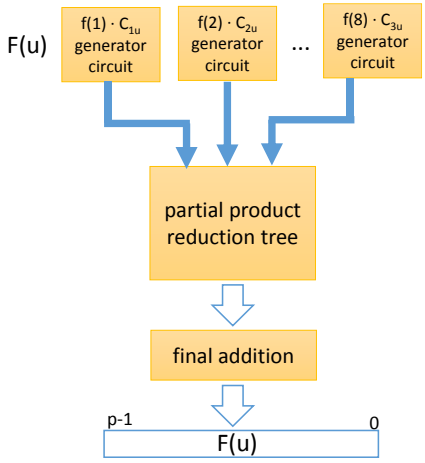


Fig. 3: Final Proposed DCT block diagram

From here, generated partial products are combined and reduced by means of a reduction tree consisting of different reduction counters and compressor devices.

With respect to the first generation method, the amount of product generated for each F(u) is $8 \cdot \lceil n/k \rceil$. The detailed configuration of the tree can be determined from precision given by n and k. For example, for n = 12 and k = 4, the configuration will have a total of 24 partial products, which may reduce according to the scheme described in the following figure.

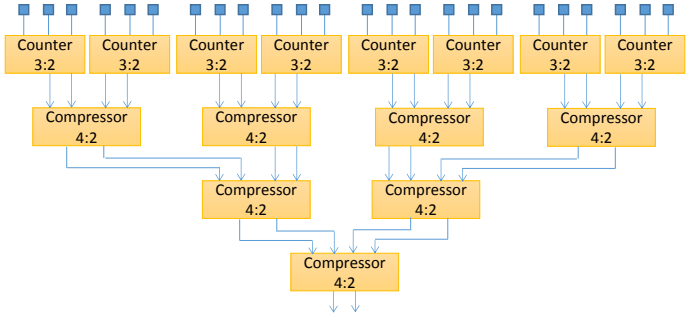


Fig. 4: 24 partial product Reduction tree

With the direct-generation method, the number of partial products produced in each F(u) is between 32 and 40; therefore, a reduction tree can be configured with the outputs from each Fi generator. As in the previous case, the signed numbers are coded in two's complement, so this representation uses 24 bits including sign. The resulting reduction tree in this case is represented in the following figure.

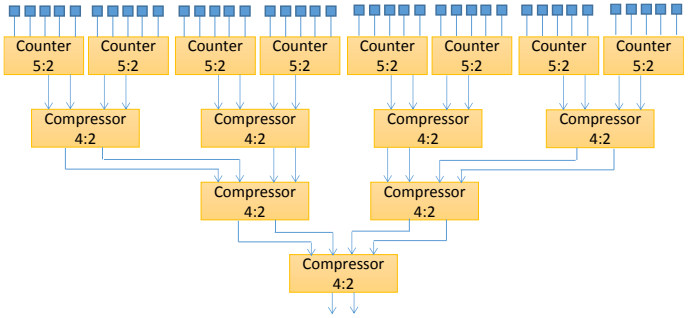


Fig. 5: 40 partial product Reduction tree

The final addition of the operation can be made by any known method, for example by a CLA adder.

V. Evaluation the Performance of Architecture

The architecture specified in the previous section allows us to compute all collected multiplications and additions of each F(u) as a whole. The composition of all addends involved in a reduction tree creates a compact structure that enhances their VLSI implementation.

To keep the highest precision as recommended by the IEEE 1180-1990 standard [35] and other recent works [12], [13], we set input signals from pixels 12 bits wide and paths between stages 16 bits wide. These precision characteristics set n and m to 12 bits in length. An additional one-bit will be added in the multiplicand to encode the sign of the terms. The partial results of the additions and reductions will be set to the most significant 16 bits, so the loss of precision in the results will be insignificant. A block size of k = 4 is established for addressing each entry.

To improve performance as much as possible, the precomputed constants are stored in look-up tables (LUTs) near the DCT operator inside the arithmetic unit; therefore, it is not necessary to lose execution cycles for memory loading for them.

The most relevant aspect of study in this work is the time delay in calculating DCT. However, we are aware of the importance of other characteristics, such as the occupied area and power consumption, which are very important for embedded applications. Therefore, the performance of the proposed technique for DCT calculation is analysed in terms of area, delay and power consumption to establish terms of comparison with other proposals.

A. Area estimation

It is common to express the area in terms of complex gates to obtain a homogeneous comparison with other designs. Let τ_a be the area occupied by a complex gate, for example an *xor* gate. Thus, it is easy to translate the performance results to another scale of comparison, such as number of transistors or occupied area in μm^2, because there are studies in the literature that report these data for several implementation technologies [36]. The following table summarize the estimated occupied area of the main elements of the architecture in homogeneous terms. These data came from well-known calculation methods in the computer arithmetic literature [3], [23], [33], [37], [38].

ISBN: 978-1-61804-290-3

TABLE III
AREA COST ESTIMATION

Device	Area (τa)
3:2 counter	3
4:2 compressor	6
5:2 counter	8
16-bit adder (CLA)	48
12-bit multiplier	109
up to 8-bit input LUT	30 Γa /Kbit
16:1 multiplexor	16
k-bit RAC (ROM accumulator)	$2^k \cdot$ precision
k-bit register	k
12-bit shifter	48

The main contributions to the required area for implementing the F(u) calculation come from inputs and stages depicted in fig. 1: constants registers bank, partial product generation, partial product reduction and final addition.

The LUT module contains the necessary seven constants of DCT. Moreover, in the case of register-based generation, implementation will also be necessary to store the odd multiples of these constants. This gives us a total estimation of occupied area to store that information of 7·12 bits = 84 bits for the constants and 7·7·12 = 588 bits for their multiples. The assumed correspondence is a complex gate to store a bit in a flip-flop, so the encoding information requires 84 τ_a for constants and 588 τ_a for their multiples.

The partial product generation stage has two implementation modes: the register-based design selects the right register constant using a multiplexer, whereas the control signals are addressed by each k-block of the multiplicand. The number of complex gates required by each 16:1 multiplexer is 16 τ_a [22]. Fragmenting the multipliers into three blocks produces an estimation for each multiplication of $3 \cdot 12 \cdot 16 = 576$ τ_a. Therefore, computing in parallel all multiplications involved in each F(u) needs 4608 τ_a.

In the case of implementing direct generation, the estimation of area of this stage is given only for the combinational circuit described by Fig. 5. This configuration requires 24 τ_a to compute each partial product and, consequently, 120 τ_a to process each multiplication. The F(u) generation requires a total estimation of 960 τ_a.

Regarding the partial product reduction tree, the delay cost is related to the number of addends to reduce. With full precision, the result of each multiplication is 24 bits long, where each partial product, with sign extension at the beginning and/or end, also has a length of 24 bits.

Including individualized data on the occupied area shown in table III, the partial product reduction trees depicted in figures 4 and 5 requires a total area of 1536 τ_a and 2544 τ_a to simultaneously compute 24 bits length.

The last stage is the final addition of the two addends produced by the tree. Although the reduction tree works with addends 24 bits long, the final addition will be made to the most significant 16 bits according to the recommendations of accuracy that set the standard. Using the known CLA addition method requires an area cost of 48 τa to perform a 16-bit addition.

Thus, from previous results, we can estimate a total area to build the architecture to compute each F(u) of 6732 τa when implementing a register-based generation method and 3504 τa when using direct generation.

To validate the design and cost estimations, we have implemented the proposed architecture in an FPGA synthesis platform (see Appendix A). The next table shows the area cost of an FPGA implementation of the F(u) function. The data do not include storage costs of constants because they are in the memory banks of the card and do not use the programmable logic.

TABLE IV
AREA RESOURCES COST IN FPGA IMPLEMENTATION

Circuit	Slices	4-input LUT
4:2 compressor	3	5
3:2 counter	1	2
5:2 counter	3	7
12-bit 16:1 multiplexor	7	12
16-bit adder (CLA)	12	16
12-bit multiplier	148	292
Register-based partial product generation	1512	2520
Register-based 24:2 partial product reduction 24:2 (Fig. 7)	504	888
Direct partial product generation	560	968
Direct 40: 2 partial product reduction (Fig. 9)	1016	1824
Register-based Total	2028	3424
Direct Total	1588	2808

B. Delay estimation

As in the previous subsection, a homogeneous convention in calculating response times of the elements involved is assumed to facilitate comparison with other designs or proposals. In this case, we set as our time unit τ_t, corresponding to the delay of a complex gate. Table IV summarizes delays of each device.

The delay in the processing of each F(u) is the strong point of our proposal. With the same criteria as in the previous section, we next break down the delays produced at each stage of the calculation.

The constants registers bank and their addressing by partial product generation by multiplexors do not have a delay because they are not stored in memory and it is not necessary to read them. The addressing of the registers towards the inputs of multiplexers is straightforward, and from them the right constant values are directly selected.

TABLE V
DELAY ESTIMATION

Device	Delay (τt)
3:2 counter	2
4:2 compressor	3
5:2 counter	4
16-bit adder (CLA)	8
12-bit multiplier	26
up to 8-bit LUT Access	3.5
16:1 multiplexor	2
k-bit RAC (ROM accumulator)	
k-bit register	-
12-bit shifter	4

The delay of partial product generation is also a fast stage. In register-based design, the delay introduced to select the correct value of the inputs is 4 τ_t. For direct generation, the delay is only 2 τ_t.

ISBN: 978-1-61804-290-3

The partial product reduction tree spends 11 τ_t to reduce 24 addends and 13 τ_t for 40 addends. Finally, the estimated delay of CLA addition for 16 bits-length is 8 τ_t.

The above results conclude a total cost for the calculation of each F(u) of 23 τ_t. As in the previous section, results of the FPGA implementation of the proposed design are shown to validate the obtained results.

TABLE VI
DELAY IN FPGA IMPLEMENTATION

Circuit	Delay (ns)
4:2 compressor	1.344
3:2 counter	0.882
5:2 counter	1.848
12-bit 16:1 multiplexor	1.068
16-bit adder (CLA)	9.068
12-bit multiplier	46.89
Constants Look-Up Table Access	4.520
Register-based partial product generation	9.551
Register-based 24:2 partial product reduction 24:2 (Fig. 7)	5.556
Direct partial product generation	1.068
Direct 40: 2 partial product reduction (Fig. 9)	7.154
Register-based Total	24.175
Direct Total	17.291

According to our simulation results using the programmable card, the direct method is faster due to the lower complexity associated with implementation.

C. Power consumption

The power consumption of the implementation is summarized in table VII. Moreover, the simulation brings a static consumption of 118.08 mW intended to supply power to the FPGA device.

In the next section, these estimations are compared with performance of other known designs to evaluate the proposed architecture.

TABLE VII
POWER CONSUMPTION IN FPGA IMPLEMENTATION

Circuit	Power (mW)
Register-based partial product generation	46
Register-based 24:2 partial product reduction 24:2 (Fig. 7)	16
Direct partial product generation	17
Direct 40: 2 partial product reduction (Fig. 9)	32
Register-based Total	62
Direct Total	50

VI. DISCUSSION AND COMPARISONS

Due to the nature of the applications where this algorithm is used, it is essential to evaluate and compare its performance with other methods. In this section, performance of the proposed method is compared with other designs reviewed in section III to evaluate the best way to build high-performance DCT calculators.

To estimate the computing cost of the other proposals, we use the information displayed in tables III and V about the area and time required to implement the main components involved in the designs. How these elements are combined to build each DCT calculator is obtained from detailed descriptions in corresponding references. The comparison circuits are taken from the best designs in the mentioned publications' design techniques sections. The obtained results are summarized in tables VIII and IX.

This section has not compared different implementations of the FPGA device because we preferred to take the performance data directly from the references to avoid interfering with the implementation done by us and to avoid distortions introduced by optimization techniques of a VHDL compiler.

A. Area comparisons

As a first architecture for comparison, we consider the basic design based on the direct implementation of a DCT algorithm from equation (3). According to this scheme, the elements required to calculate each F (u) consist of an LUT for storing the seven-set constants of C, a multiplier operator and an adder. These components need a combined area of 237 τ_a.

The designs according to the Distributed Arithmetic ROM-based technique mainly consist of addition-accumulation iterations of results read from memory (RAC – ROM Accumulator). The main contributors to required hardware for computing DCT by means of these methods are memory components. In addition, the necessary hardware to build accumulation records and parallel adders of the first stage must be added. Consequently, the design requires a combined area between 25 Kτ_a [11] for the standard method and 2.56 Kτ_a for efficient hardware implementations [20].

DA adder-based architectures do not need ROM components; therefore, they can be implemented with less silicon. However, they need more addition operations to compute the inner products of the formula [25]. Reusing adder operators as the specification shows, we can estimate an area cost of approximately 1.5Kτ_a [26].

On the other hand, architectures based on the FGA algorithm minimize the number of addition and multiplication operations needed for DCT calculation. Consequently, implementations based on Loeffer method require only 11 multiplications and 29 additions [8]. Considering just the resources to implement these elements, we obtain an area estimation of 237 τ_a to compute 1D DCT. In a parallel architecture, eight adder and six multiplier units would be necessary to perform all computation required. This arithmetic processing has an area cost of 1118 τ_a, apart from necessary control circuitry.

For FGA-based proposals that replace multiplications by CORDIC rotations, we consider only the smallest number of iterations to obtain the required accuracy. In addition, these proposals postpone the final compensation multiplication to the next quantization stage, once DCT processing is finished. Nevertheless, to perform a homogeneous comparison, we include this calculation within the process at the end of the operation.

According to designs outlined in the literature [4], [13], [16], between 48 and 56 additions with corresponding shift instructions are needed to perform the calculations. If execution is carried out in a serial way, a CORDIC processor and a multiplier for the final compensation will be necessary. This configuration requires 349 τ_a to compute a whole DCT. A parallel configuration needs eight adders, three CORDIC

processors and six multipliers. The resulting hardware cost is 1614 τ_a.

TABLE VIII
AREA COST ESTIMATION COMPARISON

Method	Area (τa)
Basic direct design	237
DA Rom-based	2.5K – 25K
DA Adder-based	1.5K
Loeffer FGA*	237 – 1.1K
Loeffer CORDIC FGA*	349 – 1.6K
Proposed register-based	6.7 K
Proposed direct calculation	3.5 K

* complete 1D DCT

B. Delay comparisons

We make a normalized balance in equality terms, using the homogeneous delays of the components in independent of the technology measurements. The delays correspond to the critical path of the circuits, and they are set by the delay of the slowest path through the circuit.

The basic scheme requires eight LUT accesses, eight multiplications and seven additions to compute each F(u). These operations have a combined delay cost of 1208 τ_t. This cost justifies the need to spend effort to reduce it, especially in applications with hard time constraints.

The path delay of the Distributed Arithmetic designs lies in the accumulated iterations with data obtained from the LUTs [18], [20]. The number of iterations is linear with operands precision (12 bit); in each iteration, a data access and an addition are processed. Associated delay costs associated with these components and, considering neither control iteration circuitry nor shifts associated, the time delay for calculating each F(u) is 94 τ_t.

In the case of adder-based DA architectures, nearly three times more addition operations are required than in the previous method. Because the adder operator is the most significant contributor to final delay, we can estimate a time cost of 282 τ_t to compute each F(u) of DCT.

Implementations based on the FGA Loeffler algorithm require 11 multiplications and 29 additions. The serial execution of these operations has a delay of 518 τ_t. With parallel execution, even at the cost of more hardware to duplicate addition and multiplication operators, flow-graph computing needs only two multiplications and four additions [8], and has a delay estimation of 84 τ_t. The CORDIC implementations of the FGA algorithm require a runtime estimation of 620 τ_t (464 τ_t without compensation multiplications). Finally, with hardware resources for concurrent execution, they would require a 146 τ_t time delay (120 τ_t without compensation multiplications).

The results of the estimations and simulations showed in table IX demonstrate that the proposal described in this paper significantly reduces the execution time of DA-based designs and maintains comparable times to those achieved by FGA-based methods. Additionally, our proposal, like those based on DA, allows a more regular structure when calculating all components of 1D DCT in a uniform way. It therefore allows implementation of segmented or pipelined processing strategies that could reduce combined computation.

TABLE IX
DELAY COST ESTIMATION COMPARISON

Method	Delay (τt)
Basic direct design	1208
DA Rom-based	94
DA Adder-based	282
Loeffer FGA*	84 – 518
Loeffer CORDIC FGA*	146 (120) – 620 (464)
Proposed register-based	23
Proposed direct calculation	23

* complete 1D DCT

VII. CONCLUSIONS

In this work, we propose a new calculation method for Discrete Cosine Transforms based on direct mathematical expression. The architecture provides a compact structure for performing operations that, unlike DA-based methods, does not require ROMs to perform multiplications.

We have shown that VLSI implementation through FPGA simulations can easily meet high-speed requirements for high performance applications. In comparison to the existing design, our approach offers some advantages that can be explored for high performance calculators. The results of the tests have shown that the implementation's performance is comparable to the best-known implementations based on both DA and FGA methods.

For future research, we are working to extend the implementation ideas to other transforms intensive in hard arithmetic operations such as DFT. Our objective is to be able to build full arithmetic units for signal processing with a core of primitives based upon the same calculation techniques.

APPENDIX A: EXPERIMENTAL ENVIRONMENT

Hardware design, simulation techniques and their implementation in reconfigurable systems such as FPGAs, enable valid designs and high productivity in the development of hardware systems. These devices have wide acceptance and are frequently used by the scientific community. The results allow checking theoretical performance estimations and correction of designs.

In this work, we are using the FPGA *xc3sd1800a*. This device belongs to the family *Xilinx Spartan Spartan3adsp*, which is optimized for digital signal processing applications for the highest system integration (http://www.xilinx.com/products/silicon-devices/fpga/xa-spartan-3a-dsp/). The synthesis tool used is *Xilinks: ISE Design Suite 14.7 (Webpack license)*, and the programming language is VHDL. We acquired data about power consumption of the circuits from the Xilink XPower Estimator (XPE) tool (http://www.xilinx.com/support/documentation/user_guides/ug440.pdf).

Circuit implementations used in the performance evaluation are based on known designs referenced in the literature. Standard circuits are taken from tested VHDL codes of the Dept. of Computer Science & Engineering, College of Engineering, University of California, Riverside. (http://esd.cs.ucr.edu/labs/tutorial/).

ISBN: 978-1-61804-290-3

REFERENCES

[1] K.R. Rao, J.J. Hwang, "Techniques and Standards for Image", *Video and Audio Coding*. Englenwood Cliffs, NJ: Prentice –Hall, 1996.

[2] M. R. Pillmeier, M. J.Schulte; E.G. Walters, "Design alternatives for barrel shifters", *Advanced Signal Processing Algorithms, Architectures, and Implementations Proceedings of the SPIE*, Volume 4791, pp. 436-447, 2002.

[3] Y. Sungwook, E.E. Swartzlander, Jr. "A scaled DCT architecture with the CORDIC algorithm", *IEEE Transactions on Signal Processing*, vol. 50 (1), pp. 160-167, 2002.

[4] W. H. Chen, C. Smith, S. Fralick, "A fast computational algorithm for the discrete cosine transform", *IEEE Trans. Commun.*, vol. 25, pp. 1004-1009, 1977.

[5] Z. Wang, "Fast algorithms for the discrete W transform and for the discrete fourier transform", *IEEE Trans. Acoust. Speech. Signal Processing*, vol. 32, pp. 803-816, 1984.

[6] B. Lee, "A new Algorithm to Compute the Discrete Cosine Transform", *IEEE Transaction on Acoustic, Speech and Signal Processing*, Vol. 32, (6), pp. 1243-1245, 1984.

[7] M. Vetterli, J. Kovacevic, *Wavelets and subband coding*. Englewood Cliffs, NJ: Prentice-Hall, 1995.

[8] C. Loeffler, A. Lightenberg, G. S. Moschytz, "Practical fast 1-D DCT algorithms with 11-multiplications", *Proc. of ICASSP*, Glagow, vol. 2, pp. 988-991, 1989.

[9] B.G. Lee, "A New Algorithm to Compute the Discrete Cosine Transform", *IEEE Trans. On Acoustics, Speech, and Signal Processing*, vol. 32, pp. 1243-1245, 1984.

[10] H.S. Hou, "A Fast Recursive Algorithm for Computing the Discrete Cosine Transform," *IEEE Trans. On Acoustics, Speech, and Signal Processing*, vol. 35, pp. 1455-1461, 1987.

[11] S.A. White, "Applications of distributed arithmetic to digital signal processing: a tutorial review", *IEEE ASSP Magazine*, Vol. 6, nº 3, pp. 4-19, 1989.

[12] El Aakif, M.; Belkouch, S.; Chabini, N.; Hassani, M.M. "Low power and fast DCT architecture using multiplier-less method", *Faible Tension Faible Consommation (FTFC)*, pp. 63 – 66, 2011.

[13] Z. Wu, J. Sha, Z. Wang, L. Li, M. Gao, "An improved scaled DCT architecture", *IEEE Transactions on Consumer Electronics*, vol.55-2, pp. 685-689, 2009.

[14] T.D. Tran, "The binDCT: fast multiplierless approximation of the DCT," *IEEE Signal Process. Letters*, vol. 7, pp. 141–144, 2000.

[15] J. Liang, T.D. Tran, "Fast multiplierless approximations of the DCT with the lifting scheme," *IEEE Trans. Signal Process.*, vol. 49, nº 12, pp. 3032–3044, 2001.

[16] C.-C. Sun, S.-J. Ruan, B. Heyne and J. Goetze, "Low-power and high quality CORDIC-based Loeffler DCT for signal processing", *IET Proc. Circuits, Devices & Systems*, vol. 1, no. 6, pp. 453–461, 2007.

[17] H. Huang; L. Xiao, "CORDIC Based Fast Radix-2 DCT Algorithm", *IEEE Signal Processing Letters*, vol. 20, nº. 5, pp. 483 - 486, 2013.

[18] A. Madisetti, A. N. Willson, "A 100 MHz 2-D 8 x 8 DCT/IDCT Processor for HDTV Applications", *IEEE Trans. On Circuits and Systems for Video Technology*. Vol. 5, No. 4, pp. 158- 165. 1995.

[19] MT. Signes, JM García, and H. Mora. Improvement of the discrete cosine transform calculation by means of a recursive method. Mathematical and Computer Modelling, 50(5–6):750 – 764. Mathematical Models in Medicine and Engineering, 2009.

[20] Y. Sungwook; E.E Swartziander, "DCT implementation with distributed arithmetic", *IEEE Transactions on Computers*, Vol. 50, nº 9, pp. 985-991, 2001.

[21] H.-C. Chen, J.-I. Guo, T.-S. Chang, C.-W. Jen, "Memory-Efficient Realization of Cyclic Convolution and its Application to Discrete Cosine Transform", *IEEE Trans. Circuits Syst. Video Technol.*, vol. 15, no. 3, pp. 445–453. 2005.

[22] P. Balasubramanian, D.A. Edwards, "Power, delay and area efficient self-timed multiplexer and demultiplexer designs", *International Conference on Design & Technology of Integrated Systems in Nanoscale Era*, pp. 173-178, 2009.

[23] R. Menon, D. Radhakrishnan, "High performance 5:2 compressor architectures", *IEE Proc. Circuits Devices Syst.*, Vol. 153, No. 5, 2006.

[24] T-S Chang, C-S. Kung, C-W Jen, "A simple processor core design for DCT/IDCT", *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 10, No. 3, pp. 439-447, 2000.

[25] Shams, A.M., Chidanandan, A., Pan, W., Bayoumi, M.A., "NEDA: a low-power high-performance DCT architecture", *IEEE Transactions on Signal Processing*, Vol. 54, No 3,pp. 955-964, 2006.

[26] Sharma, V.K.; Mahapatra, K.K.; Pati, U.C. "An Efficient Distributed Arithmetic Based VLSI Architecture for DCT", *International Conference on Devices and Communications*, pp. 1-5, 2011.

[27] J. Xie, P. K. Meher, J. He, "Hardware-Efficient Realization of Prime-Length DCT Based on Distributed Arithmetic", *IEEE Transactions on Computers*, vol. 62, no. 6, pp. 1170-1178, 2013.

[28] A. Kinane, N. O'Connor, "Energy-efficient Hardware Accelerators for the SA-DCT and Its Inverse", *Journal of VLSI Signal Processing*, Vol. 47, pp. 127-152, 2007.

[29] V. Srinivasan, K.J.R. Liu, "VLSI design of high-speed time-recursive 2-D DCT/IDCT processor for video applications", *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 6 (1), pp. 87-96, 1996.

[30] D. Slawecki and W. Li, "DCT/IDCT processor design for high-data rate image coding", *IEEE Trans. Circuits Syst. Video Technol*., Vol. 2, pp.135-146 1992.

[31] P. Subramanian, A.S.C. Reddy, "VLSI Implementation of fully pipelined multiplierless 2D DCT/IDCT architecture for JPEG", *IEEE International Conference on Signal Processing (ICSP)*, pp. 401-404, 2010.

[32] Mora Mora H., Mora Pascual J., Sánchez Romero JL., Pujol López F, Partial product reduction based on look-up tables, International Conference on VLSI Design, (2006).

[33] R. Rajsuman, "Design and test of large embedded memories: an overview", *IEEE Design and Test of Computers* 18 (3), pp.16-27, 2001.

[34] W. Yeh, C. Jen, "High speed booth encoded parallel multiplier design", *IEEE Trans. Comput*. 49 (7), pp. 692–700, 2000.

[35] IEEE Std. 1180-1990, *IEEE standard specification for the implementation of 8x8 inverse cosine transform*. Institute of Electrical and Electronics Engineers, New York USA, International Standard 1990.

[36] S. S. Mishra, A. K. Agrawal, R.K. Nagaria, "A comparative performance analysis of various CMOS design techniques for XOR and XNOR circuits", *International Journal on Emerging Technologies*, vol. 1, nº 1, 2010.

[37] M.A Song et al., "A low-error and area-time efficient fixed-width booth multiplier", *IEEE Midwest Symposium on Circuits and Systems*, pp.590-593, 2003.

[38] Mora-Mora H, Mora-Pascual J, García-Chamizo JM, Jimeno-Morenilla A, Real-time arithmetic unit, Real-Time Systems 34 (1), 53-79, (2006).

[39] G. A. Ruiz, M. Granda, "Efficient canonic signed digit recoding", *Microelectronics Journal*, no. 42, pp. 1090-1097, 2011.

Hybrid Directional Weight-based Demosaicking for Bayer Color Filter Array

Yonghoon Kim and Jechang Jeong

***Abstract*—** This paper proposes and improves gradient estimation techniques in order to suppress the common demosaicking artifacts and introduces hybrid directional weights for adaptive directional interpolation. The hybrid weights are calculated using color correlation and intensity differences, and these weights apply to the whole interpolation process. Experimental results demonstrate that the proposed algorithm outperforms recent demosaicking methods on 42 images of Kodak and McM datasets.

***Keywords*—** Bayer pattern CFA, CFA interpolation, demosaicking, directional interpolation.

I. Introduction

CONVENTIONAL digital cameras, which use single chip, exploit color filter array (CFA) to capture different spectral components. CFA interpolation is the process that reconstructs the full color image from the CFA captured image. CFA interpolation is a key algorithm for low-cost digital cameras which uses a single-chip. These cameras take images only one color per a pixel position although three color values are required for a full color image. Therefore to record the color channel effectively, single-chip digital cameras use a CFA. The resulting image is given as a gray-scale mosaic-like image. The CFA interpolation, which is called demosaicking, is the process that reconstructs a three color layer form CFA pattern image. The most commonly exploited pattern is Bayer pattern [1], which is shown in Fig. 1. The reason why green (G) values have a higher sampling rate than red (R) and blue (B) color channels is that the human visual system is sensitive to medium wavelength ranges which match with green color spectrum.

There are two main ideas which are exploited in a modern CFA interpolation algorithm. The first idea is interpolating the green channel first because it has twice as many numbers as other color channels. Therefore, a green channel suffers less from aliasing and an accurately interpolated green channel helps reconstruction of other color channels. The second idea is using the correlation of green-red and green-blue. Based on this concept, several demosaicking solutions have been proposed. In [2], the spectral correlation is modeled as constant color ratio rule and other concept, which is constant color difference rule, is proposed [3-4].

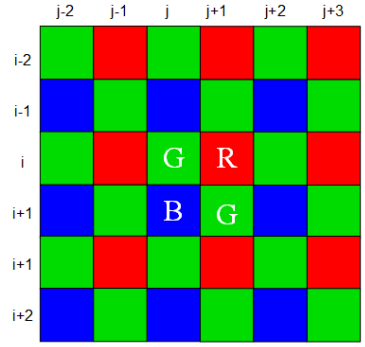


Fig. 1. Bayer color filter array pattern.

In [5], inter-channel correlation is used for color interpolation and one-step iteration algorithm is exploited to reduce the complexity. Chung *et al*. [6] tried to preserve details in texture region by using hard direction decision based on color difference variance. In [7], Paliy *et al*. modified the algorithm of [6] by proposing scale adaptive filtering based on local polynomial approximation (LPA). In [8], authors adopted the concept of directional demosaicking noise and proposed directional linear minimum mean square-error estimation (DLMMSE) algorithm.

Several methods proposed more precise gradient estimation algorithms. For example, Menon *et al*. [9] proposed demosaicking with directional filtering and a posteriori decision using color gradients over a window. In [10], Leung *et al.* tried to solve the demosaicking problem in the frequency domain. Chen and Chang [11] calculated the weight for interpolation by detecting the edge characteristics of a digital image, and Dengwen *et al*. [12] proposed color demosaicking algorithm using directional filtering and weighting. Pekkucuksen *et al*. proposed orientation-free edge strength filter (ESF) [13] and iteratively refined the green channel using neighboring color difference values. They also proposed multi-scale gradient (MSG) [14] based CFA interpolation algorithm which achieves objective performance improvement. In [15], Chen *et al*.

Y. Kim and J.Jeong are with Department of Electronics and Computer Engineering, Hanyang University, Haengdangdong, Sungdonggu, Seoul, South Korea (e-mail: hoonykim85@gmail.com, jjeong@hanyang.ac.kr*).
"This research was supported by the MSIP(Ministry of Science, ICT & Future Planning), Korea, under the project for technical development of information communication & broadcasting supervised by IITP (Institute for Information & Communications Technology Promotion)(IITP 2014-044-057-001)

proposed voting strategy to interpolate color component more accurately.

In this paper, we present gradient estimation methods to suppress color interpolation artifacts and introduce hybrid directional weights (HDW) for adaptive directional interpolation. The hybrid weights are calculated using color correlation and intensity differences, and these weights apply to the whole demosaicking process. The remaining of this paper is described as following. In Section II, the proposed hybrid directional weight-based demosaicking algorithm is explained. Experimental results are described in Section III. Conclusion remarks are reported in Section IV.

II. Proposed Algorithm

The gradient information is important for adaptive directional interpolation. The proposed algorithm uses gradient information which is calculated from the color difference value. The estimated horizontal (H) and vertical (V) color difference are combined with the ratio of integrated gradient of each direction over the window. The color differences are calculated as follows:

$$d_{i,j}^{H}=\begin{cases} Z_{i,j}-Z_{i,j}^{H}, & Z_{i,j}\in\{G\} \\ Z_{i,j}^{H}-Z_{i,j}, & Z_{i,j}\in\{R,B\} \end{cases},$$
$$d_{i,j}^{V}=\begin{cases} Z_{i,j}-Z_{i,j}^{V}, & Z_{i,j}\in\{G\} \\ Z_{i,j}^{V}-Z_{i,j}, & Z_{i,j}\in\{R,B\} \end{cases}, \quad (1)$$

where d^H and d^V represent horizontal and vertical color difference estimates, respectively. Parameters Z^H and Z^V are defined as follows:

$$Z_{i,j}^{H}=\frac{2Z_{i,j}-Z_{i,j-2}-Z_{i,j+2}}{4}+\frac{Z_{i,j-3}+9\left(Z_{i,j-1}+Z_{i,j+1}\right)+Z_{i,j+3}}{20},$$
$$Z_{i,j}^{V}=\frac{2Z_{i,j}-Z_{i,j-2}-Z_{i+2,j}}{4}+\frac{Z_{i-3,j}+9\left(Z_{i-1,j}+Z_{i+1,j}\right)+Z_{i+3,j}}{20}.$$

Here, Z denotes the Bayer pattern image plane. Using these values, the gradients are calculated as follows:

$$\Delta_{i,j}^{H}=\left|d_{i,j+1}^{H}-d_{i,j-1}^{H}\right|,$$
$$\Delta_{i,j}^{V}=\left|d_{i+1,j}^{V}-d_{i-1,j}^{V}\right|, \quad (3)$$

where Δ^H and Δ^V represent horizontal and vertical gradient. Gradient maps of each direction are produced during this process.

A. Hybrid Directional Weights (HDW)

For the directional interpolation, two sets of weights are used. The first set is vertical and horizontal weights for the initial green channel and R/B interpolation. The second set is four directional weights, up (u), down (d), left (l), and right (r), for green channel refinement step and R/B interpolation. To increase the accuracy, not only the gradient values of color difference but also gradient of pixel intensity of same color plane are exploited. This is because color difference is good measure for estimating gradient while it cannot fully cover the intensity changes. The gradient of same color plane is defined as follows:

$$\Phi_{i,j}^{H}=\left|Z_{i,j+1}-Z_{i,j-1}\right|,$$
$$\Phi_{i,j}^{V}=\left|Z_{i+1,j}-Z_{i-1,j}\right|, \quad (4)$$

where Φ^H and Φ^V denote the horizontal and vertical gradients of same color plane, respectively. For the directional weight, Δ^H and Δ^V are used as main factors with window size 5x5, and Φ^H and Φ^V are used as control parameter with window size 3x3. The hybrid directional weights are given as follows:

$$w_{i,j}^{H}=1/\left[\left(\sum_{m=-2}^{2}\sum_{l=-2}^{2}\Delta_{i+m,j+l}^{H}\times\sum_{m=-1}^{1}\sum_{l=-1}^{1}\Phi_{i+m,j+l}^{H}\right)^{2}+1\right],$$
$$w_{i,j}^{V}=1/\left[\left(\sum_{m=-2}^{2}\sum_{l=-2}^{2}\Delta_{i+m,j+l}^{V}\times\sum_{m=-1}^{1}\sum_{l=-1}^{1}\Phi_{i+m,j+l}^{V}\right)^{2}+1\right], \quad (5)$$

where w^H and w^V denote the horizontal and vertical weights and 1 is added to void division by zero. The four hybrid directional weights are calculated similarly. The four weights are calculated as follows:

$$w_{i,j}^{u}=1/\left[\sum_{m=-2}^{0}\sum_{l=-1}^{1}\Delta_{i+m,j+l}^{V}\times\sum_{m=-2}^{0}\sum_{l=-1}^{1}\Phi_{i+m,j+l}^{V}+1\right],$$
$$w_{i,j}^{d}=1/\left[\sum_{m=0}^{2}\sum_{l=-1}^{1}\Delta_{i+m,j+l}^{V}\times\sum_{m=0}^{2}\sum_{l=-1}^{1}\Phi_{i+m,j+l}^{V}+1\right],$$
$$w_{i,j}^{l}=1/\left[\sum_{m=-1}^{1}\sum_{l=-2}^{0}\Delta_{i+m,j+l}^{H}\times\sum_{m=-1}^{1}\sum_{l=-2}^{0}\Phi_{i+m,j+l}^{H}+1\right],$$
$$w_{i,j}^{r}=1/\left[\sum_{m=-1}^{1}\sum_{l=0}^{2}\Delta_{i+m,j+l}^{H}\times\sum_{m=-1}^{1}\sum_{l=0}^{2}\Phi_{i+m,j+l}^{H}+1\right], \quad (6)$$

where w^u, w^d, w^l, and w^r denote weight of up, down, left, and right, respectively.

B. Green Channel Interpolation

In this stage, color difference value which is located at the missing green pixel position is reconstructed by blending directional color difference calculated using Eq. (1). The initial green channel interpolation process is given as follows:

ISBN: 978-1-61804-290-3

$$\tilde{d}_{i,j} = \frac{d_{i,j}^{H} \cdot w_{i,j}^{H} + d_{i,j}^{V} \cdot w_{i,j}^{V}}{w_{i,j}^{H} + w_{i,j}^{V}}, \tag{7}$$

where $\tilde{d}$ denotes estimated color difference value. In this stage, green value can be calculated from estimated color difference, but estimated values are kept for the next process.

C. Green Channel Update

After the initial green channel interpolation step, color difference estimates are fully reconstructed and they can be further refined with neighboring correlations based on constant color difference assumption. The four close neighbors are combined with estimate color difference with its own weights. The updating process is given as follows:

$$\hat{d}_{i,j} = \tilde{d}_{i,j} \cdot \varepsilon + \frac{\left[w_{i,j}^{u} \cdot \tilde{d}_{i-2,j} + w_{i,j}^{d} \cdot \tilde{d}_{i+2,j} + w_{i,j}^{l} \cdot \tilde{d}_{i,j-2} + w_{i,j}^{r} \cdot \tilde{d}_{i,j+2} \right]}{w_{i,j}^{u} + w_{i,j}^{d} + w_{i,j}^{l} + w_{i,j}^{r}} \cdot (1-\varepsilon), \tag{8}$$

where parameter $\hat{d}$ represents the refined color difference value, and ε is the parameter to control the contribution of original and update part. Parameter ε is obtained empirically, which is given as 0.4 in this paper. This process improves PSNR and reduces the color artifacts, and it performs only once. After complete the process, final green value is calculated as follows:

$$G_{i,j}^{'} = Z_{i,j} + \hat{d}_{i,j}, \tag{9}$$

where G' denotes the interpolated green value.

D. Red and Blue Channel Interpolation

The R/B channel interpolation consists of two steps. Red pixels at the blue pixel location and blue pixels at the red pixel location are interpolated first, then red and blue pixels at the green pixel location are performed. The R/B pixels at the B/R pixel location are estimated by using the 7 by 7 filter. For better estimation with avoiding heavy complexity, the directional weights defined in Eq. (6) are reused to make four diagonal weights. The interpolation filter is given as follows:

$$f_{(i,j)}^{RB} = \frac{1}{6(w_{i,j}^{ul} + w_{i,j}^{ur} + w_{i,j}^{dl} + w_{i,j}^{dr})} \times \begin{bmatrix} 0 & 0 & -w_{i,j}^{ul} & 0 & -w_{i,j}^{ur} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -w_{i,j}^{ul} & 0 & 8 \cdot w_{i,j}^{ul} & 0 & 8 \cdot w_{i,j}^{ur} & 0 & -w_{i,j}^{ur} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -w_{i,j}^{dl} & 0 & 8 \cdot w_{i,j}^{dl} & 0 & 8 \cdot w_{i,j}^{dr} & 0 & -w_{i,j}^{dr} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -w_{i,j}^{dl} & 0 & -w_{i,j}^{dr} & 0 & 0 \end{bmatrix}, \tag{10}$$

where $w_{i,j}^{ul}, w_{i,j}^{ur}, w_{i,j}^{dl}$, and $w_{i,j}^{dr}$ are computed as follows:

$$\begin{aligned} w_{i,j}^{ul} &= w_{i,j}^{u} + w_{i,j}^{l}, \\ w_{i,j}^{ur} &= w_{i,j}^{u} + w_{i,j}^{r}, \\ w_{i,j}^{dl} &= w_{i,j}^{d} + w_{i,j}^{l}, \\ w_{i,j}^{dr} &= w_{i,j}^{d} + w_{i,j}^{r}. \end{aligned} \tag{11}$$

The red and blue pixel estimates at blue and red pixel location is calculated as follows:

$$\begin{aligned} R_{i,j}^{'} &= G_{i,j}^{'} - (\hat{d}_{i,j} \otimes f_{i,j}^{RB}), \\ B_{i,j}^{'} &= G_{i,j}^{'} - (\hat{d}_{i,j} \otimes f_{i,j}^{RB}), \end{aligned} \tag{12}$$

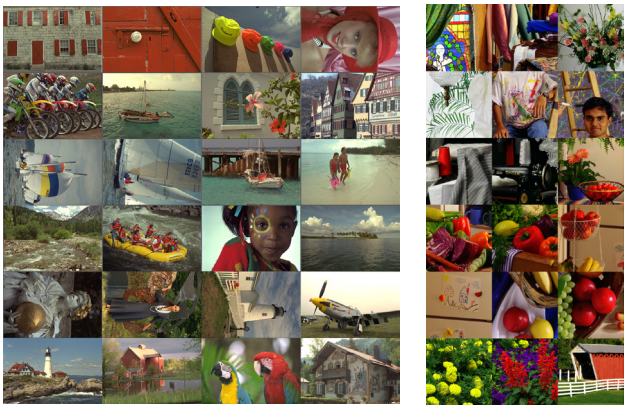
where $\otimes$ denotes the sum of all elements after element-wise matrix multiplication.

The red and blue pixels at green pixel location are interpolated only using weight calculated in Eq. (5). The reason that four hybrid directional weights are not used for this interpolation step is that it does not provide any performance gain. The red and blue pixel estimation is calculated as follows:

$$\begin{aligned} R_{i,j}^{'} &= G_{i,j}^{'} - \left[\frac{w_{i,j}^{V} \cdot \left(d_{i-1,j}^{GR} + d_{i+1,j}^{GR} \right) + w_{i,j}^{H} \cdot \left(d_{i,j-1}^{GR} + d_{i,j+1}^{GR} \right)}{2(w_{i,j}^{V} + w_{i,j}^{H})} \right], \\ B_{i,j}^{'} &= G_{i,j}^{'} - \left[\frac{w_{i,j}^{V} \cdot \left(d_{i-1,j}^{GB} + d_{i+1,j}^{GB} \right) + w_{i,j}^{H} \cdot \left(d_{i,j-1}^{GB} + d_{i,j+1}^{GB} \right)}{2(w_{i,j}^{V} + w_{i,j}^{H})} \right], \end{aligned} \tag{13}$$

where d^{GB} and d^{GR} represent green-blue color difference and green-red color difference.When you submit your final version, after your paper has been accepted, prepare it in two-column format, including figures and tables.

III. EXPERIMENTAL RESULTS



(a) (b)

Fig. 2. (a) Kodak and (b) McM test images.

ISBN: 978-1-61804-290-3

TABLE I
CPSNR Comparison for Different Demosaicking Algorithms (dB)

Test set	No.	DLMMSE	LSLCD	ESF	EDAEP	MSG	VDI	Proposed
Kodak	1	38.52	38.68	39.91	34.91	39.79	35.34	39.36
	2	40.93	40.81	40.82	39.70	41.59	39.64	41.92
	3	42.75	42.42	42.55	41.69	43.63	41.67	43.90
	4	41.09	41.37	40.45	39.99	41.42	39.56	41.84
	5	38.10	38.38	37.55	35.77	38.94	36.70	39.31
	6	40.27	40.47	41.22	36.83	41.15	37.02	41.15
	7	42.39	42.89	42.15	41.23	43.39	41.86	43.77
	8	36.08	35.75	37.19	32.82	37.39	34.07	37.20
	9	42.86	42.87	42.96	40.85	43.71	41.72	43.83
	10	42.61	42.85	42.60	40.93	43.30	41.25	43.42
	11	40.09	40.35	40.69	37.45	41.23	37.74	41.19
	12	43.53	43.48	43.80	41.61	44.32	41.70	44.55
	13	34.81	35.17	36.11	31.48	35.96	31.35	35.58
	14	37.03	36.99	36.11	36.25	37.77	36.58	38.35
	15	39.87	40.08	39.27	38.78	40.22	38.08	40.46
	16	43.83	44.12	44.77	40.40	44.58	40.62	44.70
	17	41.86	42.04	41.95	39.59	42.46	39.81	42.44
	18	37.45	37.80	37.72	35.37	38.22	35.18	38.21
	19	40.90	40.89	41.49	37.73	41.99	38.97	41.90
	20	41.27	41.28	41.52	39.24	42.13	39.96	42.14
	21	39.17	39.33	40.30	36.43	40.26	37.03	40.00
	22	38.46	38.70	38.41	37.15	39.00	37.46	39.14
	23	43.30	43.16	42.45	42.29	43.89	41.42	44.16
	24	35.52	35.59	35.34	33.72	35.72	33.07	35.67
Average		40.11	40.23	40.31	38.01	40.92	38.24	41.01
McM	1	26.98	26.05	26.38	27.60	27.05	28.01	27.61
	2	33.68	33.05	33.48	33.99	33.67	34.18	34.12
	3	32.59	32.34	32.56	32.07	32.93	32.64	33.24
	4	34.32	35.00	34.97	34.36	35.49	36.00	36.09
	5	31.27	30.61	30.64	32.10	31.12	32.63	31.89
	6	33.84	33.08	32.57	35.03	33.56	35.63	34.59
	7	38.64	38.33	39.10	36.22	39.17	36.03	38.77
	8	37.45	36.70	37.85	37.12	37.61	37.41	38.24
	9	34.41	33.67	34.39	35.23	34.69	35.95	35.51
	10	36.34	35.48	35.78	37.01	36.47	37.28	37.08
	11	37.25	36.22	36.61	37.83	37.28	37.98	37.86
	12	36.60	36.12	36.16	37.16	36.80	37.09	37.33
	13	38.79	38.21	38.67	39.34	38.83	39.40	39.45
	14	37.23	36.52	37.21	37.65	37.13	37.32	37.56
	15	37.27	36.55	37.01	37.76	37.19	37.86	37.62
	16	30.46	29.05	29.24	31.39	30.18	31.40	30.85
	17	29.31	28.65	28.57	30.59	29.30	31.15	30.20
	18	33.92	33.05	33.69	34.05	34.10	34.23	34.50
Average		34.46	33.82	34.16	34.81	34.59	35.12	35.14
Total average		37.69	37.48	37.67	36.64	38.21	36.90	38.49

To evaluate the performance of the proposed HDW algorithm, it is tested on the Kodak and McM of color test-images shown in Fig. 2. The Kodak images consist of 24 images with size of 768x512 and McM consist of 18 images with size of 500x500. The proposed HDW has been compared with directional linear minimum mean square-error estimation (DLMMSE), least-squares luma-chorma demultiplexing (LSLCD), effective demosaicking algorithm based on edge property (EDAEP), edge strength filter (ESF), multi-scale gradient (MSG), and voting-based directional interpolation (VDI). The interpolated images with various methods are compared excluding the border within 10 pixels, and we conducted simulations using MATLAB with an Intel(R) Core(TM) i7-4770K CPU at 3.50-GHZ quad-core processor.

The CPSNR results are presented in Table I for each 42 images. In case of Kodak image set, the proposed HDW shows the best average CPSNR among all algorithms and its average CPSNR was 0.09dB and 1.77dB higher than MSG and VDI, respectively. The proposed algorithm also shows the best

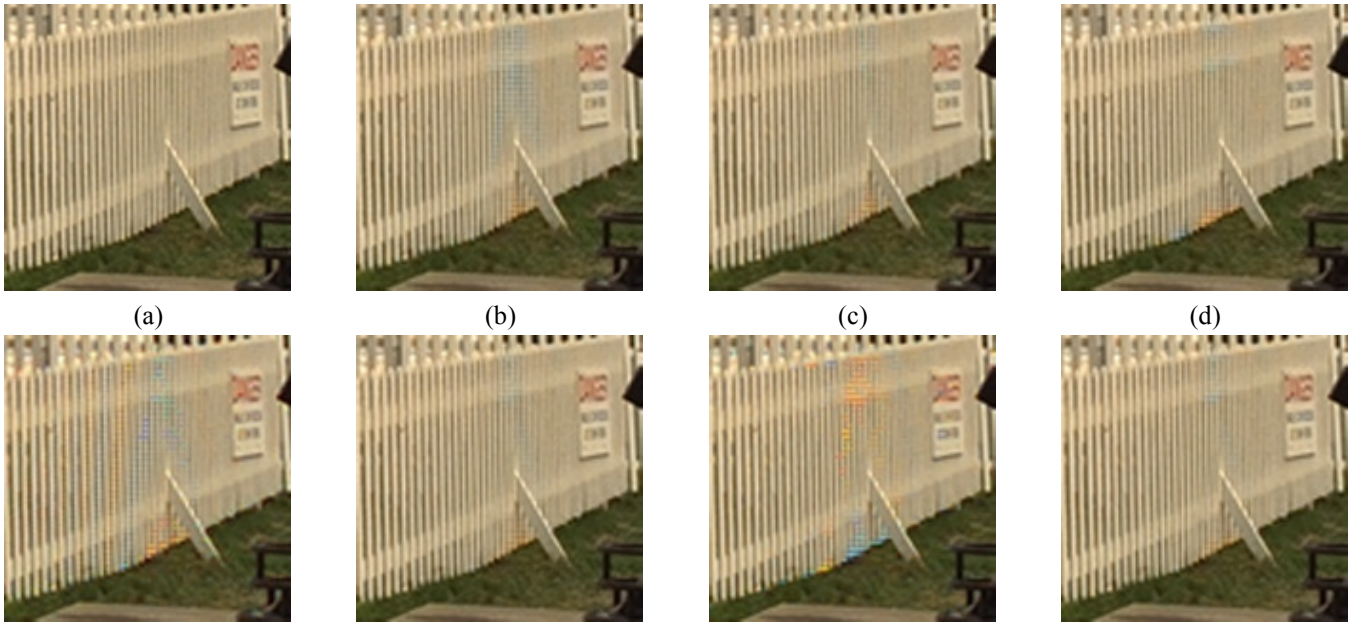
ISBN: 978-1-61804-290-3

TABLE II
S-CIELAB ΔE^* Comparison for Different Demosaicking Algorithms

Test set	No.	DLMMSE	LSLCD	ESF	EDAEP	MSG	VDI	Proposed
Kodak	1	1.108	1.133	0.995	1.474	0.960	1.368	0.969
	2	0.649	0.642	0.663	0.703	0.601	0.728	0.582
	3	0.479	0.508	0.510	0.524	0.450	0.541	0.439
	4	0.672	0.643	0.727	0.729	0.639	0.795	0.614
	5	1.070	1.011	1.242	1.271	1.011	1.107	0.902
	6	0.774	0.766	0.741	0.989	0.727	0.950	0.689
	7	0.574	0.523	0.584	0.664	0.489	0.564	0.479
	8	1.332	1.443	1.190	1.825	1.138	1.477	1.118
	9	0.569	0.581	0.585	0.653	0.534	0.624	0.520
	10	0.544	0.541	0.569	0.610	0.518	0.617	0.506
	11	0.737	0.730	0.722	0.923	0.670	0.859	0.649
	12	0.461	0.478	0.466	0.550	0.432	0.529	0.422
	13	1.560	1.539	1.477	2.012	1.448	2.042	1.422
	14	0.981	0.971	1.043	1.138	0.871	1.091	0.836
	15	0.657	0.633	0.722	0.691	0.636	0.726	0.610
	16	0.531	0.525	0.508	0.704	0.509	0.689	0.490
	17	0.517	0.506	0.543	0.641	0.499	0.618	0.483
	18	1.008	0.964	1.092	1.160	0.969	1.147	0.930
	19	0.752	0.753	0.736	0.957	0.674	0.847	0.669
	20	0.545	0.564	0.555	0.647	0.502	0.579	0.497
	21	0.901	0.889	0.861	1.091	0.831	1.019	0.812
	22	0.964	0.907	1.023	1.056	0.918	0.995	0.889
	23	0.495	0.479	0.560	0.536	0.465	0.544	0.457
	24	1.018	0.969	1.102	1.156	1.019	1.118	0.970
Average		0.787	0.779	0.801	0.946	0.730	0.899	0.706
McM	1	3.204	3.493	3.784	2.903	3.170	2.620	2.922
	2	1.315	1.295	1.453	1.246	1.252	1.144	1.152
	3	2.083	1.942	2.076	2.290	1.825	1.893	1.784
	4	1.443	1.236	1.461	1.484	1.086	1.113	1.021
	5	1.615	1.714	1.771	1.526	1.551	1.374	1.430
	6	1.511	1.594	1.857	1.278	1.502	1.127	1.311
	7	0.941	1.001	0.944	1.228	0.897	1.199	0.942
	8	0.631	0.637	0.659	0.723	0.589	0.613	0.516
	9	1.377	1.313	1.396	1.305	1.203	1.058	1.084
	10	1.083	1.075	1.192	0.999	1.025	0.932	0.940
	11	0.793	0.827	0.907	0.725	0.783	0.675	0.725
	12	1.089	1.098	1.257	0.940	1.081	0.956	0.993
	13	0.711	0.783	0.739	0.683	0.692	0.650	0.658
	14	0.797	0.807	0.860	0.789	0.784	0.742	0.749
	15	0.855	0.848	0.913	0.819	0.836	0.773	0.795
	16	2.064	2.572	2.839	1.673	2.317	1.499	2.079
	17	2.826	2.647	3.090	2.290	2.577	1.960	2.219
	18	1.510	1.582	1.601	1.451	1.460	1.287	1.358
Average		1.436	1.470	1.600	1.353	1.368	1.201	1.260
Total average		1.065	1.075	1.143	1.120	1.003	1.028	0.944

performance on McM datasets, and the average CPSNR of proposed algorithm is better than VDI by 0.02dB and EDAEP by 0.33dB. Interestingly, most algorithms work well only for one dataset, but do not for the other dataset. For the Kodak images, DLMMSE, LSLCD, ESF, and MSG show fine results but they gave relatively worse results on McM images. On the other hand, EDAEP and VDI only perform well on McM datasets. It is remarkable that the proposed algorithm gives the best performance on both datasets. In terms of total average CPSNR, HDW was better than the MSG, which shows the second best results on Kodak dataset, by 0.28dB, and VDI, which gives second best results on McM dataset, by 1.59dB.

To further evaluate the objective performance, we exploited another metric S-CIELAB ΔE^*, which measures how accurate the reproduction of a color is to the original when viewed by a human observer, and the results are provided in Table II. It can be seen that the proposed algorithm outperforms other algorithms in terms of total average S-CIELAB ΔE^*. MSG



(a) (b) (c) (d)

(e) (f) (g) (h)

Fig. 3. Picket fence region from the #19 Kodak image. (a) Original, (b) DLMMSE, (c) LSLCD, (d) ESF, (e) EDAEP, (f) MSG, (g) VDI, and (h) HDW (proposed).

TABLE III
Average Computational Time Comparison (Tested on Images with Size of 768x512)

	DLMMSE	**LSLCD**	**ESF**	**EDAEP**	**MSG**	**VDI**	**Proposed (HDW)**
Time (s)	12.89	4.52	9.31	0.46	8.29	5.31	3.09

scores the second best performance followed by VDI. The VDI shows fine results in terms of total average of S-CIELAB ΔE^* even it gives the second worst total average CPSNR.

The visual comparison of the picket fence region of the lighthouse (#19 in Kodak dataset) is provided in Fig. 3. It can be seen that DLMMSE, VDI and EDAEP show severe rainbow artifacts at the high frequency region. The proposed algorithm gives the clear image and successfully removes the most of the demosaicking artifact. In Fig. 4, demosaicking results of partial zoomed image (#18 in McM dataset) are presented. It is obvious that EDAEP and VDI produce more artifacts around the character and edge of the roof, and DLMMSE, LSLCD, and ESF show zipper effect at the region where chrominance component abruptly changes across the boundary.

In Table III, the computational complexity of HDW and the other demosaicking methods were evaluated using computational time. The execution times have been evaluated using the source codes which were originally distributed from the authors under the same condition. To evaluate computational complexity, the Kodak dataset was used to calculate the computational time. From the results, we can see that the proposed algorithm gives approximately 1.6 times faster than VDI which gives the second best results on McM dataset and 2.6 times faster than MSG which shows the second best results on Kodak dataset.

From the above objective and subjective evaluations results, it can be seen that the proposed algorithm outperformances other state-of-the-art algorithms with low complexity burden.

IV. Conclusion

In this paper, a hybrid directional weight for CFA interpolation algorithm is proposed, which exploits both gradient of color difference and intensity difference for calculating weights. The whole interpolation process is adaptively combined with hybrid directional weights and it leads to reducing the common demosaicking artifacts and improving the interpolation performance in terms of average CPSNR, S-CIELAB ΔE^*.

References

[1] B. E. Bayer, "Color imaging array," U.S. Patent 3 971 065, Jul. 1976.

[2] R. Kimmel, "Demosaicing: Image reconstruction from color CCD samples," *IEEE Trans. Image Process*, vol. 8, no. 9, pp. 1221–1228, Sep. 1999.

[3] J. F. Hamilton and J. E. Adams, "Adaptive color plane interpolation in single sensor color electronic camera," U.S. Patent 5 629 734, May 1997.

[4] C. A. Laroche andM. A. Prescott, "Apparatus and method for adaptively interpolating a full color image utilizing chrominance gradients," U.S. Patent 5 373 322, Dec. 1994.

[5] Y.M. Lu, M. Karzand, M. Vetterli, 'Demosaicking by alternating projections: theory and fast one-step implementation', *IEEE Trans. Image Process*., vol. 19, no. 8, pp. 2085–2098, Aug. 2010.

[6] K. H. Chung and Y. H. Chan, "Color demosaicing using variance of color differences," *IEEE Trans. Image Process.,* vol. 15, no. 10, pp. 2944–2955, Oct. 2006.

[7] D. Paliy, V. Katkovnik, R. Bilcu, S. Alenius, and K. Egiazarian, "Spatially adaptive color filter array interpolation for noiseless and noisy data," *Int. J. Imag. Syst. Technol.,* vol. 17, no. 3, pp. 105–122, 2007.

[8] L. Zhang and X. Wu, "Color demosaicking via directional linear minimum mean square-error estimation," *IEEE Trans. Image Process.,* vol. 14, no. 12, pp. 2167–2178, Dec. 2005.

[9] D. Menon, S. Andriani, and G. Calvagno, "Demosaicing with directional filtering and a posteriori decision," *IEEE Trans. Image Process.*, vol. 16, no. 1, pp. 132–141, Jan. 2007.

[10] B. Leung, G. Jeon and E. Dubois, "Least-squares luma-chroma demultiplexing algorithm for Bayer demosaicking," *IEEE Trans. Image Process*., vol. 20, no. 7, pp. 1885–1894, Jul. 2011.

[11] W. Chen, P. Chang, "Effective demosaicking algorithm based on edge property for color filter arrays," *Digit. Signal Process.,* vol. 22, no. 1, pp. 163–169, Jan. 2012.

[12] Z. Dengwen, S. Xiaoliu and D. Weiming, "Colour demosaicking with directional filtering and weighting," *IET Image Process.,* vol. 6, no. 8, pp. 1084–1092, Nov. 2012.

[13] I. Pekkucuksen and Y. Altunbasak. "Edge strength filter based color filter array interpolation," *IEEE Trans. Image Process.,* vol. 21, no. 1 pp. 393–397, Jan. 2012.

[14] I. Pekkucuksen and Y. Altunbasak, "Multiscale Gradients-Based Color Filter Array Interpolation," *IEEE Trans. Image Process.*, vol. 22, no. 1, pp. 157–165, Jan. 2013.

[15] X. Chen, G. Jeon, and J. Jeong, "Voting-based directional interpolation method and its application to still color image demosaicking," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 24, no. 2, pp. 255-262, Feb. 2014.

[16] L. Zhang, L, Zhang, X. Mou, and D. Zhang, "FSIM: a feature similarity index for image quality assessment," *IEEE Trans. Image Process.,* vol. 20, no. 8 pp. 2378–2386, Aug. 2011.

Novel concept of Power Management Architecture based on Smart EV Learning DataBase

Chokri MAHMOUDI
Email: Chokri.mahmoudi@gmail.com
Cell: (+216) 23 022 655
SPEG Research Unit
ENIG National School of Engineering of Gabès, University Of Gabès, TUNISIA

Dr. Aymen FLAH
Email: dr.aymen.flah@ieee.org
Cell: (+216) 21 104 838
SPEG Research Unit
ENIG National School of Engineering of Gabès, University Of Gabès, TUNISIA

Pr. Lassaad SBITA
Email: lassaad.sbita@enig.rnu.tn
Cell: (+216) 98 660 023
SPEG Research Unit Director
ENIG National School of Engineering of Gabès , University Of Gabès, TUNISIA

Abstract **-** A new architecture for Power Management in Electric Vehicle is proposed in this paper and a new concept of Smart/learning architecture is provided. Structure concept is explained in details describing different layers. This work also introduces a new learning technique concept based on cloud experience exchange between Electric Vehicles. This enhancement aims to build a better vehicle experience in power management through Energy Experience sharing.

Index terms—Electric Vehicle; Power Management; PMC algorithms; Energy Experience, Cloud DataBase.

I. INTRODUCTION

Intelligent Transportation Systems (ITSs) have been widely studied for different purposes, especially traffic safety. Collision warning, avoidance or even autonomous correction and driving are now industrialized. Many intelligent systems such as Collision Mitigation Braking System (CMBS), Adaptive Cruise Control (ACC) and Lane Keeping Assist System (LKAS) have been introduced during last decade and received updates for more accuracy and efficiency [1]-[4].

ITSs can improuve power management and take it to the next level. Vehicles are equipped with many comfort features, driving assist and security systems that can provide useful informations about driver behavior, vehicle position, weather conditions, etc. to optimize power management [5].

In 2007, some results were introduced on this concept. Toyota lunched her first GPS-Assisted ACC system for HEV, result proved that, compared to normal driving; fuel consumption was reduced by 6.3% [6]. A predictive ACC version was proposed by Van Keulen from Eindhoven University for HEVs [7]. This enhancement improved fuel efficiency in a reasonable way.

In 2012, Keqiang LI and his team from Tsinghua University proposed an Intelligent Environment-Friendly Vehicle concept taking advantage of ITSs [8]. It integrates clean-energy powertrain, electrified chassis and intelligent information interaction devices. Vehicle sensors provide generous informations and through structure sharing, data fusion and control coordination technologies, great results were achieved.

Today, it is clear that current state of the art is reaching a point where new and significantly different architecture are needed. This work has as objective to provide a new point of view for power management based on different architecture[4,9]. The new architecture is designed for EV operating with Smart learning PMC Algorithm to enhance power management and improve electric vehicles' reliability, but can also be applied for different purposes and take advantage of experience sharing.

The remainder of this paper proceeds as follows. Section II is devoted to detail the intelligent power management architecture. Section III introduces the concept of energy experience in Electric Vehicle. In section IV, a general layout of Cloud Power Management DataBase is presented. The last section concludes and discusses future work.

II. INTELLIGENT POWER MANAGEMENT ARCHITECTURE

A. Conventionnal Learning Power Management Control Algorithm

To date, many works contributed to give a better understanding for Power Management in Electric Vehicle. Depending on powertrain architecture, various PMC algorithms have been introduced [5, 9-11]. This work focuses in improving learning algorithms and enhancing EV ability to master power management over time and experience. Driving factors and driver behavior has major impact on fuel efficiency, vehicle responsiveness and driving autonomy[12].

Up to now, the research reported few potential works in this category; a multi-modes PMC strategy for parallel HEV proposed in 2002 by Jeon using driving pattern recognition to select a control algorithm automatically from six optimized driving modes using Artificial Neural Networks (ANNs) [13,14]. Online prediction of future driving cycle based on recorded data was introduced by Ichikawa in 2004 [15].

ISBN: 978-1-61804-290-3

The learning strategy was introduced in 2005 by Chen and Salman for parallel HEVs to maximize fuel economy. A learning technique is applied to the cost function in order to adjust parameters in real time[16]. Neurodynamic programming was also used in power management control algorithms. Few years later, Kolmanovsky and Dextreit introduced Game Theory Algorithms in power management and experimental results on Land Rover Freelander HEV mark-2 were remarkable [17,18].

Thus, it is clear that to obtain the best energy consumption, a real-time controller must adapt itself to varying driving circumstances and conditions. In 2011, including driving conditions in power management has been though out[19,20]. Global Positioning System (GPS) retrieved informations were used to determine upcoming topography of the road and adjust PMC parameters. An environment friendly architecture is proposed for HEVs by Li in 2012 to improve efficiency[8]. This concept integrates three modules; a clean energy powertrain, and electrified chassis and intelligent information interaction devices.

B. New Smart Learning Archichitecture

In EV Power Management, main introduced algorithms, whether they are Offline or Online, are generally applied locally to optimize energy efficiency[21]. No potential studies were introduced in this category except safety through Vehicle to Vehicle communication[22,23]. The new architecture introduced in this work aims to explore vehicles' communication to improve power management. It is based on two levels;

- The first level; where vehicle optimizes power management using self-learning techniques and builds its own Energy Experience (E.EX). This unique E.EX will take advantage of various driving situations. The way the vehicle adjusts PMC algorithm parameters is related to its driving experience; the vehicle learns from driver's behavior and mood, GPS position, road conditions, time conditions, weather conditions. All these informations will help to build a rich energy experience and to set fine selection criteria for the next level. Now according to each driving situation, a new set of adjustment parameters is provided. Through time, Vehicle will develop a unique understanding for energy management. Its experience will grow, while adaptation and responsiveness will get better.
- The second level in this new architecture will take advantage of single vehicle achievements in energy experience to build a Smart Learning Database (SLDB). In this architecture, a cloud based Database will collect EV Energy Experiences from different connected vehicles to be sort by events.

Energy experiences are uploaded to the base and sort by very fine event criterion. Through a comprehensive power management data base solution, a vehicle detecting similar conditions during its path sends an assist request to the DB. This request will be identified and feedback will be downloaded to vehicle if available. Therefore, it gets advantage of valuable previous energy experience of another vehicle. By downloading adjustment algorithm parameters for power management, it obtains optimal, know to date, energy economy and vehicle responsiveness immediately.

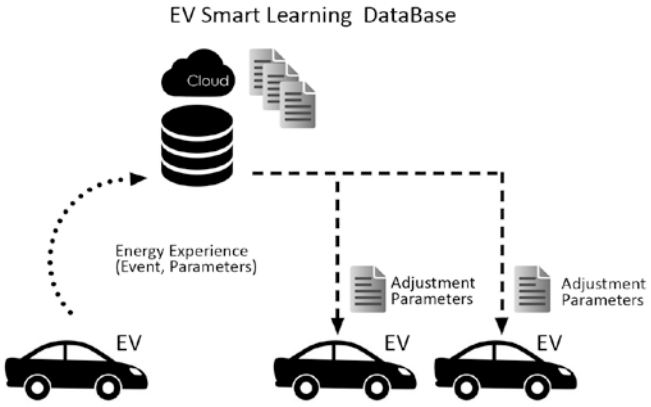


Fig1. Smart learning architecture concept

In figure (1), a basic representation of Electric Vehicles interaction according to the proposed architecture is illustrated.

Description of a sequence:

In this architecture, we examine all informations collected by vehicle. This contributes in building vehicle decision. These informations constitute an event; e.g. An electric vehicle being driven by a female driver, having a happy mood, from a GPS position A to B, in specific road conditions, in summer, in specific weather condition and temperature, the ride was in a specific date & Time a weekend, a feast day, in the morning, traffic conditions during that day a traffic jam occurred. The vehicle status indicates fully charged batteries, with half fuel tank left. According to this event in space and time, with this gender of driver, in this mood, working with all mentioned conditions, we generate a unique set of adjustment parameters. In vehicle recorded data this event may occurs probably in the future. Drivers generally select unconsciously same itinerary to link between two points A and B. some drivers may know shortcuts during peak time in order to avoid jam and improve fuel economy. All these informations can bring assistance to new vehicles in this area to define, in a comprehensive way, the decision to make, itinerary to suggest and parameters to set in its PMC algorithm.

III. NEW ENERGY EXPERIENCE CONCEPT

Adaptive and Smart Learning algorithms estimate their optimal values in real-time in order to satisfy the charge-sustaining constraint and to achieve best performance[24,25].

Intelligent decision taken seems to be only related to vehicle. However, different circumstances may lead to complete different decisions and various results. An energy Experience is introduced in this section to define in a better perspective driver and environment important role in Vehicle self-learning.

ISBN: 978-1-61804-290-3

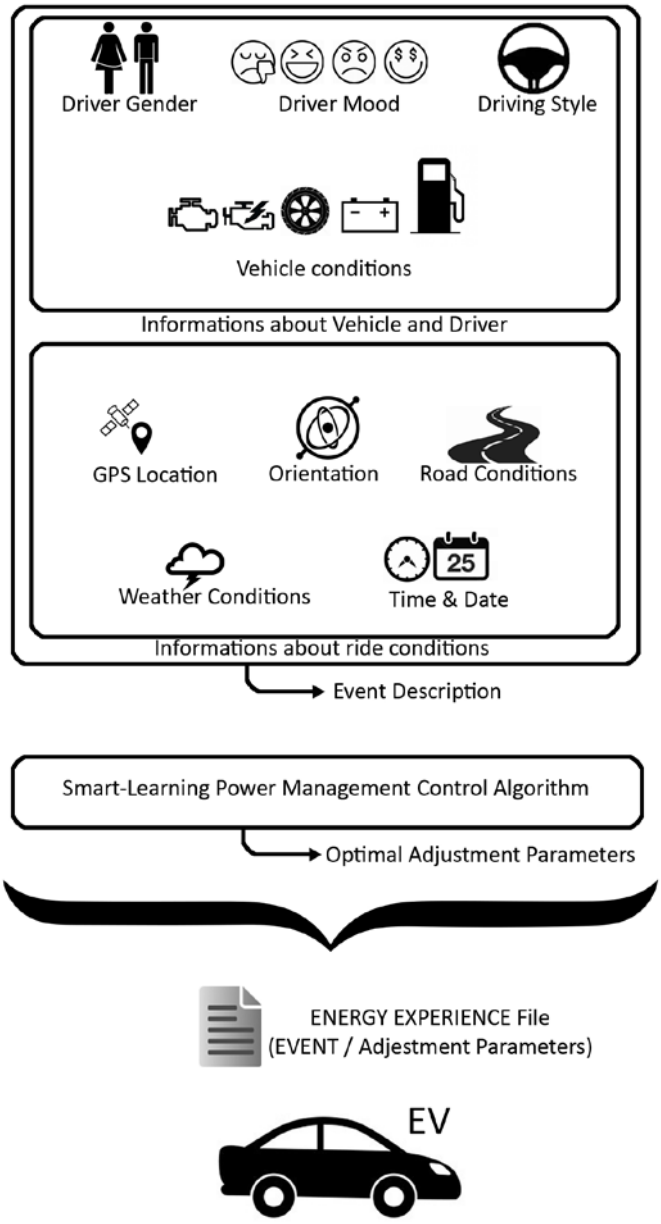


Fig2. Smart learning architecture concept

In figure(2), Energy Experience layers are introduced; Informations such as engine status, batteries' rate, etc. are provided as conventional from the Electric vehicle. To build a rich experience, informations from driver and environment such as mood, gender, GPS location, weather and road conditions are added to create an Event.

The Smart learning Power Management Control Algorithm runs separately to predict new adjustment parameters which are suitable for this event. The combination between the event and the parameters constitutes the energy experience.

To get DataBase assistance, a request mentioning the current event is formulated and sent. The vehicle should have an inexistent or poor experience related to the occurred situation. Basically, two scenarios are faced. In first case, according to the sent request, no similar situations are found (Figure 3). The DataBase cannot afford adjustment parameters for the Electric Vehicle. Thus, no assistance is offered. The vehicle generates its own adaptive parameters for the faced situation aiming to optimize power management. These Parameters, coupled with the event description are uploaded as an experience for future use.

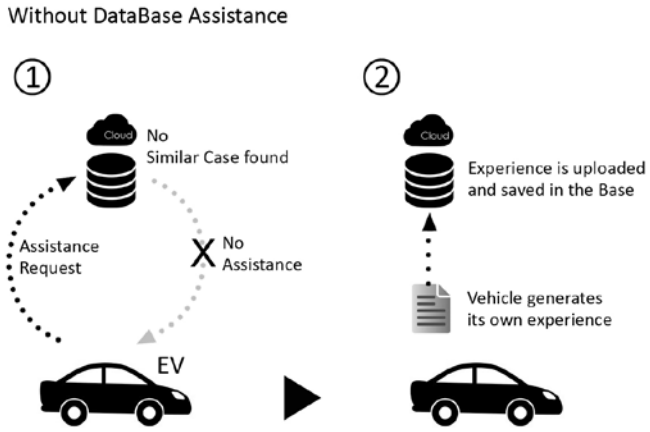


Fig3. PM without Smart DataBase assistance

In second case, a same or similar situation is found (Figure 4). The DataBase provides Electric Vehicle with adjustment parameters. The vehicle utilizes optimized parameters for the faced situation. These Parameters will allow to obtain immediate optimal PM, better vehicle responsiveness and to improve immediately fuel economy for a longer ride range.

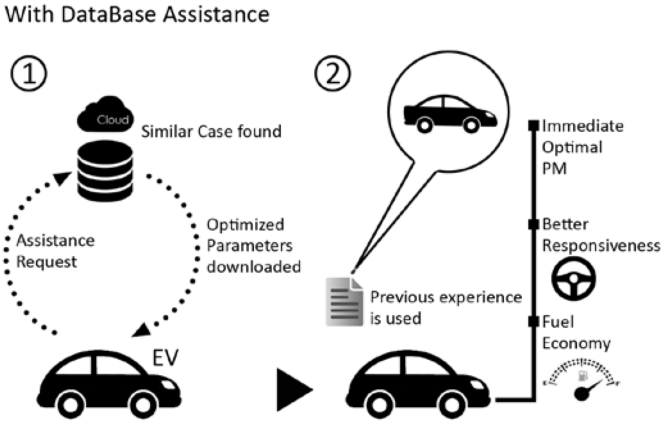


Fig4. PM with Smart DataBase assistance

As well as Electric Vehicles intend to improve knowledge through local and cloud experience sharing, the DataBase operates identically. Each time the uploaded parameters define a better efficiency for an existing experience compared to the saved data, the new parameters replace the old ones.

ISBN: 978-1-61804-290-3

TABLE I. REQUIRED INFORMATIONS TO BUILD AN EVENT

Information	***Description***
Vehicle Identification	Informations about Manifacturer[a] / Class: HEV, PHEV, BEV, FCEV, Technologies
Vehicle Status	Informations about Batteries, Capacitors, Fuel, engines, sensors
GPS Location	A GPS antenna provides coordinates
Vehicle Orientation	A Gyroscope calculates yaw, pitch and roll rates in real time
Weather Conditions	Weather informations can be provided by onboard probe and sensors or downloaded from online weather services[b]
Terrain/Road Conditions	Toporgaphy, road type and driving conditions can be downloded according to GPS location
Traffic Conditions	Traffic conditions and jam locations can be downloaded from online traffic services
Driver Gender	A camera with face analysing software can determine if driver is male or female
Driver Mood	A camera with face analysing software and steering-wheel heartbeat sensors can determine driver's mood (happy, angry, nervous or excited).
Driving Style	Acceleration and deceleration rates provided by speedometer and gyroscope can identify smooth, modrate or sport driving style.
Time	It identifies when event happens and helps to create chronological order.
Date	It identifies general timing: weekends, holidays, feast days, special events.

[a.] Architecture considers manufacturer independency as first sort criterion.

[b.] Online informations need data connection through cellular networks or wireless hotspots.

In Table1, we enumerate various informations delivered by vehicle sensors to create the event. In order to simplify systems and reduce complexity, we use structure sharing strategy. Which involves that same physical structure or sensor can be shared by different functions and provide needed information.

E.g. same camera can be used in Drowsiness Detection System (DDS) to alert driver sleepiness and as a face analyzer to determine driver mood. Same sensors used in Lane Keeping Assist System (LKAS) [23] to oblige driver to keep hands on steering wheel, can provide heartbeats rate and determine if the driver is nervous. Previous researches [26-29] have highlighted the prominence of structure sharing in Electric Vehicle.

Therefore, for complex system design, few studies were introduced. In 2012, Keqiang Li proposed a comprehensive design for structure sharing; In order to minimize sensors onboard and simplify architecture, same physical equipment can provide informations for different functions or features in a vehicle. This improves efficiency for complex architecture at the lowest cost [8].

In our case, we are more interested in multisensory principle, which aims to reduce the number of onboard sensors. As shown in figure (3), sensors are shared among ordinary functions such as comfort or security and Power Management Learning process. E.g. GPS antenna provides location informations to onboard navigation system for turn-by-turn assistance and, simultaneously, calculates coordinates which are useful to build the energy experience or to retrieve assistance from the Smart Database. Informations come from vehicle as well as driver and environment. The multisensory structure sharing should be controlled. Depending on valuation, to reduce cost, further redundancy or reliability, the structure will be defined and designed[30,31].

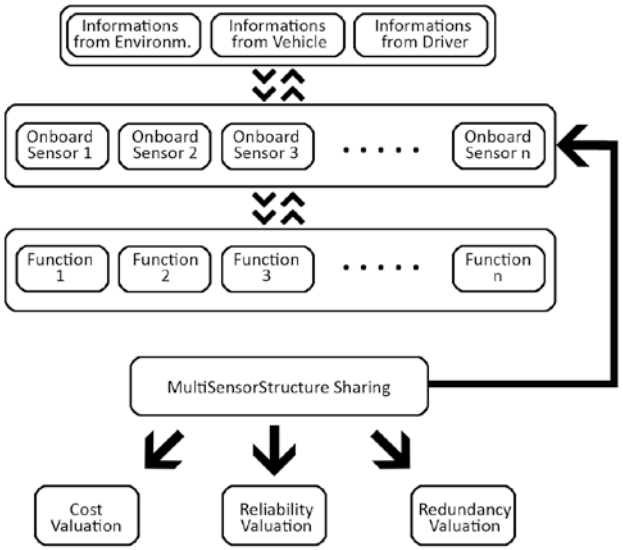


Fig.5. Multi-Sensor Structure Sharing

ISBN: 978-1-61804-290-3

IV. CLOUD POWER MANAGEMENT

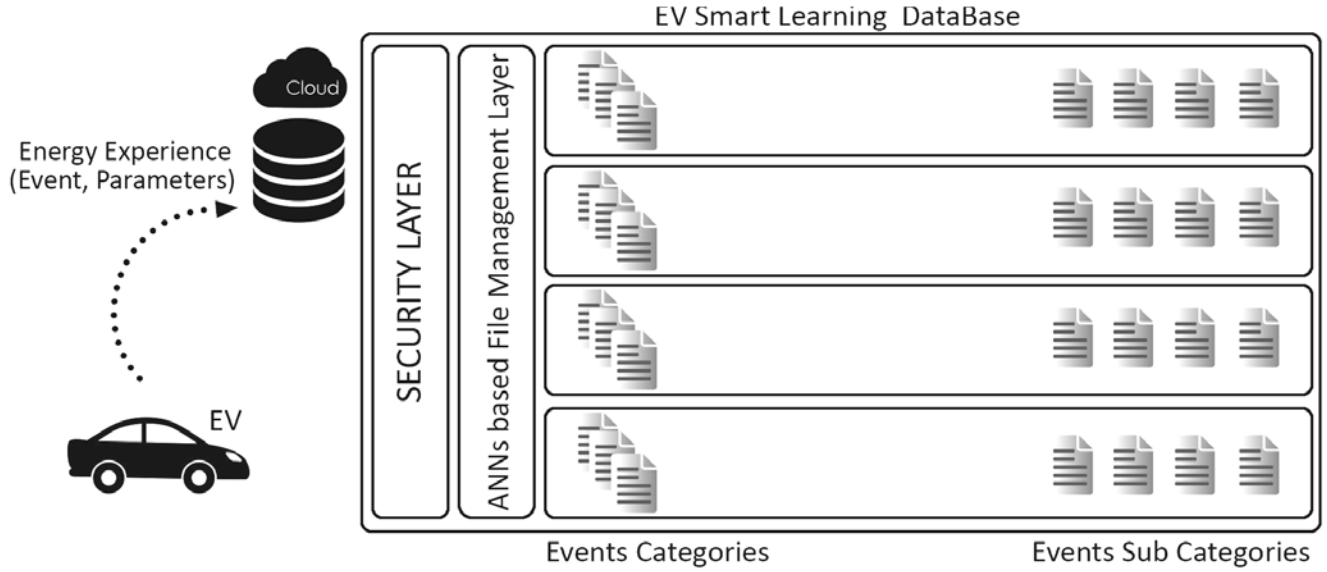


Fig6. Smart learning DataBase structure

A cloud database typically runs on a cloud computing platform, such as Salesforce, Go Grid, and Microsoft Azure. As deployment, the choice of Cloud databases is the independency due to different manufacturers' competition and professional secret anonymity. By using secured and independent virtual machines, decision about data model will be taken in further work whether it should SQL-based NoSQL data model.

As a general layout, Energy experiences are uploaded systematically to the base. When informations are requested in normal or predictive approach, database seeks the convenient experience base on similarity in events. Research and experience selection will be optimized with Artificial Neuronal Networks (ANNs). These Software functions aim to take advantages of Cloud Computing revolution

V. CONCLUSION

Providing a new green driving experience with environmental friendly goals should not compromise in vehicle responsiveness or driver's style and experience. To obtain such optimal behavior, many power management algorithms have been introduced and various high level supervisory architectures have been developed. In this work we introduced a new perspective for power management in electric vehicles based on a complete new architecture.

Connecting vehicles represents an important trend for the next generation. The key technology in the Smart Learning Architecture is experience sharing between vehicles to improve Self learning via Cloud Learning and enhance power management.

VI. REFERENCES

[1] C. Manzie, H. C. Watson, S. Halgamuge, and K. Lim, "A comparison of fuel consumption between hybrid and intelligent vehicles during urban driving," inProc. Inst. Mech. Eng. D, J. Automobile Eng., 2006, vol. 1,pp. 67–76.

[2] C. C. Chan, "The challenges and opportunity in the new century-clean, efficient and intelligent electric vehicles," inProc. Elect. Mach. Syst., 2003, vol. 1, pp. XIII–XXXI.

[3] C. Musardo, G. Rizzoni, and B. Staccia, "A-ECMS: An adaptive algorithm for hybrid electric vehicle energy management, decision and control, 2005.

and 2005 European control conference," inProc. 44th IEEE Conf. CDCECC, 2005, pp. 1816–1823.

[4] A. A. Malikopoulos "Supervisory Power Management Control Algorithms for Hybrid Electric Vehicles: A Survey". IEEE transactions on intelligent transportation systems. Digital Object Identifier 10.1109/TITS.2014.2309674, 2014 , Page(s): 1869 - 1885

[5] C. MAHMOUDI, A.FLAH, L.SBITA, "AN OVERVIEW OF ELECTRIC VEHICLE CONCEPT AND POWER MANAGEMENT STRATEGIES". IEEE Journal, Accepted under reference 978-1-4799-7300-2/14/©2014 IEEE. Paper accepted for inclusion in a future issue of this journal.

[6] A. Zlocki and P. Themann, “Improved energy efficiency by model based predictive ACC in hybrid vehicles based on map data,” inProc. AVEC, Loughborough, U.K., 2010.

[7] T. van Keulen, G. Naus, A. de Jager, M. J. G. van de Molengraft, and M. Steinbuch, “Predictive cruise control in hybrid electric vehicles,” in Proc. EVS24, Stavanger, Norway, 2009, pp. 1–11.

[8] K. Li, T. Chen, Y. Luo, and J. Wang, “Intelligent environment-friendly vehicles: Concept and case studies,”IEEE Trans. Intell. Transp. Syst., vol. 13, no. 1, pp. 318–328, Mar. 2012.

[9] Siang Fui Tie, Chee Wei Tan, « A Review of Power and Energy Management Strategies in Electric Vehicles », 2012 4th International Conference on Intelligent and Advanced Systems (ICIAS2012), 2012, pp. 412-417.

[10] Hongjun Chen, Fei Lu, Fujuan Guo, “Power Management System Design for Small Size Solar-Electric Vehicle”, 2012 IEEE 7th International Power Electronics and Motion Control Conference - ECCE Asia, 2012, pp. 2658-2662.

[11] B. Ganji and A. Z. Kouzani, "A study on look-ahead control and energy management strategies in hybrid electric vehicles," 2010 8th IEEE International Conference on Control and Automation (ICCA), 2010, pp. 388-392.

[12] A. A. Malikopoulos and J. P. Aguilar, “Optimization of driving styles for fuel economy improvement,” in Proc. 15th Int. IEEE Conf. Intell. Transp. Syst., 2012, pp. 194–199.

[13] S.-I. Jeon, S.-t. Jo, Y.-I. Park, and J.-M. Lee, “Multi-mode driving control of a parallel hybrid electric vehicle using driving pattern recognition,”J. Dyn. Syst., Meas., Control, vol. 124, no. 1, pp. 141–149, Aug. 2002.

[14] Y. Lin, P. Tang, W. J. Zhang, and Q. Yu, “Artificial neural network modeling of driver handling behavior in a driver-vehicle-environment system,” Int. J. Veh. Des., vol. 37, no. 1, pp. 24–25, 2005.

[15] S. Ichikawa, Y. Yokoi, S. Doki, S. Okuma, T. Naitou, T. Shiimado, and N. Miki, “Novel energy management system for hybrid electric vehicles utilizing car navigation over a commuting route,” inProc. IEEE Intell, 2004, pp. 161–166.

[16] J.-S. Chen and M. Salman, “Learning energy management strategy for hybrid electric vehicles,” inProc. IEEE Conf. Veh. Power Propulsion, 2005, pp. 68–73.

[17] A. Boyali and L. Guvenc, “Real-time controller design for a parallel hybrid electric vehicle using neuro-dynamic programming method,” in Proc. IEEE Int. Conf. SMC, 2010, pp. 4318–4324.

[18] C. Dextreit and I. Kolmanovsky, “Approaches to energy management of hybrid electric vehicles: Experimental comparison,” inProc. UKACC Int. Conf. Control, 2010, pp. 1–6.

[19] X. Huang, Y. Tan, and X. He, “An intelligent multifeature statistical approach for the discrimination of driving conditions of a hybrid electric vehicle,”IEEE Trans. Intell. Transp. Syst., vol. 12, no. 2, pp. 453–465, Jun. 2011.

[20] D. Ambuhl and L. Guzzella, “Predictive reference signal generator for hybrid electric vehicles,”IEEE Trans. Veh. Technol., vol. 58, no. 9, pp. 4730–4740, Nov. 2009.

[21] J. Chunyu, Z. Qu, E. Pollak, and M. Falash, “A new multi-objective control design for autonomous vehicles,” Optimization Cooperative Control Strategies, vol. 381, Lecture Notes in Control and Information Sciences, pp. 81–102, 2009.

[22] M. Nakaoka, P. Raksincharoensak, and M. Nagai, “Study on forward collision warning system adapted to driver characteristics and road environment,” inProc. ICCAS, Oct. 14–17, 2008, pp. 2890–2895.

[23] S. Mammar, S. Glaser, and M. Netto, “Time to line crossing for lane departure avoidance: A theoretical study and an experimental setting,” IEEE Trans. Intell. Transp. Syst., vol. 7, no. 2, pp. 226–241, Jun. 2006.

[24] C. Zhang, A. Vahidi, P. Pisu, X. Li, and K. Tennant, “Role of terrain preview in energy management of hybrid electric vehicles,”IEEE Trans. Veh. Technol., vol. 59, no. 3, pp. 1139–1147, Mar. 2010.

[25] K. Li, Y. Wang, F. Gao, L. Zhang, and L. Guo, “Vehicle driving safety assistant systems based on ITS technologies,” inProc. Automobile Technol., 2006, vol. S1, pp. 32–35.

[26] C. Stiller, F. P. León, and M. Kruse, “Information fusion for automotive applications—An overview,”Inf. Fusion, vol. 12, no. 4, pp. 244–252.

[27] K. Ulrich, “Computational and pre-parametric design,” Ph.D. dissertation, Artif. Intell. Lab., Mass. Inst. Technol., Cambridge, MA, 1988.

[28] R. Mahler, Statistical Multisource-Multitarget Information Fusion. Norwood, MA: Artech House, 2007. Oct. 2011.

[29] A. Chakrabarti, “A new approach to structure sharing,” J. Comput. Inf. Sci. Eng., vol. 4, no. 1, pp. 11–19, Mar. 2004.

[30] T. Chen, Y. Luo, and K. Li, “Multi-objective adaptive cruise control based on nonlinear model predictive algorithm,” inProc. IEEE Int. Conf. Veh.Electron. Saf., Beijing, China, 2011, pp. 274–279.

[31] A. A. Malikopoulos, Real-Time, Self-Learning Identification and Stochastic Optimal Control of Advanced Powertrain Systems. Ann Arbor,MI, USA: ProQuest, Sep. 2011.

ACRONYMS AND ABBREVIATIONS NOMENCLATURE

ACC Adaptive Cruise Control
ANNs Artificial Neuronal Networks
CMBS Collision Mitigation Braking System
DDS Drowsiness Detection System
EV Electric Vehicle
GPS Global Positioning System
ITSs Intelligent Transportation Systems
LKAS Lane Keeping Assist System
PM Power Management
PMC Power Management controller

ACKNOWLEDGMENT

We gratefully acknowledge support from the following companies:

ATB: Arab Tunisian Bank, a Tunisian commercial bank, created on June 30, 1982. Its mission was contribution to the economic and financial development of the country.

SOMEF TUNISIE: Founded in May 1988, SOMEF TUNISIE is a leading Tunisian company manufacturing electrical equipment for both domestic and industrial use.

Think Electrical is a Tunisian engineering company providing electrical engineering solutions for industry, health and education.

Dual band CPW-Fed Antenna Based on Metamaterial

Mohamed Lashab[1], Chemss-Eddine[2] ,Fatiha Benabdelaziz[3]
[1]Skikda university, Algeria.[3]Constantine university, Algeria
[2]Setif university, Algeria.
lashabmoh@yahoo.fr
zebiri@ymail.com

Abstract— **In this paper a dual band antenna realized by CPW-Fed antenna (coplanar waveguide), the antenna contains three slots as monopole bars, the metamaterial is loaded with split ring resonator (SRR) inserted between the slots. The aim of this work is to exhibit the improvement of the gain and the bandwidth by metamaterial insertion. The obtained results from HFSS simulation concerning the constitutive parameters of the (SRR), show that there is a DNG (Double Negative) permeability and permittivity in the frequency of interest. In this work the operating bandwidth of the proposed antenna (dual band) is in the range of 0.41GHz to 0.92 GHz as DVB-T band, and 1.45 GHz to 2.45 GHz as Wi-Fi application. Here in this paper the inserted SRR has a remarkable effect in the second band.**

***Keywords*- Metamaterial; Coplanar Waveguide; SRR; HFSS; Broadband**

I. INTRODUCTION

Coplanar waveguide antennas have been widely used in radio frequency application for instance small resonators are one of these applications [1]. Metamaterial are known as artificial materials having negative permittivity, negative permeability or both negative (DNG), for such values of the constitutive parameters the material offers excellent properties of the antenna especially for coplanar waveguide [2]. Coplanar waveguide are generally used for planar antennas, the properties required are the wideband and high gain [3]. Here in this paper CPW is loaded with (SRR) metamaterial operating in the DVB-T frequency Band for application in terrestrial digital television broad casting.

Coplanar waveguide loaded with metamaterial can find application in beam steering with low cost [4], many other research work on CPW with metamaterial are seeking the enhancement of the bandwidth [5,6], also some research work were dealing with the gain improvement [7].

The proposed antenna has a geometrical dimension of 150mmx70mm, the substrate is an FR4, and the metamaterial unit cell has a dimension of 34mm x 34mm. Here in this paper we consider the case of 4 SRR as to show more explicitly the effect of metamaterial placement. The antenna is simulated with and without metamaterial in order to exhibit the antennas characteristics.

This paper deals firstly with the design of unit cell made of SRR as metamaterial which is operating in the range of 1.4 GHz to 3 GHz, the unit cell is designed by HFSS and the obtained S-parameters are used for the extraction of the constitutive parameters of the unit cell. The SRR cells are inserted at the bottom of the CPW antenna, the obtained reflection coefficient by simulation show that the resonant frequency of the CPW antenna is shifted towards the unit cell as the metamaterial part, an enlargement of the bandwidth is observed, and also the gain of the antenna is improved slightly.

II. PALANAR METAMATERIALS

A. Theoretical aspect

The extraction of the constitutive parameters of the SRR considered as metamaterial usually needs experimental tests or analytical models [8]. Drude-Lorentz model [9], known as dispersion model is very accurate, in which the magnetic permeability and electric permittivity are extracted analytically using mathematical model [10]. Here in this paper the constitutive parameters are obtained using the well-developed characterization method of metamaterials known as the standard retrieval procedure [11], where the assigned effective refractive index (n) and relative impedance (z) values of the metamaterial or (SRR) can be extracted from the S-parameters assuming that the unit cell test is symmetric with respect to the (x-y) plane, which means S11 = S22 and S21 = S12.

B. Unit Cell

The unit cell is realized by utilizing the well-developed circuit board (SRR) technology. Figure 1, shows the geometrical dimension of the SRR metamaterial. The SRR dimensions are as follow: the first ring is 14mm radius, the second one is 11mm radius, the third one is 8 mm radius, the fourth one is 5 mm radius, the fifth on is 3 mm radius.

ISBN: 978-1-61804-290-3

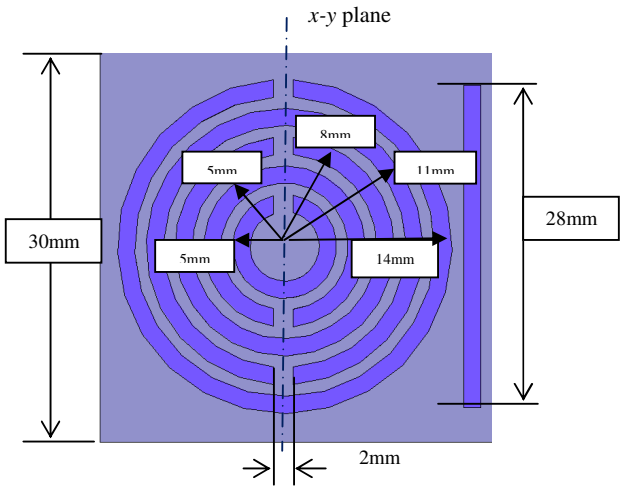


Fig. 1, SRR metamaterial cell,

The gap of the squares is 2 mm and the thickness of all the rings is 3 mm, the spacing between rings is 2mm. The substrate is an FR4 with relative permittivity ε_r = 4.4 and dielectric loss tangent δ = 0.02, the thickness of the substrate is 1.6 mm.

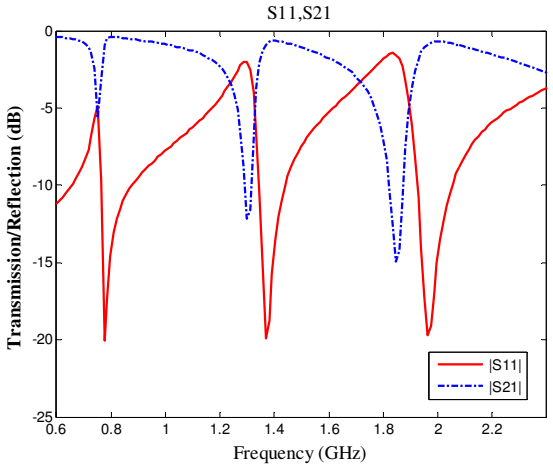
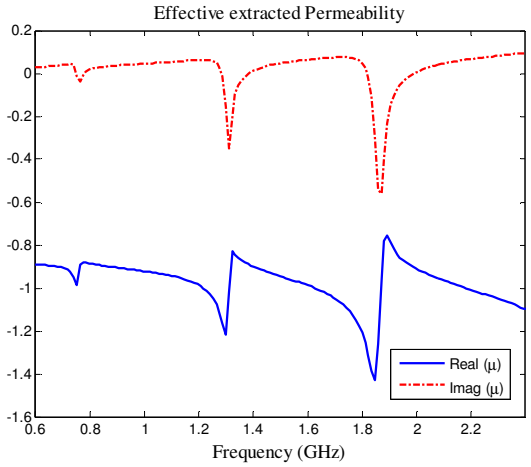
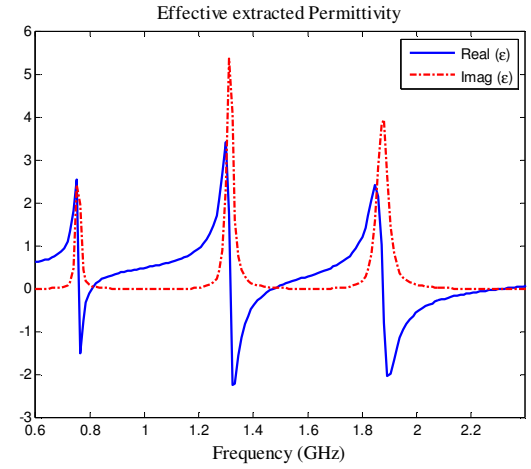


Fig. 2, Transmission and Reflexion coefficients

The transmission and reflection coefficients of the SRR metamaterial are presented by figure 2, two resonant frequencies are observed, the first one is on 1.84 GHz and the second one on 2.5 GHz. The simulated unit cell on HFSS has shown very interesting results, figure 3.a and figure3.b present the imaginary and real parts of respectively the permeability and the permittivity, both the permittivity and the permeability have negative real parts (DNG) from 1.96 GHz up to 2.2 GHz.



(a)



(b)

Fig. 3, (a) Real and Imaginary parts of the unit cell permeability. (b) Real and Imaginary parts of the unit cell permittivity.

III. COPLANAR WAVEGUIDE

The CPW antenna used in this work is constituted by two rectangular rings in which it is cut a small ring 28mmx28mm as given in figure 4.a, the SRR metamaterial cells are placed at the top of the antenna as shown by figure 4.b.

The total gain shown in figure 5, is recorded at the frequency 1.65 GHz as the second resonant frequency of the CPW antenna, figure 5 presents the case of the antenna without SRR, the radiation pattern for xz-plane and yz-plane has a maximum gain of 5.25 dBi. Figure 6 presents the radiation pattern of the antenna loaded with 4 SRR at the frequency 1.65 GHz, the xz-plane and yz-plane radiation pattern has almost the same shape as without SRR, but the maximum gain has changed to 6.85 dBi.

z

ISBN: 978-1-61804-290-3

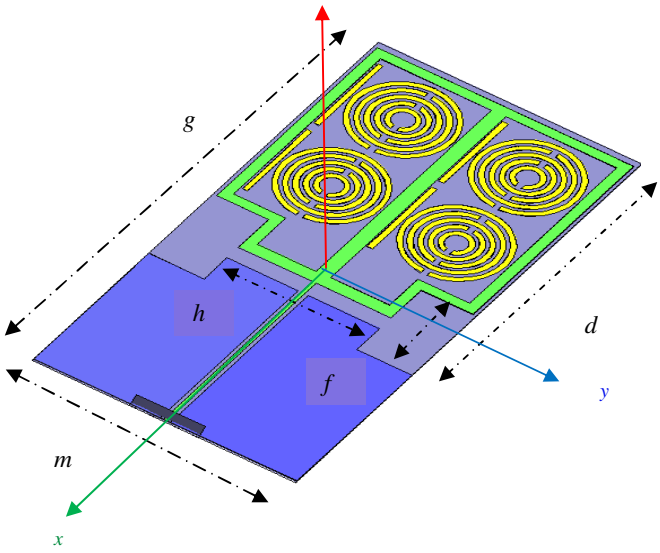


Fig. 4, a) Top view of the CPW-fed antenna
b) Bottom view of the CPW-fed antenna
m=76mm, g=150mm, f=28mm, d=100mm, h= 44mm

From figure 5 and figure 6, it is clear that there is an increase of the total gain from 5.25 dBi to 6.85 dBi, considered as increase of 20 %.

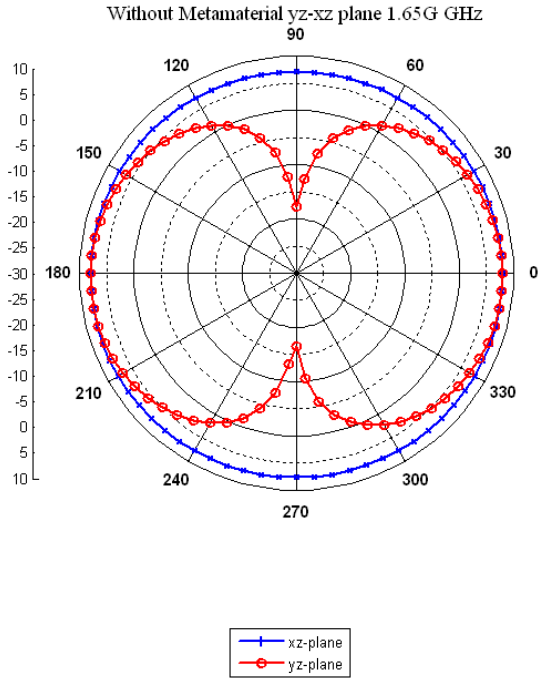


Fig. 5, Total gain radiation pattern without SRR, F= 1.65 GHz

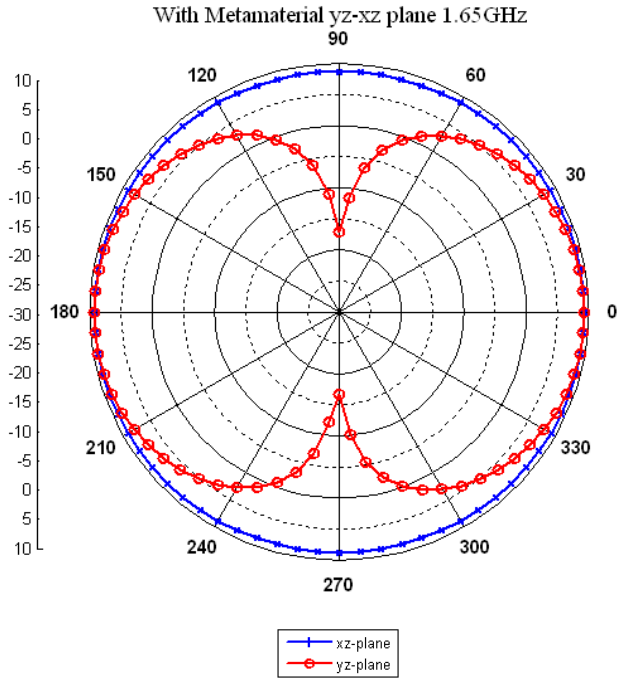


Fig. 6, Total gain radiation pattern with 4 SRR, F= 1.65 GHz

Figure 7 shows the reflection coefficient with and without SRR, we can see that the first and second bandwidth is shifted up by almost 7 %, when using the SRR as metamaterial compared to the case without SRR.

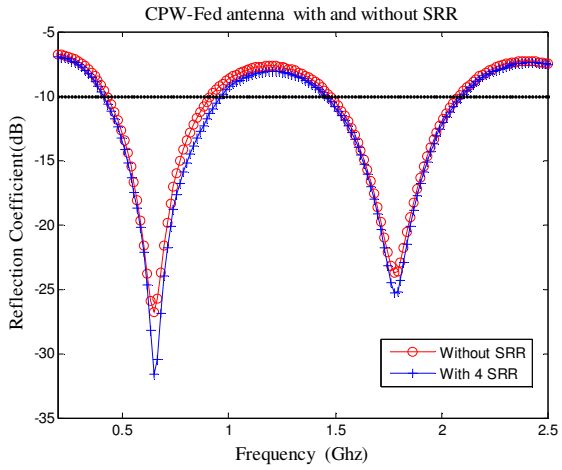


Fig. 7, Reflection coefficient with and without metamaterial.

Figure 8 shows the effect of the SRR metameterial cells on the total gain radiation pattern, the figure shows also that the SRR cells have an effect which starts from 0.45GHz to 0.8 GHz and from 1.4 GHz to 2.5 GHz.

ISBN: 978-1-61804-290-3

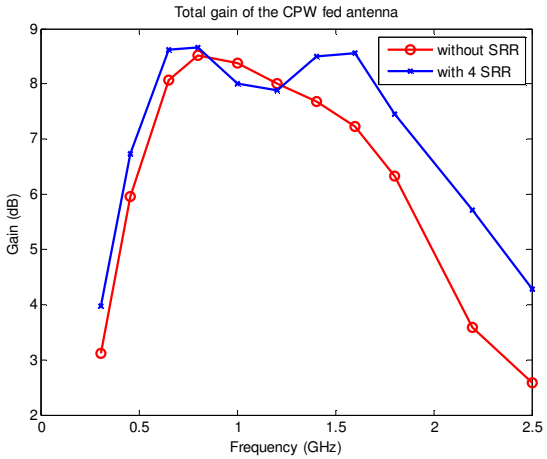


Fig. 8, Effect of the SRR on the total gain

IV. CONCLUSION

In this paper, a dual band antenna as CPW antenna loaded with SRR Metamaterial is designed for DVB-T and Wifi bands applications. The antenna is designed by HFSS and the constitutive parameters are retrieved by the standard procedure. The obtained results show that both the permeability and the permittivity have negative real parts; this means that the SRR cells can produce resonant frequency in any medium placed in. The results obtained by simulation, show that, when loading the coplanar antennas with metamaterial, the resonant frequency is shifted of 7% towards lower frequencies and the bandwidth is almost the same, whereas the gain is increased by 20 %.

V. REFERENCES

[1] Miguel Durán-Sindreu, Jordi Naqui, "Electricall Small Resonators for planar metamaterial, Microwave circuit and Antenna Design: A comparative analysis," *Applied Science, Vol. 2, pp.375-395, 2012.*

[2] L.-M. Si and X. Lvon, " CPW-Fed Multi-Band Omni-Directional Planar Microstrip Antenna Using Composite Metamaterial resonator for wireless communications," *Progress In Electromagnetic Research PIER 83, 133-146, 2008* .

[3] A. L. Borja, A. Belenguer, J. Cascón, H. Esteban, and V. E. Boria, "Wideband Passband Transmission Line Based on Metamaterial-Inspired CPW Balanced Cells," *IEEE antennasand wirelessPropagation Letters, Vol. 10,*pp. 1421-1427, *2011.*

[4] Yue Li, Magdy F. Iskander, Zhijun Zhang, "A New Low Cost Leaky Wave Coplanar Waveguide Continuous Transverse Stub Antenna Array Using Metamaterial-Based Phase Shifters for Beam Steering," *IEEE Transaction on antennas and Propagation, Vol. 61, No. 7, July 2013* .

[5] Li-Ming Si, Weiren Zhu, and Hou-Jun Sun, "A Compact, Planar, and CPW-Fed Metamaterial-Inspired Dual-Band Antenna," *IEEE Antennas and Wireless Propagation Letters, Vol. 12, 2013.*

[6] Bing-Jian Niu and Quan-Yuan Feng, "Bandwidth Enhancement of CPW-Fed Antenna Based on Epsilon Negative Zerothand First-Order Resonators," *IEEE Antennas and Wireless Propagation Letters, Vol. 12, 2013.*

[7] Gonghan Wang and Quanyuan Feng, "A Novel Coplanar Waveguide Feed Zeroth-Order Resonant Antenna With Resonant Ring", *IEEE Antennas and Wireless Propagation Letters, Vol. 13, 2014.*

[8] Lubkowski, G., R. Schuhmann, and T. Weiland, "Extraction of effective metamaterial parameters by parameter fitting of dispersive models", *Microw. Opt. Technol. Lett.*, Vol. 49, No. 2, 285-288, 2007.

[9] Kamil Boratay Alici and Ekmel Ozbay, "Theoretical Study and Experimental Realization of a Low-Loss Metamaterial Operating at the Millimeter-Wave Regime: Demonstrations of Flat- and Prism-Shaped Samples", *IEEE Iournal of selected topics in quantum electronics,* vol. 16, no. 2, march/april 2010.

ISBN: 978-1-61804-290-3

Bounded Control based on Norm Differential Game for Three-Player Conflict

Mao Su, Yongji Wang, and Lei Liu

***Abstract*—**A Three-player conflict with bounded controls is considered in this paper. Optimal pursuit-evasion strategy based on differential game with bounded control is derived assuming that the attacker, target and it's defender have linearized kinematics, arbitrary-order linear adversaries' dynamics, and perfect information. The obtained strategy is dependent on the zero-effort miss distance of two pursuer-evader pairs: attacker with target and defender with attacker. For adversaries with first-order dynamics, this paper presents algebraic conditions for the three-player problem and gives an analytical hybrid strategy for the attacker to capture the target while evading the defender is obtained. The simulation results shows that, by using the hybrid strategy, the attacker will evade form the defender successfully and guarantee the miss distance from the target.

***Keywords*—**Bounded control, Norm differential game, Three-player conflict, Optimal strategies.

I. Introduction

In this paper, a differential game of three players with bounded controls is investigated. In this game, an attacker denoted by *M* pursues a target denoted by *T*, who has a defender denoted by *D* to protect itself from the attacker. Differential game theory is a natural setup to discuss problems such as this one[1]. The most common pursuit and evasion game called "zero-sum differential game" deals with two vehicles with respect to miss distance. In generating guidance laws, a common practice is to linearize with respect to collision course, which implies linearized kinematics. There are two formulations[2]: the first is the "linear quadratic differential game"(LQDG), and the second is the "Norm differential game"(NDG). In the LQDG, the controls are unbounded, the cost function is the weighted sum of three quadratic terms: the square of the miss distance and two penalty terms: the integrals of the respective control energy of the players[3, 4]. The optimal solution of this formulation is linear. In the NDG[5], the controls have hard bounds and the cost is purely terminal to account for imposed on the miss distance. Contrary to the LQDG, the optimal strategies are nonlinear, at a certain time before termination, the guidance law becomes pure bang-bang. The typical engagement is a one-side engagement of missile and aircraft, however in these days, a missile-missile engagement is concerned, particularly an exoatmospheric engagement. In this scenario, the guaranteed miss distance for the interceptor must be very small for hit-to-kill, especially for evasive maneuvers.

Recently, an interest in defending aircraft from an attacking missile is produced. In such a scenario, the target launching a defending missile to protect itself from an attacking missile, it is a three-body pursuit and evasion problem. In[6], a closed-form relation was derived for the initial missile–target range ratio, under the assumption of a constant collision course. In[7] the three-body game was presented as a two-team dynamic game: the lady, the bandit, and the bodyguard. The objective of the bandit is to capture the lady, while the objective of the lady and her bodyguard is to prevent this. Perelman[8] has presented a cooperative target–defender guidance strategy based on a two team LQDG against a pursuing missile provided an optimal analytic solution for the target–defender pair. Shaferman[9] and Shima[10] have presented a multiple-model adaptive guidance strategy to defend the target from the missile. Yamasaki[11], Ratnoo[12] and Shima[13] have also made some noticeable contributions on this problem. However, the obtained guidance laws in these papers are still linear and suffer the same drawbacks mentioned before. Contrary to the research before, Rubinsky and Gutman[14, 15] have focused on the strategies for the attacker and gave algebraic conditions for the attacker to capture the evader while evading the defender.

In this paper, a three-player conflict problem has been investigated based on differential game with bounded controls, an optimal strategy has been derived. Under assumption of first-order dynamics, linear kinematics and perfect information, the optimal strategies for the attacker, target and defender are obtained. Moreover, for the attacker a useful strategy based on differential game has been proposed to perform an evasion maneuver with respect to the defender, without losing its pursuit capabilities and finally an analytical solution for the game has been obtained.

This paper is organized as follows: the engagement formulation of this problem is outlined in Section II; a simple solution for the three-player game is derived and some simulation results are shown in section III; and a conclusion follows in Section IV.

II. Engagement Formulation

In this problem, there are three entities: an evading target denoted as *T*, an attacker for intercepting the target denoted as

This work was supported in part by the National Nature Science Foundation of China (NO. 61203081 and 61174079), Doctoral Fund of Ministry of Education of China (NO. 20120142120091), Fundamental Research Funds for the Central Universities of HUST (NO. 2013054), and Precision Manufacturing Technology and Equipment for Metal Parts (NO. 2012DFG70640)

Mao Su is with the School of Automation, Huazhong University of Science and Technology,Wuhan, China. (e-mail: sumao1988@ hust.edu.cn).

Yongji Wang is with School of Automation, Huazhong University of Science and Technology, Wuhan, China. (wangyjch@mail.hust.edu.cn)

Lei Liu is with School of Automation, Huazhong University of Science and Technology, Wuhan, China. (liulei@mail.hust.edu.cn)

M, and a defender for protecting the target denoted as *D*. The strategy of each player in the game is known by others. The target launches the defender to intercept the incoming attacker. It is obviously that, to protecting the target the engagement between the defender and attacker should be planned to terminate before that between the attacker and the target. On the other hand, the attacker should evade the defender and pursue the target by appropriate strategy. In this section we first present the nonlinear kinematic equations of the problem, then we will present the linearized equations used for the optimal strategies derivation. The engagement is analyzed and simulated in two dimensions.

A. Nonlinear Kinematics

A schematic view of the three-body terminal engagement geometry in the planar is shown in

Fig. 1. The notations *MT* and *MD* denote the attacker with target and attacker with defender respectively. The speed, normal acceleration and flight-path angles are denoted by *V*, *a* and *γ*, respectively. The range between the adversaries is *r*, and *λ* is the angle between the line of sight (LOS) and the reference frame axis.

Also given the hard bounds on the accelerations,

$$|a_M| \le a_M^{\max},\ |a_D| \le a_D^{\max},\ |a_T| \le a_T^{\max} \tag{1}$$

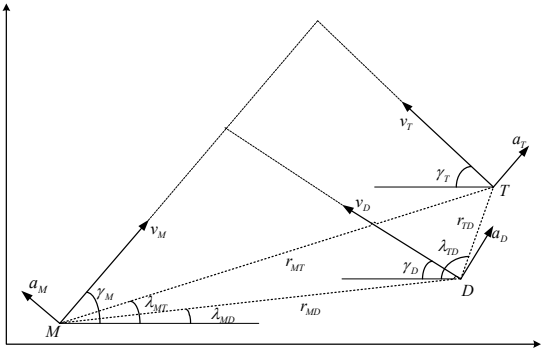


Fig. 1 Target-attacker-defender engagement geometry.

Neglecting the gravitational force during the terminal guidance phase, the engagement kinematics between the attacker and the target can be expressed as:

$$\dot{r}_{MT} = -V_M \cos(\gamma_M - \lambda_{MT}) - V_T \cos(\gamma_T + \lambda_{MT}) \tag{2}$$

$$\dot{\lambda}_{MT} = \left[V_T \sin(\gamma_T + \lambda_{MT}) - V_M \sin(\gamma_M - \lambda_{MT})\right] / r_{MT} \tag{3}$$

Similarly, the engagement kinematics equations between the defender and the attacker are

$$\dot{r}_{MD} = -V_M \cos(\gamma_M - \lambda_{MD}) - V_D \cos(\gamma_D + \lambda_{MD}) \tag{4}$$

$$\dot{\lambda}_{MD} = \left[V_D \sin(\gamma_D + \lambda_{MD}) - V_M \sin(\gamma_M - \lambda_{MD})\right] / r_{MD} \tag{5}$$

During the terminal guidance phase, the speeds of the adversaries are assumed constant.

The path angles of the adversaries are

$$\dot{\gamma}_i = a_i / V_i \quad i = \{M, T, D\} \tag{6}$$

The dynamics of each agent during the endgame can be represented by arbitrary-order linear systems:

$$\begin{cases} \dot{\boldsymbol{x}}_i = A_i \boldsymbol{x}_i + \boldsymbol{B}_i u_i \\ a_i = \boldsymbol{C}_i \boldsymbol{x}_i + d_i u_i \quad i = \{M, T, D\} \\ |u_i| \le \rho_i \end{cases} \tag{7}$$

where $\boldsymbol{x}_i$ is the state vector of an player's internal state variables with $\dim(\boldsymbol{x}_i)=n_i$ and u_i is the controller.

B. Linearized Kinematics

During the terminal guidance phase, we have two collision triangles: one is the scenario between the attacker and target and the other in the scenario between the defender and the attacker. We can assume that the trajectories of the three bodies can be linearized near the collision triangles.

Denote y_{MT} as the relative displacement between the attacker *M* and the target *T*, normal to the initial LOS is denoted as LOS_{MT0} of attacker and the target; similarly, y_{MD} is the relative displacement between *M* and *D*, normal to LOS_{MD0}, which is the initial LOS attached with *M* and *D*.

It is worth noting that in the engagement problem, *a*, the acceleration, is perpendicular to the LOS. The controller u_M, u_T and u_D are also normal to the corresponding LOS.

The state vector of the linearized problem is

$$\boldsymbol{x} = \begin{bmatrix} \boldsymbol{x}_{MT}^T & \boldsymbol{x}_M^T & \boldsymbol{x}_{MD}^T \end{bmatrix}^T \tag{8}$$

where

$$\boldsymbol{x}_{MT} = \begin{bmatrix} y_{MT} & \dot{y}_{MT} & \boldsymbol{x}_T^T \end{bmatrix}^T \tag{9}$$

$$\boldsymbol{x}_{MD} = \begin{bmatrix} y_{MD} & \dot{y}_{MD} & \boldsymbol{x}_D^T \end{bmatrix}^T \tag{10}$$

and $\dim(\boldsymbol{x})=4+n_M+n_T+n_D$.

The state x_1 and x_{n_T+3} are the differences between the target and the attacker positions and between the attacker and the defender positions normal to the initial line of sight; x_2 and x_{n_T+4} are therefore the relative respective lateral speeds, and their derivatives are the relative lateral accelerations of attacker-target and attacker-defender. Thus, the equations of motion that represent the engagement's kinematics can be written as

$$\dot{\boldsymbol{x}} = \begin{cases} \dot{\boldsymbol{x}}_{MT} = \begin{cases} \dot{x}_1 = x_2 \\ \dot{x}_2 = a_T - a_M \\ \dot{\boldsymbol{x}}_T = \boldsymbol{A}_T \boldsymbol{x}_T + \boldsymbol{B}_T u_T \end{cases} \\ \dot{\boldsymbol{x}}_M = \boldsymbol{A}_M \boldsymbol{x}_M + \boldsymbol{B}_M u_M \\ \dot{\boldsymbol{x}}_{MD} = \begin{cases} \dot{x}_{n_T+n_M+3} = x_{n_T+n_M+4} \\ \dot{x}_{n_T+n_M+4} = a_M - a_D \\ \dot{\boldsymbol{x}}_D = \boldsymbol{A}_D \boldsymbol{x}_D + \boldsymbol{B}_D u_D \end{cases} \end{cases} \tag{11}$$

The vector form of these equations can be written as

$$\dot{\boldsymbol{x}} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\begin{bmatrix} u_T & u_D \end{bmatrix}^T + \boldsymbol{C}u_M \tag{12}$$

C. Timeline

The initial range between the attacker and the target is r_{MT0}. Similarly, the initial range between the attacker and the defender is r_{MD0}. Under the linearization assumption of small deviations from a collision triangle, the interception time is fixed, satisfying

$$t_{fMT} = -r_{MT_0}/\dot{r}_{MT_0} = -r_{MT_0}\Big/\Big[V_M\cos\left(\gamma_{M_0}+\lambda_{MT_0}\right) + V_T\cos\left(\gamma_{T_0}-\lambda_{MT_0}\right)\Big] \quad (13)$$

Similarly,

$$t_{fMD} = -r_{MD_0}/\dot{r}_{MD_0} = -r_{MD_0}\Big/\Big[V_M\cos\left(\gamma_{M_0}+\lambda_{MD_0}\right) + V_D\cos\left(\gamma_{D_0}-\lambda_{MD_0}\right)\Big] \quad (14)$$

We define Δt as the time difference between interceptions,

$$\Delta t = t_{fMT} - t_{fMD} \quad (15)$$

and we require that the *MD* engagement terminates before that of *MT*. Thus, $\Delta t>0$, and the defender ceases to exist after $t=t_{fMD}$, we enforce $u_D = 0 \forall t \geq t_{fMD}$.

D. Cost Function

The interception scenario can be considered as a zero-sum differential game for the system(12) with bounded controls(1). We define two miss distance as the norm cost functions

$$J_{MT} = \left\| Z_{MT}\left(t_{fMT}\right)\right\| \quad (16)$$

$$J_{MD} = \left\| Z_{MD}\left(t_{fMD}\right)\right\| \quad (17)$$

The problem involving the three agents is posed as a norm differential game between two teams. One team is composed of the target and its defender to maximize the cost function J_{MD} and minimize J_{MT} , and the attacker belongs to the other team to minimize J_{MT} and maximize J_{MD} . Thus ,we can rewrite the cost functions as

$$\begin{aligned} &\max_{[u_T \quad u_D]} \min_{u_M} J_{MT} \\ &\min_{[u_T \quad u_D]} \max_{u_M} J_{MD} \end{aligned} \quad (18)$$

E. Order Reduction

To simplify the solution and to reduce the problem's order, we will use the following transformation by introducing a new variable:

$$\begin{cases} Z_{MT}(t) = \boldsymbol{D}_{MT}\boldsymbol{\Phi}\left(t_{fMT},t\right)\boldsymbol{x}(t) \\ Z_{MD}(t) = \boldsymbol{D}_{MD}\boldsymbol{\Phi}\left(t_{fMD},t\right)\boldsymbol{x}(t) \end{cases} \quad (19)$$

where $\boldsymbol{\Phi}$ is the transition matrix associated with Eq.(12), and $\boldsymbol{D}_{MT}$, $\boldsymbol{D}_{MD}$ are constant vectors,

$$\begin{cases} \boldsymbol{D}_{MT} = \left[1 \quad [0]_{1\times(n_T+n_D+n_M+3)}\right] \\ \boldsymbol{D}_{MD} = \left[[0]_{1\times(n_T+n_M+2)} \quad 1 \quad [0]_{1\times(n_D+1)}\right] \end{cases} \quad (20)$$

and $Z_{MT}(t)$ and $Z_{MD}(t)$ define the zero-effort miss(ZEM) vector:

$$\boldsymbol{Z}(t) = \left[Z_{MT}(t) \quad Z_{MD}(t)\right]^T \quad (21)$$

The derivative with respect to time of Z_{MT} is

$$\begin{aligned} \dot{Z}_{MT}(t) &= \boldsymbol{D}_{MT}\left[\dot{\boldsymbol{\Phi}}\left(t_{fMT},t\right)\boldsymbol{x} + \boldsymbol{\Phi}\left(t_{fMT},t\right)\dot{\boldsymbol{x}}\right] \\ &= \boldsymbol{D}_{MT}\boldsymbol{\Phi}\left(t_{fMT},t\right)\left(\boldsymbol{B}\left[u_T \quad u_D\right]^T + \boldsymbol{C}u_M\right) \end{aligned} \quad (22)$$

Similarly,

$$\begin{aligned} \dot{Z}_{MD}(t) &= \boldsymbol{D}_{MD}\left[\dot{\boldsymbol{\Phi}}\left(t_{fMD},t\right)\boldsymbol{x} + \boldsymbol{\Phi}\left(t_{fMD},t\right)\dot{\boldsymbol{x}}\right] \\ &= \boldsymbol{D}_{MD}\boldsymbol{\Phi}\left(t_{fMD},t\right)\left(\boldsymbol{B}\left[u_T \quad u_D\right]^T + \boldsymbol{C}u_M\right) \end{aligned} \quad (23)$$

Note that in Eq.(22) and Eq.(23), it is reasonable to find the explicit form of the ZEM variables. Consider the transition matrix from Eq.(19):

$$\dot{\boldsymbol{\Phi}}\left(t_f,t\right) = -\boldsymbol{\Phi}\left(t_f,t\right)\boldsymbol{A},\ \boldsymbol{\Phi}\left(t_f,t_f\right) = \boldsymbol{I} \quad (24)$$

Change the running time t to the time to go t_{go}

$$\begin{aligned} &t \to t_f - t = t_{go},\ \ dt \to -dt_{go} \\ &\dot{\boldsymbol{\Phi}}\left(t_{go}\right) = \boldsymbol{\Phi}\left(t_{go}\right)\boldsymbol{A},\ \ \boldsymbol{\Phi}(0) = \boldsymbol{I} \end{aligned} \quad (25)$$

Thus, we can obtain an equivalent reduced-order problem of finding the optimal pursue and evasion strategy for the target, it's defender and the attacker respectively. The cost function is Eq.(18) and the problem is subject to the scalar dynamics of Eq.(22)-Eq.(23) and the constraints of Eq.(1).

F. Simple Solution for Three-player Game

The defined engagement problem could be divided into two phases: the first phase is before the termination of the engagement between the defender and the attacker ($t < t_{fMD}$), the second phase is from that time onward ($t_{fMD} < t < t_{fMT}$).

G. General solution of the differential game

Differentiate the ZEM norm $\left|Z_{MT}(t)\right|$ and $\left|Z_{MD}(t)\right|$ and obtain

$$\frac{d}{dt}\left|Z_{MT}(t)\right| = \operatorname{sign}\left(Z_{MT}\right)\left(P_{MT}u_T + Q_{MT}u_D + R_{MT}u_M\right) \quad (26)$$

$$\frac{d}{dt}\left|Z_{MD}(t)\right| = \operatorname{sign}\left(Z_{MD}\right)\left(P_{MD}u_T + Q_{MD}u_D + R_{MD}u_M\right) \quad (27)$$

where

$$\begin{aligned} P_{MT} &= \phi_{12}d_T + \boldsymbol{\phi}_{1T}\boldsymbol{B}_T \\ Q_{MT} &= -\phi_{16}d_D + \boldsymbol{\phi}_{1D}\boldsymbol{B}_D \\ R_{MT} &= -\phi_{12}d_M + \phi_{16}d_D + \boldsymbol{\phi}_{1M}\boldsymbol{B}_M \\ P_{MD} &= \phi_{52}d_T + \boldsymbol{\phi}_{5T}\boldsymbol{B}_T \\ Q_{MD} &= -\phi_{56}d_D + \boldsymbol{\phi}_{5D}\boldsymbol{B}_D \\ R_{MD} &= -\phi_{52}d_M + \phi_{56}d_M + \boldsymbol{\phi}_{5M}\boldsymbol{B}_M \end{aligned} \quad (28)$$

The target has interest to maximize the variable

$$\frac{d}{dt}\left|Z_{MT}(t)\right|$$

with its controller $u_T(t)$. Therefore, the optimal strategy for $u_T(t)$ would be

$$u_T^*(t) = \rho_T \operatorname{sign}\left(Z_{MT}\right)\operatorname{sign}\left(P_{MT}\right) \quad (29)$$

The defender, on the other hand, will try to minimize the variable

$$\frac{d}{dt}\left|Z_{MD}(t)\right|$$

with its controller $u_D(t)$. Thus. The optimal strategy for it would be

$$u_D^*(t) = -\rho_D \operatorname{sign}\left(Z_{MD}\right)\operatorname{sign}\left(Q_{MD}\right) \quad (30)$$

There are two extreme situations to be assumed to derive the bounds for the attacker.

1) The attacker will evade the defender and ignore the target. In such a case, the strategy of the attacker would be

$$u_{Me}^{*} = \rho_M \operatorname{sign}\left(R_{MD}\right)\operatorname{sign}\left(Z_{MD}\right) \tag{31}$$

2) The attacker will pursue the target and ignore

$$u_{Mp}^{*} = -\rho_M \operatorname{sign}\left(R_{MT}\right)\operatorname{sign}\left(Z_{MT}\right) \tag{32}$$

III. Simple Solution for Three-Player Game

The defined engagement problem could be divided into two phases: the first phase is before the termination of the engagement between the defender and the attacker ($t < t_{fMD}$), the second phase is from that time onward ($t_{fMD} < t < t_{fMT}$).

A. Example of first-order dynamics

When all three entities have first-order acceleration dynamics, the switch functions of the attacker, target and defender can be written as

$$\begin{aligned} P_{MD}\left(t_{go}^{MD}\right) &= 0 \\ Q_{MD}\left(t_{go}^{MD}\right) &= -\tau_D \psi\left(t_{go}^{MD}/\tau_D\right) \\ R_{MD}\left(t_{go}^{MD}\right) &= \tau_M \psi\left(t_{go}^{MD}/\tau_M\right) \end{aligned} \tag{33}$$

$$\begin{aligned} P_{MT}\left(t_{go}^{MT}\right) &= \tau_T \psi\left(t_{go}^{MT}/\tau_T\right) \\ Q_{MT}\left(t_{go}^{MT}\right) &= 0 \\ R_{MT}\left(t_{go}^{MT}\right) &= -\tau_M \psi\left(t_{go}^{MT}/\tau_M\right) \end{aligned} \tag{34}$$

where $\psi(\zeta) = \exp(-\zeta) + \zeta - 1$, $\tau_i \ \{i = T, D, M\}$ is the time constant of each player in the game.

Using the optimal strategies, the optimal ZEM dynamics can be computed in reverse time (time-to-go), satisfying

$$\begin{aligned} \frac{dZ^{*}}{dt_{go}} &= \frac{dZ^{*}}{dt}\frac{dt}{dt_{go}} = \Gamma \operatorname{sign}\left\{Z\left(t_{go}=0\right)\right\} \\ Z\left(t_{go}=0\right) &\neq 0 \end{aligned} \tag{35}$$

B. Game space decomposition

Consider the normalization miss distance (NZEM) z_{pe}^{*} in two-player game problem, the pursuer will use the optimal pursuit law and the evader will use the optimal evasion law,

For the pursuer and the evader, we assume that, the maneuverability ratio and dynamics ration, denoted as μ and ε, respectively:

$$\begin{aligned} \mu &= \rho_p / \rho_e \\ \varepsilon &= \tau_p / \tau_e \end{aligned} \tag{36}$$

Define z as the normalization zero miss distance and ξ as the normalization time-to-go.

$$z = Z/\left(\tau_e^2 \rho_e\right) \tag{37}$$

$$\xi_{go} = t_{go}/\tau_e \tag{38}$$

Therefore,

$$\frac{d\left|z_{pe}^{*}\right|}{d\xi_{go}} = \mu\varepsilon\psi\left(\xi_{go}/\varepsilon\right) - \psi\left(\xi_{go}\right) \tag{39}$$

For the maneuverability ratio and dynamics ration, there are several cases to be discussed.

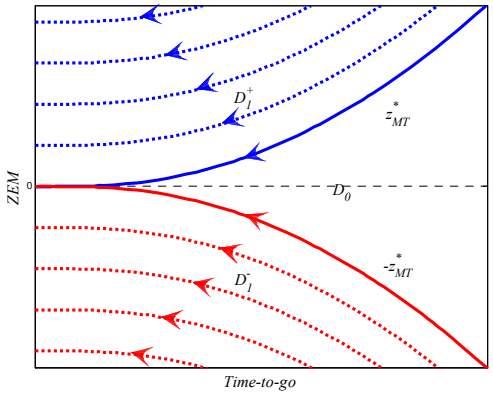


Fig. 2 Game space decomposition for pursuer-evader (case 1)

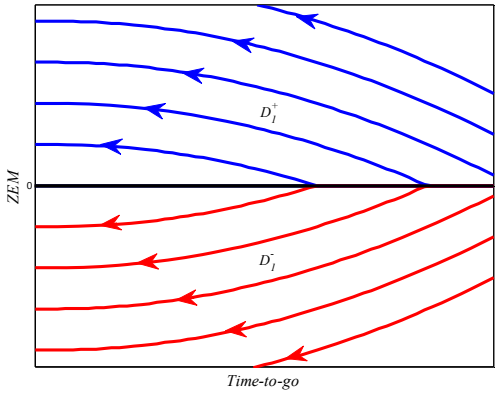


Fig. 3 Game space decomposition for pursuer-evader (case 2)

1) $\mu > 1, \varepsilon < 1$

In this case, it means that the acceleration of the pursuer is bigger than the evader and the time constant is smaller. It is obviously that, the game space decomposition is shown in Fig.2. z_{pe}^{*} and $-z_{pe}^{*}$ are the normal border trajectories and the region between the border trajectories is the singular region, in this region, any strategy used by the pursuer and the evader is optimal, as eventually these border trajectories will be reached and maintained. Thus, all initial conditions in this region will lead to a zero miss distance in the attacker-target engagement. Outside the singular region the pursuer and the evader must apply an optimal strategy.

In the region defined by $\left|z_{pe}\right| > z_{pe}^{*}$ we obtain, if the attacker and the target play optimal, the NZEM $\left|z_{pe}\right|$ will go parallel to $\left|z_{pe}^{*}\right|$ .In the region between the two border trajectory z_{pe}^{*} and $-z_{pe}^{*}$, the evader and the pursuer can use arbitrary strategies to the border trajectories z_{pe}^{*} or $-z_{pe}^{*}$. We denote this region as being singular.

ISBN: 978-1-61804-290-3

2) $\mu<1,\varepsilon<1$

In this case, it means that the acceleration of the pursuer is smaller than the evader but the time constant is smaller. The game space decomposition for pursuer-evader is shown in Fig.3. We can easily find that in this case, the miss distance increase by the time-to-go.

3) $\mu<1,\varepsilon>1$

In this case, the time constant of the defender is bigger than the attacker, thus, like the case 2, the miss distance will increase by the time-to-go, and the game space decomposition is the same as shown in Fig.3.

C. Normalization zero miss distance

Consider the two extreme situations mentioned in the Section II, the attacker has two optimal strategies in different situation during the first phase.

1) The attacker will use the optimal evasion strategy, therefore,

$$\begin{aligned}\frac{d\left|Z_{MD}^{*}\right|}{dt_{go}^{MD}}&=\Gamma_{MD}^{*}=\rho_D\left|Q_{MD}\right|-\rho_M\left|R_{MD}\right|\\&=\rho_D\tau_D\psi\left(t_{go}^{MD}/\tau_D\right)-\rho_M\tau_M\psi\left(t_{go}^{MD}/\tau_M\right)\end{aligned}\tag{40}$$

2) The attacker will use the optimal pursuit strategy, therefore,

$$\begin{aligned}\frac{d\left|Z_{MD}^{**}\right|}{dt_{go}^{MD}}&=\Gamma_{MD}^{**}=\rho_D\left|Q_{MD}\right|+\rho_M\left|R_{MT}\right|\operatorname{sign}\left(Z_{MT}\right)\operatorname{sign}\left(Z_{MD}\right)\\&=\rho_D\tau_D\psi\left(t_{go}^{MD}/\tau_D\right)+\rho_M\tau_M\psi\left(t_{go}^{MD}/\tau_M\right)\operatorname{sign}\left(Z_{MT}\right)\operatorname{sign}\left(Z_{MD}\right)\end{aligned}\tag{41}$$

Similarly, in the second phase, we have

$$\begin{aligned}\frac{d\left|Z_{MT}^{*}\right|}{dt_{go}^{MT}}&=\Gamma_{MT}^{*}=-\rho_T\left|Q_{MT}\right|+\rho_M\left|R_{MT}\right|\\&=-\rho_T\tau_T\psi\left(t_{go}^{MT}/\tau_T\right)+\rho_M\tau_M\psi\left(t_{go}^{MT}/\tau_M\right)\end{aligned}\tag{42}$$

Integrating Eq.(35) backward from any given end condition $Z\left(t_{go}=0\right)$ generates a candidate optimal trajectory. Define the optimal border trajectory:

$$Z^{*}\left(t_{go}\right)=\int_0^{t_{go}}\Gamma^{*}d\zeta\tag{43}$$

From Eq.(43), we can easily find that the family of the optimal trajectory determines the structure of the game solution.

Under the assumption that all these three players in the endgame have similar maneuverability, we can obtain a conclusion without proving that, Γ_{MD}^{*} , Γ_{MD}^{**} and Γ_{MT}^{*} are monotonic, thus, there are three cases discussed in this paper.

For the attacker and the target, denote the maneuverability ratio μ_T and dynamics ration ε_T, respectively:

$$\begin{aligned}\mu_T&=\rho_T/\rho_M\\\varepsilon_T&=\tau_T/\tau_M\end{aligned}\tag{44}$$

and for the defender

$$\begin{aligned}\mu_D&=\rho_D/\rho_M\\\varepsilon_D&=\tau_D/\tau_M\end{aligned}\tag{45}$$

As mentioned in game space decomposition for two-player game, different maneuver ratio and dynamics ratio lead to different game space decomposition. In this three-player game, we assume that

$$\begin{aligned}\mu_T&<1\\\mu_D&<1\end{aligned}\tag{46}$$

otherwise the game will terminate in the first phase and the problem will be a two-player game.

Substitute and obtain

$$\frac{d\left|z_{MD}^{*}\right|}{d\xi_{go}^{MD}}=\mu_D\varepsilon_D\psi\left(\xi_{go}^{MD}/\varepsilon_D\right)-\psi\left(\xi_{go}^{MD}\right)\tag{47}$$

$$\frac{d\left|z_{MD}^{**}\right|}{d\xi_{go}^{MD}}=\mu_D\varepsilon_D\psi\left(\xi_{go}^{MD}/\varepsilon_D\right)+\psi\left(\xi_{go}^{MD}\right)\operatorname{sign}\left(z_{MT}\right)\operatorname{sign}\left(z_{MD}\right)\tag{48}$$

$$\frac{d\left|z_{MT}^{*}\right|}{d\xi_{go}^{MT}}=-\mu_T\varepsilon_T\psi\left(\xi_{go}^{MT}/\varepsilon_T\right)+\psi\left(\xi_{go}^{MT}\right)\tag{49}$$

From Eqs.(48) it can be found that during the first phase, $\left|z_{MD}^{**}\right|$ will depend on the sign of the z_{MT} and z_{MD} ,there are two separate situations are to be discussed.

1) In the case of opposite rotation, the miss distance z_{MT} and z_{MD} have the opposite signs:

$$\operatorname{sign}\left(z_{MT}\right)=-\operatorname{sign}\left(z_{MD}\right)\tag{50}$$

From this, it is readily seen that

$$u_{Me}^{*}=u_{Mp}^{*}\tag{51}$$

Eqs.(48) can be rewrite as

$$\frac{d\left|z_{MD}^{**}\right|}{d\xi_{go}^{MD}}=\mu_D\varepsilon_D\psi\left(\xi_{go}^{MD}/\varepsilon_D\right)-\psi\left(\xi_{go}^{MD}\right)\tag{52}$$

Thus,

$$\left|z_{MD}^{**}\right|=\left|z_{MD}^{*}\right|\tag{53}$$

In this case, the optimal evasion law is the same as the pursuit law. It is readily seen that in both phase of the endgame, the attacker use the only one optimal strategy to evade from the defender while pursue the target. It is the simplest case because the obtained law holds for every initial condition. If the attacker has an ideal condition, when using the optimal pursuit strategy it can not only pursue the target, but also evade from the defender successfully. Case 1 is a product of initial conditions and the others' strategy, so that the attacker can't enforce it.

2) In the case of the same rotation, both line of sight rotate in the same direction, the miss distance z_{MT} and z_{MD} have the opposite signs:

$$\operatorname{sign}\left(z_{MT}\right)=\operatorname{sign}\left(z_{MD}\right)\tag{54}$$

Therefore,

$$u_{Me}^{*}=-u_{Mp}^{*}\tag{55}$$

Eqs.(48) can be rewrite as

$$\frac{d\left|z_{MD}^{**}\right|}{d\xi_{go}^{MD}}=\mu_D\varepsilon_D\psi\left(\xi_{go}^{MD}/\varepsilon_D\right)+\psi\left(\xi_{go}^{MD}\right)\tag{56}$$

ISBN: 978-1-61804-290-3

Thus, there are three trajectory bounds denoted as $\left|z_{MD}^{*}\right|$, $\left|z_{MD}^{**}\right|$ and $\left|z_{MT}^{*}\right|$ in this three-player game.

D. Simulation for the three-player game

Given the player maneuver capabilities(ρ_M, ρ_T, ρ_D), the normalization time to go(ξ_{MD}, ξ_{MT}), the desired miss distance $\left|z_{MT}^{*}\left(\xi_{go}^{MT}=0\right)\right|$ and $\left|z_{MD}^{*}\left(\xi_{go}^{MD}=0\right)\right|$, and integrate Eqs.(47), Eqs.(49) and Eqs.(56) respectively, yields, the trajectory bounds are described in Fig. 4.

$\left|z_{MD}^{*}\right|$ is the evasion bound and $\left|z_{MT}^{*}\right|$ is the pursuit bound, it is essential for the attacker to keep both the of them inside the bounds, so that

$$\left|z_{MT}\left(\xi_{go}^{MT}\right)\right|<\left|z_{MT}^{*}\left(\xi_{go}^{MT}\right)\right|,\left|z_{MD}\left(\xi_{go}^{MD}\right)\right|>\left|z_{MD}^{*}\left(\xi_{go}^{MD}\right)\right| \quad (57)$$

$\left|z_{MD}^{**}\right|$ is the failsafe bound for the attacker in the case 2.

During the first phase, when the attacker use the pursuit strategy, it must guarantee the miss distance $\left|z_{MD}\right|>\left|z_{MD}^{*}\right|$, or at $\xi_{go}^{MD}=0$, the miss distance will smaller than the desired miss distance, which the attacker cannot endure.

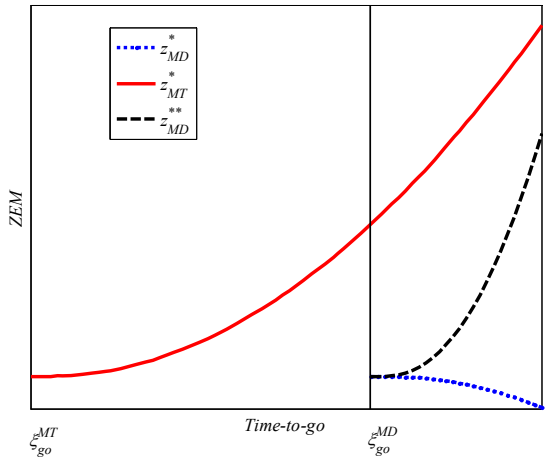


Fig. 4 Defender-attacker and attacker-target ZEM bounds

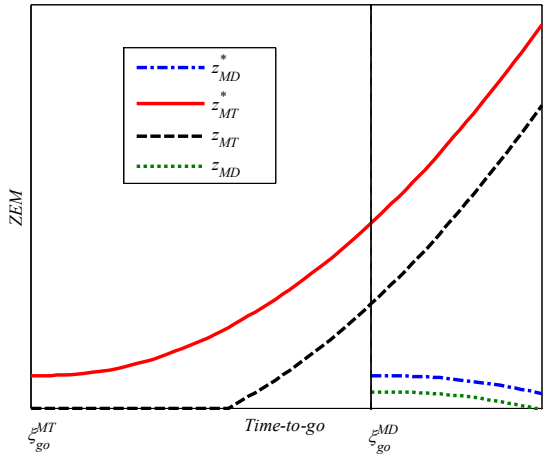


Fig. 5 Case 1 linear simulation

In the case 1 in part C in this section, when the attacker use the optimal strategy, the trajectory are both in the bounds as described in Fig. 5. The attacker can guarantee the desired miss distance and the first case is trivial.

By using the failsafe bound, the attacker can use the hybrid strategy in the first phase in the case 2. For a given desired miss distance $\left|z_{MD}^{*}\left(\xi_{go}^{MD}=0\right)\right|$, there is a failsafe bound for the attacker, so that if $\left|z_{MD}\left(\xi_{go}^{MD}\right)\right|=\left|z_{MD}^{*}\left(\xi_{go}^{MD}\right)\right|$ for any $\xi_{go}^{MD}>0$, the attacker's strategy can be safely switched form u_{Me}^{*} to u_{Mp}^{*} at that point and the miss distance of $\left|z_{MD}^{*}\left(\xi_{go}^{MD}=0\right)\right|$ can be guaranteed.

To reach $\left|z_{MD}^{**}\right|$ the attacker can use a variety of controls. Define the switch time ξ^{*}. The endgame can be divided into another two phase named as phase A, and phase B. In phase A, the attacker will use the evasion strategy (perhaps not optimal) to reach $\left|z_{MD}^{**}\right|$, while in phase B, it will use the optimal pursuit strategy to pursue the target and guarantee the miss distance.

Given the initial conditions $\left|z_{0}^{MD}\right|=\left|z_{MD}\left(\xi=0\right)\right|$ and $\left|z_{0}^{MT}\right|=\left|z_{MT}\left(\xi=0\right)\right|$, where, $\xi=t/\tau_M$. Rename some of our variables to work with a single time-to-go variable. Define

$$\xi_{go}=\xi_{go}^{MD},\xi_{f}=\xi_{f}^{MD},\xi_{go}+\Delta\xi=\xi_{go}^{MT},\xi_{f}+\Delta\xi=\xi_{f}^{MT} \quad (58)$$

To verify the optimal strategy in the phase A, consider the attacker uses an evasive maneuver $u_M=a\text{sign}\left(z_{MD}\right)$, $a\le\rho_M$ to evade the defender, as the target uses its optimal evasion law $u_T=\rho_T\text{sign}\left(z_{MT}\right)$ and the defender uses its optimal pursuit law $u_D=-\rho_D\text{sign}\left(z_{MD}\right)$

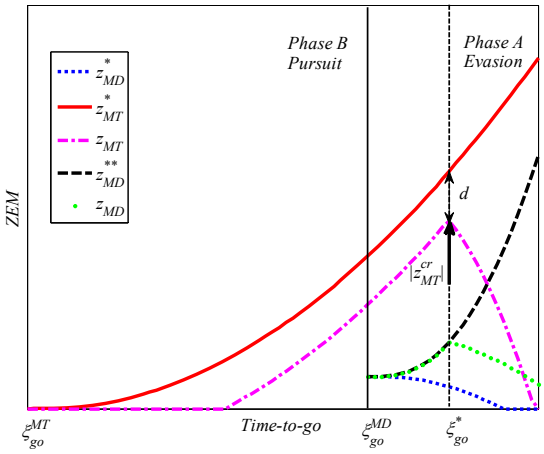


Fig. 6 Case 2 linear simulation

The $\left|z_{MT}\left(\xi_{go}\right)\right|$ equation becomes

$$\left|z_{MT}\left(\xi_{go}\right)\right|=\left|z_{0}^{MT}\right|+\frac{a}{\rho_M}\int_{0}^{\xi}\psi\left(\xi_{f}^{MT}-\eta\right)d\eta +\mu_T\varepsilon_T\int_{0}^{\xi}\psi\left(\left(\xi_{f}^{MT}-\eta\right)/\varepsilon_T\right)d\eta \quad (59)$$

Similarly, for the second ZEM variable,

$$\left|z_{MD}\left(\xi_{go}\right)\right|=\left|z_0^{MD}\right|+\frac{a}{\rho_M}\int_0^{\xi}\psi\left(\xi_f^{MD}-\eta\right)d\eta \\ -\mu_D\varepsilon_D\int_0^{\xi}\psi\left(\left(\xi_f^{MD}-\eta\right)/\varepsilon_D\right)d\eta \tag{60}$$

The most aggressive strategy uses u_{Me}^* to reach the $\left|z_{MD}^{**}\right|$ and, then, switch to u_{Mp}^* .The trajectory is shown in **错误！未找到引用源。**.

Define

$$\left|z_{MT}^{cr}\right|=\left|z_{MT}\left(\xi_{go}^*\right)\right| \tag{61}$$

$$d\left(\xi_{go}^*\right)=\left|z_{MT}^*\left(\xi_{go}^*\right)\right|-\left|z_{MT}^{cr}\right| \tag{62}$$

From **错误！未找到引用源。**, it is readily found that, when the attacker uses pursuit law in the phase B, the miss distance $\left|z_{MT}^*\left(\xi_{go}^{MT}=0\right)\right|$ depends on the value of $d\left(\xi_{go}^*\right)$, so to find the optimal strategy, it is necessary to find the pair $\left(\xi_{go}^*,a^*\right)$ for which the value of the defined $d\left(\xi_{go}^*\right)$ is maximal in the appropriate interval. The cost is

$$d^*=\max_{\xi_{go}^*}d\left(\xi_{go}^*\right) \tag{63}$$

Define

$$l=\left|z_{MD}^*\left(\xi_{go}^{MD}=0\right)\right|,m=\left|z_{MT}^*\left(\xi_{go}^{MT}=0\right)\right| \tag{64}$$

Integration yields

$$\begin{aligned}\left|z_{MD}^{**}\left(\xi_{go}\right)\right|&=l+\mu_D\varepsilon_D\int_0^{\xi_{go}}\psi\left(\eta/\varepsilon_D\right)d\eta+\int_0^{\xi_{go}}\psi\left(\eta\right)d\eta\\ \left|z_{MD}\left(\xi_{go}\right)\right|&=\left|z_0^{MD}\right|-\mu_D\varepsilon_D\int_0^{\xi}\psi\left(\left(\xi_f^{MD}-\eta\right)/\varepsilon_D\right)d\eta\\ &\quad+k\int_0^{\xi}\psi\left(\xi_f^{MD}-\eta\right)d\eta\\ \left|z_{MT}^*\left(\xi_{go}\right)\right|&=m-\mu_T\varepsilon_T\int_0^{\xi_{go}+\Delta\xi}\psi\left(\eta/\varepsilon_T\right)d\eta+\int_0^{\xi_{go}+\Delta\xi}\psi\left(\eta\right)d\eta\\ \left|z_{MT}\left(\xi_{go}\right)\right|&=\left|z_0^{MT}\right|+\mu_T\varepsilon_T\int_0^{\xi}\psi\left(\left(\xi_f^{MT}-\eta\right)/\varepsilon_T\right)d\eta\\ &\quad+k\int_0^{\xi}\psi\left(\xi_f^{MT}-\eta\right)d\eta\end{aligned} \tag{65}$$

where $k=a/\rho_M$.

Denote $\theta\left(\eta\right)=\int\psi\left(\eta\right)d\eta$, and $\theta\left(0\right)=-1$

To find the intersection points of the functions $\left|z_{MD}\left(\xi_{go}\right)\right|$ and $\left|z_{MD}^{**}\left(\xi_{go}\right)\right|$, equate them and obtain

$$\theta\left(\xi_{go}^*\right)=\frac{\left|z_0^{MD}\right|-l-\mu_D\varepsilon_D^2\theta\left(\xi_f/\varepsilon_D\right)+k\theta\left(\xi_f\right)+\left(\mu_D\varepsilon_D^2+1\right)\theta\left(0\right)}{k+1} \tag{66}$$

or alternatively

$$k^*\left(\xi_{go}\right)=\frac{l-\left|z_0^{MD}\right|+\mu_D\varepsilon_D^2\theta\left(\xi_f/\varepsilon_D\right)-\left(\mu_D\varepsilon_D^2+1\right)\theta\left(0\right)+\theta\left(\xi_{go}\right)}{\theta\left(\xi_f\right)-\theta\left(\xi_{go}\right)} \tag{67}$$

Consider that $\theta\left(\xi_{go}\right)\ge 0$, so that

$$k\ge\frac{l-\left|z_0^{MD}\right|+\mu_D\varepsilon_D^2\theta\left(\xi_f/\varepsilon_D\right)-\left(\mu_D\varepsilon_D^2+1\right)\theta\left(0\right)}{\theta\left(\xi_f\right)} \tag{68}$$

This means that, the ratio k is bounded

$$\frac{l-\left|z_0^{MD}\right|+\mu_D\varepsilon_D^2\theta\left(\xi_f/\varepsilon_D\right)-\left(\mu_D\varepsilon_D^2+1\right)\theta\left(0\right)}{\theta\left(\xi_f\right)}=k_{\min}\le k\le 1 \tag{69}$$

If

$$\frac{l-\left|z_0^{MD}\right|+\mu_D\varepsilon_D^2\theta\left(\xi_f/\varepsilon_D\right)-\left(\mu_D\varepsilon_D^2+1\right)\theta\left(0\right)}{\theta\left(\xi_f\right)}\ge 1 \tag{70}$$

the defender can guarantee a miss distance smaller than l . The $\theta\left(\xi_{go}^*\right)$ is also bounded:

$$0\le\theta\left(\xi_{go}^*\right)\le\frac{\left|z_0^{MD}\right|-l-\mu_D\varepsilon_D^2\theta\left(\xi_f/\varepsilon_D\right)+\theta\left(\xi_f\right)+\left(\mu_D\varepsilon_D^2+1\right)\theta\left(0\right)}{2} \tag{71}$$

Furthermore,

$$\begin{aligned}d\left(\xi_{go}^*\right)&=m-\left|z_0^{MT}\right|-\mu_T\varepsilon_T^2\left(\theta\left(\left(\xi_f+\Delta\xi\right)/\varepsilon_T\right)-\theta\left(0\right)\right)\\ &\quad+\left(\theta\left(\xi_f+\Delta\xi\right)-\theta\left(0\right)\right)+\left(k+1\right)\left(\theta\left(\xi_f+\Delta\xi\right)-\theta\left(\xi_{go}+\Delta\xi\right)\right)\end{aligned} \tag{72}$$

Substituting $k^*\left(\xi_{go}\right)$ into the $d\left(\xi_{go}^*\right)$ and obtain

$$\begin{aligned}d\left(\xi_{go}^*\right)&=m-\left|z_0^{MT}\right|-\mu_T\varepsilon_T^2\left(\theta\left(\left(\xi_f+\Delta\xi\right)/\varepsilon_T\right)-\theta\left(0\right)\right)\\ &\quad+\left(\theta\left(\xi_f+\Delta\xi\right)-\theta\left(0\right)\right)\\ &\quad+\frac{l-\left|z_0^{MD}\right|+\mu_D\varepsilon_D^2\theta\left(\xi_f/\varepsilon_D\right)-\left(\mu_D\varepsilon_D^2+1\right)\theta\left(0\right)+\theta\left(\xi_f\right)}{\left(\theta\left(\xi_f\right)-\theta\left(\xi_{go}\right)\right)/\left(\theta\left(\xi_f+\Delta\xi\right)-\theta\left(\xi_{go}+\Delta\xi\right)\right)}\end{aligned} \tag{73}$$

When given the desired miss distance m, initial condition $\left|z_0^{MT}\right|$, final time ξ_f^{MT} and ξ_f^{MD}, the value of the cost function depends on the ξ_{go} .

Define

$$f\left(\xi_{go}\right)=\frac{\theta\left(\xi_f+\Delta\xi\right)-\theta\left(\xi_{go}+\Delta\xi\right)}{\theta\left(\xi_f\right)-\theta\left(\xi_{go}\right)} \tag{74}$$

Thus, to maximize the function $d\left(\xi_{go}\right)$, the function $f\left(\xi_{go}\right)$ should be maximal.

Differentiate $\theta\left(\xi_{go}\right)$ with respect to ξ_{go} , simplify, and obtain for any $\xi_{go}\ge 0$

$$\frac{d\theta}{d\xi_{go}}=\psi\left(\xi_{go}\right)=e^{-\xi_{go}}+\xi_{go}-1\ge 0 \tag{75}$$

$$\frac{d^2\theta}{d\xi_{go}^2}=-e^{-\xi_{go}}+1\ge 0 \tag{76}$$

So the function $\theta\left(\xi_{go}\right)$ is monotonic increasing and it is a concave function, the function $f\left(\xi_{go}\right)>0$ and it is also a monotonic increasing function. From this ,it is readily seen that to maximize the function $d\left(\xi_{go}\right)$,the switch time should be chosen as $\xi_{go\max}$, which corresponds to $a^*=\rho_M$.

ISBN: 978-1-61804-290-3

Finally, the strategy that maximizes $d\left(\xi_{go}\right)$ is

$$\begin{aligned}&\text{if } \operatorname{sign}\left(Z_{MD}\right)=\text{-}\operatorname{sign}\left(Z_{MT}\right)\\&u_M=u_{Me}^*\left(t\right)=u_{Mp}^*\left(t\right)\\&\text{if } \operatorname{sign}\left(Z_{MD}\right)=\operatorname{sign}\left(Z_{MT}\right)\\&u_M=u_e^*\left(t\right)\ \text{ until}\left|Z_{MD}\left(t_{go}\right)\right|=\left|Z_{MD}^{**}\left(t_{go}\right)\right|,\text{then},u_M=u_p^*\left(t\right)\end{aligned}\tag{77}$$

IV. Conclusion

Optimal evasion and pursuit strategies for a three-player conflict have been derived in this paper. This problem is analyzed for a general arbitrary-order linear dynamics, under the assumption that the information is perfect and the control is bounded, a linearized model is derived. Based on the norm differential game (NDG), a general solution for an arbitrary-order linear model is obtained, by using the first-order system for example, the optimal evasion and pursuit strategies for the attacker, the target and its defender are derived, and the game space decomposition for attacker-target and defender-attacker is analyzed, then the closed-form solution for first-order system is obtained. For the attacker, the strategies are more complicated than the target and the defender, because, it must capture the target while evading from the defender, the maneuver ratio and dynamics ratio are very important in the game. As the case 1 described in the paper, the initial conditions are very important too, but the player in the game cannot enforce it. By using the NDG method, the optimal evasion strategy for the target and the optimal pursuit strategy for the defender have been obtained. To pursue the target and evade from the defender a hybrid evasion-pursuit strategy for the attacker has been investigated, and an analytical solution is obtained. The switch time t_{go}^* is a key parameter in this problem. In this paper the optimal t_{go}^* is obtained by analysis, and the optimal switch strategy is obtained too. In the future research nonlinear kinematics would be considered, and in the real scenario, the information is not perfect, it is necessary to provide a useful estimation method considering the nonlinear kinematics of the problem, and consider the influence of the noise in the endgame.

References

[1] Ho, Y., Bryson, A., Baron, S. Differential games and optimal pursuit-evasion strategies. Automatic Control, IEEE Transactions on. 1965, 10(4): 385-389.

[2] Turetsky, V., Shinar, J. Missile guidance laws based on pursuit–evasion game formulations. Automatica. 2003, 39(4): 607-618.

[3] Shima, T., Golan, O. M. Linear Quadratic Differential Games Guidance Law for Dual Controlled Missiles. Aerospace and Electronic Systems, IEEE Transactions on. 2007, 43(3): 834-842.

[4] Prokopov, O., Shima, T. Linear Quadratic Optimal Cooperative Strategies for Active Aircraft Protection. Journal of Guidance, Control, and Dynamics. 2013: 1-12.

[5] Gutman, S. On Optimal Guidance for Homing Missiles. Journal of Guidance, Control, and Dynamics. 1979, 2(4): 296-300.

[6] Boyell, R. L. Defending a Moving Target Against Missile or Torpedo Attack. Aerospace and Electronic Systems, IEEE Transactions on. 1976, AES-12(4): 522-526.

[7] Rusnak, I. The Lady, The Bandits, and The Bodyguards–A Two Team Dynamic Game. in: Proceedings of the 16th World IFAC Congress: 2005.

[8] Perelman, A., Shima, T., Rusnak, I. Cooperative differential games strategies for active aircraft protection from a homing missile. Journal of Guidance, Control, and Dynamics. 2011, 34(3): 761-773.

[9] Shaferman, V., Shima, T. Cooperative Multiple-Model Adaptive Guidance for an Aircraft Defending Missile. Journal of Guidance, Control, and Dynamics. 2010, 33(6): 1801-1813.

[10] Shima, T. Optimal Cooperative Pursuit and Evasion Strategies Against a Homing Missile. Journal of Guidance, Control, and Dynamics. 2011, 34(2): 414-425.

[11] Yamasaki, T., Balakrishnan, S. N., Takano, H. Modified Command to Line-of-Sight Intercept Guidance for Aircraft Defense. Journal of Guidance, Control, and Dynamics. 2013, 36(3): 898-902.

[12] Ratnoo, A., Shima, T. Line-of-sight interceptor guidance for defending an aircraft. Journal of Guidance, Control, and Dynamics. 2011, 34(2): 522-532.

[13] Ratnoo, A., Shima, T. Guidance Strategies Against Defended Aerial Targets. Journal of Guidance, Control, and Dynamics. 2012, 35(4): 1059-1068.

[14] Rubinsky, S., Gutman, S. Three Body Guaranteed Pursuit and Evasion. in: AIAA Guidance, Navigation, and Control Conference: American Institute of Aeronautics and Astronautics, 2012.

[15] Rubinsky, S., Gutman, S. Three-Player Pursuit and Evasion Conflict. Journal of Guidance, Control, and Dynamics. 2014, 37(1): 98-110.

INVITED-DIMITROVA - Short CV of Yongji Wang

Yongji Wang received his PhD in Power Plant Engineering from Huazhong University of Science and Technology, China in 1990. He is currently a Professor in School of Automation at Huazhong University of Science and Technology. His research interests include neural network, system identification and control, flight vehicle control.

Model of Resources Requirements for Software Product Quality Using ISO Standards

Kenza Meridji, Khalid T. Al-Sarayreh and Tatiana Balikhina

Abstract— **Resources requirements according to ISO standards describe all requirements related software product quality including resources list of the hardware environment in which the software is specified to operate and resource utilization requirements list of the sizing and timing requirements applicable to the software item under specification and computer software requirements description of the computer software to be used with the software under specification or incorporated into the software item; for instance operating system and software items to be reused. This paper presents a proposed model of resources requirements on the basis of ISO standards for measuring the software resources product quality; whether the software it has already been delivered or has yet to be built**.

***Keywords*—**Resources Requirements, Software Product Quality, ISO Standards

I. INTRODUCTION

Resource collection and task scheme are fundamental function in software system environments, for instance, the cloud computing tasks try to win system resources. The choices made by the parallel algorithms ought to be judged based not only on measurements related to customer satisfaction, such as the proportion of tasks hand out without affecting their quality requirements, but also stands on resources-related performance measurements, such as the number of resources used to hand out the tasks and their exploitation competence.

Developers of software products are in charge for identifying the requirements of any product, developing software that put into practices the requirements, and for allocating suitable resources such as processors and communication networks. Improvement such quality software systems has challenge for software developers. In practice, identifying the non-functional resources requirements are often captured at high level while the focusing only on the functionality of the system [1-7]. Several products have failed because of neglect of such non-functional requirements

Resources Requirements describes what the component needs from its environment to perform its function and define the limits of software budgets associated with computer resources such as: CPU load and maximum memory size to be considered by the supplier as well as [8] and [9] Indicates to computer hardware resource requirements on the utilization (e.g. processor capacity and memory capacity) available for the software item (e.g. sizing and timing). Moreover indicates to Computer software resource requirements on the software items to be used by or incorporated into the system (or constituent software product) (e.g. a specific real time operating system).

Resources requirements are considered as an important part in the software life cycle to assure from the suitability and for the availability of resources to implement it by all functions involved in its application [1-7]. Consequently ISO standards [8] describe the resources requirements as the capability of the software product to use appropriate amounts and types of resources when the software performs its function under stated conditions.

This paper will report the design measurement method to identify of the software resources based on international standards as an autonomous method to identify the size of the software resources independently of the software languages types, which avoids the weaknesses observed in the resources measures currently available.

The paper scope is to identify separately the all functionality allocated to software resources as a piece of the application in the requirements for embedded and real time software, whether it has yet to be built or it has already been delivered.

Furthermore, the main contribution of this paper is the proposed a standard based model of software resources requirements. The proposed generic model is considered as kind of a ‘reference model’ in the sense of an ‘etalon’ standard that is being used for the measurement of resources.

This paper is organized as follows. Section 2 presents the related works. Section 3 presents design the measurements method for resources requirements as defined in ISO. Section 4 presents design a Meta model of resources requirements. Section 5 presents Design numerical rules of software resources requirements. Discussion and a conclusion are presented in section 6.

F. A. Kenza Meridji is now with the University of Petra, Collage of Information Technology, Department of Software Engineering, 11196 Amman, Jordan. (e-mail: kmeridji@uop.edu.jo).

S. B. Khalid T. Al-Sarayreh is now with the Hashemite University, Prince Hussein Bin Abdullah II for Information Technology, Department of Software Engineering, 13115 Zarqa, Jordan. (e-mail: khalidt@hu.edu.jo).

S. C. Tatiana Balikhina is now with the University of Petra, Collage of Information Technology, Department of Computer Science, 11196 Amman, Jordan. (e-mail: tbalikhina@uop.edu.jo).

II. Related Work

Much of the work done up to date on resource non functional requirements was considering resources in general without dealing with them in detail. For instance a model in [10] suggest the assignment of tasks to resources to be able to reduce the problems related the tasks' time requirements at the same time increasing the resources' utilization efficiency for a given number of resources. The proposed method takes concepts derived from graph partitioning, and collects tasks together to be able to reduce the overlapping time of the task that is assigned to a given resource and to be able to increase the time overlapping with tasks assigned to dissimilar resources [10].

Furthermore, [11] outlined five steps concerning resources quality requirements defined by equipment resources consumption, function and structure environmental impacts. The five steps are described and encourage the following first, to decrease the dependence of equipment on non-sustainable resources: this will allow decreasing the non-sustainable resource usage and consumption; in addition it will allow utilizing vigorously the sustainable resource as alternative supply. Second, is to decrease wastes and to improve the use of resources efficiently. Third, renew and modernize equipments, to be able to use entirely the equipment potential and to effectively decrease the retired equipment. Fourth, reuses and retrieve the equipment resources, to be able to improve the recycling of resources. Fifth, decrease environment damage.

While in [12] the authors proposed a model to tackle the Stake Cloud community platform. This model has ability to work as a cloud resources marketplace. By permitting the users to input their resource needs and give them the matching cloud services.

Moreover, [13] defined eleven steps by conducting an empirical study on the role that requirements and resources play in the building a software product quality. This will allow observing and defining how software quality is constructed in software development organizations. Therefore eleven software programmers, testers, quality control personnel, requirement managers and research and development personnel were interviewed and common practices of quality construction were analyzed. The result showed that quality construction practices differ significantly among different organizations. Differences were mentioned about and the degree of involvement of the customer in the software development, methods used for requirements elicitation, and objectives of software testing.

For instance, [14] studied the impact of non-functional requirements on requirements evolution; this paper listed and analyzed different approaches, available in the literature related to non-functional requirements during software development. This paper focused on three issues: Different views on non-functional requirements, Representation of non-functional requirements and how to deal with non-functional requirements.

Whilst [15] proposed a definition and a discussion of the most used agile software development methods and they investigate the software SMEs challenges and for comparison purposes formulate it into criteria. In addition these methods were compared against the defined criteria and as a result their similarities and differences were outlined.

Finally, [16] introduced a datacenter resources integrated provisioning (DRIP) architecture using synchronized virtualization of distributed datacenters and operate multi-domain software defined optical networks. The DRIP architecture objective is to achieve the integration and allocation of IT resources and optical network resources. In order to examine the feasibility and efficiency of the anticipated architecture, two IT resources allocation strategies and two virtual networks composition strategies are evaluated [17] and [18].

The motivation of this research paper is to contribute to better define, describe and measure some of the NFR inputs required for the adequate *a priori* cost estimation of software projects. The measurement scope in this paper is to identify separately all functionalities allocated to software resources requirements for software product quality.

The focus of this paper is on a single type of NFR that is, resources requirements. This paper reports on the work carried out to define an integrated view of software functional user requirements for resources requirements for the software product on the basis of ISO international standards.

III. Design Measurement Method of Resources Requirements as Defined in ISO

Based on the resources requirements definitions stated by ISO standards the design measures steps for resources requirements as follows:

A. Determination of measurement objectives for Software Product

This section illustrates the measurement objectives of resources requirements as a piece of a software product quality, followed by the measurement point of view and the intended uses of the measurement results.

1) **The objective**: is to measure the size of the resources requirements as defined in ISO.
2) **Measurement point of view**: Software perspective.
3) **Intended uses of the measurement results**: throughout the software life cycle: the size of the resources for a software product, whether it has yet to be built or it has already been delivered.

B. Characterization Resources Concepts to Measured

This section illustrates the resources requirements concepts and the identified resources to be measured

1) **Definition of the concept to be measured:** the resources measurements can be internal or external. The proposed measurement method is to be applicable for non-embedded software resources.

   ***External resources Measures*:** should be able to measure such attributes as the utilised resources behaviour of computer system including software during testing or operating and can be measured based on the following resource utilization: (I/O resource

measurements, Memory resource measurements and Transmission resource measurements).

***Internal resources Measures*:** indicate a set of attributes for predicting the utilization of hardware resources by the computer system including the software product during testing or operating.

2) **The Resource entities to be measured**

- External Resources Entities
  1) I/O resource measurements
     - I/O Devices Utilization
     - User Waiting Time of I/O Devices Utilization
  2) Memory resource measurements
     - Memory Utilization
  3) Transmission resource measurements.
     - Maximum Transmission Utilization
     - Transmission Capacity Utilization
     - Media Device Utilization
- Internal Resources Entities
  1) I/O Related Errors
  2) I/O loading

IV. Design A Meta Model of Resources Requirements

This section presents the meta model of the software resources requirements on the basis of the previous section.

A. *I/O devices resources*

In the following design of the Meta models- see Figure 1:

- Entity type 1 can be used to measure the e external software resources throughout executing concurrently a large number of tasks and record I/O device utilization for one functional process.
- Entity type 2 can be used to measure the internal software resources throughout calibrating the test conditions and emulate a condition whereby the system reaches a situation of maximum I/O loading to define the I/O errors for one functional process.
- Entity type 3 can be used to measure the internal software resources throughout calibrating the test condition to define maximum I/O loading for one functional process.
- Entity type 4 can be used to measure the external software resources throughout run the application of record of errors due to I/O failures and warning for one functional process.

B. *Memory resources*

In the following design of the Meta models- see Figure 2:

- Entity type 5 can be used to measure external software resources throughout executing concurrently a large number of tasks and run the application and record number of errors due to memory failures and warnings for one functional process.

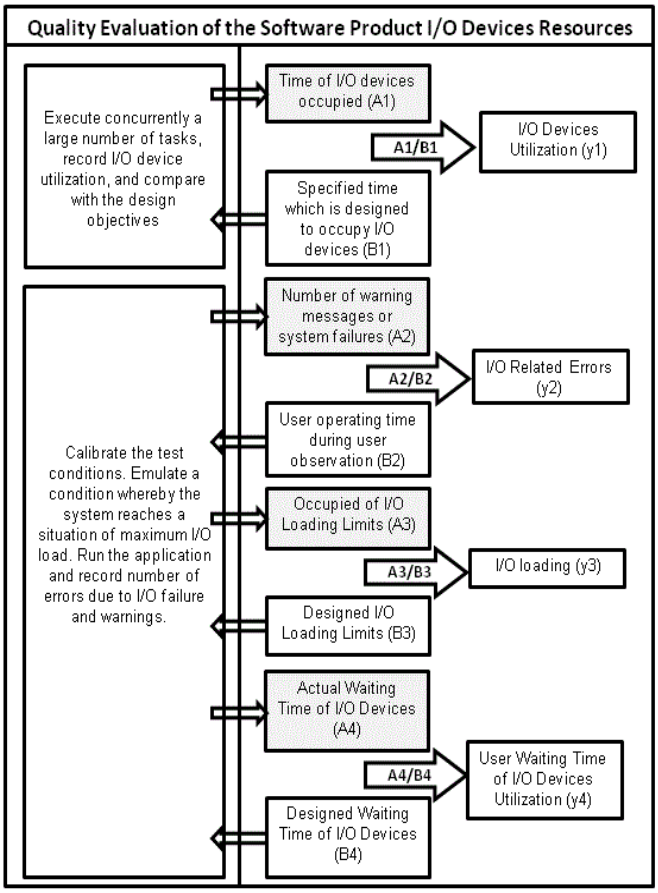


Fig. 1: I/O Devices Resources Meta Model

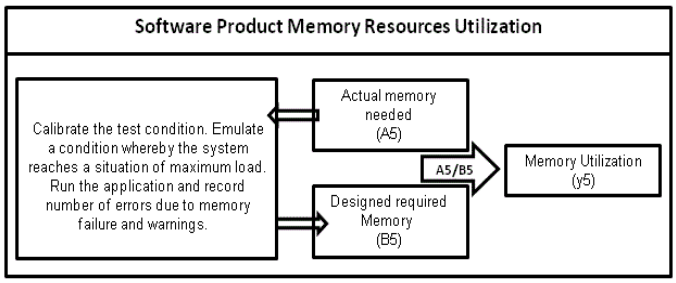


Fig. 2: I/O Memory Resources Meta Model

C. *Transmission resources*

In the following design of the Meta models- see Figure 3:

- Entity type 6 can be used to measure the external software resources throughout evaluate what is required for the system to reach a situation of maximum load for one functional process.
- Entity type 7 can be used to measure the external software resources throughout observe transmission capacity and compare specified one for one functional process.
- Entity type 8 can be used to measure the external software resources throughout execute concurrently specified tasks with multiple user for one functional process.

ISBN: 978-1-61804-290-3

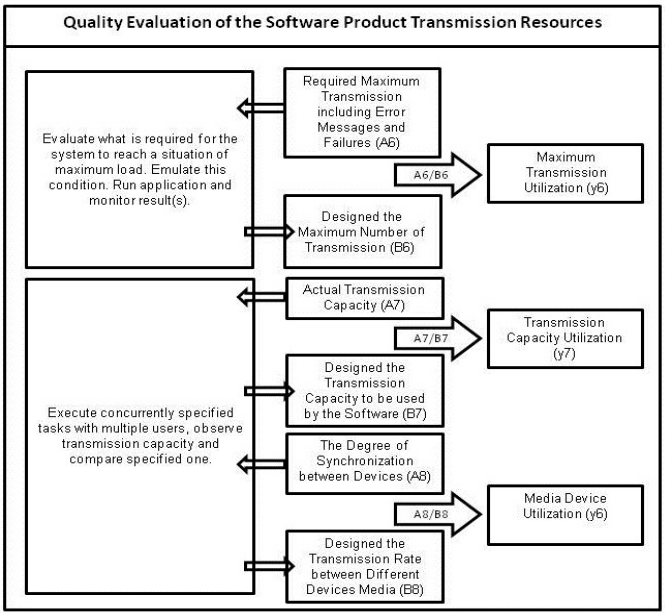


Fig. 3: Transmission Resources Meta Model

V. Numerical Assignment Roles for Resources Requirements

In this section the basis for these numerical assignment rules are the proposed meta-models as defined in Figures 1, 2 and 3. Numerical assignments rules can be described through a descriptive text or what so called (a practitioner's description) or through using mathematical expressions or as defined as (a formal theoretical viewpoint). In this paper the numerical assignments rules are built based on mathematical expressions while the descriptive text rules are defined in ISO.

A. Identification of data Resources Groups

This section illustrates data resources groups as defined in ISO (i.e. Input and output) form sources and/or to data destinations for Resources Requirements for details see-table 1 and table 2.

Table 1 : Resources Requirements Data Sources

Categories	Data Sources	Objects of Interest
I/O Devices Resources	• Specified time which is designed to occupy I/O devices. • Actual time of I/O devices occupied. • User operating time during user observation. • The number of warning messages or system failures. • Designed I/O loading limits. • The occupied of I/O loading limits • Designed waiting time of I/O devices. • The actual waiting time of I/O devices.	• Time • Time • Number • Number • Loading limit • Loading limit • Time • Time

Table 1 : Resources Requirements Data Sources (Contd)

Categories	Data Sources	Objects of Interest
Memory Resources	• Designed required memory. • The actual memory needed	• Size • Size
Transmission Resources	• Designed maximum number of transmission. • The required maximum transmission including error messages and failures. • The designed the transmission capacity. • The actual transmission capacity. • The design of the transmission rate between different devices media. • The degree of synchronization between devices.	• Transmission no. • Transmission no. • Transmission Capacity • Transmission Capacity • Transmission Rate • Transmission Rate

Table 2 : Resources Data Destinations

Categories	Data Destinations
I/O devices resources	• I/O devices utilization • I/O related errors • I/O loading • User waiting time of I/O devices
Memory resources	• Memory utilization
Transmission resources	• Maximum transmission utilization • Transmission capacity utilization • Media devices utilization

B. Numerical Roles for Resources

Regards to three meta models, resources requirements (internally and externally) for **I/O devices resources**: measures the executing concurrently the tasks, record maximum I/O loading and errors due I/O failures., **Memory resources**: measures the memory maximum load and the errors due to memory failures and warnings and **Transmission resources**: measures maximum transmission, transmission capacity and synchronization between media devices.

1) Mathematical Rules of Resources I/O device:

This section presents mathematical assignments rules for resources I/O devices follows:

- The measurement size of the I/O device (externally) for one process

  $\sum$ Data Movement (Data Recourses Group)

  $\sum$ (I/O Devices Utilization + User Waiting Time of I/O Devices).

- The measurement size of the I/O devices (internally) for one process.

∑ Data Movement (Data Resource Group)

∑ (I/O related errors + I/O Loading).

- The measurement size of the I/O devices (internally and externally)

(∑ (I/O Devices Utilization + User Waiting Time of I/O Devices) + ∑ (I/O related errors + I/O Loading))

- The **T**otal Measurement **S**ize of the I/O devices [for the all functional processes]

N x (∑ (I/O Devices Utilization + User Waiting Time of I/O Devices) +∑ (I/O related errors + I/O Loading))

N: number of functional processes for the I/O devices.

2) Mathematical Rules for Memory Resources:

- The measurement size of the memory resources (externally) for one process

∑ Data Movement (Data Resource Group)

∑ (memory Utilization).

Note: There are no internal measures as defined in ISO.

- The **T**otal **S**ize of the memory resources [for the all functional processes]

N x ∑ (memory resources)

N: number of functional processes for the memory resources.

3) Mathematical Rules of Transmission Resources:

- The measurement size of the transmission resources (externally) for one process

∑ (Data Movement (Data Resource Group)

∑ (Maximum Transmission + Transmission Capacity + Media Devices Utilization)

- The **T**otal **F**unctional **S**ize of the transmission resources [for the all functional processes]

N x ∑ (Maximum Transmission + Transmission Capacity + Media Devices Utilization)

N: number of functional processes for the transmission resources.

4) Total measurement Size of Resources Requirements

The **T**otal **M**easurement **S**ize of the resources [for the all functional processes]

N x (∑ (I/O Devices Utilization + User Waiting Time of I/O Devices) +∑ (I/O related errors + I/O Loading))

+

N x ∑ (memory resources)

+

N x ∑ (Maximum Transmission + Transmission Capacity + Media Devices Utilization)

N: number of functional processes for the resources.

VI. CONCLUSION

This paper introduced a new design method to measure the resources non-functional requirements internally and externally a . As well as proposed three meta models for resources requirement as defined in ISO 9126 standards independently of the software type or languages used .

Moreover, the design of the measurement method is defined to specify the strategy of the measurement rules. This will allow performing the mapping between the concepts of ISO 19761 and the concepts of the suggesting design of the generic resources Meta models and rules and then identification of the data movements and the performance of the measurement process.

It is important to mention that the design of the measurement procedure for resources requirements for the software product quality have been developed to apply the ISO 19761 and to apply a measurement method to the resources requirements in order to obtain the functional size of the resources as a separate piece of a software in early stages of the software development process.

The future work will concentrate on experimental test for the proposed resources model, comparison this model with the previous ones and list all the strength and weaknesses after experiment process as well as mapping our proposed model with the definitions of standard Etalon.

REFERENCES

[1] Abran, A., K. T. Al-Sarayreh, and J. J. Cuadrado-Gallego, " A Standards-based Reference Framework for System Portability Requirements", Computer Standards and Interface, Elsevier, 2013. http://dx.doi.org/10.1016/j.csi.2012.11.003

[2] Al-Sarayreh, K. T., A. Abran and and J. J. Cuadrado-Gallego," A Standards-based model of system maintainability requirements", Journal of Software: Evolution and Process, John Wiley & Sons, Ltd, 2013. http://dx.doi.org/10.1002/smr.1553

[3] Meridji, Kenza, Khalid T. Al-Sarayreh, and Ahmad Al-Khasawneh. "A generic model for the specification of software reliability requirements and measurement of their functional size." *International Journal of Information Quality* 3, no. 2 (2013): 139-163.

[4] Al-Sarayreh, Khalid T., Ibrahim Al-Oqily, and Kenza Meridji. "A standard-based reference framework for system operations

requirements." *International Journal of Computer Applications in Technology* 47, no. 4 (2013): 351-363.

[5] Al-Sarayreh, Khalid T., Ibrahim Al-Oqily, and Kenza Meridji. "A standard based reference framework for system adaptation and installation requirements." In *Next Generation Mobile Applications, Services and Technologies (NGMAST), 2012 6th International Conference on*, pp. 7-12. IEEE, 2012.

[6] Al-Sarayreh, Khalid T., Kenza Meridji, Ebaa Fayyoumi, and Sahar Idwan. "A Novel Approach to Build a Generic Model of Photovoltaic Solar System Using Sound Biometric Techniques." *International Journal of Information Technology and Web Engineering (IJITWE)* 9, no. 1 (2014): 31-44.

[7] K. T. Al-Sarayreh and A. Abran, "A Generic Model for the Specification of Software Interface Requirements and Measurement of Their Functional Size," 8th ACIS International Conference on Software Engineering Research, Management and Applications, SERA 2010, Montreal, Canada, pp. 217-222, 2010.

[8] ISO/IEC-9126, "Software Engineering - Product Quality - Parts 1-4: Quality Model ", International Organization for Standardization, Geneva (Switzerland), 2008.

[9] ISO/IEC-19761, "Software Engineering - COSMIC v 3.1 - A Functional Size Measurement Method", International Organization for Standardization, Geneva (Switzerland), 2011

[10] Doulamis, N.D.; Kokkinos, P.; Varvarigos, E., "Resource Selection for Tasks with Time Requirements Using Spectral Clustering," *Computers, IEEE Transactions on* , vol.63, no.2, pp.461,474, Feb. 2014

[11] Shilun Liu; Mingfang Ni; Kaikai Hu; Ma Yu, "Study on the main requirements of equipment resource-ability design," *Quality, Reliability, Risk, Maintenance, and Safety Engineering (ICQR2MSE), 2011 International Conference on* , vol., no., pp.709,713, 17-19 June 2011

[12] Todoran, I., "StakeCloud: Stakeholder requirements communication and resource identification in the cloud," *Requirements Engineering Conference (RE), 2012 20th IEEE International* , vol., no., pp.353,356, 24-28 Sept. 2012

[13] Seth, F.P.; Mustonen-Ollila, E.; Taipale, O.; Smolander, K., "Software Quality Construction: Empirical Study on the Role of Requirements, Stakeholders and Resources," *Software Engineering Conference (APSEC), 2012 19th Asia-Pacific* , vol.2, no., pp.17,26, 4-7 Dec. 2012

[14] Khatter, K.; Kalia, A., "Impact of Non-functional Requirements on Requirements Evolution," *Emerging Trends in Engineering and Technology (ICETET), 2013 6th International Conference on* , vol., no., pp.61,68, 16-18 Dec. 2013

[15] Hamed, A.M.M.; Abushama, H., "Popular agile approaches in software development: Review and analysis," *Computing, Electrical and Electronics Engineering (ICCEEE), 2013 International Conference on* , vol., no., pp.160,166, 26-28 Aug. 2013

[16] Chen, H.; Zhang, J.; Zhao, Y.; Deng, J.; Wang, W.; He, R.; Yu, X.; Ji, Y.; Zheng, H.; Lin, Y.; Yang, H., "Experimental Demonstration of Datacenter Resources Integrated Provisioning over Multi-Domain Software Defined Optical Networks," *Lightwave Technology, Journal of* , vol.PP, no.99, pp.1,1. 2015.

[17] Penzenstadler, B.; Raturi, A.; Richardson, D.; Tomlinson, B., "Safety, Security, Now Sustainability: The Nonfunctional Requirement for the 21st Century," *Software, IEEE* , vol.31, no.3, pp.40,47, May-June 2014

[18] Dal Bianco, V.; Myllarniemi, V.; Komssi, M.; Raatikainen, M., "The Role of Platform Boundary Resources in Software Ecosystems: A Case Study," *Software Architecture (WICSA), 2014 IEEE/IFIP Conference on* , vol., no., pp.11,20, 7-11 April 2014

BW Variation and MCLCombination for the Operation of HAPS at 5.8 GHz

[1]Mastaneh Mokayef, [2]Yasser Zahedi, [3]Razali Ngah

[1]Faculty of Engineering technology and Built environment UCSI University, [2]Wireless Communication Center (WCC) Universiti Teknologi Malaysia

Abstract—**Adjacent channel interference makes a severe degradation in the Fixed Services (FS) when High Altitude platform System (HAPS) deployed in the congested band of 5.8 GHz. Hence this paper introduces the combination of tow typical techniques i.e. Minimum Coupling Loss (MCL) and Bandwidth (BW) variation of victim, is introduced as an interference mitigation technique. The Spectrum Emission Mask (SEM) and Interference to Noise Ratio (INR) as interference criterion are considered. Consequently, effective separation distances and guard bands are derived for the feasible coexistence between systems.**

Keywords—***Separation distance; interference; MCL, MICL, SEM.***

I. INTRODUCTION

Radio spectrum is a priceless resource relied upon most portions of our private and public lives.As demand for the congested frequency bands such as 5.8 GHz grows rapidly specially in the densely occupied urban areas, the spectrum sharing and interference mitigation techniques became the center of attenuation in most research centers.Hence, the spectrum is sharing in several dimensions like time, space and geography[1, 2]. In order to utilize spectrum more efficiently, the spectrum sharing access is studied in [3-9]. More over the early stage of spectrum planning which is the interference mitigation technique is of the important considerations in spectrum management approach. Consequently several studies such as [2, 5, 10, 11] have been focused on interference mitigation methods in early stage of spectrum sharing when the link budget is under calculation. Consequently, one of the most easy methods known as Minimum Coupling Loss (MCL) is considered in studies. MCL approach is a straight forward method that models only a solo interferer-victim pair. Consequently the MCL easily acts as a guard against the interference. In contrast, this method is spectrally ineffective therefore to overcome this inefficiency, the MCL and the spectrally effective method of BW variation are combined in the shape of Interference Coupling Loss (ICL) to evaluate the minimum frequency and distance separations between two systems operating in adjacent frequency bands.

Technology developments since 1986 that High Altitude Platform system (HAPS) has been first introduced by International Telecommunication Union (ITU); has increased speedily.So advanced approaches such as dynamic spectrum access were then introduced to the spectrum sharing approach toinvolve the spectrum in space and time slots[1].

Due to the modulation process or non- linearity in the transmitter, the out of band emissions will occur. These unwanted emissions occur instantly freestanding the assigned channel bandwidth. Hence the Spectrum Emission Mask (SEM) and Adjacent Channel Leakage power Ratio are defined by European Telecommunication Standard Institute (ETSI)to specify the limitations ofsuch emission.

In view of the above mentioned reasons the graphical representation of a set of rules applied to the radio transmitter and receiver sides are defined by European Telecommunication Standard Institute (ETSI) as Spectrum Emission Mask (SEM) and Channel Selectivity (CS) respectively.

Accordingly, this paper has shed light on the new combination of mitigation techniques to reduce the interference level as low as possible. Therefore the effect of combination of MICL with additional losses due to the bandwidth of victim is highlighted in this paper.

The rest of this paper is organized as follows: section II represents the interference model and sharing scenarios following the system parameters in section III. Results are presented and discussed in section IV. And finally the paper is concluded in section V.

II. INTERFERENCE MODEL AND SHARING SCENARIO

To assign the radio link coordinated with already existing services in a specific frequency spectrum, the interference from the newly activated system to the existing services has to be taken in to the consideration. Hence the receiver selectivity, antenna gain, path loss and bandwidth of receiver are as important factors that must be highlighted.

In this paper, the interference to noise ratio (INR) as an unwanted to wanted ratio and effective limitation levels are considered.Since the HAPS and FS have different symbol rates, the Net Filter Discrimination (NFD) as mentioned in ITU-RF.746 is considered to show the exact amount of ratio of HAPS transmitter and overall receiver filtering of FS. The

SEM of Type F is applied to show the relative losses and frequency separation due to the bandwidth variations of FS. Consequently, the combination of MCL and additional losses due to the BW of FS is introduced as follows:

$$ICL_{new} = MCL + NFD + S_{BW} \quad (1)$$

Where $MCL(dB)$ represents the minimum required coupling isolation, $NFD\ (dB)$ represents the actual ratio of transmitted power and filtered received power; while the $S_{BW}(dB)$shows the attenuation of signal through its bandwidth.

III. SYSTEM PARAMETERS

The intersystem interference between communication systems depends on several factors such as interferer and victim receiver, propagation path loss between systems, antenna gain and feeder loss. Consequently, having information on technical parameters of systems are the first and the most important factor. Accordingly, Table 1 shows the required technical parameters of systems.

TABLE I. TECHNICAL PARAMETERS OF SYSTEMS

Parameters	HAPSGS	Wi-MAX
Transmitter power (dBm)	-22	-19
Antenna gain (dBi)	47	45
Coverage (km)	36	0.5
Bandwidth (MHz)	11	1.75
Simulation environment	Urban	
Frequency band (MHz)	5850-7075	
INR(dB)	-17.5	
Received thermal noise (dB)	---------	-130

The C band is considered as a reference frequency for operation of systems. More over in order to assure the protection of the existing services in this congested band, the Worst Case interference scenario is considered in this paper.

IV. RESULTS AND DISCUSSION

Subsequently the maximum interference level of -54.8dBW at the FS receiver is achieved. Encrypting the later value along with the parameters of Table 1 will approximately result in 118.5 dB of minimum coupling loss isolation. Hence the following results obtained from the mentioned methodology.

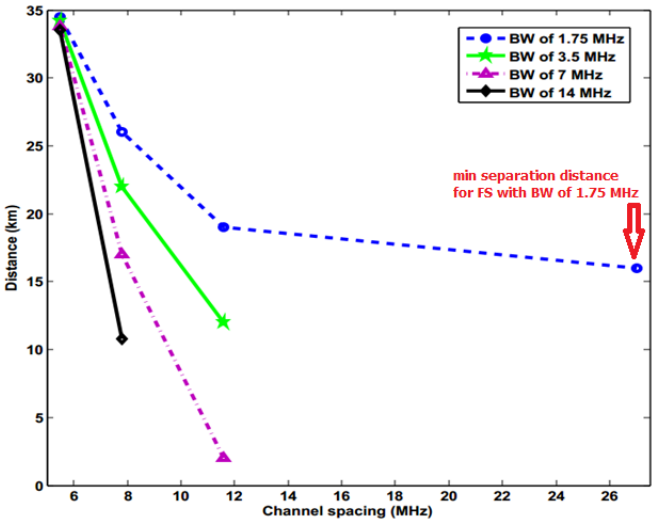


Fig. 1. Required physical and frequency isolation between systems for a)1.75 MHz b)3.5 MHz, c)7 MHz and d) 14 MHzvictim bandwidth

The general view on effect of bandwidth variation on both separation distance and frequency separation is shown in Figure1. It can be observed that HAPS and FS systems can be facilitated for simultaneous operation if the proper channel spacing applies between them. Moreover, less spectral offset will cause extremely larger physical separation, where distance starts increasing rapidly.

V. CONCLUSION

The significant theoretical analysis of the level of transmitter emission up to the receiver is done; hence, a proper generic approach resulted. Investigation on MICL as a combined mitigation technique resulted in the mandatory geographical and frequency spectrum separation rise as interference victim bandwidth decreases and the opposing is correct. The minimum interference coupling loss is resulted as a prediction model to estimate the required separation distance.

REFERENCE

1.Karapantazis, S. and F.-N. Pavlidou, Impact of imperfect power control multiuser detection on the uplink of a WCDMA high altitude platform system. Communications Letters, IEEE, 2005. 9(5): p. 414-416.

2. Mokayef, M., et al., *Spectrum Management for Coexistence of High Altitude Platform System (HAPS) and Fixed Services in 5.8 GHz Band.* Life Science Journal, 2013. **10**(4).

3. Grace, D., et al., *Improving spectrum utilisation for broadband services in mm-wave bands using multiple High Altitude Platforms.* 2002.

4. Kyriazakos, S.A., *Practical radio resource management in wireless systems*. 2004: Artech House.

5. Mokayef, M., et al., *Spectrum Sharing Model for Coexistence between High Altitude Platform System and Fixed Services at 5.8 GHz.* International Journal of Multimedia & Ubiquitous Engineering, 2013. **8**(5).

6. Shamsan, Z.A., A.M. Al-Hetar, and T.A. Rahman, *Spectrum sharing studies of IMT-advanced and FWA services under different clutter loss and channel bandwidths effects.* Progress In Electromagnetics Research, 2008. **87**: p. 331-344.

7. Peha, J.M., *Approaches to spectrum sharing.* Communications Magazine, IEEE, 2005. **43**(2): p. 10-12.

8. Etkin, R., A. Parekh, and D. Tse, *Spectrum sharing for unlicensed bands.* Selected Areas in Communications, IEEE Journal on, 2007. **25**(3): p. 517-528.

9. Duan, L., L. Gao, and J. Huang, *Cooperative spectrum sharing: a contract-based approach.* Mobile Computing, IEEE Transactions on, 2014. **13**(1): p. 174-187.

10. Seidenberg, P., et al. *Statistics of the minimum coupling loss in UMTS/IMT-2000 reference scenarios.* in *Vehicular Technology Conference, 1999. VTC 1999-Fall. IEEE VTS 50th*. 1999. IEEE.

11. Ahmed, M., et al., *Interference Coupling Loss Between Highaltitude Platform Gateway and Fixed Satellite Service*

Earth Station at 5850–7075 MHz. Journal of Electromagnetic Waves and Applications, 2011. **25**(2-3): p. 339-350.

Yang-Baxter Equations, Informatics and Unifying Theories

Radu Iordanescu, Florin F. Nichita and Ion M. Nichita*

April 11, 2015

Abstract

The quantum mechanics had an important influence on building computers; nowadays, the quantum mechanics principles are used for the processing and transmission of information. The Yang-Baxter equation is related to the universal gates from quantum computing and it realizes a unification of certain non-associative structures. Unifying structures could be seen as structures which comprise the information contained in other (algebraic) structures. Recently, we gave the axioms of a structure which unifies associative algebras, Lie algebras and Jordan algebras. Our presentation is a review and a continuation of that approach.

Keywords: universal gate, quantum computer, Yang-Baxter equation, Jordan algebras, Lie algebras, associative algebras

1 Introduction

The importance of computers in our days is that big that we could call our times the "computers era". The quantum mechanics had an important influence on building computers; for example, it led to the production of transistors. At present, the quantum mechanics laws are used for the processing and transmission of information. The first quantum computer (which uses principles of quantum mechanics) was sold to the aerospace and security of defense company Lockheed Martin. The manufacturing company, D-Wave, founded in 1999 and called "a company of quantum computing" promised to perform professional services for the computer maintenance as well. The quantum computer can address issues related to number theory and optimization, which require large computational power. An example is the Shor's algorithm, a quantum algorithm that determines quickly and effectively the prime factors of a big number. With enough qubii, such a computer could use the Shor's algorithm to break algorithms encryption used today.

Non-associative algebras are currently a research direction in fashion (see [1], and the references therein). There are two important classes of non-associative structures: *Lie structures* and *Jordan structures* (see [2]). Various Jordan structures play an important role in quantum group theory and in fundamental physical theories (see [3]). Associative algebras and Lie algebras can be unified at the level of Yang-Baxter structures. A new unification for associative al-

*The first two authors are researchers at the Simion Stoilow Institute of Mathematics of the Romanian Academy, 21 Calea Grivitei Street, 010702 Bucharest, Romania, while the third author is a computer scientist.

ISBN: 978-1-61804-290-3

gebras, Jordan algebras and Lie algebras was obtained recently (see [4]), and we present further results in this paper.

Several papers published in the open access journal AXIOMS deal with the Yang-Baxter equation (see [5] and the references therein). The Yang-Baxter equation can be interpreted in terms of combinatorial logical circuits, and, in logic, it represents some kind of compatibility condition, when working with many logical sentences in the same time. This equation is also related to the theory of universal quantum gates and to the quantum computers (see, for example, [6]). It has many applications in quantum groups and knot theory.

The organization of our paper is the following. In the next section we give the preliminaries and some interpretations of the Yang-Baxter equation in geometry. Section 3 deals with algorithms and interpretations of this equation in computer science. In Section 4, we discuss about the applications of the Yang-Baxter equation in quantum groups and knot theory (with few remarks about universal gates). Sections 5 and 6 are about unification theories for non-associative algebras, and their connections with the previous sections. Sections 7 and 8 are about transcendental numbers and some of their applications. A conclusions section ends our paper.

2 The Yang-Baxter Equations

All tensor products will be defined over the field k, and for V a k-space, we denote by $\tau : V \otimes V \to V \otimes V$ the twist map defined by $\tau(v \otimes w) = w \otimes v$, and by $I : V \to V$ the identity map of the space V.

For $R : V \otimes V \to V \otimes V$ a k-linear map, let $R^{12} = R \otimes I,\ R^{23} = I \otimes R,\ R^{13} = (I \otimes \tau)(R \otimes I)(I \otimes \tau)$.

Definition 2.1 *A Yang-Baxter operator is k-linear map $R : V \otimes V \to V \otimes V$, which is invertible, and it satisfies the braid condition (the Yang-Baxter equation):*

$$R^{12} \circ R^{23} \circ R^{12} = R^{23} \circ R^{12} \circ R^{23}. \quad (1)$$

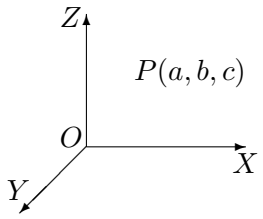
An important observation is that if R satisfies (1) then both $R \circ \tau$ and $\tau \circ R$ satisfy the QYBE:

$$R^{12} \circ R^{13} \circ R^{23} = R^{23} \circ R^{13} \circ R^{12}. \quad (2)$$

Thus, the equations (1) and (2) are equivalent.

There is a similar terminology for the set-theoretical Yang-Baxter equation, for which V is replaced by a set and the tensor product by the Cartesian product (see for example [7, 8]).

Let us now consider the interpretations of the Yang-Baxter equation in geometry.



The symmetries of the point $P(a, b, c)$ about the axes $OX,\ OY,\ OZ$ are defined as follows:
$S_{OX}(a, b, c) = (a,\ -b,\ -c)$,
$S_{OY}(a, b, c) = (\ -a,\ b,\ -c)$,
$S_{OZ}(a, b, c) = (\ -a,\ -b,\ c)$.
They form a group isomorphic with Klein's group: $\{I, S_{OX},\ S_{OY},\ S_{OZ}\}$.

The symmetries of the point $P(a, b, c)$ about the planes $XOY,\ XOZ,\ YOZ$ are defined as follows:

ISBN: 978-1-61804-290-3

$S_{XOY}(a,b,c) = (a,\ b,\ -c)$,
$S_{XOZ}(a,b,c) = (a,\ -b,\ c)$,
$S_{YOZ}(a,b,c) = (\ -a,\ b,\ c)$.

One could check the following instances of the Yang-Baxter equation.

$$S_{XOY} \circ S_{XOZ} \circ S_{YOZ} = S_{YOZ} \circ S_{XOZ} \circ S_{XOY} \quad (3)$$

$$S_{OX} \circ S_{OY} \circ S_{OZ} = S_{OZ} \circ S_{OY} \circ S_{OX} \quad (4)$$

Remark 2.2 *Let us observe that* $S_{OX} \circ S_{OY} \circ S_{OZ} = Id_{k^3}$*, and we can generalize the symmetries about the axes as follows:*
$S'_{OX}(a,b,c) = (a,\ pb,\ qc)$,
$S'_{OY}(a,b,c) = (\ pa,\ b,\ qc)$,
$S'_{OZ}(a,b,c) = (\ pa,\ qb,\ c)$*, for* $p,q \in k$*, such that*
$S'_{OX} \circ S'_{OY} \circ S'_{OZ} = S'_{OZ} \circ S'_{OY} \circ S'_{OX}$.
This is a generalization for the formula (3) as well.
It can be proved that the only rotation operators R *which satisfy (2) are the identity and the operator related to* S_{OX}, S_{OY} *and* S_{OZ}.

3 The Yang-Baxter Equations in Informatics

The Yang-Baxter equation can be interpreted in terms of combinatorial logical circuits (see [9]). It is also related to the theory of universal quantum gates and to the quantum computers (see [6, 10]).

In logic, it represents some kind of compatibility condition, when working with many logical sentences in the same time. Let us consider three logical sentences $p,\ q,\ r$. Let us suppose that if all of them are true, then the conclusion A could be drawn, and if $p,\ q,\ r$ are all false then the conclusion C can be drawn; in other cases, we say that the conclusion B is true. Modeling this situation by the map R, defined by $(p,\ q) \mapsto (p' = p \vee q,\ q' = p \wedge q)$, helps to comprise our analysis: we can apply R to the pair $(p,\ q)$, then to $(q',\ r)$, and, finally to $(p',\ q'')$. The Yang-Baxter equation explains that the order in which we start this analysis is not important; more explicitly, in this case, it states that $(\ (p')',\ q''',\ r') = (\ p',\ q''',\ (r')')$.

Another interpretation of the Yang-Baxter equation is related to the algorithms which order sequences of numbers (see, for example, a recent paper on sorting: [11]). For example, the core of the following algorithm is related to the left hand side of (2).

```
♯ include ⟨iostream⟩
int L,n,j,aux,i, sir[20],a,b;
int main()
{
std::cout ≪ "You may choose how many numbers will be compared";
std::cin ≫L;
int sir[L];
int sir2[L];
for (n=1;n≤L;n++)
{
std::cout≪"Please, give the numbers A["≪n≪"]=";
std::cin≫sir[n];
std::cout≪" "≪endl;
}
std::cout≪"We are now ordering the given numbers!";
std::cout≪"  "≪std::endl;
for (i=1;i≤L-1;i++)
for (j=i+1;j≤L;j++)
```

```
if (sir[i]≥sir[j])
{
aux=sir[i];
sir[i]=sir[j];
sir[j]=aux;
}
for (n=1;n≤L;n++)
std::cout≪ “  ”≪sir[n];
std::cout≪ “  ”≪std::endl;
system(“PAUSE”);
return EXIT–SUCCESS;
}
```

The following “Bubble sort” algorithm is related to the right hand side of (1).

```
  int m, aux;
m=L;
while (m)
{
for (int i=1; i≤L-1; i++)
if (a[L-i] ≥ a[L+1-i])
{
aux = a[L+1-i];
a[L+1-i] = a[L-i];
a[L-i] = aux;
}
m - -;
}
```

Ordering three numbers is related to the following common solution of the equations (1) and (2): $R(a,b) = (min(a,b), max(a,b))$

Since R can be extended to a braiding in a certain monoidal category, we obtain an interpretation for the case when we order more numbers.

The “divide et impera” algorithm for finding the maximum of sequence of numbers could be related to Yang-Baxter systems and to the gluing procedure from [12].

4 The Yang-Baxter Equation in Quantum Groups and Knot Theory

For A be a (unitary) associative k-algebra, and $\alpha, \beta, \gamma \in k$, the authors of [13] defined the k-linear map $R^A_{\alpha,\beta,\gamma} : A \otimes A \to A \otimes A$,

$$a \otimes b \mapsto \alpha ab \otimes 1 + \beta 1 \otimes ab - \gamma a \otimes b \quad (5)$$

which is a Yang-Baxter operator if and only if one of the following cases holds:
(i) $\alpha = \gamma \neq 0, \ \ \beta \neq 0$; (ii) $\beta = \gamma \neq 0, \ \ \alpha \neq 0$; (iii) $\alpha = \beta = 0, \ \ \gamma \neq 0$.

The link invariant associated to (5) is the Alexander polynomial of knots (cf. [14, 15]).

For $(L, [,])$ a Lie super-algebra over k, $z \in Z(L) = \{z \in L : [z, x] = 0 \ \ \forall \ x \in L\}$, $|z| = 0$ and $\alpha \in k$, the authors of the papers [16] and [17] defined the following Yang-Baxter operator: $\phi^L_\alpha \ : \ L \otimes L \ \longrightarrow \ L \otimes L$,

$$x \otimes y \mapsto \alpha [x, y] \otimes z + (-1)^{|x||y|} y \otimes x \ . \quad (6)$$

Remark 4.1 *In dimension two, $R^A_{\alpha,\beta,\alpha} \circ \tau$, can be expressed as:*

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1-q & q & 0 \\ \eta & 0 & 0 & -q \end{pmatrix} \quad (7)$$

where $\eta \in \{0, \ 1\}$, and $q \in k - \{0\}$. For $\eta = 0$ and $q = 1$, $R^A_{\alpha,\beta,\alpha}$ becomes:

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} \quad (8)$$

which is a universal gate (according to [10]), and it is related to the CNOT gate:

ISBN: 978-1-61804-290-3

$$CNOT = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix} \qquad (9)$$

Remark 4.2 *The matrix (8) can be interpreted as a sum of Yang-Baxter operators, using the techniques of [12].*

Remark 4.3 *Using Theorem 3.1 (i) and Remark 3.3 from [18], we can construct a bialgebra structure associated to the operator* $R^A_{\alpha,\beta,\gamma}(a \otimes b) = \alpha ab \otimes 1 + \beta 1 \otimes ab - \gamma a \otimes b$, *if one of the following cases holds: (i)* $\alpha = \gamma \neq 0, \ \beta \neq 0$; *(ii)* $\beta = \gamma \neq 0, \ \alpha \neq 0$; *(iii)* $\alpha = \beta = 0, \ \gamma \neq 0$.

For $\gamma = -1$ *and* $\alpha = \beta = 0$, *this is the tensor algebra* $T(A)$ *associated to the underlying vector space of the algebra* A.

5 Nonassociative Algebras

Jordan algebras emerged in the early thirties, and their applications are in physics, differential geometry, ring geometries, quantum groups, analysis, biology, etc (see [3, 19, 20]).

One of our main results is the following theorem, which explains when the Jordan identity implies associativity. It is an intrinsic result.

Theorem 5.1 *Let* V *be a vector space spanned by* a *and* b, *which are linearly independent. Let* $\theta : V \otimes V \to V, \ \ \theta(x \otimes y) = xy$, *be a linear map which is a commutative operation with the property*

$$a^2 = b \ , \quad b^2 = a \ . \qquad (10)$$

Then: (V, θ) *is a Jordan algebra* $\Longleftrightarrow$ (V, θ) *is a non-unital commutative (associative) algebra.*

The next remark finds a relationship between Jordan algebras, Lie algebras and associative algebras. In this case, we have an extrinsic result about non-associative structures.

Remark 5.2 *For the vector space* V, *let* $\eta : V \otimes V \to V, \ \ \eta(x \otimes y) = xy \ ,$ *be a linear map such that for any* $a, b, c \in V$ *we have:*

$$(ab)c + (bc)a + (ca)b \ = \ a(bc) + b(ca) + c(ab) \ ; \qquad (11)$$

$$(a^2 b)a \ = \ a^2(ba) \ . \qquad (12)$$

Then, (V, η) *is a structure which unifies (non-unital) associative algebras, Lie algebras and Jordan algebras.*

Indeed, the associativity and the Lie identity are unified by relation (11). Also, the commutativity of a Jordan algebra implies (11). But, the Jordan identity, (12), which appears in the definition of Jordan algebras, is verified in any associative algebra and Lie algebra.

6 Unification of Nonassociative Structures

The formulas (5) and (6) lead to the unification of associative algebras and Lie (super)algebras in the framework of Yang-Baxter structures (see [2, 21]). On the other hand, for the invertible elements in a Jordan algebra, one can associate a

ISBN: 978-1-61804-290-3

symmetric space (see [19], page 58), and, therefore, a Yang-Baxter operator. Thus, the Yang-Baxter equation can be thought as a unifying equation.

The first isomorfism theorem for groups (algebras) and the first isomorfism theorem for Lie algebras, can be unified as an isomorphism theorem for Yang-Baxter structures (see [25]).

Definition 6.1 *For the vector space V, let $\eta : V \otimes V \to V$, $\eta(a \otimes b) = ab$, be a linear map which satisfies:*

$$(ab)c+(bc)a+(ca)b = a(bc)+b(ca)+c(ab), \quad (13)$$

$$(a^2b)a \;=\; a^2(ba), \quad (14)$$

$\forall\ a, b, c \in V$.

Then, (V, η) is called a "UJLA structure".

Remark 6.2 *The UJLA structures unify Jordan algebras, Lie algebras and (non-unital) associative algebras; results for UJLA structures could be "decoded" in properties of Jordan algebras, Lie algebras or (non-unital) associative algebras.*

Remark 6.3 *An anti-commutative UJLA structure is a Lie algebra.*

Obviously, a commutative UJLA structure is a Jordan algebra.

Remark 6.4 *Let W be a vector space spanned by a and b, which are linearly independent. Let $\theta : W \otimes W \to W$, $\theta(x \otimes y) = xy$, be a linear map with the property: $a^2 = b$, $b^2 = a$.*

If θ satisfies also the relations (13) and (14), then $(W,\ \theta)$ is a (non-unital) associative algebra.

Remark 6.5 *If $(A,\ \theta)$, where $\theta : A \otimes A \to A$, $\theta(a \otimes b) = ab$, is a (non-unital) associative algebra, then we define $(A,\ \theta')$, where $\theta'(a \otimes b) = \alpha ab \ + \ \beta ba$.*

If $\alpha = \frac{1}{2}$ and $\beta = \frac{1}{2}$, then $(A,\ \theta')$ is a Jordan algebra.

If $\alpha = 1$ and $\beta = -1$, then $(A,\ \theta')$ is a Lie algebra.

If $\alpha = 0$ and $\beta = 1$, then $(A,\ \theta')$ is the opposite algebra of $(A,\ \theta)$.

Obviously, if $\alpha = 1$ and $\beta = 0$, then $(A,\ \theta')$ is the algebra $(A,\ \theta)$.

If we put no restrictions on α and β, then $(V,\ \theta)$ is a UJLA structure.

7 Transcendental Numbers

The following identities which contain the transcendental numbers e and π are well-known:

$$e^{i\pi} + 1 = 0, \quad (15)$$

$$\int_{-\infty}^{+\infty} e^{-x^2} dx = \sqrt{\pi}\ , \quad (16)$$

$$\int_{-\infty}^{+\infty} e^{-ix^2} dx = \sqrt{\frac{\pi}{2}}(1-i)\ . \quad (17)$$

Other inequalities for e and π are quite new:

$$|\ e^{1-z} + e^{\bar{z}}\ | > \pi \qquad \forall z \in \mathbb{C}\ , \quad (18)$$

$$\int_a^b e^{-x^2} dx \ <\ \frac{e^e}{\pi}(\frac{1}{e^{\pi a}} - \frac{1}{e^{\pi b}})\ .$$

Numerical / experimental results are very important in studying these new results. The use

ISBN: 978-1-61804-290-3

of TI graphing calculators could be the first step in this approach.

There exist real solutions for the equations $x^2-\pi x+(1+\frac{1}{r})^r=0$, $r\in\mathbb{Q}^*$, for r sufficiently small, but there are no real solutions for the "limit" equation $x^2-\pi x+e=0$, because $\Delta=\pi^2-4e<0$.

The question if $\Delta=\pi^2-4e$ is a transcendental number is an open problem! (Numerical / experimental results could give a partial answer for this problem.)

Resembling the problem of squaring the circle, the geometrical interpretation of the formula $\pi^2<4e$ could be stated as: "The length of the circle with diameter π is almost equal (and less) to the perimeter of a square with edges of length e". In this case, the area of the above circle is greater than the area of the above square, because $\pi^3>4e^2$.

OPEN PROBLEMS. For an arbitrary closed curve, we consider the smallest diameter (d) and the maximum diameter (D). (These can be found by considering the center of mass of a body which corresponds to the domain inside the given curve.)

(i) If L is the length of the given curve and the domain inside the given curve is a convex set, then we conjecture that:

$$\frac{L}{D}\leq\pi\leq\frac{L}{d}\quad .$$

(ii) Moreover, the first inequality becomes equality if and only if the second inequality becomes equality if and only if the given curve is a circle.

(iii) If the area of the domain inside the given curve is A, then $d\ D\ >A$.

(iv) The equation $x^2-\frac{L}{2}x+A=0$ *and its implications are not completely understood. For example, if the given curve is an ellipse, solving this equation in terms of the semi-axes of the ellipse is an unsolved problem.*

Remark 7.1 *Graphics for arbitrary closed convex curves related to the above open-problems could be represented using graphing calculators and computers. Thus, some numerical (experimental) results can be obtained. This direction seems to be a challenging one for computer scientists.*

Remark 7.2 *One could consider the equation* $x^i=i^x\qquad x\in\mathbb{R}_+^*$, *which is equivalent to* $e^{\frac{\pi}{2}}=x^{\frac{1}{x}}\quad x\in\mathbb{R}_+^*$, *and it has no real solution, because* $\frac{\pi}{2}>\frac{1}{e}$.

At this moment we do not have convincing numerical / experimental results for the following generalization of the above equation.

$$z^i=i^z\qquad z\in\mathbb{C}^*$$

This is work in progress. (We tried to solve it with MathLab.)

OPEN PROBLEM. Prove the inequality: $\sum_1^n\frac{1}{k^2}<\frac{2}{3}\ (\frac{n+1}{n})^n\quad\forall n\in\mathbb{N}^*$.

8 Transcendental Numbers in Mathematical Physics

The Yang-Baxter equation first appeared in theoretical physics, in a paper by the Nobel laureate C.N. Yang, and in statistical mechanics, in

ISBN: 978-1-61804-290-3

R.J. Baxter's work. It has applications in many areas of physics, informatics and mathematics. Many scientists have used computer calculations or the axioms of various algebraic structures in order to solve this equation, but the full classification of its solutions remains an open problem (see [22, 23, 24, 25]). Below, we describe its connection with transcendental numbers.

Let V be a complex vector space, and $I_j : V^{\otimes j} \to V^{\otimes j} \quad \forall j \in \{1,2\}$ identity maps. We consider $J : V^{\otimes 2} \to V^{\otimes 2}$ a linear map which satisfies $J \circ J = -I_2$ and $J^{12} \circ J^{23} = J^{23} \circ J^{12}$, where $J^{12} = J \otimes I_1$, $J^{23} = I_1 \otimes J$.

$R(x) = \cos x I_2 + \sin x J$ satisfies the colored Yang-Baxter equation:

$$R^{12}(x) \circ R^{23}(x+y) \circ R^{12}(y) = = R^{23}(y) \circ R^{12}(x+y) \circ R^{23}(x) \ . \qquad (19)$$

The proof of (19) could be done by writing $R(x) = e^{x\ J}$, and checking that (19) reduces to $x\ J^{12} + (x+y)\ J^{23} + yJ^{12} = y\ J^{23} + (x+y)\ J^{12} + xJ^{23}$.

Such an operator J could have, in dimension two, the following matrix form (for $\alpha \in \mathbb{R}$):

$$\begin{pmatrix} 0 & 0 & 0 & \frac{1}{\alpha} i \\ 0 & 0 & i & 0 \\ 0 & i & 0 & 0 \\ \alpha i & 0 & 0 & 0 \end{pmatrix} \qquad (20)$$

Based on results from the previous section, a counterpart for the formula

$$e^{\pi\ J} + I_4 = 0_4 \qquad J,\ I_4, 0_4 \in \mathcal{M}_4(\mathbb{C})$$

could be the following inequality:

$$X^2 + eI_2 > \pi X \ , \qquad (21)$$

$$\forall\ X \in \mathcal{M}_2(\mathbb{R}_+^*),\ trace(X) > \pi \ .$$

9 Conclusions

Many ideas of the current paper emerged after the International Conference "Mathematics Days in Sofia", July 7-10, 2014, Sofia, Bulgaria.

Dr. Violeta Ivanova ([26]) was interested in the applications of these problems in informatics. The Yang-Baxter equation can be interpreted in terms of combinatorial logical circuits, and, in logic, it represents some kind of compatibility condition, when working with many logical sentences in the same time. This equation is also related to the theory of universal quantum gates and to the quantum computers (see, for example, [6]). The first quantum computer (which uses principles of quantum mechanics) was sold to the aerospace and security of defense company Lockheed Martin. It can address issues related to number theory and optimization, which require large computational power. An example is the Shor's algorithm, a quantum algorithm that determines quickly and effectively the prime factors of a large number. With enough qubii, such a computer could use the Shor's algorithm to break algorithms encryption used today.

An explanation of the fact that the study of Jordan structures and their applications is at present a wide-ranging field of mathematical research could be the following: at the beginning, mathematics was associative and commutative, then (after the invention of matrices) it became associative and non-commutative, and now (after the invention of non-associative structures) it becomes non-associative and non-commutative (see [20]).

Our talk will follow results from [28, 29]. Then, it will present other applications in informatics and system theory, and implications in art. Also, we will refer to the work of math-

ISBN: 978-1-61804-290-3

ematicians from Barcelona (see [8]), and other scientists from Spain (see [30]), France (see [31]), Hungary, Romania (see, for example, [32, 33]), etc.

Acknowledgment

The authors would like to thank the Simion Stoilow Institute of Mathematics of the Romanian Academy.

References

[1] Wills-Toro, L. A. *Classification of some graded not necessarily associative division algebras I.* Communications in Algebra 2014, 42: 5019-5049.

[2] Iordanescu, R. *The associativity in present mathematics and present physics.* Presentation, Bucharest, 2014.

[3] Iordanescu, R. *Jordan structures in mathematics and physics.* Mathematics 2011, http://arxiv.org/abs/1106.4415.

[4] Iordanescu, R.; Nichita, F.F.; Nichita, I.M. *Non-associative algebras, Yang-Baxter equations and quantum computers.* Bulgarian Journal of Physics **2014**, vol.41 n.2, 71-76.

[5] Nichita, F.F. *On Transcendental Numbers.* Axioms 2014, 3, 64-69.

[6] Alagic, G.; Bapat, A.; Jordan, S. *Classical simulation of Yang-Baxter gates.* Mathematics 2014 http://arxiv.org/abs/1407.1361.

[7] T. Gateva-Ivanova (2014) *Quadratic algebras, Yang-Baxter equation,and Artin-Schelter regularity*, presentation - Sofia.

[8] David Bachiller, Ferran Cedo, Eric Jespers *Solutions of the Yang-Baxter equation associated with a left brace*, arXiv:1503.02814.

[9] Nichita, F.F.; Nichita, I.M. *Some Problems On Combinational Logical Circuits.* Acta Universitatis Apulensis 2002, 4, 139-144.

[10] Kauffman, L.H.; Lomonaco, S.J. *Braiding Operators are Universal Quantum Gates.* New Journal of Physics 2004, Volume 6, 134.

[11] Adjeroh, D.; Nan, F. *Suffix-Sorting via Shannon-Fano-Elias Codes.* Algorithms 2010, 3, 145-167.

[12] T. Brzezinski, T.; Nichita, F.F. *Yang-Baxter systems and entwined structures*, Communications in Algebra 2005, vol. 33(4), 1083-1093.

[13] S. Dăscălescu, F. F. Nichita, *Yang-Baxter operators arising from (co)algebra structures.* Comm. Algebra 1999, 27, 5833–5845.

[14] Turaev, V., *The Yang-Baxter equation and invariants of links*, Invent. Math. 1988 92 527-553.

[15] Massuyeau, G.; Nichita, F.F. *Yang-Baxter operators arising from algebra structures and the Alexander polynomial of knots*, Comm. Algebra 2005, 33 (7) 2375-2385.

[16] Majid S. *Solutions of the Yang-Baxter equation from braided-Lie algebras and braided groups.* J. Knot Theory and Its Ramifications 1995, 4, 673-697.

[17] Nichita, F.F.; Popovici, B.P. *Yang-Baxter operators from (G, θ)-Lie algebras.* Romanian Reports in Physics 2011, 63(3), 641-650.

[18] Nichita, F.F. *Yang-Baxter systems, algebra factorizations and braided categories.* Axioms 2013, 2(3), 437-442.

[19] Iordanescu, R. *Jordan structures in geometry and physics with an Appendix on Jordan structures in analysis*, Romanian Academy Press, 2003.

[20] Iordanescu, R. *Romanian contributions to the study of Jordan structures and their applications*, Mitteilungen des Humboldt-Clubs Rumanien 2004-2005, No. 8-9, Bukarest, 29-35.

[21] Nichita, F.F. *Lie algebras and Yang-Baxter equations.* Bulletin of the Transilvania University of Brasov, Series III: Mathematics, Informatics, Physics **2012**, 5(54), 195-208.

[22] Nichita, F.F. (Editor), *Special Issue "Hopf Algebras, Quantum Groups and Yang-Baxter Equations 2014"*, Axioms Open Access Journal, http://www.mdpi.com/journal/axioms/special_issues/hopf_algebras_2014.

[23] B. Abdesselam, A. Chakrabarti, V. K. Dobrev, S. G. Mihov, *Exotic bialgebras from 9 9 unitary braid matrices*, Physics of Atomic Nuclei 74, 2011, 824-831.

[24] Wang, G., Xue, K., Sun, C., Du, G. *Yang-Baxter R matrix, Entanglement and Yangian*, arXiv: 1012.1519.

[25] Nichita, F.F. *Non-linear Equations, Quantum Groups and Duality Theorems*, VDM Verlag, 2009.

[26] V. N. Ivanova (2014) *Approaches to the parallelization of data mining algorithms with the aim of improving the accuracy*, presentation - Sofia.

[27] Nichita, F.F. *Introduction to the Yang-Baxter Equation with Open Problems.* Axioms 2012, 1(1), 33-37.

[28] Iordanescu, R.; Nichita, F.F.; Nichita, I.M. *The Yang-Baxter Equation, (Quantum) Computers and Unifying Theories.* Axioms 2014, 3, 360-368.

[29] Solomon Marcus, Florin F. Nichita, *On transcendental numbers: new results and a little history*, arXiv:1502.05637.

[30] Samuel G. Moreno, *A One-Sentence and Truly Elementary Proof of the Basel Problem*, arXiv:1502.07667 [math.HO].

[31] Filippo Bonchi and Fabio Zanasi, *Bialgebraic Semantics for Logic Programming*, http://arxiv.org/abs/1502.06095; Indexed Feb 21, 2015.

[32] Nicolescu, B. *Manifesto of Transdisciplinarity*, State University of New York (SUNY) Press, New York, 2002 , translation in English by Karen-Claire Voss.

[33] Florin F. Nichita, *On Models for Transdisciplinarity*, Transdisciplinary Journal of Engineering and Science, Vol. 2011, 42-46.

ISBN: 978-1-61804-290-3

Randomized poly-encrypted image exploiting chaotic beheviour

Bouslehi Hamdi
Université des Sciences et Techniques de Tunis
Email: hamdouchb@gmail.com

Seddik Hassen
Université des Sciences et Techniques de Tunis
Email: seddikhassene@yahoo.fr

Amaria Wael
Université des Sciences et Techniques de Tunis
Email: amaria.wael@hotmail.fr

Ezzedine Ben Braiek
Université des Sciences et Techniques de Tunis
Email: ebenbraiek@yahoo.fr

***Abstract*—The security of digital data has became an essential need. Since the storage, transmission of data social networking has became inevitable, the need to ensure that data is no longer a luxury but an absolute necessity. The picture takes a large part of the data and its presence every day is more important. In this document, we used a new encryption technique: The poly encryption, this technique is based on the Decomposition of the image in a random manner, and each block will be encrypted by an algorithm that uses different algorithm.**

Keywords—poly-encryptiont; chaos; permutation; random; block

I. INTRODUCTION:

Along with the fast progression of data exchange in electronic way, it is important to protect the confidentiality of the data from unauthorized access. Security breaches may affect user's privacy and reputation. So, data encryption is widely used to ensure security in open networks such as Internet. Due to the substantial increase in digital data transmission via internet, the security of digital images has become more prominent and attracted many attentions in the digital world today. Also, the extension of multimedia technology in our society has promoted digital images to play a more significant role than the traditional texts, which demand serious protection of users' privacy for all applications. Each type of data has its own features; therefore, different techniques should be used to protect confidential image data from unauthorized access. Most of the available encryption algorithms are used for text data. However, due to the large data size and real time requirement, the algorithms that are appropriate for textual data may not be suitable for multimedia data.

Classical cryptographic algorithms such as RSA, DES,… are inefficient for image encryption due to inherent features, especially high volume image data. Many researchers proposed different image encryption schemes to overcome image encryption problems [1, 2, 3, and 4].

In this research we tried to find a simple, fast and secure algorithm for image encryption using a random permutation function and the characteristics of chaotic functions in changing pixel intensity and in encryption by using a logic operator . According to key's large space in the chaotic functions, this method is very robust. Finally, this algorithm is very sensitive to small changes so even with the knowledge of the key approximate values; there is no possibility for the attacker to break the cipher.

II. THE LOGISTIC FUNCTION:

The logistic map [5] is a basic mapping polynomial, which has chaotic behavior, and it can be obtained by a very simple nonlinear dynamical equation [6].

Recurrence logistics is an example where the recurrence is not linear. This recurrence was popularized by the biologist Robert May in 1976. Its recurrence relation is.

$$\tau(x_n) = x_{n+1} = \lambda\, x_n (1 - x_n) \qquad (1)$$

The control parameter "λ" is fixed and chosen so that equation (1) has a chaotic behavior $(3.57 < \lambda < 4)$ [7]. However, if we study the map with a different value of "λ", it shows that it is a trigger for the chaos. Mathematically, the "Logistic map" is written with

"x" is a number between 0 and 1, and represents the initial condition $0 < x_0 < 1$.

"λ" is a positive number [8].

III. ENCRYPTION ALGORITHM BASED ON ITERATING THE IMPROVED "LOGISTIC MAP":

Once the "Logistic map" function is iterated N times, we obtain N value x_n between 0 and 1 and x_0: the initial value and $0 < x_0 < 1$ and λ: control parameter.

In our encryption algorithm we took x_0=0.4582 [0, 1] and λ =3.759889 [3.57, 4], we obtained a chaotic signal and the value generated chaotic sweep the entire range of value between 0 and 1. After 70 iterations, the signal from equation 1 is summarized in (figure 1).

Indeed, the chaotic function "Logistic Map" has several properties, such as frequency and sensitivity to initial conditions (this is a characteristic of all chaotic systems: if we take a different value which is very close to x_0 then the values from the iteration change completely. If we take a different value which is very close to the values of "λ "the iteration changes dramatically: this can be seen by a simulation tool Matlab, in particular by the value of these

functions which are completely random. Although they are limited from a few bands, the iterative values never give the impression to converge even after an infinite number of iterations. The change of control parameters (λ) and the initial condition (x_0) by very close values in order to know the decryption algorithm always gives cryptograms so radically different that it is interesting to use the function in logistic encryption).

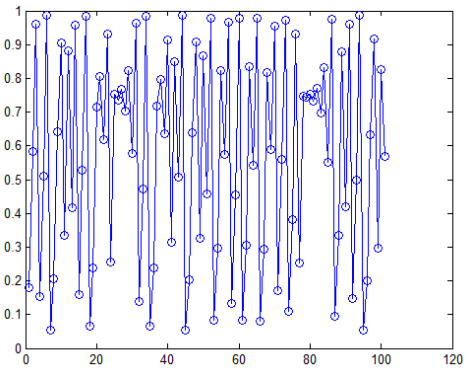


Fig. 1. The logistic map outputs after 100 iterations.

IV. PROPOSED ENCRYPTION APPROACH:

This approach is consisted to decomposed our image to many blocks and decrypt them with different algorithms, in this document we decomposed our image into four blocks encrypted with the following functions : a random function of vector permutation, a random function of diagonal permutation with change pixels intensity based on logistic function and the encryption by logical function "XOR" with using the chaotic function, all these algorithms will be used in the same image, but each algorithm will be used to encrypt one block of the image decomposed in advance with a random manner.

Fig. 2. Original image and its histogram

A. *First block*

The first block is encrypted with the random function of permutation vector combined with the function of change pixels intensity.

- **Permutation function**

Mathematically our permutation function P can be defined by the following equation: $\forall x \in N$ such as p(x) = y so $\exists\ P^{-1}(y) = x$ with $P \circ P^{-1}(x) = I$ and $P^{-1}(y) - x = \varepsilon$ with $\varepsilon \to 0$.

The permutation has a paramount importance in this approach. It is used to make the system more safe and reliable. An image having pixels swapped does not make sense and becomes misunderstood.

We suppose "I" the image to encrypt and L the permutation key that can be represented by string or decimal value chain. If our key is string we convert it to ASCII as follows:

ASCII: X → Y with X: string and Y: vector of integers l = ASCII (l_1) with $l_1 \in$ X and Y ∈ l.

Our function contains "n" round, each round uses a key l_i where n is the key length. The permutation function is applied iteratively in each round with a different key in each step.

First round. The process of permutation is partitioned into several rounds. The round number corresponds to the length of the key

$$l = \{l_1, l_2, \ldots\ldots\ldots\ldots, l_n\} \quad (1)$$

l is a key string which contains a set of different elements l_i In each iteration an independent key, different from the previous, is introduced. With $l_i \in \mathbb{N}^*$, i∈ [1, n] and "n" is a variable number representing the length of the encryption key string. So we start by the decomposition of our image into a set of vectors whose length is equal to the value of l_1 as presented by the following equation:

$$I = \sum_{i=1}^{n} C_i^{l_1} \quad (2)$$

I: is the original image,

$C_1^{l_1}$: The first vector of the decomposed image I of length l_1. After the image decomposition, we permute each bloc by using the permuted function described as follows:

$$P(C_1^{l_1} = \sum_{i=1}^{l_1} b_i) = C_{P1}^{l_1} = \sum_{i=0}^{l_1} b_{(l_1 - i)} \quad (3)$$

Where $C_{P1}^{l_1}$ is the first vector from the image I after permutation.

$$P(C_2^{l_1} = \sum_{i=1}^{l_1} b_i) = C_{P2}^{l_1} = \sum_{i=0}^{l_1} b_{(l_1 - i)} \quad (4)$$

$C_{P2}^{L_1}$ is The 2nd extracted vector from the image I after permutation. Finaly we have an image permutated l_1 times described as follows:

ISBN: 978-1-61804-290-3

$$P_{l_1}(I) = I_{(P,l_1)} = \sum_{i=1}^{n} C_{Pi}^{l_1} \quad (5)$$

Where $P(I) = I_{(P,l_1)}$: Image after the first permutation.

Second round: $I_{(P,l_1)}$: is the output of the first round and the input of the second round which start by a new decomposition of the image permuted by l_1 into many blocs whose length are equal to the value of l_2 as mentioned in the following equation

$$I_{(P,l_1)} = C_{P1}^{l_1} C_{P2}^{l_1} \ldots\ldots C_{Pn}^{l_1} = C_1^{l_2} C_2^{l_2} \ldots\ldots C_n^{l_2} \quad (6)$$

Once the decomposition is done, we permute each vector by using the permuted function as presented by the following equation:

$$P(C_1^{l_2} = \sum_{i=1}^{l_2} b_i) = C_{P1}^{l_2} = \sum_{i=0}^{l_2} b_{(l_2 - i)} \quad (7)$$

$$P(C_2^{l_1} = \sum_{i=1}^{l_1} b_i) = C_{P2}^{l_1} = \sum_{i=0}^{l_1} b_{(l_1 - i)}$$

. (8)

.

$$P_{l_1}(I) = I_{(P,l_1)} = \sum_{i=1}^{n} C_{Pi}^{l_1}$$

To obtain finally we have an image permutated by C_2 :

$$P_{l_2}(I) = I_{(P,l_2)} = C_{P1}^{l_2} C_{P2}^{l_2} \ldots\ldots\ldots\ldots C_{Pn}^{l_2} \quad (9)$$

The final output of this function iterations is a matrix permuted in a random manner so there is not a mathematical model that represents a relationship between permuted image I (P, l_n) to original image I .

Therefore, we can resume this permutation function by the following figure:

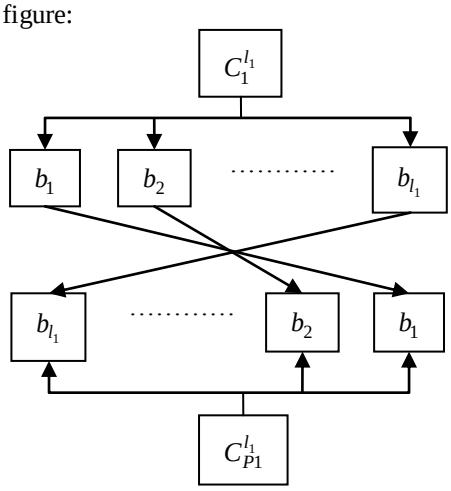


Fig. 3. Permutation function

1) ***Complexity study of the permutation function.***

As illustrated by the previous equations, the permutation function introduces a random behavior to the pixels position in the image matrix. To illustrate this random iterative permutation, the trajectory of three different pixels located in the following spatial coordinates [(87; 45), (48; 154), (156; 81)] are plotted in each iteration step and illustrated by the following figure:

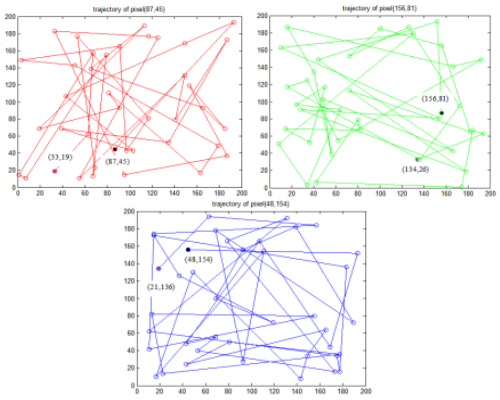


Fig. 4. Different random trajectories of three different pixels belonging to the image map

As presented in the figure above, a random trajectories corresponding to three permuted pixels is illustrated. It clear that the three trajectories are completely different and random. Using this function each pixel is changed over the image matrix by an individual random trajectory. Finally we can affirm that the permutation function is random and differs iteratively for each pixel of the image.

- **Changing pixels intensites:**

After permutation performed on the message to be encrypted, which is in our case an image, we change the intensity to complicate the task. This procedure is to change the intensity of all image pixels by reducing or adding a random value between 0 and *d*/ 2, generated by a logistic function "*F*".

With $x_0 = 0.4777$ and $\lambda = 3.598895$

"d" denotes the image dynamic, d = 256 for the case of images coded on 8 bits.

$$F = \sum_{i=1}^{n} d_i \quad (11)$$

With d_1 : a binary value

$$F = \sum_{i=1}^{k} D_i \quad (12)$$

With $D_1 = d_1 d_2 \ldots . d_7$: the first 7 binary values

ISBN: 978-1-61804-290-3

$I(i, j)$ Is the pixel coordinate (i, j) avec $i \in [1...h]$ and with $j \in [1...l]$ such that h is the number of line and l is the number of columns.

And $k = l * k$

If $I(i, j) < d/2$ then $I_c(i, j) = I(i, j) + V_k$ (13)

else $I_c(i, j) = I(i, j) - V_k$ (14)

V_k : The decimal value of D_k from the logistic function.

$$V_k = \sum_{i=1}^{n} 2^{i-(k(i-1))} \times d_{i-(k(i-1))} \quad (15)$$

"V_k" is a random value belong to the interval [0, d/2].

Fig. 5. Original block and the correspondent encrypted one

B. Second block

The second block will be encrypted with the XOR function combined with changing pixel intensity.

- **Encryption process**

For a more secure encryption system, we try to complicate the task. Each pixel intensity obtained after the previous steps "position and intensities variation" is modified by the XOR function adding the obtained scrambled matrix to a random values generated by the logistic function.

The XOR logic operator is a common function used in many complex ciphers techniques. Its primary merit is its simplicity to implement and non inexpensive computational time. A simple XOR operator is therefore sometimes used for hiding information in cases where no particular security is required. The use of the XOR logic operators between image pixels converted to binary and binary value of the logistic function is detailed by the following equations:

- Image binarisation

$$I_{pc} = \begin{pmatrix} i_{(1,1)}\, i_{(1,2)} \cdots\cdots i_{(1,M)} \\ i_{(2,1)} \cdots\cdots\cdots i_{(2,M)} \\ \cdots\cdots\cdots\cdots\cdots \\ \cdots\cdots\cdots\cdots\cdots \\ i_{(N,1)} \cdots\cdots\cdots i_{(N,M)} \end{pmatrix} \quad (16)$$

So I_{PC} is the image after permutation and change intensity will be converted to binary with $M = h*8$ and $N = w*8$

- The binarisation procedure is described by the following equations:

The output of the logistic function is binirized as described by the following equation to obtain a binary sequence F:
The value of x is represented by:

$$x = 0.a_1(x)a_2(x).....a_i(x) \quad (17)$$

With $x \in [0,1] et a_i \in \{0,1\}$

$$a_i = \sum_{r=1}^{2^i - 1} (-1)^{r-1} \beta_{(r/2^i)}(x) \quad (18)$$

$a_i(x)$: is the ith bit

$$\beta_t = \begin{cases} 0, & six < t \\ 1, & six \geq t \end{cases} \quad (19)$$

β_t : is the function increases

$$F = \begin{pmatrix} a_{(1,1)}\, a_{(1,2)} \cdots\cdots a_{(1,M)} \\ a_{(2,1)} \cdots\cdots\cdots a_{(2,M)} \\ \cdots\cdots\cdots\cdots\cdots \\ \cdots\cdots\cdots\cdots\cdots \\ a_{(N,1)} \cdots\cdots\cdots a_{(N,M)} \end{pmatrix} \quad (20)$$

$$I_{cry} = \begin{pmatrix} i_{(1,1)}\, i_{(1,2)} \cdots\cdots i_{(1,M)} \\ i_{(2,1)} \cdots\cdots\cdots i_{(2,M)} \\ \cdots\cdots\cdots\cdots\cdots \\ \cdots\cdots\cdots\cdots\cdots \\ i_{(N,1)} \cdots\cdots\cdots i_{(N,M)} \end{pmatrix} \oplus \begin{pmatrix} a_{(1,1)}\, a_{(1,2)} \cdots\cdots a_{(1,M)} \\ a_{(2,1)} \cdots\cdots\cdots a_{(2,M)} \\ \cdots\cdots\cdots\cdots\cdots \\ \cdots\cdots\cdots\cdots\cdots \\ a_{(N,1)} \cdots\cdots\cdots a_{(N,M)} \end{pmatrix} \quad (21)$$

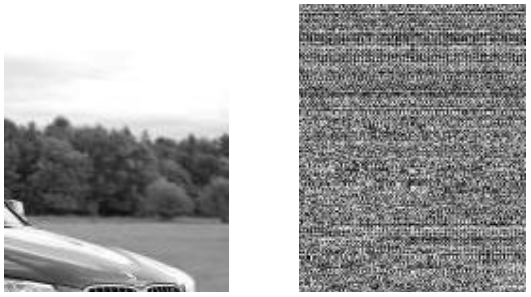


Fig. 6. Original block and the correspondent encrypted one

C. Third block

The third block will be encrypted by the random function of diagonal permutation combined with change pixel intensity.

ISBN: 978-1-61804-290-3

The obtained image is processed to break the correlation between its three channels. We pretend that in color image this step is mandatory. In fact to avoid any possibility to recover the content of the image, we have to disperse the relationship between pixels belonging to the three channels. This relationship is the base of the forms and content generation.

TABLE I. IMAGE MATRIX

P(1,1)	P(1,2)	P(1,3)	…	P(1,w)
P(2,1)	P(2,2)	P(2,3)	…	P(2,w)
P(3,1)	P(3,2)	P(3,3)	…	P(3,w)
…	…	…	…	…
P(h,1)	P(h,2)	P(h,3)	…	P(h,w)

After encrypting the three matrices corresponding to the red, green and blue, we recover a new image M (i,j) which has h lines and w columns.

A transposition function noted P_d "Pd (M) $\rightarrow MP$" inter-channels is developed to break the correlation between the three image channels. Each diagonal line of the resulting image is processed by performing cyclic translations. The number of translation is randomly generated using a logistic function.

For example:

When i = 1 we take only the first pixel

When $i = 2$ we take the permuted diagonal line that contains the two pixels M (2, 1) and M (1, 2) that are transposed using the following function:

$$P_s^n(M_{2,1}, M_{1,2}) \tag{19}$$

Such as n: the number of permutations that will be made and " s" is the permutation way (from right to left or left to right). If the index i = k we take the permuted diagonal line that contains the following pixels *(M (k, 1), M (k−1, 2), ...M (1, k))* to be transposed as follows:

$$P_s^n(\sum_{i=0}^{K} M_{((K-i),(i+1))}) \tag{20}$$

Fig. 7. Original block and the correspondent encrypted one

D. Fourth block

The fourth block will be encrypted by using all functions using in the previous blocks (permutation function, change pixel intensity and encryption with logic function).

encryption algorithm of the fourth block: to encrypt this block we choose to use the all algorithm that we used before in the same time. So the encryption's algorithm of the fourth block can be present by the following diagram.

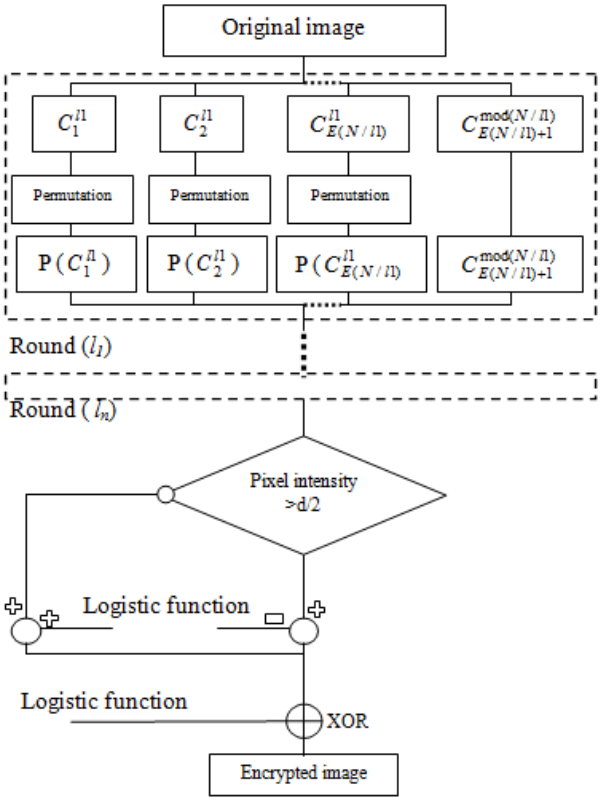


Fig. 8. Diagram of a new encryption approach

The following figure can show as the fourth block and its encryption block correspondent.

Fig. 9. Original block and the correspondent encrypted one

Now we can combine all these blocks in the same image.

ISBN: 978-1-61804-290-3

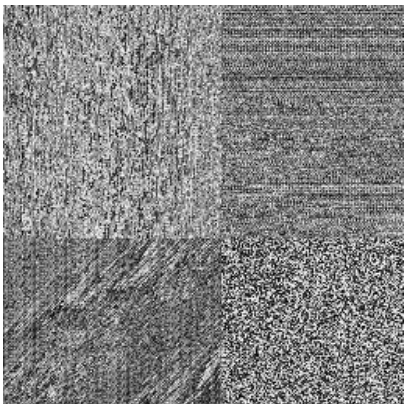


Fig. 10. Encrypted image

The security operation does not stop here we will also use a simple technique but very efficient, we'll split the image into *N* blocks and we going to swapped them randomly by the following permutation function:

- *Permutation*

The disorder is crucial in cryptography. It is used to remove the meaning of the image. The permutation is divided in few steps: first of all the plain image is divided in many blocks and we divided them in tow symmetrical sets. The output of the permutation process is the mixture of these two sets. The blocks sequence is modified in such a way that the image will be scrambled. we take the first block of each set and put them one behind the other in the output image. In the same way, we continue this gait until the end of the plain image.

$$M = \{B_1, B_2, \ldots\ldots\ldots\ldots B_n\} \quad (5)$$

$$P(B) = \{B_{(n/2)+1}, B_1, B_{(n/2)+2}, B_2, \ldots, B_n, B_{(n/2)}\} \quad (6)$$

$$P(B) = \{P_1, P_2, \ldots\ldots\ldots\ldots P_n\} \quad (7)$$

Where "*M*" is the plain image, "n" is the number of the blocks after reshaping and "P" is the permutation function which returns a set of integers put into a row.

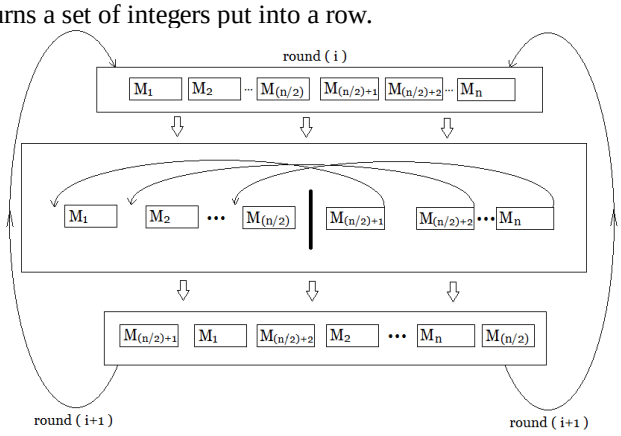


Fig. 11. Permutation diagram

We applied this technique on a (20×20) images. The following figure shows the different positions in the image taken by 2 randomly selected pixels. The pixels trajectories are random, and illustrated by white dots while red and green circle represent respectively the coordinates of departure and arrival of the pixel trajectory. We see in the figure below that the trajectory of a pixel is entirely random. There is no correlation between the different pixels; each of them has its own trajectory.

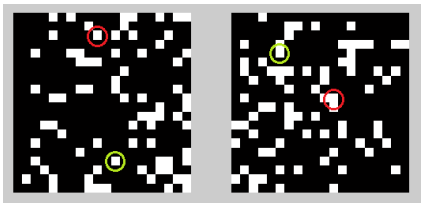


Fig. 12. Trajectory of two permuted different pixels

Finally and after the permutation function we can see the following image ciphered by block and permuted:

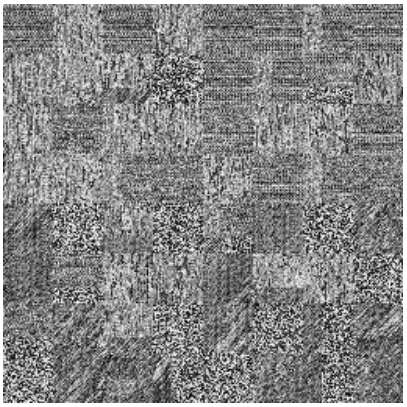


Fig. 13. Encrypted and permuted image and its histogram

V. IMAGE DECRYPTION STEPS:

A. *Decryption function*

Once the image is encrypted and transmitted just go the opposite way to recover the original image. The last procedure performed on the image during encryption is the XOR function, which has the original image as input whose pixels are switched and having intensities modified and logistics values. This procedure will be the first to perform during decryption by introducing the same parameters that are sent as "key".

$$I = I_{cry} \oplus F \quad (20)$$

B. *Changing intensity:*

Original intensity of the image are modified, therefore they must be restored by adding (or subtracting) the same value subtracted (or added) during the process of changing intensity. As indicated previously the choice of the operator "+" or "-" is stored in a vector that is sent as a key.

im_b: is a binary image formed during encryption process.

If im_b (i ,j)=0 then $I(i, j) = I_c(i, j) - V_k$ (21)

ISBN: 978-1-61804-290-3

Else $$I(i,j) = I_c(i,j) + V_k \quad (22)$$

V_k : The value decimal derived from the logistic function.

$$V_k = \sum_{i=1}^{n} 2^{i-(k(i-1))} \times d_{i-(k(i-1))} \quad (23)$$

C. Permutation:

The last operation to be done to recover the original image is permutation one more time.

$$P_{l_n}^{-1}(I_{(P,l_n)}) = P_{l_n}^{-1}(C_{P_1}^{l_n} C_{P_2}^{l_n} \ldots\ldots\ldots\ldots C_{P_n}^{l_n}) \quad (24)$$

$$P_{l_{(n-1)}}^{-1}(I_{(P,l_{(n\_1)})}) = P_{l_{(n-1)}}^{-1}(C_{P1}^{l_{n-1}} C_{P2}^{l_{n-1}} \ldots\ldots C_{Pn}^{l_{n-1}}) = I_{(P,l_{n-2})} \quad (25)$$

.

.

$$I = C_1^{l_1} C_2^{l_1} \ldots\ldots\ldots\ldots C_n^{l_1}$$

Once completed, the original image is covered. For a better interpretation of the results, we calculate the difference between the original image and the image decrypted. Losses for this approach is zero, no data is lost.

Fig. 14. decrypted image

VI. Results and Interpretation:

Evaluation tools are used to evaluate the encryption performance so we must quantify its performance and characteristics.

A. Mean square error (MSE):

We use to quantify the error between the original sequence and the encrypted

$$MSE = \frac{\sum_{i=1}^{n} (I_i - I_i^*)}{n} \quad (26)$$

Such as *n* is the length of the sequence.
*I** and *I* represent respectively the original image and the encrypted image.

B. Peak Signal Noise Ratio (PSNR):

It is a function derived from the MSE, it allows complete degradation of image, and it measures offset (in dB) of the original image by contributing to the encrypted image.

$$PSNR = 10Log_{10}\left(\frac{255^2}{MSE}\right) \quad (27)$$

Result of simulation:
PSNR = Inf

C. Similitude rates between adjacent pixels:

For an ordinary image each pixel is highly correlated with its adjacent pixels in the horizontal or vertical direction. An efficient encryption algorithm should break this correlation and disperse it over the whole matrix. This dispersion proves that the forms and content of the image is no longer visible. The following figures show the correlation between the tested horizontally or vertically adjacent pixels of the image.

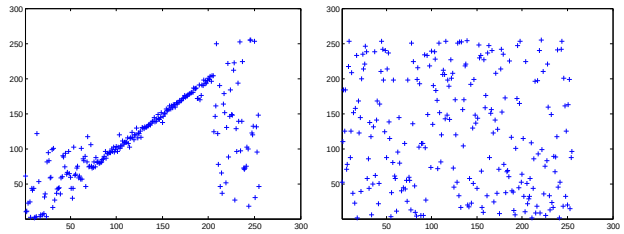


Fig. 15. Horizontal pixels correlation of our image before and after encryption

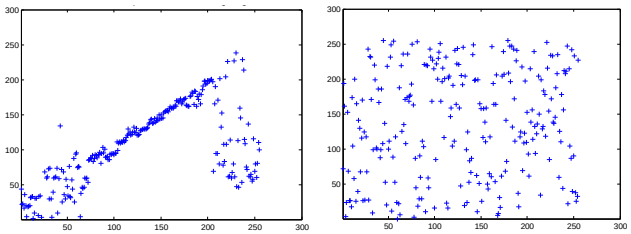


Fig. 16. Vertical pixels correlation of the processed image before and after encryption

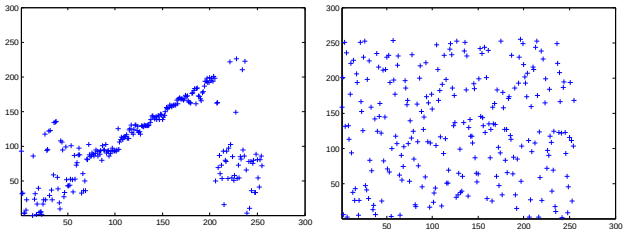


Fig. 17. Diagonal pixels correlation of the processed image before and after encryption

It is clear that the correlation between adjacent pixels is dispersed in the image by the encryption process. That means that our algorithm is efficient because the pixels relation is broken and no relation can be established between them to identify the algorithm structure or image content

ISBN: 978-1-61804-290-3

D. Similitude rates between encrypted and original image

A correlation is a number between 0 and 1 which measures the degree of association between two signals. A positive value for the correlation implies a positive association. A negative value for the correlation implies a negative association. In our case we studied the correlation between the original signal and the signal decoded (see figure (2) and figure (6)).

The correlation between the original image and decryption image is equal to 1 deciphered. So it can be concluded from results obtained in this code the correlation between the transmitted image and the received image is perfect, which means that our algorithm is reversible.

E. Test by subtraction:

You can also test our signal by calculating deference between the original signal and the signal is decrypted gives us the following result:

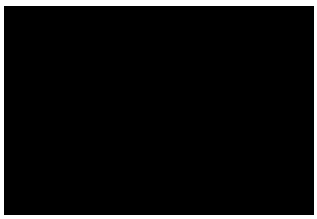


Fig. 18. Subtraction result

The previous image has showed a high similarity between the two image (original, decrypted) since the difference between them is 0 so we haven't any loss of information , this prove that our algorithm reversible.

VII. CONCLUSION

In this document we presented a new approach to encryption that contain all the criterion of an encryption algorithm that is robust: the randomness of the function used which is provided by the chaotic function and the function of the permutation, so that complexity of the key in our case is ensured by the choice of key (value integer or character string), although the choice of encryption methods have to be conservative and reversible, which is the case in our work.

REFERENCES

[1] M. Sharma and M.K. Kowar, "Image Encryption Techniques Using Chaotic Schemes: a Review," International Journal of Engineering Science and Technology, vol. 2, no. 6, 2010, pp. 2359–2363.

[2] A. Jolfaei and A. Mirghadri, "An Applied Imagery Encryption Algorithm Based on Shuffling and Baker's Map, "Proceedings of the 2010 International Conference on Artificial Intelligence and Pattern Recognition (AIPR-10), Florida, USA, 2010, pp. 279–285.

[3] A. Jolfaei and A. Mirghadri, "A Novel Image Encryption Scheme Using Pixel Shuffler and A5/1,"Proceedings of The 2010 International Conference on Artificial Intelligence and Computational Intelligence (AICI10), Sanya, China, 2010.

[4] L. Xiangdong, Z. Junxing, Z. Jinhai, and H. Xiqin, "Image Scrambling Algorithm Based on Chaos Theory and Sorting Transformation, "IJCSNS International Journal of Computer Science and Network Security, vol. 8, no. 1, 2008, pp. 64–68.

[5] Lin, C., Kang, J., Han, D., Tian, D., Wang, W., Zhang, J., et al. (2003). Electrical properties of Al2O3 gate dielectrics. Microelectronic Engineering, 66, 830-834

[6] Hirtle, P. B. (2008, July-August). Copyright renewal, copyright restoration, and the difficulty of determining copyright status. D-Lib Magazine, 14(7), 14-23. doi:10.1045/july2008-hirtle

[7] Roosevelt, F. (1997). Childhood acquisition of Pig Latin by native speakers of English. Manuscript submitted for publication.

[8] Nala, A. (1998). Teaching vocabulary: Evidence from research in Pig Latin. Unpublished manuscript, Brigham Young University, Provo, UT.

Model of Early Specifications of Performance Requirements at Functional Levels

Khalid T. Al-Sarayreh

***Abstract*— This paper presents an integrated standards-based model that helps in early identification, specification and measurement for a single type of NFR, which is the performance requirement. The development stages of the standards-based framework have passed by two main steps: the first step is constituted in identifying and analyzing the system performance requirements and their allocated software performance requirements that are dispersed into the IEEE and ECSS international standards, the second step is modeling the identified system/software performance requirements using the Soft-goal Interdependency Graphs and clarifying the interdependency relations between these requirements.**

***Keywords*—** Performance Requirements, International Standards, Soft-goal Interdependency Graphs.

I. INTRODUCTION

The proper identification, specification and measurement of the system requirements at early development phases constitute the most significant factor to build a successful system that satisfies the stakeholder expectations and needs. In software engineering, the requirements are classified under two types: the functional requirements (FR) which are defined as the functionality that is required to be provided by the system (for instance: “The system shall be able to transfer data via internet”), and the non-functional requirements (NFR) are defined as the restrictions that should be applied on the required functions (for instance: “The system shall be able to transfer data via internet with low response time”).

In the academic field, several researchers have referred in their reports to the difficulties and challenges that the developers are faced to handle with NFR, for instance: taking NFR as a quantitative input to be measured and involved in the project budget estimation alongside with the FR [1-2]. Several approaches and methods are proposed from different researcher's perspectives to facilitate dealing with these challenges; nevertheless, there is currently a lack of generic models for early addressing and measuring these requirements at the system level and their related functional requirement at the software/ hardware level [3-4].

In parallel with the academic field, international standards organizations (such as the ECSS and the IEEE) are interested in describing and categorizing the NFR types. Since the European Cooperation for Space Standardization (ECSS) and the Institute of Electrical and Electronics Engineer (IEEE) categorized the performance requirements as a single type of NFR and discussed them by various terminologies and views.

This paper will account a new model for early specifications of performance requirements at functional levels based on the finding of international standards in parallel with academic previous work of some of the respected models regarding non functional performance requirements as an self-sufficient model to identify the size of the software performance separately of the languages types, whereas keep away from the limitations viewed in the performance measures presently offered.

The paper scope is to classify independently the all functionality allocated to software performance as a part of set pieces of the system application in the requirements phase for any software applications, whether the application has been built or it has already to be delivered.

In addition, the main contribution of this paper is the proposed model of software performance requirements. The proposed nonspecific model is considered as type of a orientation model in the common sense of an etalon standard that is being used for the measurement of software performance.

This paper is organized as follows. Section 2 presents the related works. Section 3 presents Performance REQUIREMENTS as defined in International Standards. Section 4 presents The Foundations of the proposed model of performance Requirements. A conclusion is presented in section 5.

II. RELATED WORK

Many early efforts have been concerned with defining, specifying and modeling NFR. For instance: [5] this paper proposed a performance requirements model; it joins together a multiplicity of types of knowledge of information systems and performance. The proposed framework includes the following performance conceptions, software performance

This work was supported in part by the U.S. Department of Commerce under Grant BS123456 (sponsor and financial support acknowledgment goes here). Paper titles should be written in uppercase and lowercase letters, not all uppercase. Avoid writing long formulas with subscripts in the title; short formulas that identify the elements are fine (e.g., "Nd–Fe–B"). Do not write "(Invited)" in the title. Full names of authors are preferred in the author field, but are not required. Put a space between authors' initials.

F. A. Khalid T. Al-Sarayreh is now with the Hashemite University, Prince Hussein Bin Abdullah II for Information Technology, Department of Software Engineering, 13115 Zarqa, Jordan. (e-mail: khalidt@hu.edu.jo).

principles for construction the performance into systems, and development knowledge. The performance and development process build by using goal-oriented approach, the performance NFR framework, which suggests a developer-directed graphical handling for stating NFRs, analyzing and connecting them, and identifying the impact of judgments leading NFRs. This move toward to built a customized solution of the domain.

For instance, [6] proposed a model for performance requirements specifications and consequential a validation testing. The model can be incorporated into agile development approaches. The performance requirements can be specified incrementally, without analysis.

More instance, [7] present a new algorithm for passive testing approach in specifying the performance time communicated protocol properties to test the real execution traces and evaluated the proposed algorithm using experimental testing on the basis of the software performance requirements through a set of properties for real execution traces.

Moreover, [8] proposed an approach to elicit performance requirement from customers for software banking system using ontology. This model divided the performance requirements into three parts: system, subsystem and component levels; between these parts ontology's inference function is used to validate and complete the requirements.

Furthermore, [9] focused on the development of performance requirements for ionospheric effects in low-frequency SAR data set applications. The performance requirements were derived considering the data quality needs of a set of SAR applications. The proposed requirements can serve as a benchmark for a performance assessment of ionospheric correction methods to define the system suitability for the system.

In addition [10] proposed a process of safety requirements with random failure of a supply system to describe geo location petri-net for model of verification and performance analysis with the widely increasing number of location-based services. Typical consumer geolocation [11] technologies are analyzed based on performance aspect for use with location-based services. While [12] proposed a performance model of requirements with related text updates to correct inconsistencies and remove limitations introduced by IEEE Std C37.118.1(TM)-2011.

In [13] present a model of distribution transformers to realize a smarter grid. The analysis has been carried out on the performance requirements and evaluation of distribution transformers when they are integrated to grid level and [14] proposed a wide Area Measurement Systems to observe the static and dynamic performance of power systems.

Finally, [15] described a method to define the performance requirements for Airport Surface Surveillance. The key idea is making the performance specification dependent on the underlying sensor deployment and geometric definition of the scenario, which enables its extension to any operational deployment.

III. Performance Requirements As Defined in International Standards

This section presents and discusses the performance terms and views for identifying the system performance NFR and their related software performance FR that may be used for specifying and measuring the system performance requirements.

A. ECSS concepts and views for performance requirements

ECSS standards [16], [17], [18] and [19] have discussed the system/software performance FR in the context of early system development phases whereas the system performance NFR have been discussed in much later phases. In the domain of these standards, the performance requirements have been defined as the specification that the output of the system does not deviate by more than a given amount from the target output.

Moreover, [20] offered a general knowledge of the control systems engineering and its applications to space missions, such as satellite system, spacecraft system, a launcher rocket system or any other technical system involving control. Such standards are emphasized on the necessity to conduct the performance analysis during all the control system development phases to evaluate that the control system is consisted and cohesive with:

- The control objectives: which are generated by the requirement engineering process?
- The numerical requirements: which are identified by the requirement analysis?

Monitoring or evaluating the performance of the system is often assisted by improving the use of the software in the system. The performance monitor is considered a facility which is integrated into a specific processer to monitor the selected characteristics to assist in debugging and analyzing systems by determining a machine's state at a particular point in time. Often, the performance monitor [20] produces information which are related to the usage of the processor's instruction execution and storage control: for instance, the performance monitor can be used to produce information related to the period of time that has passed between events in a processing system. The information produced usually guides system architects toward ways of improving performance of a given system or developing improvements in the design of a new system.

The following terms are mentioned by ECSS standards to describe the performance requirements:

- Response to reference signals (e.g. response time, settling time, and tracking errors for command profiles).
- Accuracy and stability errors in the presence of disturbances.
- Frequency domain requirements (e.g. bandwidth).
- Measurement errors (e.g. attitude knowledge)
- Processing speed.
- Resource consumption.
- Throughput.

ISBN: 978-1-61804-290-3

B. IEEE concepts and views for performance requirements

IEEE organization defines the performance requirements as a static and dynamic numerical requirement that is located on the software or on the human interaction with the entire software [21]. These two types of performance requirements should be stated in measurable form.

The following terms are mentioned by IEEE standards to describe the performance requirements:

- Static numerical requirements (e.g. capacity and concurrency).
- Dynamic numerical requirements (e.g. workload).

IV. The Foundations of the Proposed Model

The proposed framework has been developed based on using the following two foundations:

- The soft-goal interdependency graphs: which is used to model and describe the interdependency relations between the system performance NFR and their allocated software performance FR.
- The roles of ISO 19761[22] method: which are used to measure the data movement size for the allocated software performance FR.

The next sub-sections are described the referred foundation in more details way.

A. The view of Softgoal Interdependency Graphs (SIGs)

The Soft-goal Interdependency Graphs (SIGs) is introduced by [23] for describing and modeling the non-functional requirements and the interdependencies relation between them. SIGs represents the NFR as soft-goals, each soft-goal (parent soft-goal) is decomposed into one or more specific soft-goals (child soft-goals until reaching one or more solutions that satisfy the NFR (parent soft-goal).

SIGs introduce three types of soft-goal, which are:

- Soft-goal: displays the NFR to be satisfied by the system.
- Operationalizing soft-goals: represents possible solutions (operations, processes, data representations) or design alternatives that assist to satisfy the NFR.
- Claim soft-goals: shows the refinement between soft-goals or the rationale related to a soft-goal [23].

The child-soft-goals provide two contribution types to satisfy the parent soft-goals: positive contributions and negative contributions. Table 1 shows that the positive contributions are divided into make, help and some+ contributions while the negative contributions are divided into break, hurt and some- contributions (see Table 1 and Figure 1).

Table 1: SIGs contributions types

Contribution types	Contributions	Contributions description
Positive contributions	Make contribution	A strongly positive contribution, adequate enough to satisfy the soft-goal.
	Help contribution	A positive contribution can help to satisfy the soft-goal, it's inadequate to introduce full satisfaction.
	Some+ contribution	Either make or help positive contribution to satisfy the soft-goal.
Negative contributions	Break contribution	A strongly negative contribution, adequate enough to deny the soft-goal.
	Hurt contribution	A partial negative contribution, but it's inadequate by itself to introduce fully soft-goal deny.
	Some- contribution	Either break or hurt negative contribution to satisfy the soft-goal.

The parent soft-goals may be connected with the child goals by one of the following three links (see Figure 1):

- AND decomposition: the parent soft-goal is decomposed into more than one related goal and it's satisfied if all the related goals are satisfied.
- OR decomposition: the parent soft-goal is decomposed into more than one related goal and it's satisfied if at least one related goal is satisfied.
- Equal decomposition: the parent soft-goal is decomposed into one related goal and it's satisfied if the linked goal is satisfied.

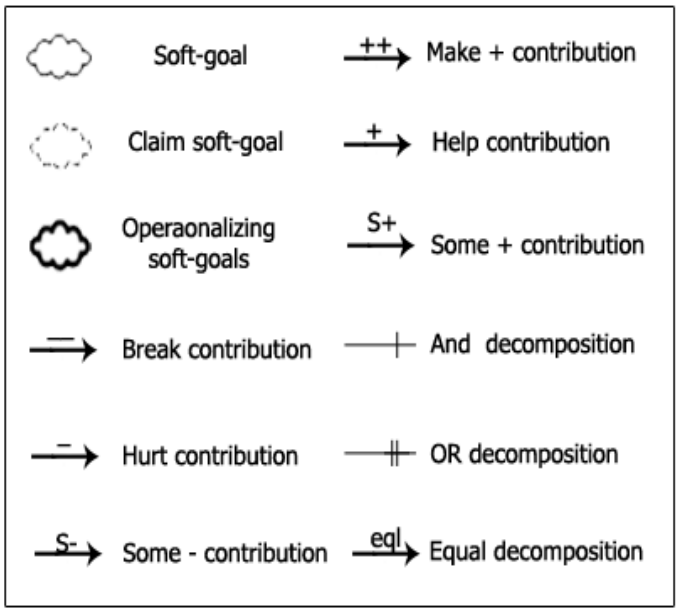


Figure 1: Soft-goal Interdependency Graphs [23]

B. Design standards-based framework for system performance requirements at functions level

At this level, the SIGs and the ISO 19761 are used for modeling and measuring the system performance requirements and their related software performance requirements. for simplicity, the proposed framework are divided into four sub-models: each sub-model clarifies one system performance requirements. Figure 2 shows a full view of the performance framework at functions level.

- The main memory time function may exchange data in a direct way with the storage device time function using COSMIC EXIT and ENTRY data movements or it may exchange data in an indirect way through the persistent storage using COSMIC READ and WRITE data movements.
- The storage device time function may exchange data in a direct way with the main memory time function and the processor instruction execution function using COSMIC EXIT and ENTRY data movements or it may exchange data in an indirect way through the persistent storage using COSMIC READ and WRITE data movements
- The processor instruction execution function may exchange data in a direct way with the storage device time function using COSMIC EXIT and ENTRY data movements or it may exchange data in an indirect way through the persistent storage using COSMIC READ and WRITE data movements
- The main memory time function, the storage device time function and the processor instruction execution function can require data from all the functions in the overall performance framework through the intermediary service using COSMIC EXIT and ENTRY data movements.
- The system scalability function may exchange data in a direct way with the concurrency function using COSMIC EXIT and ENTRY data movements or it may exchange data in an indirect way through the persistent storage using COSMIC READ and WRITE data movements.
- The concurrency function may exchange data in a direct way with the system scalability function using COSMIC EXIT and ENTRY data movements or it may exchange data in an indirect way through the persistent storage using COSMIC READ and WRITE data movements.
- The system scalability function and the concurrency function may require data from any function in the overall performance framework through the intermediary service using COSMIC EXIT and ENTRY data movements.
- The absolute performance error function may exchange data in a direct way with the performance stability error function using COSMIC EXIT and ENTRY data movements or it may exchange data in an indirect way through the persistent storage using COSMIC READ and WRITE data movements.
- The performance stability error function may exchange data in a direct way with the absolute performance error function using COSMIC EXIT and ENTRY data movements or it may exchange data in an indirect way through the persistent storage using COSMIC READ and WRITE data movements.
- The absolute performance error function and the performance stability error function may require data from any function in the overall performance framework through the intermediary service using COSMIC EXIT and ENTRY data movements.
- The absolute knowledge error may exchange data in a direct way with the relative knowledge error function using COSMIC EXIT and ENTRY data movements or it may exchange data in an indirect way through the persistent storage using COSMIC READ and WRITE data movements.
- The relative knowledge error function may exchange data in a direct way with the absolute performance error function using COSMIC EXIT and ENTRY data movements or it may exchange data in an indirect way through the persistent storage using COSMIC READ and WRITE data movements.
- The absolute knowledge error function and the relative knowledge error function may require data from any function in the overall performance framework through the intermediary service using COSMIC EXIT and ENTRY data movements.
- The response time function may exchange data in a direct way with the settling time function using COSMIC EXIT and ENTRY data movements or it may exchange data in an indirect way through the persistent storage using COSMIC READ and WRITE data movements.
- The settling time function may exchange data in a direct way with the response time function and the tracking error function using COSMIC EXIT and ENTRY data movements or it may exchange data in an indirect way through the persistent storage using COSMIC READ and WRITE data movements.
- The tracking error function may exchange data in a direct way with the settling time function using COSMIC EXIT and ENTRY data movements or exchange data in an indirect way through the persistent storage using COSMIC READ and WRITE data movements.
- The response time function, the settling time function and the tracking error function may require data from any function in the overall performance framework through the intermediary service using COSMIC EXIT and ENTRY data movements.
- The bandwidth function may exchange data in a direct way with the workload function using COSMIC EXIT and ENTRY data movements or it may exchange data in an indirect way through the persistent storage using COSMIC READ and WRITE data movements.
- The workload function may exchange data in a direct way with the bandwidth function using COSMIC EXIT and ENTRY data movements or it may exchange data in an indirect way through the persistent storage using COSMIC READ and WRITE data movements.
- The bandwidth function and the workload function may require data from any function in the overall performance framework through the intermediary service using COSMIC EXIT and ENTRY data movements.

Measurement observations

From Figure 2, the following points can be observed for measurement purposes:

- In the direct data exchange situation, each EXIT and ENTRY data movement will be assigned by 1CFP.

- In the indirect data exchange situation, each READ and WRITE data movement will be assigned by 1CFP.
- To require data through intermediary service that requires using 4 EXITS and 4 ENTRIES. Such process will be assigned by 4 CFP.

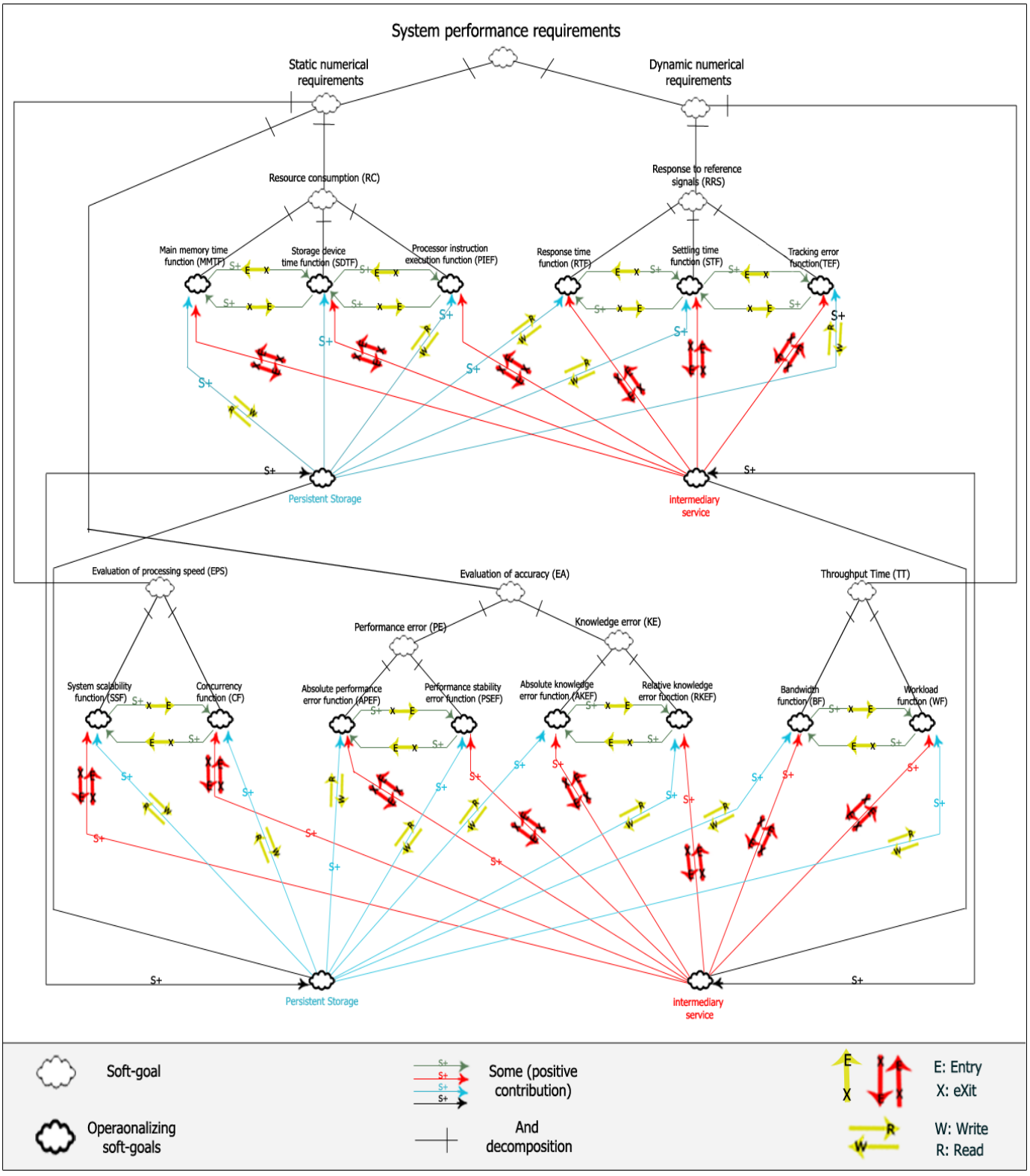


Figure 2: The full view of the system performance model at the functions level

ISBN: 978-1-61804-290-3

V. Conclusion and Future Work

Developing software systems with quality attributes level is considered a significant factor to increase the value of the system. Several researchers have referred in their reports to the difficulties and challenges that faced the developers and limited their ability in identifying, addressing and measuring the NFR during early development phases, for instance: taking such requirements as a quantitative input. Recently, the researchers have introduced an extensive work for dealing with the NFR in different development phases; nevertheless, there is no based framework that can be used to facilitate early dealing with the NFR challenges and difficulties.

In this research work, we extends our previous work on the portability and maintainability NFR reported in [1] and [4] to cover the performance NFR; Where our main contribution from this work is achieved in proposing standards-based framework for identifying, specifying and measuring the system performance requirements. The proposed standards-based framework introduce to the developers three main contributions: 1) assisting in identifying and specifying the system performance requirements, 2) allocating the system performance requirements to the related software performance requirements, 3) measuring the allocated software performance requirements with an ISO-recognized measurement unit.

Our future work aims at extending this work to introduce integrated standards-based frameworks that cover all the NFR types. Also, we would like to present an experimental study to prove the effectiveness of using our standards-based frameworks during the systems development phases.

References

[1] Al-Sarayreh, K. T., A. Abran and and J. J. Cuadrado-Gallego," A Standards-based model of system maintainability requirements", Journal of Software: Evolution and Process, John Wiley & Sons, Ltd, 2013. http://dx.doi.org/10.1002/smr.1553

[2] Meridji, Kenza, Khalid T. Al-Sarayreh, and Ahmad Al-Khasawneh. "A generic model for the specification of software reliability requirements and measurement of their functional size." *International Journal of Information Quality* 3, no. 2 (2013): 139-163.

[3] Al-Sarayreh, Khalid T., Ibrahim Al-Oqily, and Kenza Meridji. "A standard-based reference framework for system operations requirements." *International Journal of Computer Applications in Technology* 47, no. 4 (2013): 351-363.

[4] Abran, A., K. T. Al-Sarayreh, and J. J. Cuadrado-Gallego, " A Standards-based Reference Framework for System Portability Requirements", Computer Standards and Interface, Elsevier, 2013. http://dx.doi.org/10.1016/j.csi.2012.11.003

[5] Nixon, Brian A., "Management of performance requirements for information systems," *Software Engineering, IEEE Transactions on* , vol.26, no.12, pp.1122,1146, Dec 2000

[6] Chih-Wei Ho; Johnson, M.J.; Williams, L.; Maximilien, E.M., "On agile performance requirements specification and testing," *Agile Conference, 2006* , vol., no., pp.6 pp.,52, 23-28 July 2006C. J. Kaufman, Rocky Mountain Research Lab., Boulder, CO, private communication, May 1995.

[7] Xiaoping Che; Maag, S., "Passive Testing on Performance Requirements of Network Protocols," *Advanced Information Networking and Applications Workshops (WAINA), 2013 27th International Conference on* , vol., no., pp.1439,1444, 25-28 March 2013

[8] Honghong Jiang; Xiaohu Yang, "Performance requirement elicitation for financial information system based on ontology," *TENCON 2009 - 2009 IEEE Region 10 Conference* , vol., no., pp.1,5, 23-26 Jan. 2009

[9] Meyer, F.J., "Performance Requirements for Ionospheric Correction of Low-Frequency SAR Data," *Geoscience and Remote Sensing, IEEE Transactions on* , vol.49, no.10, pp.3694,3702, Oct. 2011

[10] Guo Zhou; Huibing Zhao; Weiwei Guo, "Safety requirements analysis and performance verification of hot standby system using colored Petri-net," *Industrial Electronics and Applications (ICIEA), 2013 8th IEEE Conference on* , vol., no., pp.656,661, 19-21 June 2013

[11] Brachmann, F., "About performance requirements set against consumer-grade geolocation technologies," *System Science and Engineering (ICSSE), 2013 International Conference* , vol., no., pp.309,312, 2013

[12] IEEE Standard for Synchrophasor Measurements for Power Systems -- Amendment 1: Modification of Selected Performance Requirements," *IEEE Std C37.118.1a-2014 (Amendment to IEEE Std C37.118.1-2011)* , vol., no., pp.1,25, April 30 2014

[13] Rao, N.M.; Narayanan, R.; Vasudevamurthy, B.R.; Das, S.K., "Performance requirements of present-day distribution transformers for Smart Grid," *Innovative Smart Grid Technologies - Asia (ISGT Asia), 2013 IEEE* , vol., no., pp.1,6, 10-13 Nov. 2013

[14] Jun Li; Miao Yu; Yanyan Yang, "Cement Porter's Protective Clothing Part I: Performance Requirements and Tests for the Fabric," *Biomedical Engineering and Computer Science (ICBECS), 2010 International Conference on* , vol., no., pp.1,4, 23-25 April 2010

[15] Lira, R.; Mycock, C.; Wilson, D.; Kang, H., "PMU performance requirements and validation for closed loop applications," *Innovative Smart Grid Technologies (ISGT Europe), 2011 2nd IEEE PES International Conference and Exhibition on* , vol., no., pp.1,7, 2011

[16] ECSS-E-40-Part-1B, Space Engineering: Software — Part 1 Principle and Requirements, European Cooperation for Space Standardization, the Netherlands, 2003.

[17] ECSS-E-40-Part-2B, Space Engineeing: Software-part 2 Document Requirements Definitions, European Cooperation for Space Standardization, The Netherlands, 2005.

[18] ECSS-ESA, Tailoring of ECSS, Software Engineering Standards for Ground Segments, Part C: Document Templates, ESA Board of Standardization and Control (BSSC), 2005

[19] ECSS-E-ST-10C, Space engineering: system engineering general requirements, Requirements & Standards Division Noordwijk, The Netherlands, 2009.

[20] ECSS-Q-ST-80C, Space Product Assurance: Software Product Assurance, Requirements & Standards Division Noordwijk, The Netherlands, 2009.

[21] IEEE-Std-830, IEEE Recommended Practice for Software Requirements Specifications, 1998.

[22] ISO/IEC-19761, "Software Engineering - COSMIC v 3.1 - A Functional Size Measurement Method", International Organization for Standardization, Geneva (Switzerland), 2011

[23] Chung L, et al. Nonfunctional requirements in software engineering. Kluwer Academic Publishing, 2000.

Hand vein authentication based wavelet feature extraction

Sarah BENZIANE

University of science and technology of Oran Mohamed Boudiaf

Abdelkader BENYETTOU

University of science and technology of Oran Mohamed Boudiaf

Abstract.
Biometrics is a growing scientific field. It aims to identify, through technological systems, an individual, using biological characteristics (eg details of hand, iris, ear, hand lines, fingerprints, gait, posture,). The Using of this technique is now generalized worldwide and takes an important place in everyday life. In the coming years, biometrics will probably be one of the techniques used, first to identify or authenticate individuals and also to control and manage access to material resources, particularly in the following sectors: banking, airports, bus and railway stations, hospitals, private and public institutions, homes, smart cars, museums, ...).The aim of our study is to build a dorsal hand vein database and test our approach on it. Just like any recognition system this last is composed of four steps: the acquisition, enhancement, feature extraction and classification. This paper presents the building protocol of a new database SAB11 BIOM14. Applying some enhancement on the database's image was required to get it ready for a real biometric's application. To validate our tests we proposed a new adaptive feature extraction method for the dorsal hand vein biometrics; which is the discrete wavelet transform.

Keywords: Hand vein pattern, wavelet transform, feature extraction, database, aunthentication.

1 Introduction

For getting a biometric system authentication; the user or the data has a long way before arriving to the final decision which authenticate or not. From the sensor data acquisition, image enhancement, hand vein pattern extraction and matching; these are the steps to be achieved to get the final decision. Each of them is essential for the good result of the authentication process.

Good recognition should have a good classification and a good classification should be above a perfect feature extraction phase this is where lies the strength of the biometric system, our work is focused on the dorsal hand veins feature extraction step, but the question asked is which method used to ensures a better feature extraction?

In this paper, we used our built database; which contains hand vein pattern in gray level image. The main objective of this work is to validate tests on our built databases SAB11 and BIOM14 by providing a method which allows feature extraction of veins pattern from low quality images.

They are several works about feature extraction of hand veins pattern; among them there is the Gabor filter, the Hough transform, discrete Curvelet transform, triangulation of minutiae ... etc. most of his method are preceded by a preprocessing step where in the Gabor filter [2] and the Hough transform [3] they use the Median filter, Wiener in Gabor [2] and SIFT method [5], the Mexican hat in triangulation minutiae [6], The following table summarizes some methods with their phases preprocessing, database size and performance. The table 1 summarizes most of this works.

ISBN: 978-1-61804-290-3

Method	Enhancemnt	Size of DB	Time Proces-sing	Performence
Gabor filter[2]	Median and Wiener	/	0.3631	EER=1.41% FRR=2.278%
Hough-transform [3]	Median&Gaus sien filter	400	/	FRR=0.0025%
Discret curvelet-transform [4]	Gaussien & high passfilter	400	/	EER=1.17%
SIFT [5]	high pass& Wiener filter	1020	/	RR=97.95%
Minutie triangula-tion [6]	Mexican Hat	300	/	FRR=1.14% FAR=1.14%
Across point and black box aproch [7]	/	100	/	FAR=0.1%
Ridgelet-transform [8]	/	128	/	ERR=0.13%
Invariant moment [9]	/	0.2	500	FAR=1.48%

Table 1 :works on feature extraction hand vein

2 PRIOR WORK

Extracting features from hand vein images have involved the attention of many researchers. Indeed, representing vein images by their skeletons was among the most published work [3] [4] [5] but sometimes it is difficult to extract these features due to acquisition complexities and lead to the high False Acceptance Rates (FAR>1%) [6]. To solve a problem like this, some used integration of multiples features based on some fusion rules [6] [7] [8]. Meanwhile the dorsal hand veins are different in visibility, structure and noise [8], several pre-processing methods and enhancement were applied. First, we describe these methods in next section.

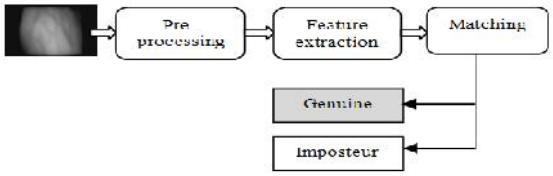


Figure 1 Dorsal hand vein biometric identification

3 DATABASE ACQUISITION SYSTEM

It was very difficult for us to find a database of dorsal hand veins. Knowing that there is no database directly available on the net; we decided to design a low-cost sensor to be able to build our own one using SAB11 device [1]. And have a hand vein images on which to test our approach. The figures below show some samples of each Database SAB11 and BIOM14.

For getting good quality of the images, we note that our hardware system needs a number of conditions:

- The day light impact on the quality of the image obtained except in the absence of IR filter.

- The temperature of the environment also influences the quality of the image must be room neither too hot nor too cold at about the temperature of the human body.

- The distance between the sensor and the object should be sufficient for good acquisition.

The following figures show nine different samples of SAB 11 vs BIOM14

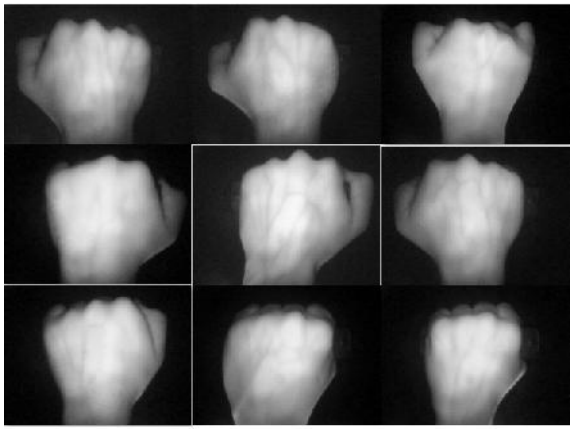


Figure 2 Samples of SAB11

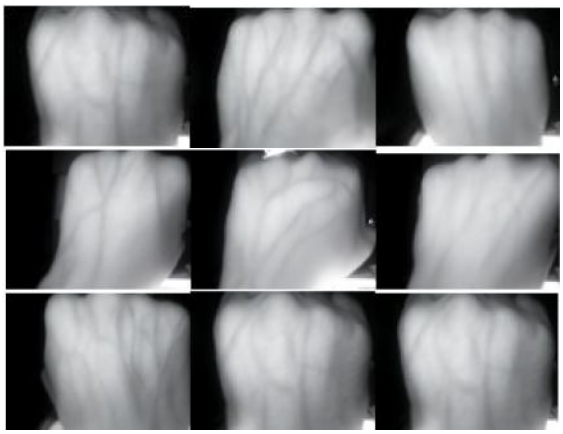


Fig.3. Samples of BIOM14

4 PREPROCESSING IMAGE

4.1 ROI extraction

For extracting the ROI, was made a small algorithm that allows an automatic extraction of the region of interest by calculating the mean of each row (column) and compared to the threshold as follows.

For each row **do**
If(avrage (oldpic(row))>=threshold)
Pic=oldpic(row :end)
Oldpic=pic
For each row **do**
If(avrage (ow)<=threshold)
Pic=oldpic(row:line)
For each colomn**do**
If(avrage (oldpic(colomn))>=threshold)
Pic=oldpic(colomn :end)
Oldpic=pic

For each colomn**do**
If(avrage (colomn)<=threshold)
Pic=oldpic(fist:colomn)

The thresholds are calculating as :

$$Threshold_t = \left(\sum_{i=1}^{Haut\ img} \sum_{j=1}^{Larg\ img} img(i,j) \right) \times Pourcntage.$$

After applying the algorithm to our image we obtain result of the Figure 2.

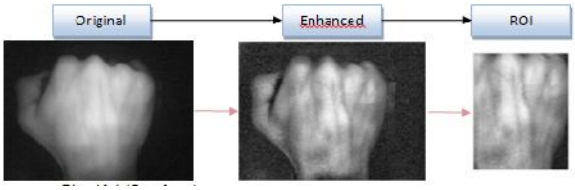


Fig. 4.ROI extraction

4.2 ENHANCEMNT

To enhance the contrast accentuations of the image, we applied a double adaptive equalization; the vein contrast the result is in the following figure. It differs from ordinary histogram equalization in the respect that the adaptive method computes several histograms, each corresponding to a distinct section of the image, and uses them to redistribute the lightness values of the image. It is therefore suitable for improving the local contrast. AHE has a tendency to over amplify noise in relatively homogeneous regions of an image. A variant of adaptive histogram equalization called contrast limited adaptive histogram equalization (CLAHE) prevents this by limiting the amplification.

ISBN: 978-1-61804-290-3

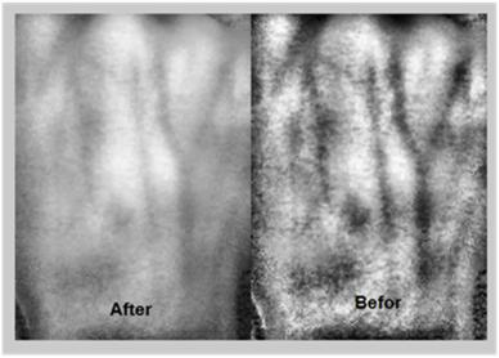


Fig. 5. Double adaptative egalization

Adaptive histogram equalization dual is applied twice to the algorithm proposed by [10], the following algorithm:

Algorithm AHE
For every pixel *i* (with grey level *l*) in image **do**
Initialize array *Hist* to zero;
For every contextual pixel *j* **do**

$$Hist[g(j)] = Hist[g(j)] + 1;$$

$$\text{Sum: CHist}_l = \sum_{k=0}^{l} \text{Hist(k)}$$

$$l' = \text{CHist}_l * L/W^2$$

is gray level of pixel j, l and l' are original and new gray level of center pixel i, l is the cumulative histogram function value in gray level. The result obtained is as follows

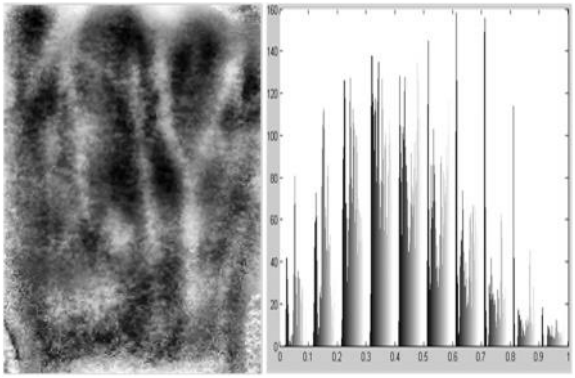


Fig. 6. Results of image equalized

4.3 FEATURE EXTRATION

The main objective of the wavelet transform is data compression, it is used for signal analysis, image compression, sound processing and geology are the main application areas for wavelet in our works we will refrain to the application images.

The **integral wavelet transform** is the integral transform defined as

$$[W_\psi f](a,b) = \frac{1}{\sqrt{|a|}} \int_{-\infty}^{\infty} \overline{\psi\left(\frac{x-b}{a}\right)} f(x)dx$$

The **wavelet coefficients** c_{jk} are then given by

$$c_{jk} = [W_\psi f]\left(2^{-j}, k2^{-j}\right)$$

Here, $a = 2^{-j}$ is called the **binary dilation** or **dyadic dilation**, and $b = k2^{-j}$ is the **binary** or **dyadic position**.

To extract dorsal hand vein pattern we have used a single 2 dimensional wavelet transform, a discrete wavelet transform in two dimensions can be achieved by running two separate one-dimensional transforms. First, the image (2D signal) is filtered horizontally (along the x axis) and divided by two. Thereafter the filtered subimage is vertically (along the y axis) and divided by two.

We then obtain an image composed of four bands after decomposition at a single level.

How is the compression of an image?

The compression is achieved by successive approximations of the initial information from the

adfa, p. 4, 2011.

coarsest to the finest. Then it reduces the size of the information by selecting a level of detail.

It exist many categories of wavelet. The most suitable for our database is the bi-orthogonal reverse wavelet family with wavelet bio 3.1. It gives us good results; which are shown in Figure 5

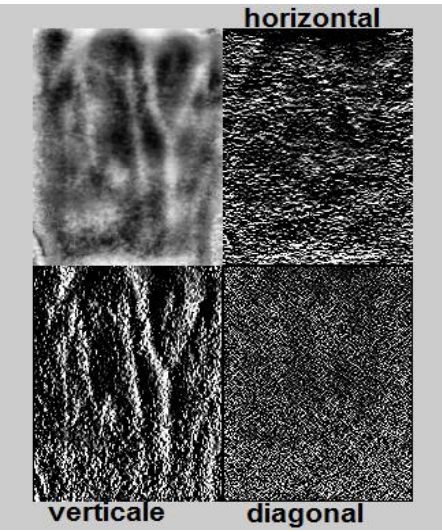
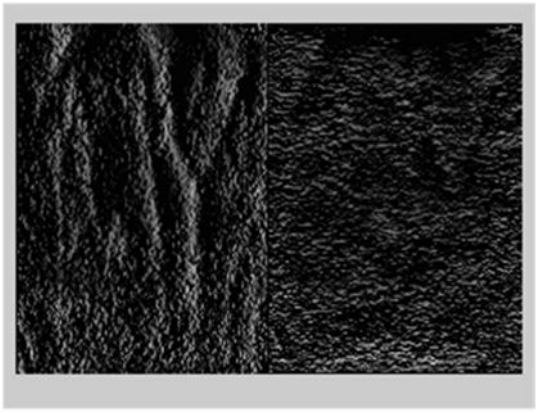


Fig.7.discret Bi-orthogonal wavelet transform

For improving the result obtained, We test an adaptive equalization on the vertical and horizontal details image for improved contours veins as in this figure7.

We used a binarization of each image to keep the pattern of the veins with using vertical threshold SV=0.12 and horizontal threshold SH=0.25 the result is in the following figure.



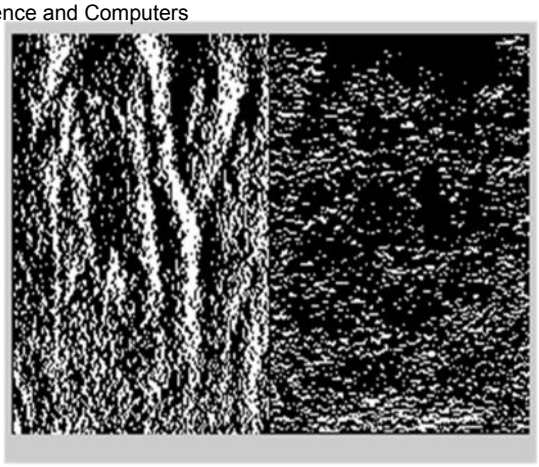


Fig.8.adaptive equalization of detail images and the binarization

The most important is to keep just the pattern of veins and eliminates noise, after a series of tests we could find a morphological operation that allows this processing called morphologically open image.

The opening of A by B is obtained by the erosion of A by B, followed by dilation of the resulting image by B:

$$A \circ B = (A \ominus B) \oplus B$$

The opening is also given by $^{\circ}B = \cup_{Bx \leq A} Bx$, which means that it is the locus of translations of the structuring element B inside the image A. In the case of the square of side 10, and a disc of radius 2 as the structuring element, the opening is a square of side 10 with rounded corners, where the corner radius is 2.

Example application: Let's assume someone has written a note on a non-soaking paper and that the writing looks as if it is growing tiny hairy roots all over. Opening essentially removes the outer tiny "hairline" leaks and restores the text. The side effect is that it rounds off things. The sharp edges start to disappear.

Where B is composition of erosion and dilatation this method is used with two parameters 4 connectivity of neighborhood and

PV=40 pixels in vertical

PH=30 pixels in horizontal

PH and PV denote the maximum size of objects to delete in a binary image.

Fig. 9.Morphologically open image

After having superimpose the two results and proceed to the deletion of noise pattern by applying morphological opening operation of the binary image on the image of the negative result with 40 pixels as the minimum object size kept, the result is in the following figure.

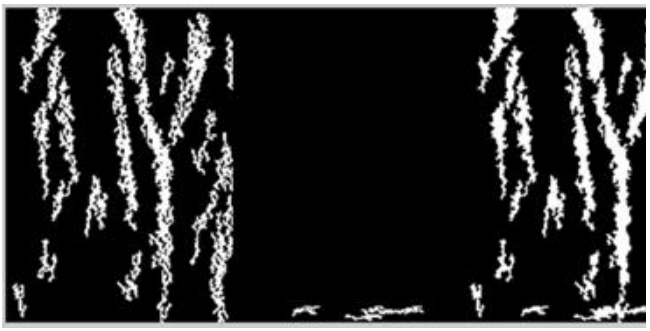


Figure 10.the result to merge two images

4.4 Authentication

Authentication uses several technics of classification. In our case we opted for the classification based on the binary image as a feature vector.

Which seem to be more appropriate is the Euclidean distance:

$$D(I,J) = \left(\sum_i |f_i(I) - f_i(J)|^2 \right)^{\frac{1}{2}}$$

Euclidean distance is to be applied in calculating the distance between two matrices (binary image) of fixed size which is not the case of all images in the ROI of each image extraction of different size with respect to another , luted to be against problem we had to do a normalization step image this step helps to make all the images in a standard size template given the choice.

According to the results we find that identification with calibration performed on detailed image provides a low error rate compared to other approaches tested.

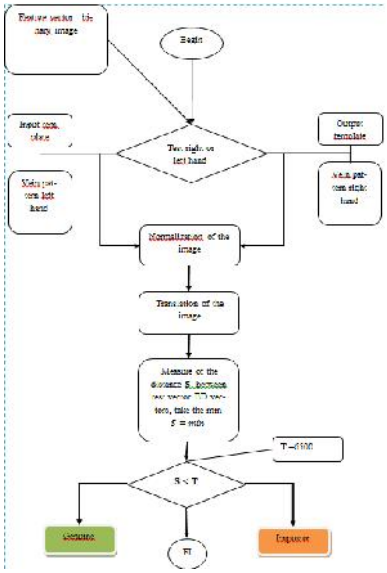


Fig. 11.Organiram of classification

5 EXPERIMENTAL RESULT

Our work has been tested on a database of 34 individuals for each individual there are 5 taken for each hand which make 3400 image database, the results of different approaches are summarized in the following:

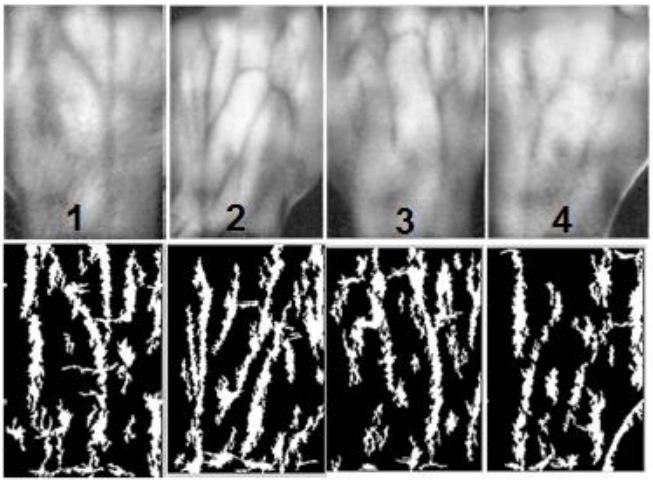


Fig.12.Four sample hand vein and extracts pattern

We see here that with automatic threshold with translation operation provides better results; Table 2, 3:

Table 2 FRR with a static thresholdd

Operation	Vertical detail	ROI
Rotat & Transl	31,74 %	48,24 %
Translation	13,43 %	19,41%

Table 3 FRR with automatic threshold

Operation	Vertical detail	ROI
Rotat & Transl	30	42,65
Translation	10,29	21,18

This result is obtained by applying the method mentioned above, the pattern is satisfactory for use to make a classification

The figure below allows you to see the histogram for each pattern extracted correspondence is by number.

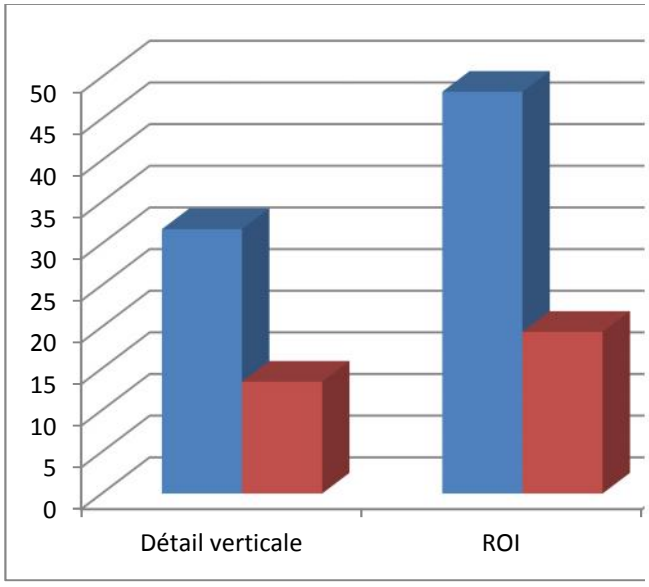


Fig. 13**.** Histogram of FRR with a static threshold binarization

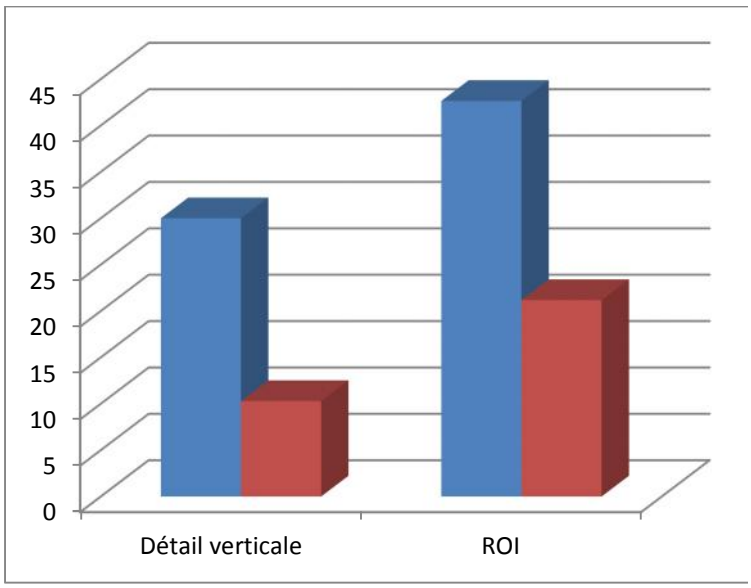


Fig. 14.Histogram of FRR with an automatic threshold binarization

6 DISSCUSION OF THE RESULTS

Based on the results obtained, we see that the sequential time was very large (about 6 second) against parallel time .Since with our approach with Akka actors, the parallel time was always smaller than the sequential time, but we found that increasing the number actors to binarize one image, speedup decreases, so we were limited to two actors who handles a single picture.

In this work, we have worked with two databases: the previously acquired and NCUT of hand-dorsa vein images which contains 2000 images from 100 individuals, 10 images of the right hand and 10 images of the left.

The figure below show the images of dorsal hand vein biometric of several persons used in this experiment with:

We have obtained the same output image in sequential version and parallel version.

7 CONCLUSION

The conception of our database was very interesting in the way that we got a dorsal hand vein database to share and test on it our actual and

future approach. This was our principal contribution with the use of the wavelet transform for the vein extraction phase. After this research, w conclude that the wavelet transform give good results for a such kind of imaging. The wavelet gives us access to the multi-resolution; it allows the manipulation and extraction of contours which helps us during the feature extraction step. Even in the case where the quality of the images is a bad.

The two built database are very efficient for future study and specially for testing some new approaches as ridglet and contourlet. Within the aspects which can be considered for future works, ca note; Generalization of the system for any acquisition system, Optimization time processing for the classification phase and method for distinguishing right from left hand.

8 REFERENCES

[1] Benziane, S., Benyettou, A., & Merouane, A. (2014). Parallel extraction hand vein biometric parameter's using a low cost IR imaging system.

[2] Ramalho, M., Correia, P. L., & Soares, L. D. (2011, April). Biometric identification through palm and dorsal hand vein patterns. In EUROCON-International Conference on Computer as a Tool (EUROCON), 2011 IEEE (pp. 1-4). IEEE.

[3] Ramsoful, P., & Heenaye-Mamode Khan, M. (2013, August). Feature extraction techniques for dorsal hand vein pattern. In Innovative Computing Technology (INTECH), 2013 Third International Conference on (pp. 49-53). IEEE.

[4] Wei, S., & Gu, X. (2011, December). A method for hand vein recognition based on curvelet transform phase feature. In Transportation, Mechanical, and Electrical Engineering (TMEE), 2011 International Conference on (pp. 1693-1696). IEEE.

[5] Wang, Y., Fan, Y., Liao, W., Li, K., Shark, L., & Varley, M. R. (2012, March). Hand vein recognition based on multiple keypoints sets. In Biometrics (ICB), 2012 5th IAPR International Conference on (pp. 367-371). IEEE.

[6] Kumar, A., & Prathyusha, K. V. (2009). Personal authentication using hand vein triangulation and knuckle shape. Image Processing, IEEE Transactions on, 18(9), 2127-2136.

[7] Kumar, A., Hanmandlu, M., Madasu, V. K., & Lovell, B. C. (2009, December). Biometric authentication based on infrared thermal hand vein patterns. In Digital Image Computing: Techniques and Applications, 2009. DICTA'09. (pp. 331-338). IEEE.

[8] Zhang, Y., Han, X., & Ma, S. L. (2006). Feature extraction of hand-vein patterns based on ridgelet transform and local interconnection structure neural network. In Intelligent Computing in Signal Processing and Pattern Recognition (pp. 870-875). Springer Berlin Heidelberg.

[9] Wang, K., Zhang, Y., Yuan, Z., & Zhuang, D. (2006, June). Hand vein recognition based on multi supplemental features of multi-classifier fusion decision. In Mechatronics and Automation, Proceedings of the 2006 IEEE International Conference on (pp. 1790-1795). IEEE.

[10] S.Benziane, A.Benyettou, «An introduction to Biometrics,» International Journal of Computer Science and Information Security, vol. 9, n° 14, pp. 40-44, 2011.

[11] Satone, M. P., & Kharate, G. K. (2012). Face Recognition Based on PCA on Wavelet Subband of Average-Half-Face. *JIPS*, *8*(3), 483-494.

[12] Birgale, L., & Kokare, M. (2012). Iris Recognition Using Ridgelets. JIPS, 8(3), 445-458.

[13] Lanitis, A. (2010). A survey of the effects of aging on biometric identity verification. *International Journal of Biometrics*, *2*(1), 34-52.

[14] Anantha kumar, T., & Premalatha, K. (2014). Personal identification using local Gaussian quadrature filter pair phase quantisation of hand vein images. International Journal of Biometrics, 6(2), 180-203.

[15] Tai, S. C., Huang, H. F., & Chung, K. C. (2007). Automatic Facial Expression Discrimination System. Far East Journal of Electronics and Communication, 1, 23-31.

Authors Index

ISBN: 978-1-61804-290-3