

# RCS

Research in Computing Science

Vol. 69

Advances  
in Computing Science, Control and  
Communications

---

Mireya García Vázquez (Eds.)  
Grigori Sidorov  
Sunil Kumar

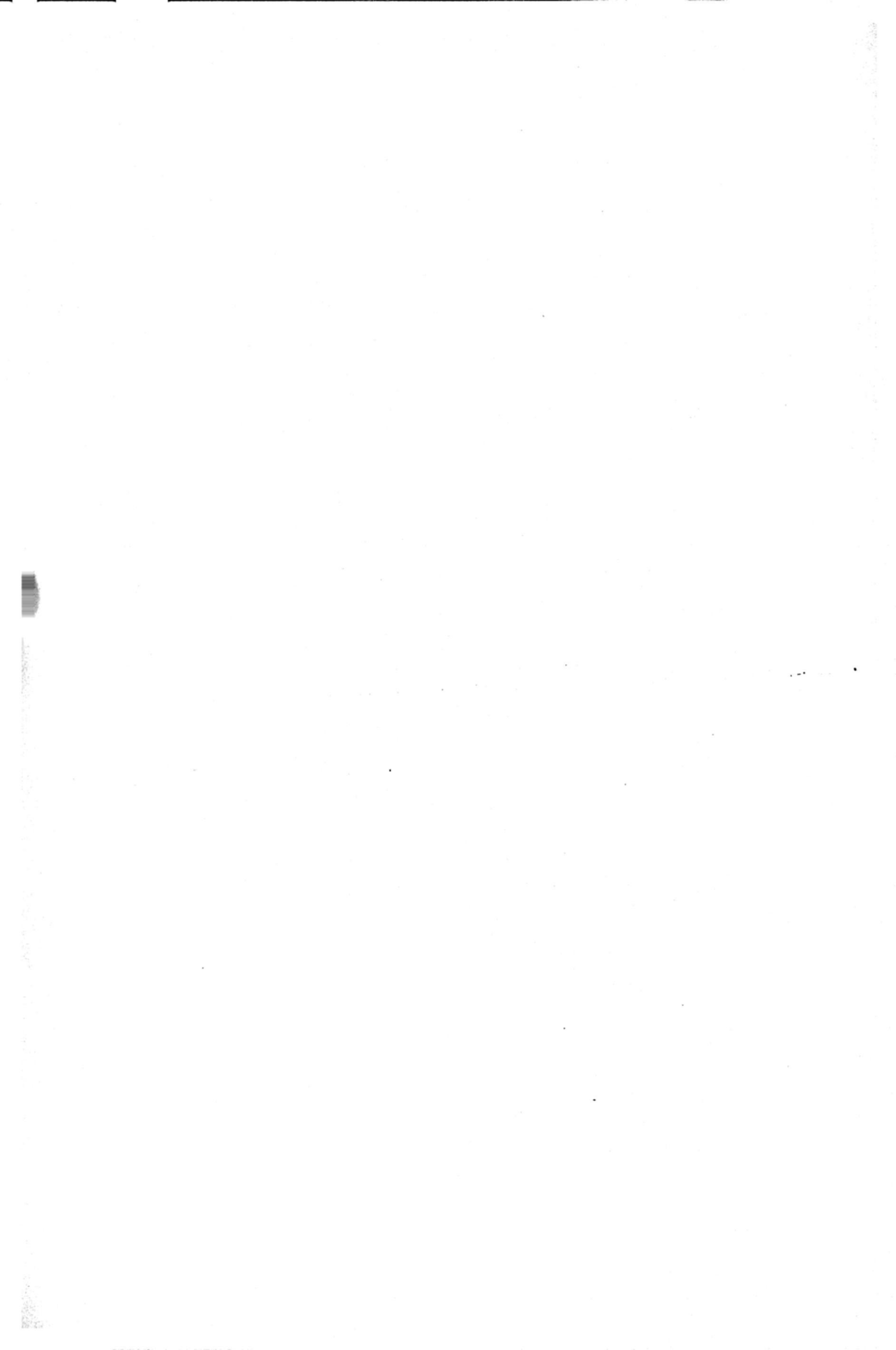














# **Advances in Computing Science, Control and Communications**

---



# Research in Computing Science

---

## Series Editorial Board

Comité Editorial de la Serie

### Editors-in-Chief:

Editores en Jefe

*Grigori Sidorov (Mexico)*

*Gerhard Ritter (USA)*

*Jean Serra (France)*

*Ulises Cortés (Spain)*

### Associate Editors:

Editores Asociados

*Jesús Angulo (France)*

*Jihad El-Sana (Israel)*

*Jesús Figueroa (Mexico)*

*Alexander Gelbukh (Russia)*

*Ioannis Kakadiaris (USA)*

*Serguei Levachkine (Russia)*

*Petros Maragos (Greece)*

*Julian Padget (UK)*

*Mateo Valero (Spain)*

### Editorial Coordination:

Coordinación Editorial

*Socorro Méndez Lemus*

### Formatting:

Formación

*Juan Carlos Sánchez Valenzuela*

*Isaura González Rubio Acosta*

*Research in Computing Science* es una publicación trimestral, de circulación internacional, editada por el Centro de Investigación en Computación del IPN, para dar a conocer los avances de investigación científica y desarrollo tecnológico de la comunidad científica internacional. Volumen 69, Abril 2014. Tiraje: 500 ejemplares. *Certificado de Reserva de Derechos al Uso Exclusivo del Título* No. : 04-2005-121611550100-102, expedido por el Instituto Nacional de Derecho de Autor. *Certificado de Licitud de Título* No. 12897, *Certificado de licitud de Contenido* No. 10470, expedidos por la Comisión Calificadora de Publicaciones y Revistas Ilustradas. El contenido de los artículos es responsabilidad exclusiva de sus respectivos autores. Queda prohibida la reproducción total o parcial, por cualquier medio, sin el permiso expreso del editor, excepto para uso personal o de estudio haciendo cita explícita en la primera página de cada documento. Impreso en la Ciudad de México, en los Talleres Gráficos del IPN – Dirección de Publicaciones, Tres Guerras 27, Centro Histórico, México, D.F. Distribuida por el Centro de Investigación en Computación, Av. Juan de Dios Bátiz S/N, Esq. Av. Miguel Othón de Mendizábal, Col. Nueva Industrial Vallejo, C.P. 07738, México, D.F. Tel. 57 29 60 00, ext. 56571.

**Editor responsable:** *Grigori Sidorov, RFC SIGR651028L69*

*Research in Computing Science* is published by the Center for Computing Research of IPN. Volume 69, April 2014. Printing 500. The authors are responsible for the contents of their articles. All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without prior permission of Centre for Computing Research. Printed in Mexico City, in the IPN Graphic Workshop – Publication Office.

---

**Volume 69**

Volumen 69

---



# **Advances in Computing Science, Control and Communications**

**Mireya García-Vázquez (Eds.)  
Grigori Sidorov  
Sunil Kumar**



**Instituto Politécnico Nacional  
"La Técnica al Servicio de la Patria"**



**Instituto Politécnico Nacional, Centro de Investigación en Computación  
México 2014**



**ISSN: 1870-4069**

---

Copyright © Instituto Politécnico Nacional 2014

Instituto Politécnico Nacional (IPN)  
Centro de Investigación en Computación (CIC)  
Av. Juan de Dios Bátiz s/n esq. M. Othón de Mendizábal  
Unidad Profesional "Adolfo López Mateos", Zacatenco  
07738, México D.F., México

<http://www.ipn.mx>  
<http://www.cic.ipn.mx>

The editors and the publisher of this journal have made their best effort in preparing this special issue, but make no warranty of any kind, expressed or implied, with regard to the information contained in this volume.

All rights reserved. No part of this publication may be reproduced, stored on a retrieval system or transmitted, in any form or by any means, including electronic, mechanical, photocopying, recording, or otherwise, without prior permission of the Instituto Politécnico Nacional, except for personal or classroom use provided that copies bear the full citation notice provided on the first page of each paper.

Indexed in LATINDEX and Periodica / Indexada en LATINDEX y Periódica

Printing: 500 / Tiraje: 500

Printed in Mexico / Impreso en México

## **Preface**

### **(Prefacio)**

This volume of the journal "Research in Computing Science" contains selected papers on the three topics: Control and automation, intelligent systems and information and communications technology. The papers were carefully chosen by the editorial board on the basis of the at least two reviews by the members of the reviewing committee. The main criteria for selection were their originality and technical quality.

This issue of the journal Research in Computing Science can be interesting for researchers and students in communications, intelligent systems, control systems, and also for persons who are interested in cutting edge themes of information technologies.

This volume contains revised version of 22 regular and 2 invited papers from different countries (USA, Spain, Israel, Canada, Mexico and France). The papers are structured into the following three sections:

- Control and automation (8 papers),
- Intelligent systems (9 papers),
- Information and communications technology (7 papers).

The main topics of the papers reflect the tendencies in the current state of art in control, communications and intelligent systems, or have major demand in the area of practical applications.

This volume is a result of work of many people. In the first place, we thank the authors of the papers included in this volume for the technical excellence of their papers that assures the high quality of this publication. We also thank the members of the International Editorial Board of the volume and the reviewer's committee for their hard work consisting in selection of the best papers out of many submissions that were received.

The submission, reviewing, and selection process was performed on the basis of the free system EasyChair, [www.EasyChair.org](http://www.EasyChair.org).

April, 2014

*Mireya García-Vázquez*  
*Grigori Sidorov*  
*Sunil Kumar*

## Table of Contents

---

### I Control and automation

---

Adaptive Generalized Synchronization of Complex Networks.....	3
<i>Juan Gonzalo Barajas Ramírez</i>	
The use of an evolutionary algorithm for the parameter selection of predictive controllers .....	14
<i>Rosalía Del Carmen Gutiérrez-Urquidez, Guillermo Valencia-Palomo and Oscar Mario Rodríguez-Elías</i>	
Model-Based Sensorless Controller of a Permanent Magnet Synchronous Motor .....	26
<i>Luis N. Coria, Paul J. Campos and Ramon Ramirez-Villalobos</i>	
Experimental Analysis of the Dynamixel AX-12 Servomotor and its Wireless Communication .....	37
<i>Eusebio Bugarin, Luis Jaciel Castañeda-García and Ana Aguilar-Bustos</i>	
Position Feedback Nonlinear H-infinity Control for Inertia Wheel Pendulum Stabilization .....	47
<i>Adrian Gomez and Luis T. Aguilar</i>	
Type-2 Fuzzy Control Lyapunov Approach for Position Trajectory Tracking .....	59
<i>Rosalío Farfán Martínez, José Antonio Ruz Hernández, José Luis Rullán Lara, William Torres Hernández and Luz Del Alba Cambrano Bravata</i>	
High-order Sliding-Mode Based Suboptimal Linear Quadratic Regulator With Application to Roll Autopilot Design .....	68
<i>Jorge Davila</i>	
Limit cycles in second order systems through sliding surface design .....	80
<i>Raul Rascon, Andres Calvillo and Luis Moreno</i>	

---

### II Intelligent systems

---

Water Distribution Systems Optimization .....	89
<i>Avi Ostfeld</i>	
Spectrum resource optimization for future cellular networks .....	99
<i>Anabel Martinez-Vargas, Ángel G. Andrade, Roberto Sepúlveda and Os- car Montiel-Ross</i>	
Segmenting supervised activities in a video sequence based on handling of artifacts towards intelligent systems .....	108
<i>Francisco E Martinez-Pérez, Hector Gerardo Pérez-González, José Ángel González-Fraga, Juan Carlos Cuevas-Tello and Sandra Edith Nava- Munoz</i>	



A Hybrid ECJ+BOINC Tool for Distributed Evolutionary Algorithms ...	120
<i>Francisco Fernández de Vega, Leonardo Trujillo, Francisco Chávez, Enrique Mediero and Luis Muñoz</i>	
GPU implementation of nonlinear anisotropic diffusion for medical image enhancement .....	131
<i>Fernando Villalbaz, Juan Tapia and Julio Rolón</i>	
Pattern Identification by an Artificial Neural Network implemented in a DSP using a touchscreen .....	143
<i>Mauricio Méndez, Juan Buendía, Julio Galindo, Mario Rodríguez, Juan Mata-Machuca and Leonardo Fonseca</i>	
Muscle Pain and Blink Classification using a Brain Computer Interface ..	151
<i>Oscar Montiel, Roberto Sepúlveda, Gerardo Díaz, Daniel Gutierrez and Oscar Castillo</i>	
Emotion recognition and emotional incentive model .....	162
<i>Adrian Rodriguez, Miguel Angel Lopez Ramirez and Arnulfo Alanis Garza</i>	
Neurofuzzy Identification Applied to a Flow Control Equipment .....	173
<i>William Torres Hernández, Rosalío Farfan Martinez, José Antonio Ruz Hernández, Ramón García Hernández and José Luis Rullán Lara</i>	

---

### III Information and communications technology

---

Remote Sensing Image Processing with Graph Cut of Binary Partition Trees .....	185
<i>Philippe Salembier and Samuel Foucher</i>	
Homogeneous Quality Video in Multi-Sources Systems .....	197
<i>Francisco De Asís López-Fuentes</i>	
Circular Monopole Antenna with defected ground plane for UWB applications .....	207
<i>Cruz Ángel Figueroa Torres, José Luis Medina Monroy, Ricardo Arturo Chávez Pérez and Andrés Calvillo Tellez</i>	
Face detection method based on nonlinear composite correlation filters...	215
<i>Everardo Santiago Ramírez, José Ángel González Fraga, Omar Álvarez Xochihua, Sergio Omar Infante Prieto and Everardo Gutiérrez López</i>	
Design and Hardware Implementation of Digital Amplitude Modulation on FPGA .....	227
<i>J.A. Galaviz Aguilar, J.R. Cárdenas-Váldez, J.A. Reynoso Hernández and J.C. Núñez Pérez</i>	

FPGA Implementation of behavioral Models for RF Power Amplifiers . . .	235
<i>José Ricardo Cárdenas Valdez, José Alejandro Galaviz Aguilar, José Cruz Núñez Pérez, Christian Gontrand and Andrés Calvillo Téllez</i>	
Adaptation of a PCA fusion stage to improve accuracy in a biometric iris recognition system for unconstrained environments . . . . .	246
<i>Juan M. Colores-Vargas, Mireya S. García-Vázquez and Alejandro Ramírez-Acosta</i>	

## **Part I**

# **Control and automation**



1964

Control and Automation

# Adaptive Generalized Synchronization of Complex Networks

J. G. Barajas-Ramírez

IPICYT, División de Matemáticas Aplicadas

Camino a la Presa de San José 2055

Lomas 4a Sec. CP 78216,

San Luis Potosí, SLP, México

Email: [jgbarajas@ipicyt.edu.mx](mailto:jgbarajas@ipicyt.edu.mx)

Phone: (52)-444-834-2000

*Paper received on 11/19/13, Accepted on 01/19/14.*

**Abstract.** We propose a synchronization strategy that adaptively adjust the coupling strength of a network of strictly different dynamical systems as a way to achieve generalized synchronization (GS). There are two basic approaches to detect GS between systems: In the first, GS is inferred from the identical synchronization of two systems under the same driving force, the so-called auxiliary system approach. Alternatively, a functional relationship between the systems can be explicitly imposed by design, this is usually called the controlled synchronization approach. In this contribution, we take the latter approach to impose GS on a linearly and diffusively coupled network, where the nodes are different dynamical systems that are fully triangularizable and the coupling strength of the network is adjusted adaptively. We illustrate our results with numerical simulations.

**Keywords:** Complex networks, Generalized synchronization, Adaptive control.

## 1 Introduction

In recent years, the synchronization of complex networks has received a great deal of attention from the scientific community (Wu, 2002; Boccaletti et al., 2006). These investigations are motivated by the necessity of an improved understanding of the synchronization phenomenon in real-world complex networks, such as the Internet, the WWW, biological and social networks, among others (Arenas et al., 2008; Newman, 2010). A particularly significant concern is to study the synchronization of chaotic systems coupled in complex topologies (Wang and Chen, 2002a; Wang and Chen, 2002b). The main concert of these investigations has been the emergence of identical synchronization in networks consisting of identical  $n$ -dimensional dynamical systems. However, in real-world situations a far more likely scenario is that the nodes be different systems with parameter uncertainties and disturbances. In these situations, identical synchronization is

unlikely. Yet, even under these situations networks exhibit some form of temporal correlation, phenomena like interdependence, auto-organization, consensus and collaboration are ever present in complex networks. Clearly these interactions go beyond simultaneous dynamical evolution as prescribed in identical synchronization, these phenomena require a more general definition of synchronization.

The concept of generalized synchronization (GS) was originally defined in the literature for the master-slave synchronization of chaotic systems (Abarbanel et al., 1996). Between two systems GS refers to the existence of a functional relationship between their dynamical states (Boccaletti et al., 2002). Different types of GS can be defined, depending how the state space of one node are mapped to the others. In this way, one can think of complete identical synchronization as a particular case of GS where the functional relationship is the identity. Another form of GS is achieved when the functional relationship is defined in terms of coordinate transformations, for example a diffeomorphism defined on a feedback linearization (Femat et al., 2005). In the literature we have basically two approaches to identify GS: An indirect method, in which synchronization in generalized terms is inferred from the identical synchronization of two systems under the same driving force, the so-called auxiliary system approach (Abarbanel et al., 1996). Alternatively, GS can be directly achieved by controllers that impose a prescribed functional relationship between the systems, this approach is usually called the controlled synchronization method. The main difference between these approaches is whether or not the description of the functional relationship between the nodes is of significance. In the case of auxiliary system approach, its existence is implied, while in controlled synchronization is a requirement. Recently the concept of GS has been extended to the case of dynamical systems coupled in complex network. In some earlier works (Hung et al., 2008; Xu et al., 2008; Liu et al., 2010) the auxiliary systems approach was considered, while others focus on the controlled synchronization approach (Guan et al., 2009).

In this contribution, we take the controlled synchronization approach to achieve GS in a network of linear and diffusively coupled exact linearizable by state feedback nonidentical  $n$ -dynamical systems. Unlike previous works on controlled GS, we propose to design an adaptive controller such that a given functional relationship between the states of different groups of nodes is imposed.

The rest of the manuscript is organized as follows. In Section II, GS problem for a network is expressed as a stability analysis problem. In Section III, we present our proposed adaptive controller designed for GS on a particular class of dynamical networks. In Section IV, the proposed design is illustrated with numerical simulations. Then, the contribution is closed with comments and conclusions.

## 2 Problem Statement

Consider a network of  $N$  non-identical nodes, with each one being a dynamical system described by

$$\dot{x}_i = f_i(x_i) \quad (1)$$



where  $x_i = (x_{i1}, x_{i2}, \dots, x_{in})^T \in \mathbf{R}^n$  are state variables of node  $i$  (all nodes are assume to have the same dimension  $n$ ); and  $f_i : \mathbf{R}^n \rightarrow \mathbf{R}^n$  is a known nonlinear function describing the dynamical evolution of node  $i$ .

The state equation of the entire dynamical network is

$$\dot{x}_i = f_i(x_i) + g_i(X) + u_i \quad (2)$$

for  $i = 1, 2, \dots, N$  where  $X = (x_1, x_2, \dots, x_N) \in \mathbf{R}^{n \times N}$  are form by the state variables of all the nodes;  $g_i : \mathbf{R}^{n \times N} \rightarrow \mathbf{R}^n$  are the coupling functions describing the connections to node  $i$  from the rest of the network; and  $u_i \in \mathbf{R}^m$  ( $m \leq n$ ) is a local controller to be designed.

A dynamical network is said to be identically synchronized, if the state solutions of every node move in unison, in the sense that

$$\lim_{t \rightarrow \infty} \|x_i - x_j\| = 0, \text{ for } i, j = 1, 2, \dots, N \quad (3)$$

The synchronization criterion for complete identical synchronization, can be interpreted as requiring that the states variables of any node in the network be exactly mapped to the state variables of any other. A generalization of this interpretation of synchronization can be introduced by considering mappings between the state variables of the nodes to be different from the identity, in this way more complicated interactions between the network components can be considered (Boccaletti et al., 2002). Then, the network in (2) will be synchronized in a generalized sense with respect to the functional relation  $H_i$  if the condition

$$\lim_{t \rightarrow \infty} \|x_i - H_i(x_j)\| = 0, \text{ for } i, j = 1, 2, \dots, N \quad (4)$$

is satisfied. Note that, the functional  $H_i$  maybe the same for all the nodes or it can be different for each pair of nodes. Additionally, potentially each system can have its own transformation  $H_{Mi}$  and  $H_{Si}$ , with a GS condition

$$\lim_{t \rightarrow \infty} \|H_{Mi}(x_i) - H_{Si}(x_j)\| = 0, \text{ for } i, j = 1, 2, \dots, N \quad (5)$$

where  $H_i = H_{Mi} \circ H_{Si}$ .

In the sense of (Abarbanel et al., 1996), GS is achieved by the existence of  $H_i$  not by its exact description. That is, if an auxiliary system is consider to experience the same driving forces as our system and they identically synchronize to each other; then, the existence of a functional relationship can be inferred for the original system. Similarly to (Hung et al., 2008; Xu et al., 2008; Liu et al., 2010), we can extend this approach to determine GS in a dynamical network. Considering the coupling to each node in the network as an external driving, an exact replica of the network dynamics

$$\dot{\hat{x}}_i = f_i(\hat{x}_i) + g_i(\hat{X}) + u_i \quad (6)$$

for  $i = 1, 2, \dots, N$  where  $\hat{X} = (\hat{x}_1, \dots, \hat{x}_N) \in \mathbf{R}^{n \times N}$  can be taken to be an auxiliary network system.

Following the auxiliary system approach the network (2) achieves GS if (Abarbanel et al., 1996)

$$\lim_{t \rightarrow \infty} |x_i - \hat{x}_i| = 0 \quad (7)$$

for  $i = 1, \dots, N$  with the initial conditions  $x_i(0) \neq \hat{x}_i(0)$ .

From (2) and (6) the dynamics of the error  $\epsilon_i = \hat{x}_i - x_i$  are given by

$$\dot{\epsilon}_i = f_i(x_i) - f_i(\hat{x}_i) + g_i(X) - g_i(\hat{X}) \quad (8)$$

for  $i = 1, \dots, N$ . The emergence of GS is equivalent to the stability of the zero equilibrium point of (8). There are different results in the literature where GS is assured for dynamical networks under some very standard assumptions like, global Lipschitz condition for all nodes with linear and diffusive couplings; e.g. in (Liu et al., 2010) adaptive coupling strengths are used to achieve GS.

The auxiliary system approach for GS can be applied to establish that the network is synchronized in a generalized sense. However, although the functional relations  $H_i$  exist, it's not possible to determine its specific form. If we want to impose a functional relationship among the nodes in the network a controlled synchronization approach is needed. We define a GS error between the  $i$  and  $j$ -th nodes as  $e_{ij} = H_{Mi}(x_i) - H_{Si}(x_j)$ , for  $i, j = 1, \dots, N$ , which has the dynamics

$$\dot{e}_{ij} = H_{Mi}(f_i(x_i) + g_i(X)) - H_{Si}(f_j(x_j) + g_j(X)) + \nu_i \quad (9)$$

for  $i, j = 1, \dots, N$  with  $\nu_i = H_{Mi}(u_i) - H_{Si}(u_j)$ .

The total number of GS errors in the network can be reduced to  $N - 1$  by defining  $j = i + 1$ , then we have

$$\dot{e}_i = H_{Mi}(f_i(x_i) + g_i(X)) - H_{Si}(f_{i+1}(x_{i+1}) + g_{i+1}(X)) + \nu_i \quad (10)$$

for  $i = 1, \dots, N - 1$ . To stabilize (10) different approaches can be undertaken. In the following Section, we use an adaptive law to adjust the coupling strength in the network in order to achieve GS.

### 3 Generalized synchronization design

The design of the local controllers  $\nu_i$  strongly depends on the nature of the nodes and the network topology. In this contribution we make a few simplifying assumptions.

In the first place, we will consider only nodes that either are or can be transformed into a triangular form, e.g. by a coordinate transformation and a feedback linearization controller. Although this may seem a very restrictive condition, a large number of chaotic systems can in fact be transformed to a triangular or at least partially triangularized form with internal dynamics by linearizing feedback. Additionally, chaotic dynamics can be generated from piecewise linear systems that are easily triangularizable (Sprott, 2000; Campos et al., 2010). It follows

from this assumption that an adequate coordinate transform  $\mathcal{T}_i$  exist for each node such that (1) can be rewritten as:

$$\dot{z}_i = A_i z_i + B \psi_i \quad (11)$$

where  $z_i = \mathcal{T}_i x_i = (z_{i1}, z_{i2}, \dots, z_{in})^\top \in \mathbf{R}^n$  are the transform state coordinates of node  $i$ ; the constant matrices  $A_i$  and  $B$  have the controller-type companion form

$$A_i = \begin{bmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 0 & 0 & 0 & \dots & 1 \\ a_{i,1} & a_{i,2} & a_{i,3} & \dots & a_{i,n} \end{bmatrix} \quad B = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix} \quad (12)$$

with  $\psi_i$  the linearizing feedback controller, if such controller is necessary.

The second simplifying assumption is that the network topology affects the vector field of each node in a linear and diffusive way, such that,

$$g_i(Z) = \gamma_i(t) \sum_{j=1}^N c_{ij} \Gamma z_j \quad (13)$$

for  $i = 1, 2, \dots, N$  where  $Z = (z_1, \dots, z_N) \in \mathbf{R}^{n \times N}$  are the node states in the transformed coordinates;  $\Gamma \in \mathbf{R}^{n \times n}$  is a 0–1 inner connection matrix describing the manner in which the state variables of node  $i$  and  $j$  are connected; and  $C = \{c_{ij}\} \in \mathbf{R}^{N \times N}$  is the 0–1 coupling matrix, which captures the topological structure of the network, if  $c_{ij} = 1$  ( $i \neq j$ ), there is a connection of strength between the nodes  $i$  and  $j$  is  $\gamma_i(t)$ , otherwise the nodes are disconnected. As a consequence of the diffusive coupling assumption, the diagonal entries of the coupling matrix satisfy the following equality

$$c_{ii} = - \sum_{j=1, j \neq i}^N c_{ij} = - \sum_{j=1, j \neq i}^N c_{ji}, \quad (14)$$

for  $i = 1, 2, \dots, N$ . Further, assuming that there are no isolated nodes in the network, the eigenvalues of  $C$  have a zero eigenvalue with multiplicity one, all other eigenvalues are real and negative. Under this connection structure in the transformed variables the network in (2) becomes:

$$\dot{z}_i = A_i z_i + B \psi_i + \gamma_i(t) \sum_{j=1}^N c_{ij} \Gamma z_j, \quad (15)$$

for  $i = 1, 2, \dots, N$ . where  $\psi_i$  are the feedback linearizing controllers. To achieve GS in the original variables, in the transformed variables we look for complete identical synchronization of the network. In particular, since the nodes in the transformed variables have a very similar structure (11)-(12), we can argue that their differences are bounded and we can define the average node as reference



for synchronization  $\bar{s} = \frac{1}{N} \sum_{j=1}^N z_j$ . The dynamics of the average node are given by

$$\dot{\bar{s}} = \frac{1}{N} \sum_{k=1}^N (A_k z_k + B \psi_k) + \frac{1}{N} \sum_{k=1}^N (\gamma_k(t) \sum_{j=1}^N c_{kj} \Gamma z_j) \quad (16)$$

Notice that under the assumption that the control actions vanish at the synchronized solution ( $z_1 = z_2 = \dots = z_N = \bar{s}$ ) and as a consequence of the diffusive nature of the coupling (14), once the network is synchronized the second term of (16) is also zero. Then, at the synchronized solution the dynamics of the reference node is the average of the nodes isolated from the network, in the transformed variables that is:

$$\dot{\bar{s}} = \frac{1}{N} \sum_{k=1}^N (A_k z_k + B \psi_k) \quad (17)$$

We define the synchronization error as  $e_i = z_i - \bar{s}$ , from (15) and (16) the error dynamics are given by:

$$\dot{e}_i = \mathcal{A}_i(z_i, \bar{s}) + \gamma_i(t) \sum_{j=1}^N c_{ij} \Gamma e_j \quad (18)$$

for  $i = 1, 2, \dots, N$  with

$$\mathcal{A}_i(z_i, \bar{s}) = A_i z_i + B \psi_i - \frac{1}{N} \sum_{k=1}^N (A_k z_k + B \psi_k)$$

Given that the term  $\mathcal{A}_i$  in (18) is the difference between the current node and the reference node ( $z_i - \bar{s}$ ). If we restrict our attention to chaotic nodes which can be triangularized. Then, is reasonable to expect that this term is bounded, that is, we assume that

$$|\mathcal{A}_i(z_i, \bar{s})| \leq \beta_i (z_i - \bar{s}) \quad (19)$$

for  $i = 1, 2, \dots, N$  with  $\beta_i$  nonnegative constants.

To stabilize the error dynamics (18) we propose to adaptively adjust the coupling strengths  $\gamma_i(t)$  as described in the following result.

**Theorem 1:** For a network of nonidentical nodes that can be transformed into a triangularized form (11), which is linearly and diffusively coupled such that the dynamics of each node in the network are given by (15). Under the assumption (19) described above, using the following adaptive law to adjust the coupling strength of the network:

$$\dot{\gamma}_i(t) = -\alpha_i \sum_{j=1}^N c_{ij} e_i^\top \Gamma e_j \quad (20)$$

where  $\alpha_i$  are positive constants, describing the adaptation speed, for  $i = 1, 2, \dots, N$ . The network in the transformed variables will identically synchronized with the reference node  $\bar{s}$ . Equivalently, in the original variables the network will synchronizes in the generalized sense of (5), in terms of the coordinate transformations  $\mathcal{T}_i$ .

**Proof:** Using the following Lyapunov function candidate

$$V = \frac{1}{2} \sum_{i=1}^N e_i^\top e_i + \frac{1}{2} \sum_{i=1}^N \frac{1}{\alpha_i} [\gamma_i(t) + \gamma^*]^2$$

The time derivative of  $V$  along the trajectories of the error dynamics and the adaptive law is given by

$$\dot{V} = \sum_{i=1}^N e_i^\top \dot{e}_i + \sum_{i=1}^N \frac{1}{\alpha_i} [\gamma_i(t) + \gamma^*] \dot{\gamma}_i(t)$$

From (18) and (20) we have

$$\dot{V} = \sum_{i=1}^N e_i^\top [A_i(z_i, \bar{s}) + \gamma_i(t) \sum_{j=1}^N c_{ij} \Gamma e_j] - \sum_{i=1}^N [\gamma_i(t) + \gamma^*] \left[ \sum_{j=1}^N c_{ij} e_i^\top \Gamma e_j \right]$$

Using (18) we get

$$\dot{V} \leq \sum_{i=1}^N e_i^\top \beta_i e_i + \sum_{i=1}^N \gamma_i(t) e_i^\top \sum_{j=1}^N c_{ij} \Gamma e_j - \sum_{i=1}^N \gamma_i(t) e_i^\top \sum_{j=1}^N c_{ij} \Gamma e_j - \sum_{i=1}^N \gamma^* \sum_{j=1}^N c_{ij} e_i^\top \Gamma e_j$$

Letting  $\beta = \max\{\beta_i | i = 1, 2, \dots, N\}$  and  $k_i = \sum_{j=1, j \neq i}^N c_{ij}$  be the largest bound and the node degree, respectively. Additionally, considering that  $C$  is a symmetric matrix, we can maneuver the indexes in the last term to get

$$\dot{V} \leq \beta \sum_{i=1}^N e_i^\top e_i - \gamma^* \sum_{i=1}^N k_i e_i^\top \Gamma e_i \quad (21)$$

Defining the error vector  $E = [e_1^\top, e_2^\top, \dots, e_N^\top]^\top$  and the matrix  $P = K \otimes \Gamma$ , with  $K = \text{Diag}(k_1, k_2, \dots, k_N)$  and  $\otimes$  the Kronecker product. The inequality in (21) can be rewritten as:

$$\dot{V} \leq \beta E^\top E - \gamma^* E^\top P E \leq E^\top \left[ \beta - \gamma^* \lambda_{\min} \left( \frac{P + P^\top}{2} \right) \right] E$$

It follows that by letting  $\gamma^*$  be a sufficiently large positive constant, we have that

$$\dot{V} \leq E^\top E$$

Then, the error dynamics in (18) are asymptotically stable about the zero fixed point ( $e_i = 0$ ) when the coupling strengths are adjusted according to (20), which implies that the network in the transformed variables achieves identical complete synchronization. In consequence, the dynamical network in the original coordinates achieves GS in the sense of (5), with respect to the coordinate transformations  $\mathcal{T}_i$ .

Q.E.D.

#### 4 Illustrative example

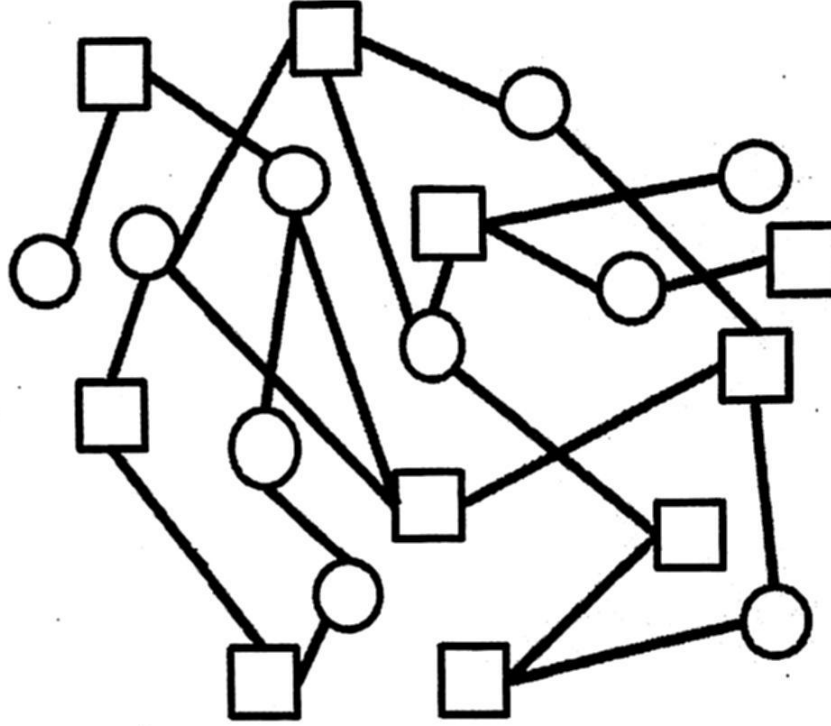
We consider a network with two different types of nodes which can be triangularized. Namely, Sprott circuits (○)(Sprott, 2000):

$$\begin{aligned}\dot{x}_1 &= x_2 \\ \dot{x}_2 &= x_3 \\ \dot{x}_3 &= -0.6x_3 - x_2 - 1.2x_1 + 2\text{sgn}(x_1) + u\end{aligned}\tag{22}$$

which is already in triangular form. The other type of nodes are Rössler systems (□) :

$$\begin{aligned}\dot{x}_1 &= -x_2 - x_3 \\ \dot{x}_2 &= x_1 + 0.1x_2 \\ \dot{x}_3 &= x_3(x_1 - 14) + 0.1 + u\end{aligned}\tag{23}$$

The different nodes are connected randomly according to the ER network model



**Fig. 1.** Network of non-identical nodes (Sprotts:○ and Rössler:□).

(Newman, 2010). A possible realization with twenty nodes is shown in Figure 1. To achieve GS in the network we use a coordinate transformation for the Rössler system. Assuming that the output of (23) is  $y = x_2$ , the following coordinate transformation takes the Rössler system to a triangular form with the transform variables  $z = \phi(x)$  (Femat et al., 2005):

$$\begin{pmatrix} z_1 \\ z_2 \\ z_3 \end{pmatrix} = \begin{pmatrix} x_2 \\ x_1 + 0.1x_2 \\ 0.1x_1 + (0.1^2 - 1)x_2 - x_3 \end{pmatrix}\tag{24}$$

This coordinate transformation is a diffeomorphism and its inverse is:

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} z_2 - 0.1z_1 \\ z_1 \\ 0.1z_2 - z_3 - z_1 \end{pmatrix} \quad (25)$$

To achieve GS in the network coupling strengths are adjusted adaptively according to Theorem 1. In Figure 2 the trajectories of the Sprott circuits in the the Rössler in its transformed coordinates are shown. In Figure 3, the synchronization error in the original coordinates is presented, as shown the adaptive coupling strength produces an identical synchronization of the network in the transformed variables. As such, GS with the mapping function  $H_i$  is obtained on the network. The error in the original coordinates is shown in Figure 3.

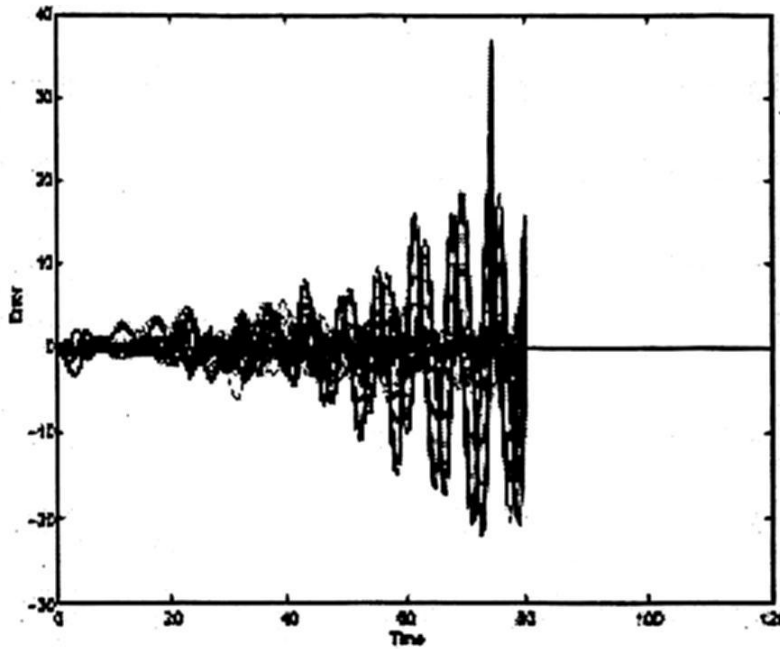


Fig. 2. Synchronization error on the transformed coordinates  $z(t)$

## 5 Conclusions

On networks with different dynamical nodes complete identical synchronization is not directly achievable. As such, alternative interpretations of the synchronization phenomena are necessary. In this contribution, we investigate the emergence of GS in a network of non identical nodes via adaptive control. In particular, we consider nodes that can be exactly linearized by state feedback. Under such conditions, GS can be achieved by imposing a functional relationship between the nodes in the network. There are limitations of the proposed method, for example the necessity of triangularizing state feedback controllers. Additionally, the imposed functional relationship between the nodes is fixed by coordinate



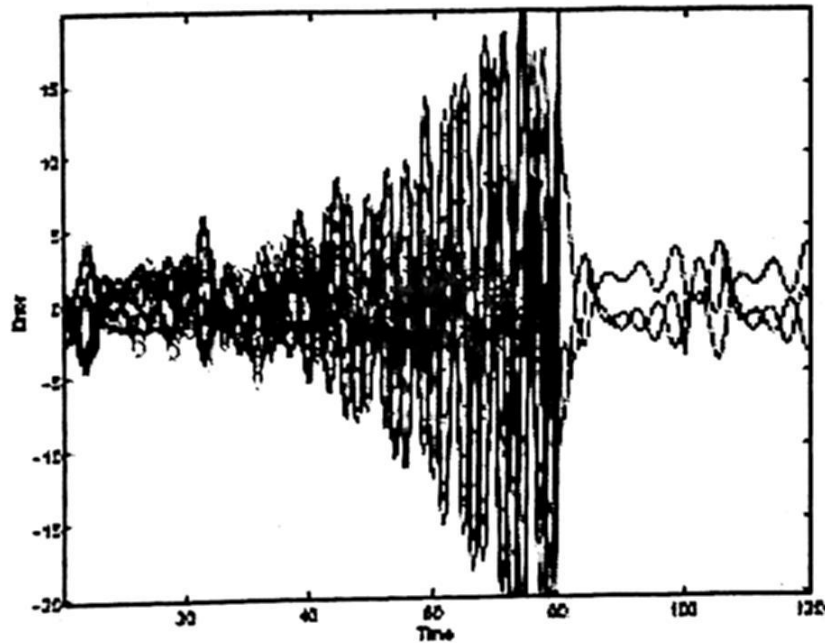


Fig. 3. Synchronization error on the original coordinates  $x(t)$

transformation. However, it seems possible to overcome these restrictions by considering alternative ways to design the synchronizing controllers. These are considerations for future work and will be reported elsewhere.

## 6 Acknowledgements

This work was supported in part by CONACYT Research Council of México under Grants CB-2008 106915-Y.

## References

- Abarbanel, H.D.I., Rulkov, N.F., Sushchik, M.M. "Generalized synchronization of chaos: The auxiliary system approach", *Phys. Rev. E*, 53(5), 4528–4535, 1996.
- Arenas, A., Díaz-Guilera, A., Kurths, J., Moreno, Y., Zhou, C. "Synchronization in complex networks" *Physics Reports*, 469, 93–153, 2008.
- Boccaletti, S., Kurths, J., Osipov, G., Valladares, D. L., Zhou, C.S. "The synchronization of chaotic systems," *Phys Rep*, 366, 1–101, 2002
- Boccaletti, S., Latora, V., Moreno, Y., Chavez, M., Hwang, D. U. *Complex networks: Structure and dynamics Physics Reports*, 424, 175–308, 2006
- Femat, R., Kocare, L., van Gerven, L., Monsivais-Perez, M. L. "Towards generalized synchronization of strictly different chaotic systems," *Phy Lett A*, 342, 247–255, 2005
- Guan, S.G., Wang, X.G., Gong, X.F., Li, K., Lai, C.H. "The development of generalized synchronization on complex networks", *Chaos*, 19, 013130, 2009.
- Hung, Y.C., Huang, Y.T., Ho, M.C., Hu, C.K. "Paths to globally generalized synchronization in scale-free networks", *Phys. Rev. E*, 77, 016202, 2008.

- Liu, H., Chen, J., Lua, J.-a., Cao, M. "Generalized synchronization in complex dynamical networks via adaptive couplings" *Physica A*, 389, 1759–1770, 2010
- Newman, M. *Networks: An introduction* Oxford University Press, USA, 2010.
- Sprott, J. C. "Simple chaotic systems and circuits", *Am. J. Phys.*, 68(8), 758–763, 2000.
- Campos-Cantón, E., Barajas-Ramírez, J. G., Solís-Perales, G., Femat, R. "Multiscroll attractors by switching systems", *Chaos*, 20, 013116, 2010.
- Wang, X., Chen, G. "Synchronization in small-world dynamical networks," *Int. J. Bifur. Chaos*, 12(1), 187-192, 2002
- Wang, X., Chen, G. "Synchronization in scale-free dynamical networks: robustness and fragility," *IEEE Trans. Circuits Syst. I*, 49(1), 54-62, 2002
- Wu, C. W. *Synchronization in coupled chaotic circuits and systems*, World Scientific, Singapore, 2002
- Xu, X., Chen, Z., Si, G., Hu, X., Luo, P. "A novel definition of generalized synchronization on networks and a numerical simulation example", *Comput. Math. Appl.*, 56(11), 2789–2794, 2008.

# The use of an evolutionary algorithm for the parameter selection of predictive controllers

R.C. Gutiérrez-Urquidez, G. Valencia-Palomo and O.M. Rodríguez-Elias

Instituto Tecnológico de Hermosillo, Av. Tecnológico S/N, Hermosillo, Mexico.  
ro\_gutierrez@ith.mx, gvalencia@ith.mx, omrodriguez@ith.mx

*Paper received on 11/19/13, Accepted on 01/19/14.*

**Abstract.** In the design of linear predictive controllers (MPC), a problem that has not yet been fully resolved, is how to determine the best strategy for the selection of the tuning parameters in order to obtain good performance and a good feasibility region while maintaining a sensible low computational burden for practical implementation. The main contribution of this paper is to achieve a systematic tuning by use of a multi-objective evolutionary algorithm (MOEA) on predictive control algorithm that has been reparameterized with Laguerre functions. Numerical simulations show that MOEA is a useful tool to obtain consistently good solutions for the selection of MPC tuning parameters.

## 1 Introduction

Although predictive control (MPC) has its background well established [1, 2], and is widely used, there are still some theoretical and practical problems to be solved. For instance, one key conflict is between feasibility and performance: if a MPC controller is well tuned to provide high performance, feasibility will have (in general) a very small region, unless one use a large number of decision variables (degrees of freedom or d.o.f.), which implies increasing the computational load of the algorithm. On the other hand, if one aims to get better feasibility with a fixed d.o.f., the result will be a detuned controller with relatively poor performance. Several authors have looked at this problem and have proposed different strategies that provide improved quality of predictions and maintain a balance between performance and computational complexity. The first MPC algorithms, such as Dynamic Matrix Control (DMC) [3] or Generalised Predictive Control (GPC) [4], will usually give reasonable performance, but only for large input and output horizons over the settling time. Furthermore, this may not be so effective when the open-loop dynamics of the system is poor or simply if there are state or output constraints. Another drawback is that they do not automatically guarantee stability, thus requiring further tuning considerations [5].

Concern for stability has been a major engine for generating different formulations of MPC. At 1990s, proposals arise to amend this problem of optimal open-loop control, so that closed-loop stability can be guaranteed. The most accepted approach for an a priori recursive feasibility, is the dual-mode MPC paradigm [5, 6], for which, the predictions have two modes: (i) a transient phase containing the d.o.f. and (ii) a terminal mode with guaranteed convergence. Following the dual-mode approach, other

work has looked at alternative ways of formulating the d.o.f. for optimisation, for instance by interpolation methods [7]. However, these methods do not currently extend well to large dimensional systems. Another interesting work considered the so-called triple mode strategies [8], where one embeds a smooth transition between a controller with good feasibility and other with good performance into a single model and use the decision variables to improve performance/feasibility further. This work is successful but relies on heavy computation and algebra and thus may be difficult for industrial implementations.

Recently, several works have been proposed for more efficient predictive control algorithms by reparametrization of the d.o.f. of the optimization problem. An alternative is to use orthogonal functions either with Laguerre/Kautz polynomials or through Generalised orthonormal functions [9–11], since they have proven to be very effective for improving the volume of the feasible region with a limited number of d.o.f. with almost no performance loss. And finally, it is also possible to reparameterize the d.o.f. using directional information for the specific control problem [12].

In summary, the recent MPC algorithms differ in the way of reparameterizing the d.o.f., but introducing one or more tuning parameters that has to be selected by the user. The original papers [9–12] fail to give guidelines in this selection, since in most cases, the selection of tuning parameters is still performed based on trial and error simulations. Therefore, the main debate is to establish the best strategy for the selection of controller tuning parameters that allow to efficiently handle the trade-off between feasibility, computational burden and controller performance. This paper provides an alternative that uses multi-objective evolutionary algorithms, in order to achieve a systematic tuning of a MPC. Specifically, the focus of this work is the tuning of a MPC algorithm whose d.o.f.s have been reparameterized using Laguerre functions. This paper is organized as follows: Sections 2 and 3 will give the necessary background about modelling, predictive control, Laguerre optimal predictive control (LOMPC) and multi-objective evolutionary algorithms (MOEAs). Section 4 presents and develops the proposed tuning algorithm for LOMPC with MOEAs. Section 5 gives a numerical example showing the efficacy of the proposed algorithm. Finally, the conclusions are presented in Section 6.

## 2 Model predictive control and Laguerre functions

### 2.1 Optimal predictive control (OMPC)

Model predictive control (MPC) has had a peculiar evolution; it was initially developed in industry, at 70's, and later was taken by academic sector. Early predictive controllers were based in heuristic algorithms until the research community established their theoretical support. Predictive Control, also known as *receding control horizon* refers to a set of control algorithms and techniques which lead to the design of controllers with a similar structure; with information based on past inputs and outputs of the plant, and making use of the process model, an optimal control input is obtained by minimizing an objective function (cost function) in a time interval named control horizon. The common elements in MPC controllers are:



- **The process model.** The mathematical model is used to generate system predictions. In particular, this model must show the dependence of the output on the current measured variable and the current/future inputs. A discrete-time state-space model is assumed, which has the following form:

$$\mathbf{x}_{k+1} = \mathbf{A}\mathbf{x}_k + \mathbf{B}\mathbf{u}_k; \quad \mathbf{y}_k = \mathbf{C}\mathbf{x}_k + \mathbf{D}\mathbf{u}_k; \quad (1)$$

with  $\mathbf{x}_k \in \mathbb{R}^n$ ,  $\mathbf{y}_k \in \mathbb{R}^l$ ,  $\mathbf{u}_k \in \mathbb{R}^m$  which are the state vectors, the measured output and the plant input respectively. This work also adopts an independent model approach with optimal feedback  $\mathbf{K}$ . Let  $\mathbf{w}_k$  the output of the independent model, hence, the estimated disturbance is  $\hat{\mathbf{d}}_k = \mathbf{y}_k - \mathbf{w}_k$  [3]. Disturbance rejection and offset free tracking will be achieved using the offset form of state feedback [13, 14], that is:

$$\mathbf{u}_k - \mathbf{u}_{ss} = -\mathbf{K}(\mathbf{x}_k - \mathbf{x}_{ss}), \quad (2)$$

where  $\mathbf{x}$  is the state of the independent model and  $\mathbf{x}_{ss}$ , are estimated values of the steady-states giving no offset; these depend upon the model parameters and the disturbance estimate [14].

- **The performance index** (objective function or cost function). This is used to quantify the deviation of the measured output with respect to the desired output and the control effort. A way of defining the cost function is by using deviation variables to penalize the deviation of absolute input on the expected steady state input:

$$\mathbf{J}_k = \sum_{i=0}^{n_y} (\mathbf{x}_{k+i} - \mathbf{x}_{ss})^T \mathbf{Q} (\mathbf{x}_{k+i} - \mathbf{x}_{ss}) + \sum_{i=0}^{n_u-1} (\mathbf{u}_{k+i} - \mathbf{u}_{ss})^T \mathbf{R} (\mathbf{u}_{k+i} - \mathbf{u}_{ss}), \quad (3)$$

with  $\mathbf{Q}$  and  $\mathbf{R}$  positive definite state and input cost weighting matrices,  $n_y$  is the prediction horizon and  $n_u$ , the control horizon. Because in practice all real processes are subject to constraints, these need to be considered in the cost function. One of the major selling points of MPC is its ability to do online constraint handling in a systematic fashion, retaining to some extent the stability margins and performance of the unconstrained law. Let the system be subject to constraints of the form:

$$\left. \begin{aligned} \mathbf{u}_{min} &\leq \mathbf{u}_k \leq \mathbf{u}_{max}; \\ \Delta \mathbf{u}_{min} &\leq \mathbf{u}_{k+1} - \mathbf{u}_k \leq \Delta \mathbf{u}_{max}; \\ \mathbf{y}_{min} &\leq \mathbf{y}_k \leq \mathbf{y}_{max}. \end{aligned} \right\} \forall k. \quad (4)$$

- **The optimization algorithm** determines the optimal control input that minimizes the cost function imposed by a set of constraints. Minimizing the cost function subject to both current and future constraints and obtaining control action, the optimization results in a quadratic program (QP).
- **The receding horizon.** Although the optimal trajectory of future control signal is completely described within the moving horizon window, the actual control input to the plant only takes the first sample of the control signal,  $\mathbf{u}_k$ , while neglecting the rest of the trajectory [15]. The horizon selected for predictions should include all significant dynamics; otherwise performance may be poor and important events may be unobserved.

The key idea of Optimal MPC (OMPC) [5] is to embed into the predictions the unconstrained optimal behaviour and handle constraints by using perturbations about this. The closed-loop paradigm uses perturbations as d.o.f. for the optimal control law without constraints. Disturbance rejection and offset free tracking will be achieved using the offset form of state feedback of (2) [2]. For convenience, d.o.f. can be reformulated in terms of a new variable  $c_k$ . Hence, assuming  $K$  is the optimal feedback, the input predictions are defined as follows:

$$u_{k+i} - u_{ss} = \begin{cases} -K(x_{k+i} - x_{ss}) + c_{k+i}; & i \in \{1, 2, \dots, n_c - 1\} \\ -K(x_{k+i} - x_{ss}); & i \in \{n_c, n_c + 1, \dots\} \end{cases}, \quad (5)$$

where the perturbations  $c_k$  are the d.o.f. for optimization; conveniently summarised in vector:  $\underline{c}_k = [c_k^T, c_{k+1}^T, \dots, c_{k+n_c-1}^T]^T$ . Input predictions (5) and state associated (1) which satisfies constraints (4):  $Mx_k + N\underline{c}_k \leq f(k)$ . In practice, if an unconstrained optimal prediction may violate a constraint defined in (4), prediction class more suitable shall be used according (5). The OMPC algorithm can be summarized [2]:

$$\underline{c}_k^* = \arg \min_{\underline{c}_k} \underline{c}_k^T W_j \underline{c}_k \quad \text{s.t.} \quad M_j x_k + N_j \underline{c}_k \leq f(k). \quad (6)$$

Use  $\underline{c}_k^*$  to construct the input (5).

OMPC algorithm has implied LQR theory and is able to find a global optimum on the objective function. If one chooses a value for  $K$  in (5) to become a optimal Linear-Quadratic-Regulator(LQR)[5], the feasible region depends only on the class of prediction and hence also the number of free movements, that is,  $n_c$ .

**Remark 21** The optimization of (6) can require a large  $n_c$  (d.o.f.) to obtain both good performance and a large feasible region.

**Definition 21** *Maximum Admissible Set (MAS).* A common method to achieve recursive feasibility is to find the region of the state space where positively invariant sets ensure the action of an unconstrained control law but satisfy all constraints in the future. This achieved using the dual-mode paradigm. And the greatest invariant set possible for use as the terminal state set is referred as Maximum Admissible Set (MAS) [6, 2]. For a linear discrete system, observable, pre-stabilized by a gain  $K$  of state feedback, associated with a set of constraints (4), there exists a set, MAS, finite and where the constraints are satisfied for all future time intervals:  $MAS = \{x_k \in \mathbb{R}^n \mid Mx_k \leq d\}$ .

**Definition 22** *Maximal Controllable Admissible Set (MCAS).* It is also possible to define a region in  $x$  in which it is possible to find a  $c_k$  such that at the future trajectory satisfying the constraints:  $MCAS = \{x_k \in \mathbb{R}^n \exists \underline{c}_k \in \mathbb{R}^{n_c m} \text{ s.t. } Mx_k + N\underline{c}_k \leq d\}$ ; and this is named Maximal Controllable Admissible Set (MCAS).

## 2.2 LOMPC: Laguerre polynomials and OMPC

The fundamental weakness of OMPC algorithms is that the d.o.f. are parameterized as individual values at specific samples and have an impact over just one sample and

thus have a limited impact on feasibility. If the initial state is far away from the MAS associated to  $c_k = 0$ , the  $n_c$  steps will be insufficient to move into the MAS. Laguerre OMPC (LOMPC) is a dual-mode MPC algorithm where the d.o.f. within the input predictions are parameterized in terms of Laguerre polynomials rather than using the more normal standard basis set. The algorithm proposes to replace common decision variables  $u_k$  and  $c_k$  by Laguerre polynomials  $L_i$  in OMPC. It has been shown that with the reparameterization of d.o.f. get an increase in feasibility region (MCAS) of controller LOMPC regarding equal number of d.o.f. in OMPC. The  $z$ -transform of discrete Laguerre polynomials are defined as follows:

$$\Gamma_i(z) = \sqrt{1-a^2} \frac{(z^{-1}-a)^{n-1}}{(1-az^{-1})^n}; \quad 0 \leq a \leq 1. \quad (7)$$

With the inverse  $z$ -transform of  $\Gamma_n(z, a)$ , denoted by  $l_{(k,n)}$ , the Laguerre functions set are the vector:  $L_k = \{l_{k,1}, l_{k,2}, \dots, l_{k,n}\}^T$ . The size of the  $A_L$  matrix, is  $n \times n$ ; and it is a function of the parameters  $a, \beta = 1 - a^2$  and initial condition  $L_0$ , so that:

$$L_{k+1} = \underbrace{\begin{bmatrix} a & 0 & 0 & 0 & \dots \\ \beta & a & 0 & 0 & \dots \\ -a\beta & \beta & a & 0 & \dots \\ a^2\beta & -a\beta & \beta & a & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}}_{A_L} L_k, \quad L_0 = \sqrt{1-a^2} [1, -a, a^2, \dots]^T. \quad (8)$$

The basic concept of OMPC is preserved [5, 2], that is the predictions take the form of (5) and optimality dynamics included in the predictions. However, a key difference is that the disturbance (terms  $c_k$ ) is defined with Laguerre polynomials instead of taking the d.o.f. individually. The relevant link between Laguerre and predicted values of  $\underline{c}_k$ , are summarized in the following equation:

$$\underline{c}_k = [c_k^T, \dots, c_{k+n_c-1}^T]^T, \quad \underline{c}_k(L) = \underbrace{[L_0^T, L_1^T, \dots]^T}_{H_L} \underline{\eta}_k = H_L \underline{\eta}_k. \quad (9)$$

Now, with  $L_{k+1} = A_L * L_k$ , the decision variable is  $\underline{\eta}_k$ ; and substituting predictions in (8), get the LOMPC optimization, which is:

$$\underline{\eta}_k^* = \arg \min_{\underline{\eta}_k} \underline{\eta}_k^T \left\{ \sum_{i=0}^{\infty} A_L^i L_0 S L_0^T A_L^{iT} \right\} \underline{\eta}_k = \underline{\eta}_k^T S_L \underline{\eta}_k \quad (10)$$

$$s.t. \quad Mx_k + N H_L \underline{\eta}_k \leq d. \quad (11)$$

### 3 Multi-Objective Evolutionary Algorithms (MOEA)

Evolutionary algorithms (EA) allows flexible representation of decision variables and performance evaluation; also are robust and methodological tool for search and optimization, with the ability to work in environments that include discontinuities, time

variance, bad behavior, multi-modality, uncertainty and noise. The EA are increasingly accepted in the control community, their applications are mainly in the off-line design and online optimization [16].

EA are systematic methods for solving search and optimization problems in which apply the same methods of the biological evolution, based on selection, reproduction and mutation of the population. These algorithms transform a set of individual mathematical objects using operations that are modeled according to the Darwinian principle of reproduction and survival of the fittest [17].

The simplest form of an optimization problem is to consider the existence of a single criterion or objective in the search for a solution. However, when there are multiple objectives to optimize, there is no single solution, but rather a set of compromise solutions together. These problems are called Multi-objective Problem (MOP). Most EA multi-objective optimization techniques use the Pareto concept. Generally, must be found a vector of decision variables,  $\vec{x}^*$ , which optimizes a vector of objective functions,  $\vec{f}(x)$ , and satisfies certain inequality constraints,  $g_i(x)$ , or equal,  $h_i(x)$  [18, 19]. Both the objective functions and constraints are functions of the decision variables. Several types of MOEAs exist; each uses different mechanisms of selection, crossover and elitism in particular, but for this paper we use the algorithm NSGA-II (Nondominated Sorting Genetic Algorithm II), proposed by [20]. NSGA-II is an improved version of NSGA that modifies the mechanism for diversity preserving and incorporates an explicit mechanism for elitism; leave the use of Sharing distance of the NSGA to use Crowding tournament selection operator as a method of diversity preservation.

## 4 Tuning LOMPC

In the work previously presented, it is shown that the reparameterization of d.o.f. has advantages over conventional approaches. However, there is still no systematic way to define the controller tuning parameters, this mainly due to the compromise between the objectives to be optimized. It is clear that the problem to be solved is a multi-objective optimization, where the search space is fairly large, which justifies the use of MOEAs. The purpose of this work is to develop a method to systematically choose the optimal values of the tuning parameters,  $a$  and  $n_c$ , of an MPC whose d.o.f. have been reparameterized with Laguerre functions; in order to guarantee the best trade-off between feasibility, performance and computational load. Therefore, there are two decision variables  $a$  and  $n_c$  and three optimizing conditions:

1. Maximize the feasibility region,  $\max f_{a,n_c}(\vartheta)$ .
2. Minimize the performance loss,  $\min f_{a,n_c}(\beta)$ .
3. Minimize the computational burden,  $\min f_{a,n_c}(\rho)$ .

Also the constraints associated with the parameter selection must be added to the multi-objective optimization problem:

$$0 \leq a \leq 1; \quad 1 \leq n_c \leq n_{c,max}; \quad n_c \text{ integer.} \quad (12)$$



#### 4.1 Feasibility evaluation ( $\vartheta$ )

In order to estimate the normalized volume, first is defined a polytope  $\mathbf{P}_{opt}$  as the global MCAS of OMPC with a large number of degrees of freedom, able to represent the largest feasible region that can be obtained by the controller (usually  $n_c \geq 20$ ) [11]:  $\mathbf{P}_{opt} = \{(x, c) \mid \mathbf{M}\mathbf{x}_k + \mathbf{N}\mathbf{c}_k \leq \mathbf{d}\}$ . Also it is defined a polytope  $\mathbf{P}_{HL}$  corresponding to proposed parameterization MCAS:  $\mathbf{P}_{HL} = \{(x, \eta) \mid \mathbf{M}_L\mathbf{x} + \mathbf{N}_L\eta \leq \mathbf{d}\}$ ; where  $\mathbf{P}_{HL}$  is the polytope sliced by the parameterization matrix  $\mathbf{H}_L$ . The volume of  $\mathbf{P}_{opt}$  and  $\mathbf{P}_{HL}$  polytopes, represent the feasible regions or feasible volumes for each type of algorithm.

The volume calculation of a high dimensional polytope is a complex task and the computing time for these polytopes can be prohibitive; consequently, this paper approximates the volume by computing the average distance from the origin to the boundary of the associated MCAS (radius). First select a large number of equi-spaced (by solid angle) or random directions in the state space i.e.  $\mathbf{x} = [x_1, \dots, x_n]$  and then, the distance from the origin to the boundary of MCAS is determined by solving a linear programming (LP) for each direction  $\mathbf{x}_i$  selected. Greater distances imply bigger feasible region. The objective function for evaluating the normalized feasible volume is then:

$$\vartheta = \frac{\text{vol}(\mathbf{P}_{HL})}{\text{vol}(\mathbf{P}_{opt})}. \quad (13)$$

#### 4.2 Performance evaluation ( $\beta$ )

The performance evaluation it is done by realizing the calculation for the  $n$ -points  $\mathbf{x}_i$  selected, they are represented by the optimized values of the associated cost function, i.e.  $J_{opt}(\mathbf{x}_i)$  and  $J_{HL}(\mathbf{x}_i)$ . To ensure fairness in comparison of these values, scaling is used (setting) in one direction  $\mathbf{x}_i$  given [11]. The objective function for evaluating the normalized performance is:

$$\beta = \frac{1}{n} \sum_{i=1}^n \frac{J_{HL}(\mathbf{x}_i)}{J_{opt}(\mathbf{x}_i)}. \quad (14)$$

#### 4.3 Computational load evaluation ( $\rho$ )

It is demonstrated that the re-parameterization of d.o.f. proposed in LOMPC algorithm is able to achieve great feasibility regions while maintaining an acceptable local optimality within a relatively low computational complexity compared to conventional OMPC approaches. However, this reduction in d.o.f. not necessarily results in a reduction of the complexity of optimization and therefore the computational load, since the resulting quadratic programming of reassignment is denser (heavier) than for OMPC [11]. So, how one can determine the minimum number of d.o.f., which gives the best performance and the largest feasible region? One alternative is to compare the online computational load for LOMPC and OMPC as a function of the number of floating point operations per second (flops) required for each algorithm. For OMPC, the computational complexity is linear with respect to the horizon length and cubic respect to

state and the input dimension, so that:

$$\varrho(OMPC) = n_c + n_x^3 + n_u^3 \quad (flops). \quad (15)$$

For LOMPC, their computational load is cubic in number of d.o.f., the state and input dimensions [11]:

$$\varrho(LOMPC) = n_c^3 + n_x^3 + n_u^3 \quad (flops). \quad (16)$$

## 5 Numerical example

Consider the following discrete-time state-space model with constraints:

$$\mathbf{x}_{k+1} = \begin{bmatrix} 0.6 & -0.4 \\ 1.0 & 1.4 \end{bmatrix} \mathbf{x}_k + \begin{bmatrix} 0.20 \\ 0.05 \end{bmatrix} \mathbf{u}_k; \quad \mathbf{y}_k = [1.0 \ -2.0] \mathbf{x}_k;$$

$$-1.5 \leq \mathbf{u}_k \leq 0.8; \quad |\Delta \mathbf{u}_k| \leq 0.4, \quad \mathbf{x}_k \leq 5; \quad \mathbf{Q} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}; \quad \mathbf{R} = 2.$$

The objective is to compare OMPC ( $n_c = 3$ ), OMPCopt ( $n_c = 20$ ) with LOMPC tuned using EA. This example utilizes the Matlab Multi-objective Optimization Toolbox that allows the use of a variant of NSGA-II [21]. The EA were run using a search space generated by varying  $n_c \in \{2, 3, 4, 5\}$  and  $a \in [0, 1]$ . The Pareto front obtained is shown in figure 1. Table 1 summarizes the LOMPC tuning parameters obtained with EA and the objective functions evaluated with these. This shows that performance and feasibility improve as  $n_c$  is increased, and as expected, the computational load too. However, within the various tuning parameters found with the EA, it can be selected any combination of these and achieve the best trade-off between the above objectives. Table 1 also includes the respective calculated values for OMPC using the same d.o.f. for comparison purposes. Figures 2, 3 and 4 show the feasibility, performance and computational load evaluation for each pair of tuning parameters with EA and indicate clearly where there are located the optimal values on the LOMPC search space.

**Table 1.** Tuning parameters,  $a$  y  $n_c$ , obtained with the NSGA-II algorithm in LOMPC

Tuning parameter		Average radius to MCAS, $J\vartheta$		Performance $J\beta$		Computational load (flops) $J\varrho$	
$n_c$	$a$	OMPC	LOMPC	OMPC	LOMPC	OMPC	LOMPC
2	0.6208	0.4522	0.6339	1.0	1.0526	11	17
3	0.7915	0.5503	0.7345	1.0	1.0528	12	36
3	0.7916	0.5503	0.7345	1.0	1.0528	12	36
4	0.6548	0.6265	0.7833	1.0	1.0528	13	73
4	0.6688	0.6265	0.7827	1.0	1.0528	13	73

Using one of the selected tuning options, i.e.  $a = 0.7915$ ,  $n_c = 3$ , figure 5 shows the plots of the controller simulation for: (i) an initial state in the MCAS of both, OMPC and

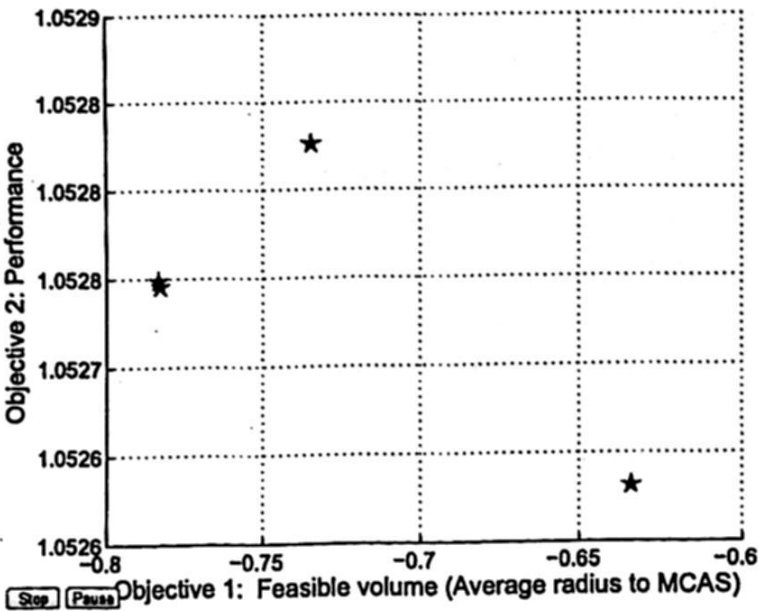


Fig. 1. Pareto front.

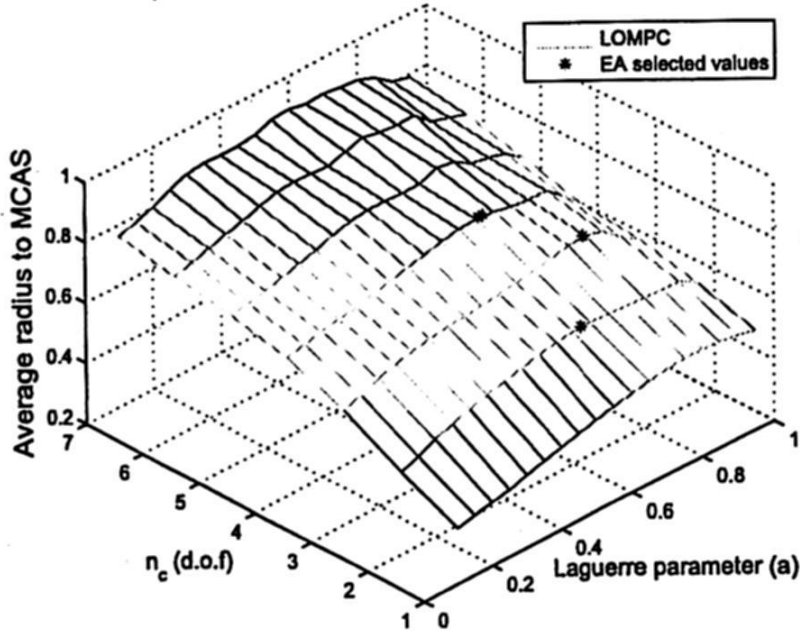


Fig. 2. Feasibility function of average radius to MCAS

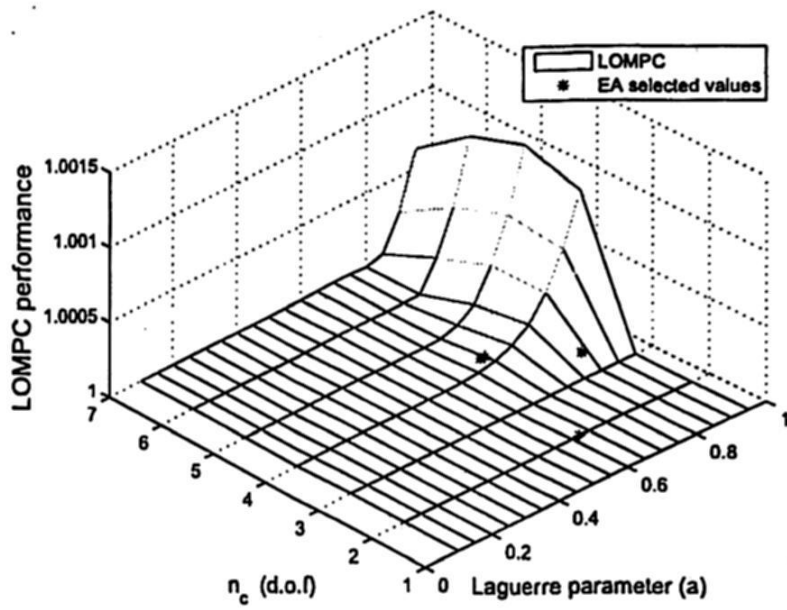


Fig.3. Performance evaluation.

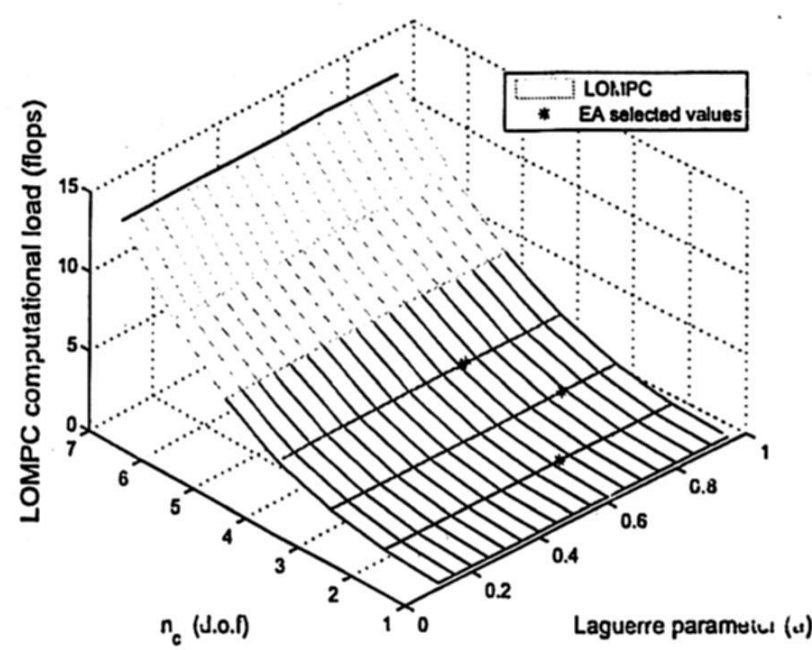


Fig.4. Computational load evaluation.



LOMPC (left); (ii) an initial state in the MCAS of LOMPC, but outside of the MCAS of OMPC (right), where the latter is infeasible. Once again, it is clear that LOMPC with EA is better than OMPC under similar conditions (d.o.f.) and very similar to the global optimum.

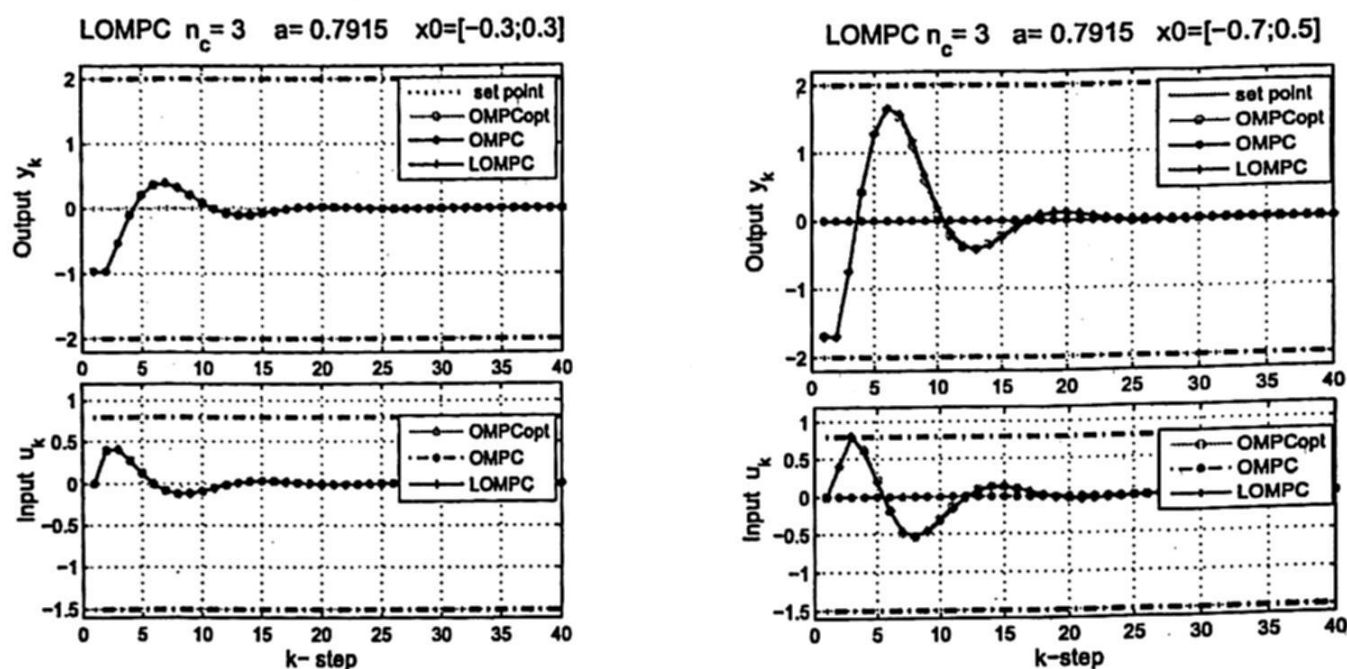


Fig. 5. Controllers simulation with different initial conditions.

## 6 Conclusions

Correct selection of the tuning parameters of the controller guarantees the best balance between closed-loop performance, feasible region and computational load. The main contribution of this work is to achieve a systematic tuning for predictive control algorithm, LOMPC. In this paper, it has been shown that MOEA can be a useful tool for obtaining one or more solutions to select the tuning parameters of an efficient MPC algorithm even when the search space is large and when there are included different objectives compromised to each other. It can be verified that this selection offer a proper balance in the trade-off between the three objectives and outperform OMPC in those objectives, under the same conditions and even for systems with constraints. However, this are just preliminary results and future work will include other efficient MPC approaches and systems with more challenging behavior.

## References

1. Camacho, E.F. and Bordons, C.: Control predictivo: Pasado, presente y futuro. *Revista Iberoamericana de Automática e Informática Industrial* 1:3, 5–28 (2004).
2. Rossiter, J.A.: *Model Predictive Control: A practical approach*. Taylor and Francis (2005)
3. Cutler, C.R. and Ramarker, B.C.: *Dynamic matrix control - a computed algorithm*. (1979).

4. Clarke, D.W. ; Mohtadi, C. and Tuffs, P.S.: Generalized predictive control: the basic algorithm. *Automatica* **23**, 137–148 (1987).
5. Scokaert, P.O.M. and Rawlings, J.B.: Constrained linear quadratic regulation. *IEEE Transactions on Automatic Control* **43**:8, 1163–1168(1998).
6. Mayne, D.Q. ; Rawlings, J.B. ; Rao, C.V. and Scokaert, P.O.: Constrained model predictive control: Stability and optimality. *Automatica* **36**:6, 789–814, (2000).
7. Bacic, M. ; Cannon M. ; Lee, Y.I. and Kouvaritakis, B.: General interpolation in mpc and its advantages. *IEEE Transactions on Automatic Control* **48**:6, 1092–1096 (2003).
8. Imsland, L. ; Rossiter, J.A. ; Pluymers, B. and Suykens, J.: Robust triple mode mpc. *International Journal of Control* **81**:4, 679–687 (2008).
9. Valencia-Palomo, G. and Rossiter, J.A.: Using laguerre functions to improve efficiency of multi-parametric predictive control. In: *Proceedings of the American Control Conference, Baltimore, USA* (2010)
10. Khan, B. ; Valencia-Palomo, G. and Rossiter, J.A.: Exploiting kautz functions to improve feasibility in mpc. In: *Automatic Control and Systems Eng. IFAC World Congress, Milán, Italia* (2011)
11. Khan, B. and Rossiter, J.A.: Alternative parameterisation within predictive control: a systematic selection. *International Journal of Control* **86**:8, 1397–1409 (2013).
12. Valencia-Palomo, G. ; Rossiter, J.A. ; Jones, C.N. and Gondhalekar, R.: Alternative parameterisations for predictive control: how and why? In: *American Control Conference, San Francisco, CA, USA* (2011)
13. Muske, K.R. and Rawlings, J.B.: Model predictive control with linear models. *American Institute of Chemical Engineers (AIChE)* **39**:2, 262–287 (1993).
14. Rossiter, J.A.: A global approach to feasibility in linear mpc. In: *Proceedings of the International Control Conference, Glasgow, Scotland* (2006)
15. Maciejowski, J.M.: *Predictive control with constraints*. Pearson Education (2006)
16. Fleming, P.J. and Purshouse, R.C.: Evolutionary algorithms in control systems engineering: a survey. *Control Engineering Practice* **10**, 1223–1241 (2002).
17. Fogel, D.B.: *Evolutionary Computation: Toward a New Philosophy of Machine Intelligence*. 3 edn. John Wiley and Sons, Inc., New York, USA (2006)
18. Coello-Coello, C.A. ; Lamont, G.B. and Van-Veldhuizen, D.A.: *Evolutionary Algorithms for Solving Multi-Objective Problems*. 2 edn. Genetic and Evolutionary Computation, New York, USA (2007)
19. Zitzler, E.: *Evolutionary Algorithms for Multiobjective Optimization: Methods and Applications*. PhD thesis, Institut für Technische Informatik und Kommunikationsnetze; Swiss Federal Institute of Technology Zurich, Computer Engineering and Networks Laboratory (1999)
20. Deb, K. ; Agrawal, S. ; Pratab, A. and Meyarivan, T.: A fast elitist non-dominated sorting geneticalgorithm for multi-objective optimization: Nsga-ii. In: *Proceedings of the Parallel Problem Solving from Nature VI Conference*. 849–858 (2000).
21. Deb, K. ; Pratap, A. ; Agarwal, S. and Meyarivan, T.: A fast and elitist multiobjective genetic algorithm nsga-ii. *IEEE Transactions on Evolutionary Computation* **6**:2, 182–197(2002).

# Model-Based Sensorless Controller of a Permanent Magnet Synchronous Motor

Luis N. Coria, Paul J. Campos, and Ramon Ramirez-Villalobos

Instituto Tecnológico de Tijuana

Blvd. Limón Padilla y Av. ITR Tijuana S/N, 22500, Mesa Otay, Tijuana, B.C., Mexico

{luis.coria, paul.javierc, ramon.rmzv}@gmail.com

*Paper received on 11/30/13, Accepted on 01/19/14.*

**Abstract.** In this work, it is presented the design and implementation of a Luenberger observer for a permanent magnet synchronous motor (PMSM). A model-based technique is used to control a nonlinear system under a Field Oriented Control scheme (FOC). Moreover, simulation and implementation both, with encoder and sensorless are realized to evaluate the performance of the control. The implementation is achieved using the MCK28335 Kit platform under a experimental environment. In addition, the error criteria IAE, ISE, ITAE and ITSE are explored to evaluate the performance of both control schemes. As a result, a better performance is achieved with sensorless scheme compared the FOC including sensor.

**Keywords:** model-based controller, permanent magnet synchronous motor, sensorless control.

## 1 Introduction

Permanent Magnet Synchronous Motor (PMSM) are widely used in many applications (e.g. printers, tape drives, hard drives, process control systems, CNC machine tools, industrial robots, aerospace, electrical vehicles, submarines), due to their efficiency and dynamic performance [7][13][17]. PMSMs have superior features such as compact size, high efficiency, high power density, wide speed ratio, high torque and absence of rotor losses [9][16]. Furthermore, advances magnetic materials having even higher power lead to wider applications of PMSM [17]. In recent years, advancements in magnetic materials and control theories have made PMSMs drives to play an important role in motion-control applications [14].

When the motor mechanical variable (rotor position or speed) are available from measurements, high-precision and robust control of a PMSM can be achieved in rotor position or speed tracking applications for PMSM. A Mechanical and optical sensors are typically used to measure the rotor position or speed for vector control of a PMSM. Nevertheless, these sensors increase hardware complexity, cost and size of the drive systems [6][10]. In addition, the reliability of the drive system is reduced, a regular maintenance is required [1][8]. The disadvantages of mechanical or optical sensors can be removed if the speed can be estimated from the terminal vectors.

Estimation of rotor position or speed without measurement of mechanical variables is a challenging problem for a PMSM. In this framework, a lot of attention has been paid by electric drives community to the “sensorless” control problem of PMSMs in which only stator current and voltage measurements are available for feedback [1][6].

There are two main categories of sensorless control: model-based estimation techniques (e.g. Luenberger and Kalman-Filter observer, full and reduced order closed-loop observer, sliding mode control, model reference adaptive system) and saliency based techniques [1][3][8][19].

Model based techniques can adopt a state observer to retrieve the rotor position information, by extracting the speed information using current quantities obtained from PMSM terminal. It is especially useful for full-state feedback control develop on the state-space theory [1][19].

The sensorless approach has several advantages: 1) only electrical connections to the machine are the main phase windings, so installation cost are minimized; 2) position-sensing function can be located with the others control electronics: it does not need to be sited adjacent to the machine, and therefore does not inhibit the operating temperature range; 3) absence of connecting leads prevents corruptions of position data by electromagnetic interference; and 4) cost of a position encoding device is avoided.

The goal of this paper is to present results for a model-based controller, in order to eliminate the speed sensor on the control law on MCK28335 Kit. The remainder of this paper is organized as follows. In Section 2, presents necessary background about observer and model-based control theory. In Section 3, the dynamical equations, variable and parameters definitions, are described. Also, The design of the proposed controller and the stability analysis are presented. Performance results of simulation and emulation on MCK28335 Kit of proposed controller, are discussed in Section 5. Finally conclusions are given in the last section.

## 2 Preliminaries

### 2.1 Luenberger observer

An observer comprises a real-time simulation of the system or plant, driven by the same input as the plant, and by a correction term derived from the difference between the real output of the plant and the estimated one derived from the observer.

Consider a continuous-time linear system expressed by the form:

$$\dot{x} = Ax + Bu; \quad (1)$$

$$y = Cx; \quad (2)$$

where  $x \in \mathbb{R}^n$ ;  $u \in \mathbb{R}^m$  and  $y \in \mathbb{R}^p$  represents the state, control and output vector;  $A$ ,  $B$  and  $C$  are matrices of corresponding dimensions.

Denoting the state vector of the observer by  $\hat{x} \in \mathbb{R}^n$ , the following state space representation defines the observer:

$$\dot{\hat{x}} = A\hat{x} + B\hat{u} = A\hat{x} + B(u + LC(x - \hat{x})); \quad (3)$$



The error vector is defined as  $\tilde{x} = x - \hat{x}$ . The error dynamics is described by the following equation:

$$\dot{\tilde{x}} = (A - BLC)\tilde{x}; \quad (4)$$

In order to guarantee convergence of the estimated states, is enough to choose a gain matrix  $L$  such that error dynamics converge asymptotically (i.e., when  $A - BLC$  is a Hurwitz matrix).

## 2.2 Model-Based Controller

The model-based control techniques have been largely extended and gained prominence during the past decades, major steps are expected in the future, especially for the non-linear case [2].

Fig. 1 shows the model-based controller, with a full order closed-loop observer's state estimate being fed back through the state feedback gain  $K$ .

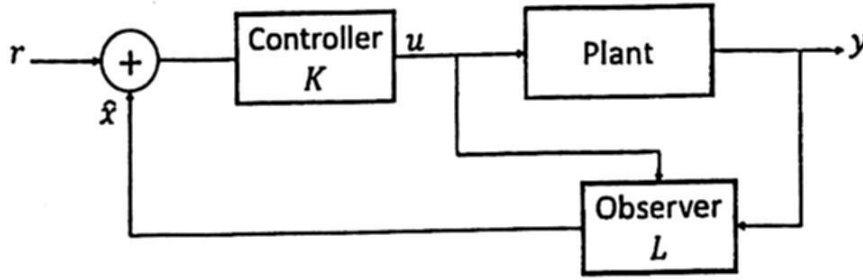


Fig. 1. Closed-loop system using the model-based output feedback controller.

Here, dimension of the plant and controller are the same. Then, the number of state variables of the closed-loop system is twice double that of the open-loop plant.

The plant is given by (1), substituting  $u = K\hat{x} + r$ :

$$\dot{x} = Ax + BK\hat{x} + Br. \quad (5)$$

To eliminate  $\hat{x}$  from (5), this equation is only in terms of the state variable  $x$  and  $\tilde{x}$ , substituting  $\hat{x} = x - \tilde{x}$ , producing:

$$\dot{x} = (A + BK)x - BK\tilde{x} + Br; \quad (6)$$

Coupling this with (4), we get the composite system's state description:

$$\frac{d}{dt} \begin{bmatrix} x \\ \tilde{x} \end{bmatrix} = \begin{bmatrix} A + BK & -BK \\ 0 & A - BLC \end{bmatrix} \begin{bmatrix} x \\ \tilde{x} \end{bmatrix} + \begin{bmatrix} B \\ 0 \end{bmatrix} r. \quad (7)$$

**Theorem 1** (Lyapunov's linearization method [15]). *If the linearized system is strictly stable (i.e, if all eigenvalues of  $A$  are strictly in the left-half complex plane), then the equilibrium point is asymptotically stable (for the actual nonlinear system).*

Since the composite system matrix is a upper triangular block, therefore the closed-loop eigenvalues are given by  $\lambda(A + BK) \cup \lambda(A - BLC)$ , where,  $\lambda(A)$  represents the set of eigenvalues of  $A$ . This fact indicates that the plant stabilization and observer design can be approached separately.



### 3 Model-Based Controller for PMSM

#### 3.1 PMSM model

The dynamical model of a brushless PMSM, according to the reference model based on the Principle of Clarke and Park transformations, is given as follows [18],[20]:

$$L_d \frac{di_d}{dt} = L_q i_q \omega - R i_d + u_d; \quad (8)$$

$$L_q \frac{di_q}{dt} = -(L_d i_d + \psi) \omega - R i_q + u_q; \quad (9)$$

$$J \frac{d\omega}{dt} = m - m_r - F \omega; \quad (10)$$

$$m = \frac{3}{2} \rho (L_d - L_q) i_d i_q + \frac{3}{2} \rho \psi i_q; \quad (11)$$

where  $i_d$  and  $i_q$  are the stator currents on  $d-q$  axis,  $L_d$  and  $L_q$  are the winding inductances on  $d-q$  axis,  $U_d$  and  $U_q$  are the  $d-q$  stator voltages,  $R$  is the stator winding resistance,  $\omega$  is the rotor speed,  $\psi$  is the permanent magnet flux,  $J$  is the rotor moment of inertia,  $F$  is the viscous friction coefficient, and  $\rho$  is the number of pole pairs. According with the model represented by (8), it is clearly seen the nonlinearity of a PMSM.

Considering that the motor has a perfect electric symmetry,  $L_d = L_q$ , and renaming the phase currents as  $x_1 = i_d$ ,  $x_2 = i_q$ ,  $x_3 = \omega$  and  $m_r = 0$ , the system represented by (8)-(11) can be expressed as:

$$\begin{aligned} \dot{x}_1 &= x_2 x_3 - \frac{R}{L} x_1 + \frac{u_d}{L}; \\ \dot{x}_2 &= -(x_1 + \frac{\psi}{L}) x_3 - \frac{R}{L} x_2 + \frac{u_q}{L}; \\ \dot{x}_3 &= \frac{3\rho\psi}{2J} x_2 - \frac{F}{J} x_3; \end{aligned} \quad (12)$$

where, assuming that  $u_d = u_q = 0$ , and the motor is running freely, the equilibrium point of the system is  $x^* = [x_1^*, x_2^*, x_3^*]^T = 0 \in \mathbb{R}^3$ . Where  $x^*$  denotes the equilibrium point of the system expressed by (12).

#### 3.2 Analysis and design of Luenberger observer

The aim of this paper is to design an observer-based speed controller, in order to eliminate the speed sensor in the control law of MCK28335 Kit. The Field Oriented Control (FOC) scheme for the control system can be seen in Fig. 2, where is shown that this scheme includes an encoder.

Now, in Fig. 3 it is shown the proposed FOC scheme for the model-based control, here the speed encoder is eliminated and substituted by an observer. The structure of the observer is given by (3). Under this scheme, the speed controller is fed with the

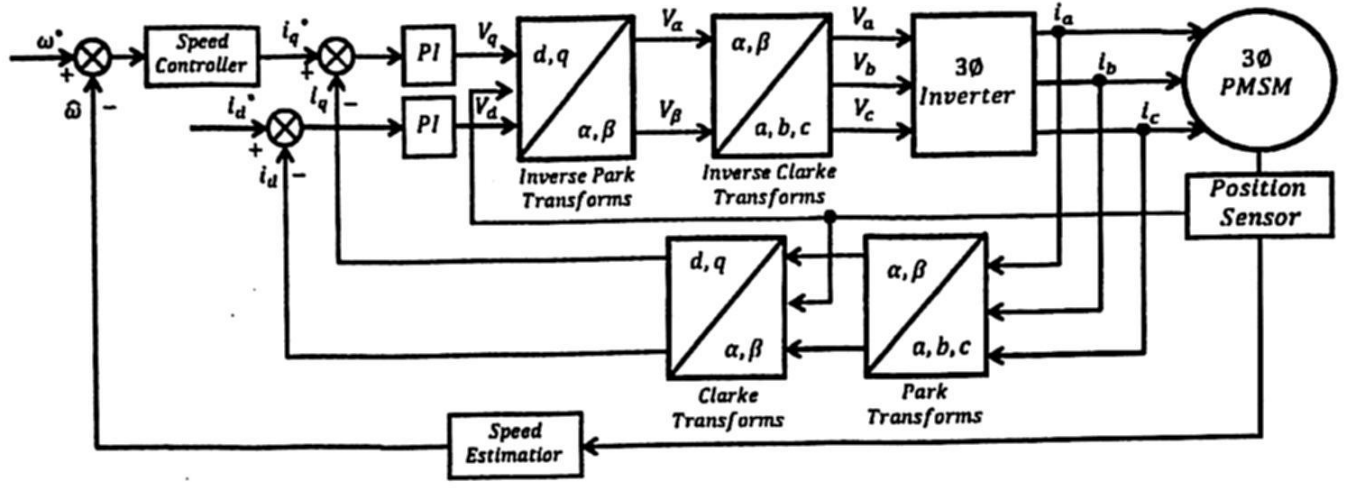


Fig. 2. Field Oriented Control Scheme for PI controller with speed encoder.

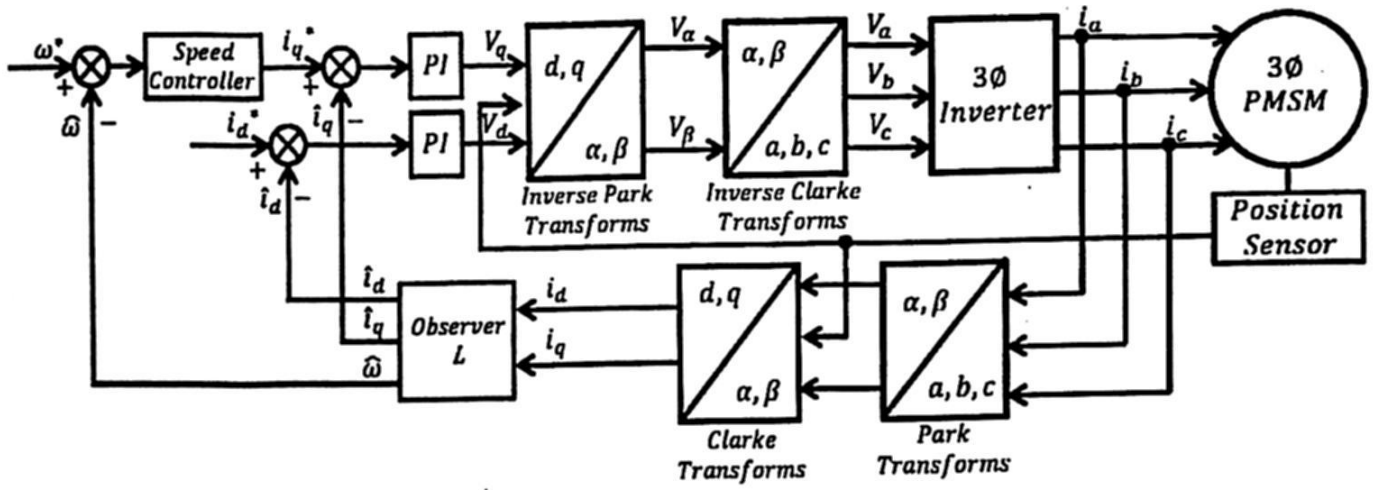


Fig. 3. Field Oriented Control Scheme for the model-based proposed controller.

estimated speed from observer, and the PIs in the inner control loop that control the direct and quadrature axes ( $d - q$ ) are estimated for  $i_d$  and  $i_q$ .

The FOC scheme of the proposed controller has a similar structure than closed-loop system showed in Fig. 1. The design of the proposed observer is based on the linearization of the dynamic model (12). Where  $x$  is the state vector defined as  $x = [x_1, x_2, x_3]^T$  and  $u$  is the input vector given by  $u = [u_d, u_q]^T$ . Matrices  $A$  and  $B$  are defined as follows:

$$A = \begin{bmatrix} -\frac{R}{L} & 0 & 0 \\ 0 & -\frac{R}{L} & -\frac{\varphi}{L} \\ 0 & \frac{3p\varphi}{J} & \frac{F}{J} \end{bmatrix}; \quad B = \begin{bmatrix} \frac{1}{L} & 0 \\ 0 & \frac{1}{L} \\ 0 & 0 \end{bmatrix}. \quad (13)$$

The output vector  $y$  consist in the measurement of the currents  $x_1$  and  $x_2$  denoted by  $y = [x_1, x_2]^T$ . Here, the error dynamics matrix  $A - BLC$  is given by:

$$A - BLC = \begin{bmatrix} -(\frac{R}{L} + \frac{l_1}{L}) & -\frac{l_3}{L} & 0 \\ -\frac{l_2}{L} & -(\frac{R}{L} + \frac{l_4}{L}) & -\frac{\varphi}{L} \\ 0 & \frac{3p\varphi}{J} & -\frac{F}{J} \end{bmatrix}; \quad (14)$$

where according with Theorem 1, in order to ensure the local asymptotic stability, the constant gains are given by:

$$l_1 > -R; \quad (15)$$

$$l_2 = l_3 = 0; \quad (16)$$

$$l_4 > -\frac{3p\varphi^2}{2F}; \quad (17)$$

therefore, with the constant gains  $l_1$  and  $l_4$  can be ensured local asymptotic stability of the observer. The dynamic model of the observer is given by:

$$\frac{d}{dt} \begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \\ \hat{x}_3 \end{bmatrix} = \begin{bmatrix} -\frac{R}{L}\hat{x}_1 + \frac{1}{L}(u_d + l_1(\hat{x}_1 - x_1)) \\ -\frac{R}{L}\hat{x}_2 + \frac{\varphi}{L}\hat{x}_3 + \frac{1}{L}(u_q + l_4(\hat{x}_2 - x_2)) \\ \frac{3p\varphi}{J}\hat{x}_2 - \frac{F}{J}\hat{x}_3 + \frac{1}{L} \end{bmatrix} \quad (18)$$

The stability conditions for the closed-loop system is not addressed due the platform already has implemented a PI controller. According with aforementioned, it is assumed that the PI controller has a good performance and only it is analyzed the observer stability.

## 4 Simulation and results

The performance of the proposed controller was validated using MATLAB-Simulink software test given by Technosoft, the real-time simulation platform consisting of:

- One three-phase Permanent Magnet Synchronous Motor. Model MBE.300.E500 with parameters shown in Table 1. Equipped with Hall sensors and 500-line encoder for direct experimentation.
- One development system model Technosoft MSK28335 with a DSP motion controller model TMS320F28335.
- One three-phase Voltage Source Inverter (VSI) model PM50 power module board with rated voltage 36 V, rated current 2.1 A, rated power 75 W, DC link voltage in a range of 1236 V, working at PWM frequencies up to 20 kHz.

Numerical simulation of the error dynamics of the proposed observer are shown in Fig. 4. The initial condition of PMSM is  $[x_1(0), x_2(0), x_3(0)]^T = 0 \in \mathbb{R}^3$  and the initial condition of (18) is  $[\hat{x}_1(0), \hat{x}_2(0), \hat{x}_3(0)]^T = 1 \in \mathbb{R}^3$ .

Fig. 4 shows convergence of the error dynamics of the observer.

In addition, numerical simulations are realized in order to make a comparison between the performance of FOC scheme with encoder and proposed sensorless FOC

**Table 1.** Parameters of the PMSM.

<b>Coil dependent parameters</b>		
Name	Value	Units
Phase-phase resistance	8.61	$\Omega$
Phase-phase inductance	0.713	$mH$
Back-EMF constant	3.86	$V/1000rpm$
Torque constant	36.8	$mNm/A$
Pole pairs	1	
<b>Dynamic parameters</b>		
Rated voltage	36	$V$
Max. Voltage	58	$V$
No-load current	73.2	$mA$
No-load speed	9170	$rpm$
Max. continuous current (at 5000 rpm)	913	$mA$
Max. continuous torque (at 5000 rpm)	30	$mNm$
Max. permissible speed	15000	$rpm$
Peak torque (stall)	154	$mNm$
<b>Mechanical parameters</b>		
Rotor inertia	$11 \text{ kgm}^2 * 10^{-7}$	
Mechanical time constant	7	$ms$
Thermal resi. housing-ambient	8.6	$C/W$
Thermal resi. winding-housing	1.0	$^{\circ}C/W$

scheme. Both schemes were evaluated considering three different set points. Fig. 5 shown FOC scheme and for FOC scheme and Fig. 6 the proposed sensorless FOC scheme.

The proposed FOC scheme presents less oscillations than FOC scheme with encoder. Also, the response of both are similar.

A comparison of the performance were realized employing the integral absolute error (IAE), the integral of squared error (ISE), the integral of time multiplied by absolute error (ITAE) and the integral of time multiplied by squared error (ITSE). The results for all these criteria are shown in Table 2.

Also, in Fig. 7 we have all these errors through time. In (a) IAE Criterion, (b) ISE Criterion, (c) ITAE Criterion and (d) ITSE Criterion are presented for FOC scheme and proposed sensorless FOC scheme.

All error criteria was evaluated considering a speed test reference of 30 rpm and 2 seconds of simulation.



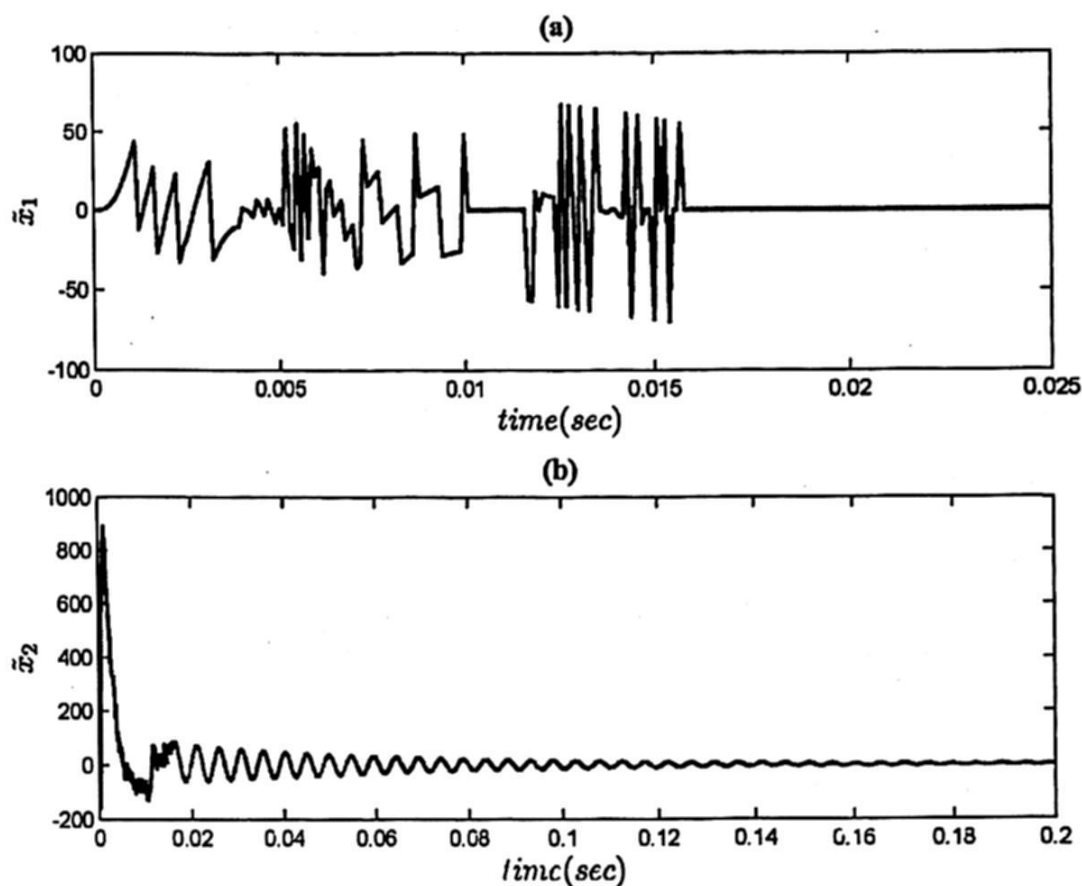


Fig. 4. Observer convergence: (a) Error estimation of  $\hat{x}_1$  and (b) Error estimation of  $\hat{x}_2$

Table 2. Error analysis.

Parameter	Error Criterion			
	IAE	ISE	ITAE	ITSE
Control with observer	0.091697	1.487097	0.00054	0.002007
Control without observer	0.60555	2.59605	0.509871	0.512472

## 5 Discussions and conclusions

This paper presents the design and simulation of a model-based speed sensorless for a PMSM. The proposed approach uses stator currents from the PMSM drive to estimate the rotor speed. A comparison of the response of the proposed model-based speed sensorless FOC scheme and a classic FOC scheme have been performed. Numerical simulations show that the proposed sensorless control presents less oscillations than FOC scheme, these oscillations appear due to sensor are sensitive to temperature, in addition, a bad estimation increase the error. The response of both systems are similar. In addition, the well performance is also demonstrated calculating the error criterion IAE, ISE, ITAE and ITSE. All the criterion of proposed FOC scheme are smaller than FOC scheme with encoder.

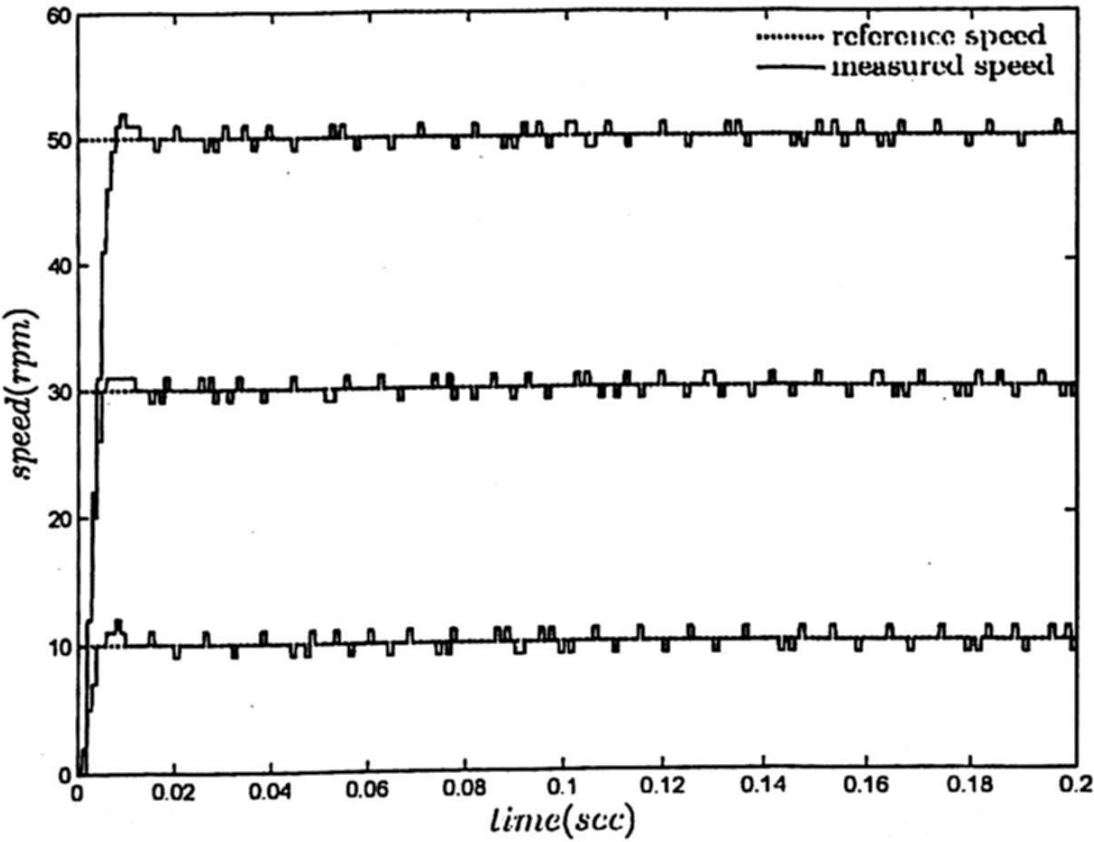


Fig. 5. Speed response of FOC.

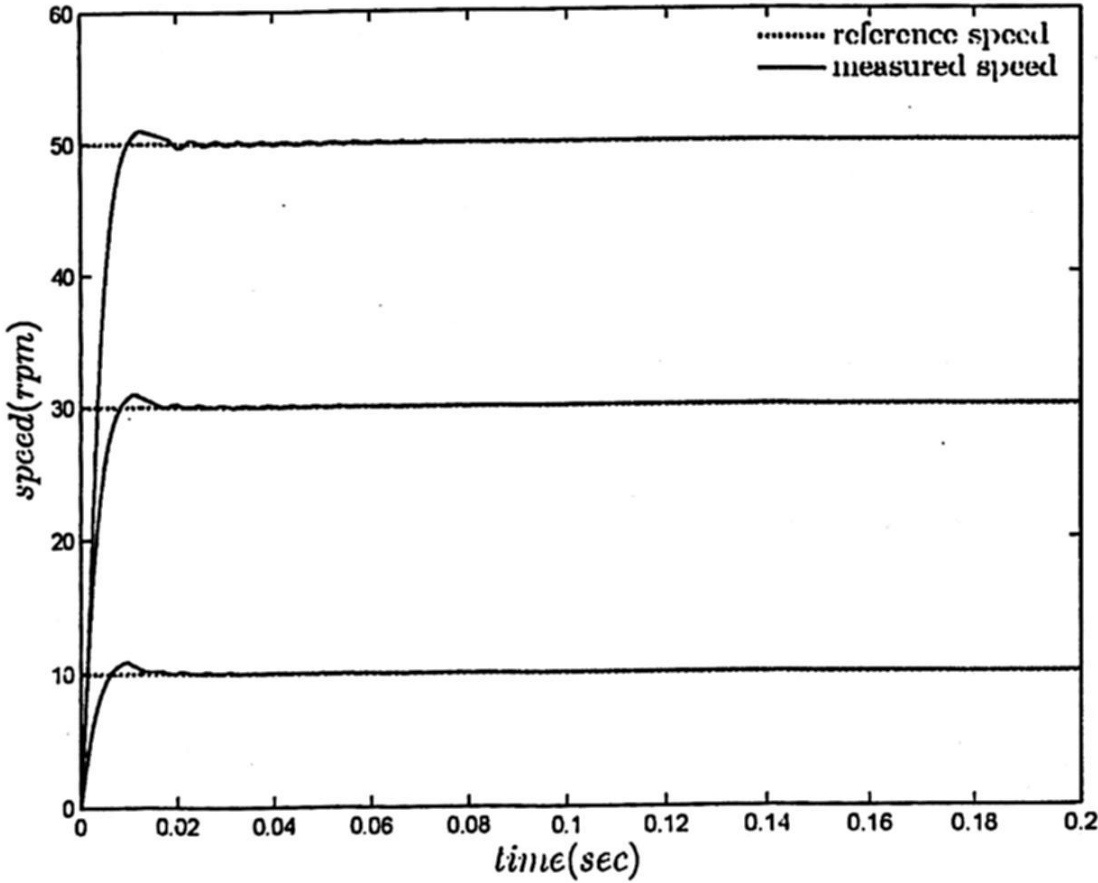
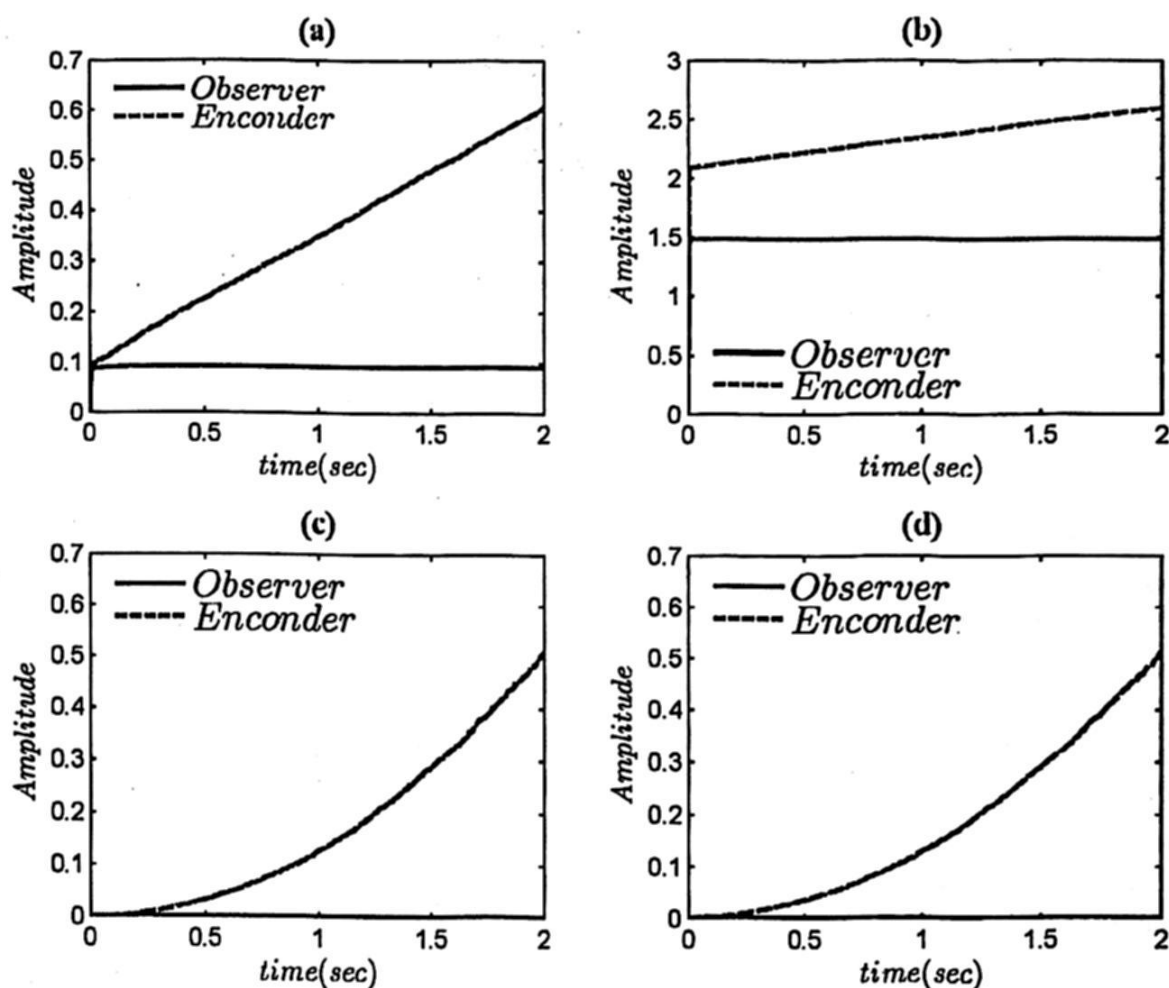


Fig. 6. Speed response of proposed sensorless FOC scheme.



**Fig. 7.** Error analysis for both FOC schemes: (a) IAE Criterion, (b) ISE Criterion, (c) ITAE Criterion and (d) ITSE Criterion.

## References

1. Accetta, A., Cirrincione, M., Pucci, M.: TLS EXIN based neuronal sensorless control of a high dynamic PMSM. *Control Engineering Practice*. 20(7), 725–73, (2012).
2. Agachi, P.S., Nagy, Z.K., Cristea, M.V.: *Model Based Control: Case Studies in Process Engineering*. Wiley, (2007).
3. Alahakoon, S., Fernando, T., Trinh, H., Sreeram, V.: Unknown input sliding mode functional observers with application to sensorless control of permanent magnet synchronous machines. *Journal of the Franklin Institute*. 350, 107–128, (2013).
4. Åström, K.J., Murray, R.M.: *Feedback Systems*. Princeton University Press, (2008).
5. Besancon, G.: *Nonlinear Observer and Applications*. Springer, (2007).
6. Bifaretti, S., Lacovone, V., Rocchi, P., Tomei, P., Verrelli, C.M.: Nonlinear speed tracking control for sensorless PMSMs with unknown load torque: From theory to practice. *Control Engineering Practice*. 20, 714–724, (2012).
7. Chai, S., Wang, L., Rogers, E.: Model predictive control of a permanent magnet synchronous motor with experimental results. *Control Engineering Practice*. 21, 1584–1593, (2013).
8. Chi, W., Cheng, M.: Implementation of a sliding-mode-based position sensorless drive for high-speed micro permanent-magnet synchronous motors. *ISA Transactions*. In Press, (2013).
9. Grouz, F., Sbita, L., Boussak, M., Khlaief, A.: FDI based on an adaptive observer for current and speed sensors of PMSM drives. *Simulation Modelling Practice and Theory*. 35, 34–39, (2013).

10. Holtz, J.: Sensorless control of induction motor machines: with or without signal injection. *IEEE Transactions on Industrial Electronic*. 18(1), 7–30, (2006).
11. Khalil, H.K.: *Nonlinear Systems*. Prentice Hall, (2002).
12. Khlaief, A., Boussak, M., Chaan, A.: A MRAS-S-based stator resistance and speed estimation for sensorless vector controlled IPMSM drive. *Electri Power Systems Research*. 08, 1–15, (2014).
13. Maria, C.: Synchronization of permanent magnet electric motors. *Nonlinear Analysis: Real Wold Applications*. 13, 395–409, (2012).
14. Öztürk, N., Celik, E.: Speed control of permanent magnet synchronous motor using fuzzy controller based on genetic algorithms. *Electrical Power and Energy Systems*. 43, 889–898, (2012).
15. Slotine, J.J.E, Li, W.: *Applied Nonlinear Control*. Prentice Hall, Inc., (1991).
16. Xu, D., Zhang, S., Liu, J.: Very-low speed control of PMSM based on EKF estimation with closed loop optimized parameters. *ISA Transactions*. 52, 835–843, (2013).
17. Yu, J., Chen, B., Yu, H., Gao, J.: Adaptive fuzzy tracking control for the chaotic permanent magnet synchronous motor drive system bya backstepping. *Nonlinear Analysis: Real Wold Applications*. 12, 671–681, (2011).
18. Zaher, A.A.: A nonlinear controller design for permanent magnet motors using a synchronization-based technique inspired from the lorenz system. *Chaos*. 18(1), 013111, (2008).
19. Zheng, S., Tang, X., Song, B., L, S., Ye, B.: Stable adaptive PI control for permanent magnet synchronous motor drive based on improved JITL technique. *ISA Trasactions*. 52(3), 539–549, (2013).
20. Zribi, M., Otafy, A., Smaoui, N.: Controlling chaos in the permanent magnet synchronous motor. *Chaos, Solitons & Fractals*. 41(3), 1266–1276, (2009).



# Experimental Analysis of the Dynamixel AX-12 Servomotor and its Wireless Communication

Eusebio Bugarin, Luis Jaciel Castañeda-García, and A. Y. Aguilar-Bustos

Instituto Tecnológico de Ensenada, Ensenada, Baja California, México  
eusebio@hotmail.com, ingluis@gmail.com, yaveni@hotmail.com

*Paper received on 12/02/13, Accepted on 01/19/14.*

**Abstract.** In this paper, it is detailed the dynamic model of the Dynamixel AX-12 servomotor, which is widely used in legged robots and in humanoid robots specially. Experimentally, it is determined its parameters and a theoretical and practical comparison of the internal position controller is performed. Additionally, in the search of applications with bipedal gait, a velocity controller is designed and implemented with satisfactory results. The experiments are done by sending the control law and receiving measurements between the servomotor and a remote PC via Zigbee wireless communication protocol.

**Keywords:** Servomotors, experimental analysis, velocity control, humanoid robots, Zigbee.

## 1 Introduction

Thanks to recent technological achievements, Advanced Robotics, who studies robots with marked characteristics of autonomy [1], has acquired an important zenith that allows not only to develop interesting theoretical proposals but also to perform experimental validation through very novel and ingenious prototypes. Today, advanced robots with better promises of performance are constituted by legged robots, mainly because they have better mobility on terrains that are not conditioned, on inclined surfaces and on environments with obstacles.

Legged robots, in general, have been frequently studied since the 1970s and, in particular, prototypes of anthropomorphic or humanoid robots have been a reality since the mid-1990s [2]; for example, the Asimo robot from Honda [3], the QRIO robot from Sony [4] and the HRP-2 robot from Kawada [5]. Currently, there are also low cost educational humanoid robots like the Nao robot from Aldebaran Robotics [6] and the Bioloid robot from Robotis (a Korean company).

The Bioloid robot from Robotis (see Fig.1) can be constructed as a humanoid robot with 18 degrees of freedom (6 per leg and 3 per arm) with a Dynamixel AX-12 servomotor in each joint. There are many and varied tasks that use this robot as a prototype for experimental tests; see for instance [7,8,9,10], among others. In this way, the main purpose of this paper is experimentally analyze the performance of, for first instance,

a single Dynamixel AX-12 servomotor with the intention to glimpse the possible applications in the field of Advanced Robotics that this Bioloid robot can perform, in particular: bipedal gait.

The Dynamixel AX-12 servomotor has also been studied in [11] where it is determined its pulse transfer function through a Box-Jenkins procedure [12], in this work it is also showed graphs of its position control performance. In [13] an exhaustive analysis of the same servomotor is described and through reverse engineering it is illustrated each one of its parts, it is determined various parameters of the dynamic model (as the viscous friction coefficient, the static friction, the motor-torque constant and the armature resistance), it is calculated the assumed PID controller gains and it is presented simulations and experiments for its position control; however, the model is not so easy to reproduce since it uses a setpoint generator trying to run, without a modification, the servomotor proper functions (which as we shall see, they are not so convenient for velocity control applications, for instance). Likewise, in [14] it is performed a calibration of the measurements provided by the servomotor; concluding that it is necessary some adjustments in each one, with the exception of the angular position measurement. In [15] the latency of this servomotor is analyzed.



Fig. 1. Bioloid humanoid robot

In this paper, the dynamic model of the Dynamixel AX-12 servomotor is detailed, its parameters are determined experimentally and a theoretical and practical comparison of the internal position controller is performed. Additionally, in the search of bipedal gait applications, a velocity controller is designed and implemented showing satisfactory experimental results. The experiments are done by sending the control law and receiving measurements between the servomotor and a remote PC via Zigbee wireless communication protocol.

## 2 Modeling and Characterization

Fig. 2 describes a block diagram of the Dynamixel AX-12 servomotor (Table 1 details its variables and parameters). As can be seen, this servomotor is composed of a DC motor, a gearbox with reduction ratio  $1:r$  (where  $r = 254$ ) and an ATmega 8 microcontroller. The power stage corresponds to an "H" bridge which is driven by the microcontroller PWM output. It has a direct temperature housing measurement for protection through a thermistor connected to the ADC2 input. To measure the angular position  $q$  of the load axis a  $10\text{ K}\Omega$  potentiometer (MuRata SV01) is connected to the ADC0 input. And the data sending to and from the servomotor is achieved via an UART TTL level half duplex serial communication at 1 Mbps.

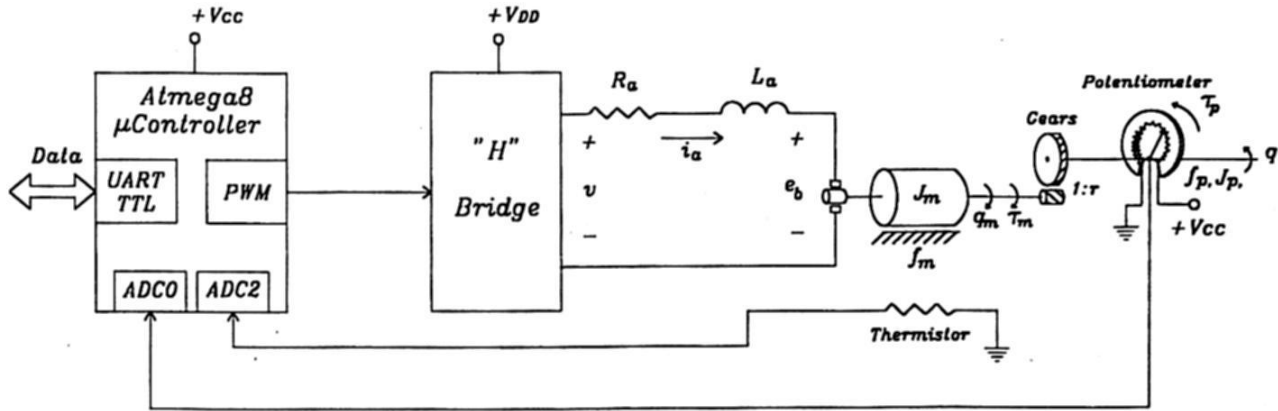


Fig. 2. Block diagram of the Dynamixel AX-12 servomotor

According to specifications, the servomotor can provide feedback measurements of position, speed and load torque. However, as noted in Fig. 2, the servomotor only has direct measurement of its position (also in its housing temperature, but this measure is for protection issues), so it is presumed that the other two given measurements are just estimated data. The microcontroller ADC0 has 10 bits of resolution and the potentiometer is used in a working interval of  $[0 - 300]\pi/180$  [rad], therefore the resolution in the angular position measurement is

$$\frac{\frac{300\pi}{180}[\text{rad}]}{(2^{10}-1)[\text{bit}]} = \frac{0.29\pi}{180} \left[\frac{\text{rad}}{\text{bit}}\right].$$

The microcontroller PWM also has a resolution of 10 bits, such that one can manipulate 1023 armature voltage levels (typically,  $V_{DD} = 12.15$  [V]).

### 2.1 Servomotor Model

To modeling the servomotor, it has been considered that the armature voltage is equal to the average voltage at the terminals of the "H" bridge after the PWM. For purposes of the actual friction on the servomotor, it has been only considered the viscous friction model. Now, to obtain the servomotor model it has been used a very similar procedure that is described in [16]. Therefore, the electrical part is governed by

$$v = R_a i_a + L_a \frac{di_a}{dt} + e_b \quad (1)$$

where  $e_b = K_b \dot{q}_m$  and the mechanic part by

$$J_m \ddot{q}_m = \tau_m - f_m \dot{q}_m - \frac{\tau_p}{r}, \quad (2)$$

$$J_p \ddot{q} = \tau_p - f_p \dot{q} \quad (3)$$

where  $\tau_m = K_a i_a$  and  $q_m = r q$ . In this formulation, it is assumed that the dynamic effects of the gearbox and the potentiometer sensor are included in (3). Finally, neglecting  $L_a$  and substituting (3) into (2), then this result into (1), we have the dynamic model of the servomotor expressed through

$$\frac{r R_a}{K_a} \left[ \frac{J_p}{r^2} + J_m \right] \ddot{q} + \frac{r R_a}{K_a} \left[ f_m + \frac{f_p}{r^2} + \frac{K_a K_b}{R_a} \right] \dot{q} = v, \quad (4)$$

$$A \ddot{q} + B \dot{q} = v \quad (5)$$

with  $A = \frac{r R_a}{K_a} \left[ \frac{J_p}{r^2} + J_m \right]$  and  $B = \frac{r R_a}{K_a} \left[ f_m + \frac{f_p}{r^2} + \frac{K_a K_b}{R_a} \right]$ .

Table 1. Variables and parameters of the AX-12 servomotor

Symbol	Description	Units
$q$	Angular position of the load axis	Rad
$v$	Armature voltage	V
$q_m$	Angular position of the motor shaft	Rad
$\tau_m$	Torque in the motor shaft	Nm
$i_a$	Armature current	A
$e_b$	Back electromotive force	V
$\tau_p$	Load torque due to potentiometer and gearbox	Nm
$R_a$	Armature resistance	$\Omega$
$L_a$	Armature inductance	H
$K_a$	Motor-torque constant	Nm/A
$K_b$	Back electromotive force constant	Vs/rad
$r$	Gear reduction ratio	
$J_m$	Motor inertia	Kg m <sup>2</sup>
$f_m$	Viscous friction coefficient of the rotor	Nm
$J_p$	Potentiometer and gearbox inertia	Kg m <sup>2</sup>
$f_p$	Viscous friction coefficient of the potentiometer and gearbox	Nm

## 2.2 Parameterization of the Servomotor Dynamic Model

The model of the servomotor (5) can be seen, in terms of the angular speed  $\omega = \dot{q}$ , in the following manner:

$$A \dot{\omega} + B \omega = v. \quad (6)$$



In such a way that for a constant input voltage  $v = v_1$  with  $\omega(0) = 0$ , we have that the solution of (6) is given by

$$\omega(t) = \frac{v_1}{B} \left[ 1 - e^{-\frac{B}{A}t} \right] \quad (7)$$

where the static gain of the system is  $k_s = \frac{1}{B}$  and its time constant is  $\tau_s = \frac{A}{B}$ .

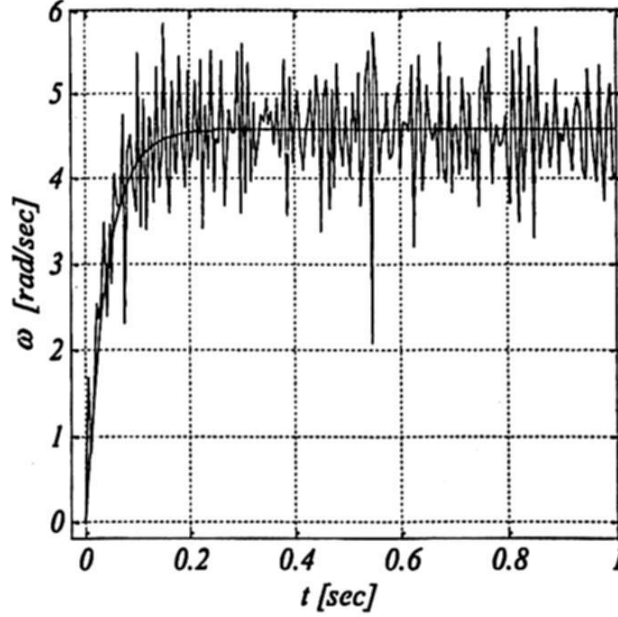


Fig. 3. Graph of  $\omega$  vs  $t$  for parameterization of the servomotor (the continuous line shows the experimental data and the discontinuous one the ideal data)

For the parameterization, the servomotor was fed, with null initial conditions, with a constant armature voltage of 8 [V] and by means of a real-time vision system (well calibrated at 200 frames per second) it was obtained his graph of angular velocity<sup>1</sup> versus time, which is shown in Fig. 3. Additionally, the steady state angular speed was measured by a tachometer resulting  $\omega_{ss} = 4.57$  [rad/sec]. In this manner

$$B = \frac{v_1}{\omega_{ss}} = \frac{8}{4.57} = 1.7505 \quad (8)$$

and, from the experimental graph  $\tau_s = 0.04$  [sec], hence

$$A = B\tau_s = 1.7505(0.04) = 0.0700. \quad (9)$$

In Fig. 3, it is presented the ideal response of (7) with the determined parameters (by a discontinuous line); as can be observed, in average, the experimental response is very similar to the ideal one.

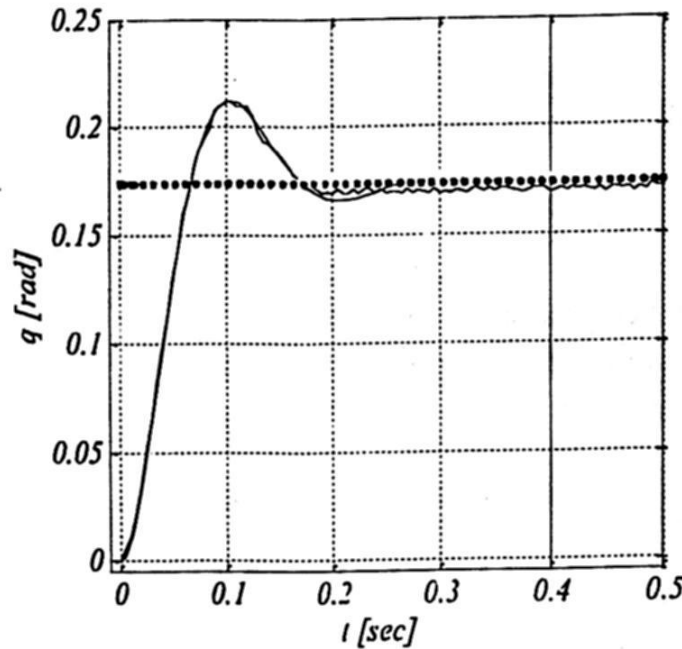
<sup>1</sup> The vision system measures its angular position and by a numeric Euler differentiation is calculated this angular velocity.

### 2.3 Internal Position Controller Gains

The studied servomotor has an internal position controller. In order to obtain the closed-loop model it is assumed that it has programmed a PID controller (typical for servomotors in general). To determine the PID gains the servomotor is subjected to a constant desired input  $q_d = \frac{10\pi}{180} = 0.1745$  [rad] (a small one with the intention that do not exist saturations in the demanded armature voltage) and with null initial conditions (with no angular speed limits and the default internal parameters for the compliance). Again, the experimental plot was obtained with the real-time vision system properly calibrated (see Fig. 4). From this plot we proceeded to compare simulations in closed loop to finally find that there exist only proportional and derivative actions (which seems reasonable since, if the output is the angular position, the servomotor possesses an integral action per se), so the control law would be

$$v = k_p(q_d - q) + k_d \frac{d}{dt}(q_d - q) \quad (10)$$

with gains  $k_p = 80$  and  $k_d = 0.3$ . It can be noted that this proportional gain is too large, in fact, in this experiment, the maximum armature voltage is demanded at the beginning and it is of the order of  $v = 80(0.1745) = 13.96$  [V] (slightly bigger than the maximum supported).



**Fig. 4.** Evolution versus time of the angular position (the continuous line shows the experimental data, the discontinuous one the simulated data and the dotted one the reference)

Fig. 4 presents the comparison between the simulated and the experimented. Observe that the transients are quite similar, however, the experiment showed a steady-state error, which is attributed to the other components of the unmodeled friction and to the quantization effects in both the control law and the angular position measurement.

### 3 Velocity Control

The tasks entrusted to legged robots and especially humanoid robots usually have to do with the movement or gait of the robot. Due to the above, there exist an interest in designing a velocity controller for the servomotor under consideration, which we must to remember that has an internal position controller. The designed controller is based on a feedforward component with the intention to deliver the servomotor torque needed to perform the desired movement. For its design, it is defined the following velocity control objective

$$\lim_{t \rightarrow \infty} \dot{e}(t) = 0 \quad (11)$$

where  $\dot{e} = \dot{q}_r - \dot{q}$  is the angular velocity error and  $\dot{q}_r$  the velocity reference.

Now, consider the servomotor model (5) with internal position control law (10), then do the next change of variables " $u = q_d$ ", where  $u$  is the new input of the system. With the above, the system (5) and (10) is

$$A\ddot{q} + B\dot{q} = k_p(u - q) + k_d \frac{d}{dt}(u - q) . \quad (12)$$

The designed control law is expressed by

$$k_d \dot{\xi} + k_p \xi = A\ddot{q}_r + B\dot{q}_r , \quad (13)$$

$$u = q_r + \xi \quad (14)$$

where  $\xi$  is a new state variable added to the system.

The closed loop equations of the system is obtained by substituting (13) and (14) into (12), being

$$A(\ddot{q}_r - \ddot{q}) + (B + k_d)(\dot{q}_r - \dot{q}) + k_p(q_r - q) = 0 ,$$

$$A\ddot{e} + (B + k_d)\dot{e} + k_p e = 0 .$$

The coefficients of this linear differential equation are all positive so that  $e(t)$  and their derivatives tends exponentially to zero as time approaches infinity. This demonstrates that the controller (13)-(14) guarantees the proposed velocity control objective (11) for the servomotor. In fact, with the particular parameters, its eigenvalues are  $-14.6464 \pm 30.4687i$  ( $i = \sqrt{-1}$ ). It is worth noticing that this design corresponds to a feedforward dynamic controller that depends entirely on the servomotor parameters (and on the internal position controller gains).

### 4 Experimental Analysis

To corroborate the developed theory, the servomotor performance and the Zigbee wireless communication (included in the Bioloid robot) it is performed experiments under the scheme illustrated in Fig. 5. The Bioloid robot, through the CM-510 control unit (composed basically by an ATmega 2561 microcontroller) can send position commands to the internal controller and receive measurements of the angular position to a

remote PC via the Zigbee protocol. The components to make this possible are shown in Fig. 5. This communication scheme allowed to carried out experiments with an average sampling period of 0.010 [sec].

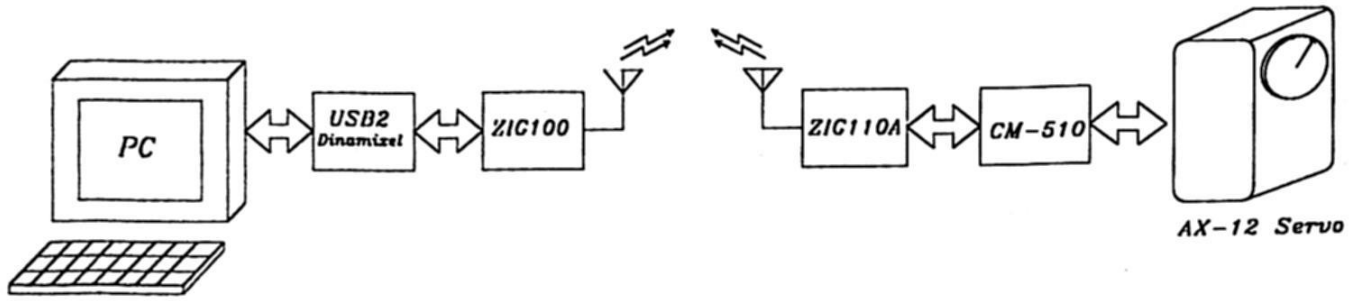


Fig. 5. Experimental scheme

For the experiment, the CM-510 control unit was only used for the sending and receiving data via Zigbee while the controller (13)-(14) is completely programmed in the remote PC (where are also stored all the variables, including the experiment time).

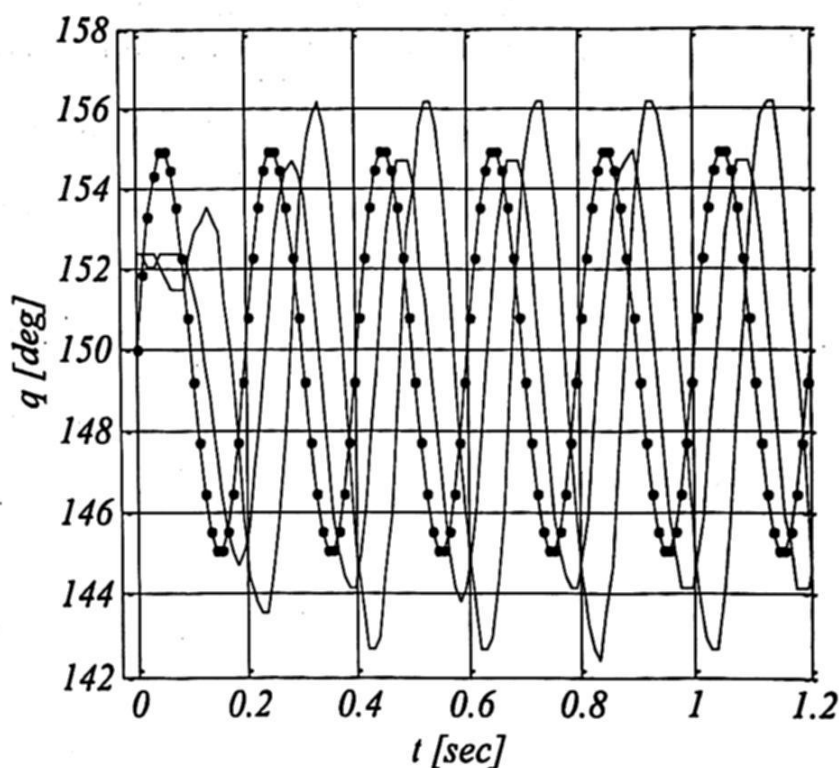
Fig. 6 illustrates the graphs of the angular position time evolution in the experiments for the velocity control with initial conditions equal to rest and with  $q_r(0) = 150$  [deg]. In the dotted line it is shown the reference angular position, expressed by

$$q_r(t) = \frac{5\pi}{180} \sin(2\pi(5)t) + \frac{150\pi}{180}$$

which corresponds to a sinusoidal signal with frequency of 5 [Hz], amplitude of 5 [deg] and with an offset of 150 [deg], a small one to do not cause saturations in the control law. In the discontinuous line is observed the graph when the internal position controller is used without modifications and in the solid line is presented the graph when the designed velocity controller (13)-(14) is applied. It should be noted a better performance with the proposed controller, both in amplitude and in phase.

## 5 Conclusions

It has been described the dynamic model of the Dynamixel AX-12 servomotor, which is widely used in legged robots and in particular for bipedal locomotion of humanoid robots. Experimentally, it has been determined its parameters and it has been performed theoretical and practical comparisons of its internal position controller. Additionally, in the search of bipedal gait applications, where periodic functions for the joint positions are used, it has been designed a velocity controller with satisfactory performance in the experimental results. The command sending and measurement receiving between the Bioloid robot and a remote PC was implemented via Zigbee wireless communication protocol, which also had an acceptable performance.



**Fig. 6.** Temporal evolution of the angular position (the continuous line shows the experimental data with the designed controller, the discontinuous one with just the internal position controller and the dotted one the reference)

It should be noted that the servomotor only offers direct measurement of its angular position in such a way that the other measurements provided are just estimated data. There exist a 10-bit resolution for both the angular position measurement and the applied armature voltage. The internal proportional controller gain is too large which provokes saturations in the armature voltage if the desired angular position is relatively large; this situation declines, in great measure, the servomotor functionality.

## 6 Acknowledgements

It is appreciated the support from CONACYT Mexico (Proyect Grant 166654), PROMEP, DGEST and from the Instituto Tecnológico de Ensenada.

## References

1. Sciavicco, L., Siciliano, B.: *Modelling and Control of Robot Manipulators*. Springer-Verlag. London. 378 pp. (2000)
2. Kajita, S, Espiau B.: *Legged Robots*. In: Siciliano, B., Khatib, O. (eds.) *Springer Handbook of Robotics*. Springer. (2008)
3. Sakagami, Y., Watanabe, R., Aoyama, C. Matasunaga, S., Higaki, N., Fujimura, K.: *The intelligen ASIMO: System overview and integration*. In: *IEEE - International Workshop on Intelligent Robots and Systems*. pp. 2478-2483.(2002)
4. Fujita, M.: *Digital creatures for Future Entertainment Robotics*. In: *IEEE - International Conference on Robotics and Automation*. pp. 801-806. (2000)



5. Kaneko, K., Kanchiro, F., Kajita, S., Yokoyama, K., Akachi, K., Kawasaki, T., Ota, S., Isozumi, T.: Design of Prototype Humanoid Robotics Platform for HRP2. In: IEEE - International Conference on Intelligent Robots and Systems. (2002)
6. Gouaillier, D., Collete, C., Kilner, C.: Omni-directional Closed-loop Walk for Nao. In: IEEE - International Conference on Humanoid Robots. (2010)
7. Kumar, A.: Optimizing walking of a humanoid robot using reinforcement learning. Master Thesis. Warsaw University of Technology. (2011)
8. Nunez, J.B., Brisco, A., Rodriguez, D. A., Ibarra, J. M., Rodriguez, V. M.: Explicit analytic solution for inverse kinematics of Bioloid humanoid robot. In: 2012 Brazilian Robotics Symposium and Latin American Robotics Symposium, pp. 33-38. (2012)
9. Wolf, J.C., Hall, P., Robinson, P., Culverhouse, P.: Bioloid based Humanoid Soccer Robot Design. In: Proc. of the Second Workshop on Humanoid Soccer Robots. (2007)
10. Akhtaruzzaman, M., Shafie, A. A.: Geometrical Analysis on BIOLOID Humanoid System Standing on Single Leg. In: 4th International Conference on Mechatronics (ICOM'11) Kuala Lumpur, Malaysia. (2011)
11. Teodoro, P.: Humanoid Robot, Development of a simulation environment of an entertainment humanoid robot. Master Thesis. Instituto Superior Tecnico de la Universidade Técnica de Lisboa. (2007)
12. Ljung, L.: System Identification: Theory for the User. Prentice Hall. Second Edition. 672 pp. (1999)
13. Mensink, A.: Characterization and modeling of a Dynamixel servo. Technical Report. Electrical Engineering, Control Engineering. University of Twente. (2008)
14. Tira-Thompson, E.: Digital Servo Calibration and Modeling. Technical Report. Robotics Institute, Carnegie Mellon University. (2009)
15. Smith, J.A., Jivraj, J.: Analysis of Robotis Dynamixel AX-12+ Actuator Latencies. In: Symposium on Brain, Body and Machine. Montreal, Canada. (2010)
16. Kelly, R., Santibáñez, V.: Control de Movimiento de Robots Manipuladores. Pearson Prentice Hall. Madrid. 344 pp. (2003)

# Position Feedback Nonlinear $\mathcal{H}_\infty$ -Control for Inertia Wheel Pendulum Stabilization

Adrián Gómez and Luis T. Aguilar

Instituto Politécnico Nacional  
Avenida del parque 1310 Mesa de Otay Tijuana, B.C., 22510 México  
agomez@citedi.mx; laguilarb@ipn.mx  
Paper received on 12/10/13, Accepted on 01/19/14.

**Abstract.** The paper deals with the position stabilization problem of the inertia wheel pendulum using a nonlinear  $\mathcal{H}_\infty$  controller via position measurements for feedback. The main objective was to stabilize the pendulum at the up-right position in spite of the external disturbances. A local  $\mathcal{H}_\infty$  controller is derived by means of a certain perturbation of the differential Riccati equations that appear while solving the corresponding  $\mathcal{H}_\infty$ -control problem for the linearized system. Since the initial conditions were far from the region of attraction of the desired equilibrium point, we construct a hybrid controller consisting of the swing-up control and the stabilization part. The performance of the proposed controller, applied to a perturbed academic wheel pendulum, was verified by simulations.

**Keywords:** Inertia wheel pendulum, nonlinear  $\mathcal{H}_\infty$ -control, output-feedback.

## 1 Introduction

The inertia wheel pendulum (IWP) is a nonlinear systems typically used to emulate postural sway such as bipedal walking motion, rocket thrust, the segway human transporter, and can be used to investigate research issues in control, autonomous navigation, group coordination, and other issues. One critical and interesting problem of the IWP is the stabilization of pendulum at the upright position because the open-loop equilibrium point is unstable and the closed-loop system can be sensitive to unmatched disturbances.

Nonlinear  $\mathcal{H}_\infty$ -control [1–3] is a tool to solve robust control problems and it has been successfully applied, among others, on the control of robot manipulators [4], underactuated mechanical systems [5], coil-power units [6], hard-disk drive servo systems [7], and hydraulic systems [8].

There are several local controllers design in the literature to stabilize the inertia wheel pendulum. For example, Ye *et al.* [9] proposed a backstepping technique for non-linear cascade systems whose driven subsystems have a feed-forward structure and include higher order terms. In their proposal, a small control is first assigned to stabilize the driven subsystem, and a simple backstepping procedure is then followed. López-Martínez *et al.* [10] considered a variable structure controller using a nonlinear sliding mode surface to deal with disturbances in the IWP. Recently, Turker *et al.* [11] developed an alternative stabilization procedure for a class of two degree-of-freedom

under-actuated mechanical systems based on a set of transformations and a Lyapunov function.

The main objective of the paper is to design a  $\mathcal{H}_\infty$  controller to stabilize the pendulum at the up-right position in spite of the external disturbances. Since the initial conditions were far from the region of attraction of the desired equilibrium point, we construct a hybrid controller consisting of the swing-up control and the stabilization part. For the swing-up control, it is proposed a desired periodic trajectory where a nonlinear  $\mathcal{H}_\infty$ -control for time-varying systems, taken from [3], drives the pendulum to the region of attraction of the desired equilibrium point. For the stabilization part, we linearize the model around that equilibrium point where a linear  $\mathcal{H}_\infty$ -control for autonomous systems, taken from [2], is applied to stabilize the pendulum.

The paper is organized as follows. The dynamic model and control objective are given in Section 2. Section 3 provides background material on output-feedback nonlinear  $\mathcal{H}_\infty$ -control for autonomous and non-autonomous systems. The nonlinear  $\mathcal{H}_\infty$ -control for swing-up and stabilization of the inertia wheel pendulum are designed in Section 4. Numerical simulations are presented in Section 5. Finally, conclusions are provided in Section 6.

## 2 Dynamic model

Dynamics of an inertia wheel pendulum, taken from [12], can be described as follows:

$$\underbrace{\begin{bmatrix} J_a + J_r & J_r \\ J_r & J_r \end{bmatrix}}_J \begin{bmatrix} \ddot{q}_1 \\ \ddot{q}_2 \end{bmatrix} + \begin{bmatrix} \bar{m}g \sin(q_1) \\ 0 \end{bmatrix} = B(\tau + w_1) \quad (1)$$

where  $q_1(t) \in \mathbb{R}$  is the angle of the pendulum,  $q_2(t) \in \mathbb{R}$  is the angle of the disk (see Fig. 1),  $\tau(t) \in \mathbb{R}$  is the control input,  $t \in \mathbb{R}^+$  is the time,  $w_1(t) \in \mathbb{R}$  is the unknown disturbance vector which is assumed square integrable over infinite-time (belongs to  $\mathcal{L}_2$ ),  $J$  is the inertia matrix which is definite positive, and  $B = [0, 1]^T$ . In the above equation of motion,  $J_a = m_1 l_{c1}^2 + m_2 l_1^2 + J_p$ , and  $\bar{m} = m_1 l_{c1} + m_2 l_1$  where  $m_1$  is the mass of the pendulum,  $m_2$  is the mass of the wheel,  $l_1$  is the length of the pendulum,  $l_{c1}$  is the distance to the center of mass of the pendulum,  $J_p$  is the moment of inertia of the pendulum, and  $J_r$  is the moment of inertia of the wheel.

The control objective pursued in this paper is to asymptotically stabilize the angle of the pendulum  $q_1$  in the upper position set-point, that is,

$$\lim_{t \rightarrow \infty} \|q_1(t) - \pi\| = 0 \quad (2)$$

starting from the initial condition  $q_1(0) = q_2(0) = \dot{q}_1(0) = \dot{q}_2(0) = 0$  in spite of the external disturbances  $w_1(t) \in \mathbb{R}$ . The position of the wheel and the pendulum are the only measurements available for feedback and these measurements are corrupted by the vector  $w_y(t) \in \mathbb{R}^2$ .

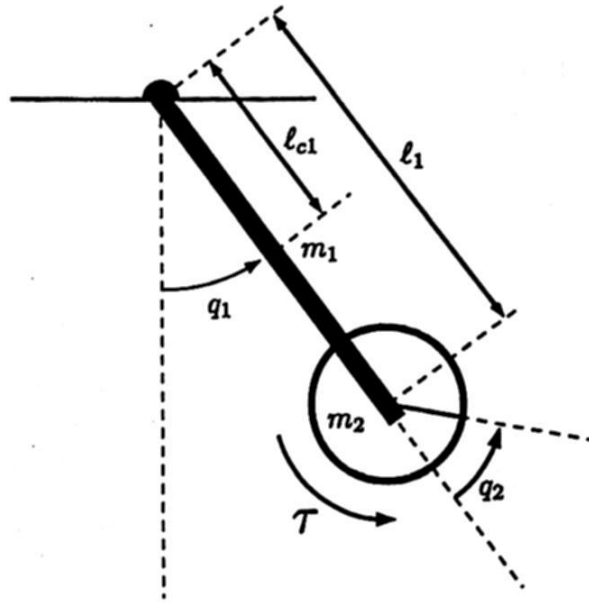


Fig. 1. Schematic representation of the inertia wheel pendulum.

### 3 Preliminaries

The present study focuses on a nonautonomous nonlinear system of the form

$$\dot{x} = f(x, t) + g_1(x, t)w + g_2(x, t)u \quad (3)$$

$$z = h_1(x, t) + k_{12}(x, t)u \quad (4)$$

$$y = h_2(x, t) + k_{21}(x, t)w \quad (5)$$

where  $x(t) \in \mathbb{R}^n$  is the state-space vector,  $u(t) \in \mathbb{R}^m$  is the control input,  $w(t) \in \mathbb{R}^r$  is the unknown disturbance,  $z(t) \in \mathbb{R}^l$  is the unknown output to be controlled, and  $y(t) \in \mathbb{R}^p$  is the only available measurement on the system.

For the underlying system, the following assumptions are made throughout.

**A1.** The functions  $f(x, t)$ ,  $g_1(x, t)$ ,  $g_2(x, t)$ ,  $h_1(x, t)$ ,  $h_2(x, t)$ ,  $k_{12}(x, t)$ ,  $k_{21}(x, t)$  are piecewise-continuous in  $t$  for all  $x$  and locally Lipschitz continuous in  $x$  for all  $t$ .

**A2.**  $f(0, t) = 0$ ,  $h_1(0, t) = 0$ , and  $h_2(0, t) = 0$  for all  $t$ .

**A3.**

$$\begin{aligned} h_1^T(x, t)k_{12}(x, t) &= 0, \quad k_{12}^T(x, t)k_{12}(x, t) = I \\ k_{21}(x, t)g_1^T(x, t) &= 0, \quad k_{21}(x, t)k_{21}^T(x, t) = I. \end{aligned} \quad (6)$$

These assumptions are made for technical reasons (cf. [13]).

The  $\mathcal{H}_\infty$ -control problem in question is stated as follows. Given the system representation (3)–(5) and a real number  $\gamma > 0$ , it required to find (if any) a causal dynamic output-feedback compensator

$$u = \mathcal{K}(\xi, t), \quad \dot{\xi} = \mathcal{F}(y, \xi) \quad (7)$$

with internal state  $\xi \in \mathbb{R}^s$ , such that the undisturbed closed-loop system is uniformly asymptotically stable around the origin and its  $\mathcal{L}_2$  gain less than  $\gamma$  if the response  $z$ , resulting from  $w$  for initial state  $x(t_0) = 0$  (and  $\xi(t_0) = 0$ ), satisfies

$$\int_{t_0}^{t_1} \|z(t)\|^2 dt < \gamma^2 \int_{t_0}^{t_1} \|w(t)\|^2 dt \quad (8)$$

for all  $t_1 > t_0$  and all piecewise-continuous functions  $w(t) = [w_1 \ w_y]^T$ . In turn, a locally admissible controller (7) constitutes a local solution of the  $\mathcal{H}_\infty$ -control problem if there exists a neighborhood  $U$  of the equilibrium such that inequality (8) is satisfied for all  $t_1 > t_0$  and all piecewise-continuous functions  $w(t)$  for which the state trajectory of the corresponding closed-loop system, starting from the initial point  $x(t_0) = 0$  (and  $\xi(t_0) = 0$ ), remains in  $U$  for all  $t \in [t_0, t_1]$ .

Under Assumptions A1-A3, coupled together, the corresponding Hamilton–Jacobi–Isaacs inequalities are subsequently linearized and a local solution of the  $\mathcal{H}_\infty$ -control problem is then obtained. The development involves the linear  $\mathcal{H}_\infty$ -control problem for the system

$$\begin{aligned} \dot{x} &= A(t)x + B_1(t)w + B_2(t)u \\ z &= C_1(t)x + D_{12}(t)u \\ y &= C_2(t)x + D_{21}(t)w \end{aligned} \quad (9)$$

where

$$\begin{aligned} A(t) &= \frac{\partial f_1}{\partial x}(0, t), \quad B_1(t) = g_1(0, t), \quad B_2(t) = g_2(0, t), \quad C_1(t) = \frac{\partial h_1}{\partial x}(0, t), \\ C_2(t) &= \frac{\partial h_2}{\partial x}(0, t), \quad D_{12}(t) = k_{12}(0, t), \quad D_{21}(t) = k_{21}(0, t). \end{aligned} \quad (10)$$

The following conditions are necessary and sufficient for a solution of the problem to exist:

C1. The equation

$$-\dot{P} = P(t)A(t) + A^T(t)P(t) + C_1^T(t)C_1(t) + P(t)\left[\frac{1}{\gamma^2}B_1B_1^T - B_2B_2^T\right](t)P(t) \quad (11)$$

possesses a uniformly bounded positive semidefinite symmetric solution  $P(t)$  such that the system

$$\dot{x} = [A - (B_2B_2^T - \gamma^{-2}B_1B_1^T)P](t)x(t) \quad (12)$$

is exponentially stable.



**C2.** Being specified with  $A_1(t) = A(t) + \frac{1}{\gamma^2} B_1(t) B_1^T(t) P(t)$ , the equation

$$\dot{Z} = A_1(t)Z(t) + Z(t)A_1^T(t) + B_1(t)B_1^T(t) + Z(t)\left[\frac{1}{\gamma^2}PB_2B_2^TP - C_2^TC_2\right](t)Z(t), \quad (13)$$

possesses a uniformly bounded positive semidefinite symmetric solution  $Z(t)$ , such that the system

$$\dot{x} = [A_1 - Z(C_2^TC_2 - \gamma^{-2}PB_2B_2^TP)](t)x(t) \quad (14)$$

is exponentially stable.

According to the time-varying strict bounded real lemma [14, p. 295], Conditions C1 and C2 ensure that there exists a positive constant  $\varepsilon_0$  such that the system of the perturbed Riccati equations

$$\begin{aligned} -\dot{P}_\varepsilon = P_\varepsilon(t)A(t) + A^T(t)P_\varepsilon(t) + P_\varepsilon(t)\left[\frac{1}{\gamma^2}B_1B_1^T - B_2B_2^T\right](t)P_\varepsilon(t) \\ + C_1^T(t)C_1(t) + \varepsilon I, \end{aligned} \quad (15)$$

$$\begin{aligned} \dot{Z}_\varepsilon = A_\varepsilon(t)Z_\varepsilon(t) + Z_\varepsilon(t)A_\varepsilon^T(t) + Z_\varepsilon(t)\left[\frac{1}{\gamma^2}P_\varepsilon B_2B_2^T P_\varepsilon - C_2^TC_2\right](t)Z_\varepsilon(t) \\ + B_1(t)B_1^T(t) + \varepsilon I \end{aligned} \quad (16)$$

has a unique uniformly bounded, positive definite symmetric solution  $(P_\varepsilon(t), Z_\varepsilon(t))$  for each  $\varepsilon \in (0, \varepsilon_0)$  where  $A_\varepsilon(t) = A(t) + \frac{1}{\gamma^2} B_1(t) B_1^T(t) P_\varepsilon(t)$ . Equations (15) and (16) are now utilized to derive a local solution of the  $\mathcal{H}_\infty$ -control problem for system (3)–(5).

**Theorem 1.** Consider system (3)–(5) with Assumptions A1–A3. Let Conditions C1 and C2 be satisfied with a certain  $\gamma > 0$  and let  $(P_\varepsilon(t), Z_\varepsilon(t))$  be a uniformly bounded positive definite symmetric solution of (15), (16) under some  $\varepsilon > 0$ . Then, the causal dynamic output-feedback compensator

$$\begin{aligned} \dot{\xi} = f(\xi, t) + \left[\frac{1}{\gamma^2}g_1(\xi, t)g_1^T(\xi, t) - g_2(\xi, t)g_2^T(\xi, t)\right]P_\varepsilon(t)\xi \\ + Z_\varepsilon(t)C_2^T(t)[y(t) - h_2(\xi, t)], \end{aligned} \quad (17)$$

$$u = -g_2^T(\xi, t)P_\varepsilon(t)\xi \quad (18)$$

is a local solution of the  $\mathcal{H}_\infty$ -control problem with the disturbance attenuation level  $\gamma$ .

In the autonomous case, the DREs (11), (13) degenerate to the algebraic Riccati equations by setting  $\dot{P} = 0$ ,  $\dot{Z} = 0$  and conditions C1 and C2 are simplified to

C1''. the equation

$$PA + A^TP + C_1^TC_1 + P\left[\frac{1}{\gamma^2}B_1B_1^T - B_2B_2^T\right]P = 0 \quad (19)$$

possesses a positive semidefinite symmetric solution  $P$  such that the matrix  $A - (B_2 B_2^T - \gamma^{-2} B_1 B_1^T)P$  has all eigenvalues with negative real part;  $C2''$ , being specified with  $A_1 = A + \frac{1}{\gamma^2} B_1 B_1^T P$ , the equation

$$A_1 Z + Z A_1^T + B_1 B_1^T + Z \left[ \frac{1}{\gamma^2} P B_2 B_2^T P - C_2^T C_2 \right] Z = 0, \quad (20)$$

possesses a positive semidefinite symmetric solution  $Z$  such that the matrix  $A_1 - Z(C_2^T C_2 - \gamma^{-2} P B_2 B_2^T P)$  has all eigenvalues with negative real part.

Conditions  $C1''$  and  $C2''$  are known from [1] to be necessary and sufficient for a solution of the linear  $\mathcal{H}_\infty$ -control problem for the time-invariant version of system (9) to exist. According to the strict bounded real lemma, Conditions  $C1''$  and  $C2''$  ensure that there exists a positive constant  $\varepsilon_0$  such that the system of the perturbed algebraic Riccati equations

$$P_\varepsilon A + A^T P_\varepsilon + C_1^T C_1 + P_\varepsilon \left[ \frac{1}{\gamma^2} B_1 B_1^T - B_2 B_2^T \right] P_\varepsilon + \varepsilon I = 0, \quad (21)$$

$$A_\varepsilon Z_\varepsilon + Z_\varepsilon A_\varepsilon^T + B_1 B_1^T + Z_\varepsilon \left[ \frac{1}{\gamma^2} P_\varepsilon B_2 B_2^T P_\varepsilon - C_2^T C_2 \right] Z_\varepsilon + \varepsilon I = 0 \quad (22)$$

has a unique positive definite symmetric solution  $(P_\varepsilon, Z_\varepsilon)$  for each  $\varepsilon \in (0, \varepsilon_0)$  where  $A_\varepsilon = A + \frac{1}{\gamma^2} B_1 B_1^T P_\varepsilon$ . Based on this solution a time-invariant  $\mathcal{H}_\infty$  controller is constructed as follows.

**Theorem 2.** *Let conditions  $C1''$  and  $C2''$  be satisfied for system (3)–(5) which is assumed to be time-invariant and let  $(P_\varepsilon, Z_\varepsilon)$  be a positive definite symmetric solution of (21), (22) under some  $\varepsilon > 0$ . Then the time-invariant output-feedback*

$$\dot{\xi} = f(\xi) + \left[ \frac{1}{\gamma^2} g_1(\xi) g_1^T(\xi) - g_2(\xi) g_2^T(\xi) \right] P_\varepsilon \xi + Z_\varepsilon C_2^T [y - h_2(\xi)], \quad (23)$$

$$u = -g_2^T(\xi) P_\varepsilon \xi \quad (24)$$

is a local solution of the  $\mathcal{H}_\infty$ -control problem in the autonomous case.

## 4 Nonlinear $\mathcal{H}_\infty$ -control synthesis

### 4.1 Swing-up control

The role of the swing-up control is to drive the pendulum inside the region of attraction of the desired equilibrium point  $[q_1^e q_2^e \dot{q}_1^e \dot{q}_2^e]^T = [\pi 0 0 0]^T$ . For this purpose, let us follow the line of reasoning of Orlov *et al.* [15] where the modified Van der Pol oscillator

$$\ddot{q}_d + \alpha \left[ \left( q_d^2 + \frac{\dot{q}_d^2}{\mu^2} \right) - \rho_v^2 \right] \dot{q}_d + \mu^2 q_d = 0, \quad \alpha, \rho_v, \mu > 0 \quad (25)$$

was used to generate a periodic reference trajectory for the pendulum. Here,  $q_d(t) \in \mathbb{R}$  stand for the desired periodic trajectory, the parameter  $\rho_v$  controls the amplitude of the limit cycle,  $\mu$  control its frequency, and  $\alpha$  controls the speed of the limit cycle transient.

For the synthesis of the  $\mathcal{H}_\infty$  tracking controller, consider the state-space vector  $x = [x_1 \ x_3]^T$  where  $x_1 = q_1 - q_d$  and  $x_3 = \dot{q}_1 - \dot{q}_d$ . Then (1), represented in terms of the state-space vector, takes the form

$$\begin{aligned}\dot{x}_1 &= x_3 \\ \dot{x}_3 &= -J_a^{-1}h \sin(x_1 + q_d) - \ddot{q}_d - J_a^{-1}w_1 - J_a^{-1}\tau,\end{aligned}\tag{26}$$

where  $h = \bar{m}g$ . It should be pointed out that it is assumed that velocity of the wheel  $\dot{q}_2$  remains bounded for all  $t \in [0, t_s)$  where  $t_s$  is the transition instant between the swing-up controller and the stabilizing controller. The objective is to design a controller of the form

$$\tau = -J_a(\ddot{q}_d + J_a^{-1}h \sin(q_1) + u)\tag{27}$$

that imposes on the disturbance-free manipulator motion desired stability properties around  $q_d(t)$  while also locally attenuating the effect of the disturbances. Thus, the controller to be constructed consists of the trajectory feedforward compensator and a disturbance attenuator  $u(t)$ ; internally stabilizing the closed-loop system around the desired trajectory.

In the sequel, we confine our design objective to position-control where

1. The output to be controlled is given by

$$z = \begin{bmatrix} u \\ \rho x_1 \end{bmatrix}\tag{28}$$

with a positive weight coefficient  $\rho$ , and

2. The position of the wheel and the pendulum are the only measurements available for feedback and these measurements are corrupted by the vector  $w_y(t) \in \mathbb{R}^2$ , that is,

$$y = \begin{bmatrix} x_1 + q_d \\ x_2 \end{bmatrix} + w_y(t).\tag{29}$$

The system (26)–(29) can be specified as in (3)–(5) with

$$\begin{aligned}f(x) &= \begin{bmatrix} x_3 \\ 0 \end{bmatrix}, \quad g_1(x) = \begin{bmatrix} 0 & 0 & 0 \\ -J_a^{-1} & 0 & 0 \end{bmatrix}, \quad g_2(x) = \begin{bmatrix} 0 \\ -J_a^{-1} \end{bmatrix}, \\ h_1(x) &= \begin{bmatrix} 0 \\ \rho x_1 \end{bmatrix}, \quad k_{12}(x) = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad h_2(x) = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \quad k_{21}(x) = [0_{2 \times 1} \ I_2].\end{aligned}\tag{30}$$

Hereinafter,  $I_n$  and  $0_{n \times m}$  stand for the  $n \times n$  identity matrix and the  $n \times m$  matrix of zeros, respectively. Thus, a solution to the  $\mathcal{H}_\infty$  output tracking controller synthesis involves the standard linear  $\mathcal{H}_\infty$ -control problem for the nonautonomous linearized

system (9) where matrices (10) are explicitly given by

$$\begin{aligned} A &= \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, & B_1 &= \begin{bmatrix} 0 & 0 & 0 \\ -J_a^{-1} & 0 & 0 \end{bmatrix}, & B_2 &= \begin{bmatrix} 0 \\ -J_a^{-1} \end{bmatrix}, \\ C_1 &= \begin{bmatrix} 0 & 0 \\ \rho & 0 \end{bmatrix}, & D_{12} &= \begin{bmatrix} 1 \\ 0 \end{bmatrix}, & C_2 &= [I_2 \ 0_{2 \times 2}], & D_{21} &= [0_{2 \times 1} \ I_2]. \end{aligned} \quad (31)$$

Finally, by applying Theorem 1 subject to (30) and (31), the  $\mathcal{H}_\infty$ -control problem is solved.

## 4.2 Stabilization control

For this stage, the nonlinear system (1) will be linearized around the desired equilibrium point therefore, the control input  $\tau$  will be injected from the  $\mathcal{H}_\infty$ -control (24) without nonlinear compensation terms, that is,  $\tau = u$ . Now, consider the state-space vector  $x = [x_1 \ x_2 \ x_3 \ x_4]^T$  where  $x_1 = q_1 - \pi$ ,  $x_2 = q_2$ ,  $x_3 = \dot{q}_1$ , and  $x_4 = \dot{q}_2$ . Then (1), represented in terms of the state-space vector, takes the form

$$\begin{aligned} \dot{x}_1 &= x_3 \\ \dot{x}_2 &= x_4 \\ \dot{x}_3 &= -J_a^{-1} h \sin(x_1 + \pi) - J_a^{-1} w_1 - J_a^{-1} u \\ \dot{x}_4 &= J_a^{-1} h \sin(x_1 + \pi) + (J_r^{-1} + J_a^{-1}) w_1 + (J_r^{-1} + J_a^{-1}) u. \end{aligned} \quad (32)$$

We confine our design objective to position regulation where the output to be controlled is given by

$$z = \begin{bmatrix} u \\ \rho x_1 \\ \rho x_2 \end{bmatrix}. \quad (33)$$

The system (29), (32)–(33) can be specified as in (3)–(5) with

$$\begin{aligned} f(x) &= \begin{bmatrix} x_3 \\ x_4 \\ -J_a^{-1} h \sin(x_1 + \pi) \\ J_a^{-1} h \sin(x_1 + \pi) \end{bmatrix}, & g_1(x) &= \begin{bmatrix} 0_{2 \times 1} & 0_{2 \times 2} \\ J^{-1} B & 0_{2 \times 2} \end{bmatrix}, & g_2(x) &= \begin{bmatrix} 0_{2 \times 1} \\ J^{-1} B \end{bmatrix}, \\ h_1(x) &= \begin{bmatrix} 0 \\ \rho x_1 \\ \rho x_2 \end{bmatrix}, & k_{12}(x) &= \begin{bmatrix} 1 \\ 0_{2 \times 1} \end{bmatrix}, & h_2(x) &= \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, & k_{21}(x) &= [0_{2 \times 1} \ I_2]. \end{aligned} \quad (34)$$

Thus, a solution to the  $\mathcal{H}_\infty$  output regulator synthesis involves the standard linear  $\mathcal{H}_\infty$ -control problem for the autonomous linearized system (9) where matrices (10) are ex-

plicitly given by

$$A = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ J_a^{-1}h & 0 & 0 & 0 \\ -J_a^{-1}h & 0 & 0 & 0 \end{bmatrix}, \quad B_1 = \begin{bmatrix} 0_{2 \times 1} & 0_{2 \times 2} \\ J^{-1}B & 0_{2 \times 2} \end{bmatrix}, \quad B_2 = \begin{bmatrix} 0_{2 \times 1} \\ J^{-1}B \end{bmatrix},$$

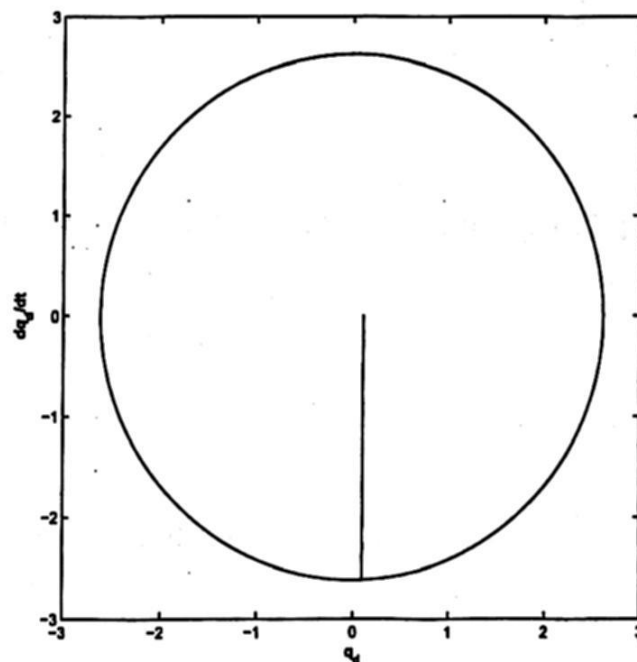
$$C_1 = \begin{bmatrix} 0_{1 \times 2} & 0_{1 \times 2} \\ \rho I_2 & 0_{2 \times 2} \end{bmatrix}, \quad D_{12} = \begin{bmatrix} 1 \\ 0_{2 \times 1} \end{bmatrix}, \quad C_2 = [I_2 \ 0_{2 \times 1}], \quad D_{21} = [0_{2 \times 1} \ I_2]. \quad (35)$$

Finally, by applying Theorem 2 subject to (34) and (35), the  $\mathcal{H}_\infty$ -control problem is solved.

## 5 Simulation results

Forthcoming result were based on the laboratory inertia wheel pendulum from Mechatronics Control Kit, prototype manufactured by QUANSER Inc., where  $J_a = 4.572 \times 10^{-3}$ ,  $J_r = 2.495 \times 10^{-5}$ ,  $h = 0.3544$ . As was specified in Section 2, the position initial conditions for the IWP were set to  $q_1(0) = q_2(0) = 0$  [rad] whereas all the velocity initial conditions were set to  $\dot{q}_1(0) = \dot{q}_2(0) = 0$  [rad/s].

The parameters specified for the modified Van der Pol equation (25), to generate a periodic trajectory with amplitude  $|q_d| = 3\pi/4$  [rad], were  $\alpha = 100$ ,  $\rho_v = 3\pi/4$ , and  $\mu = 4$  (see Fig. 2). The parameters  $\rho = 250$ ,  $\gamma = 10$ , and  $\varepsilon = 5$  were chosen for the swing-up  $\mathcal{H}_\infty$  controller and for  $\mathcal{H}_\infty$  regulator were  $\rho = 1$ ,  $\gamma = 40$ , and  $\varepsilon = 0$ .



**Fig. 2.** Phase portrait produced by the modified Van der Pol equation with parameters  $\alpha = 100$ ,  $\rho_v = 3\pi/4$ , and  $\mu = 4$  initialized at  $q_d(0) = 3\pi/4$  and  $\dot{q}_d(0) = 0$ .



Figure 3 provides the position, velocity, and torque of the pendulum without disturbances. Figure 4 shows that positions and velocities of the inertia wheel pendulum are driven to the desired equilibrium point in spite of the external disturbances

$$w_1 = 0.1 \cos(2t), \quad w_y = [1 \times 10^{-3} \cos(5t) \ 2 \times 10^{-3} \cos(3t)]^T. \quad (36)$$

The switching between controllers occur at  $t_s = 0.4$  [s]. There were no significant differences between Figs. 3 and 4 in terms of the overshoot. As can be seen, also, from these Figures, is that the velocity  $\dot{q}_2(t)$  does not escape to infinity in finite time  $t_s$ , however, high velocity of the wheel is required to satisfy the control objective. New methods for swing-up the pendulum without demanding high energy must be investigated.

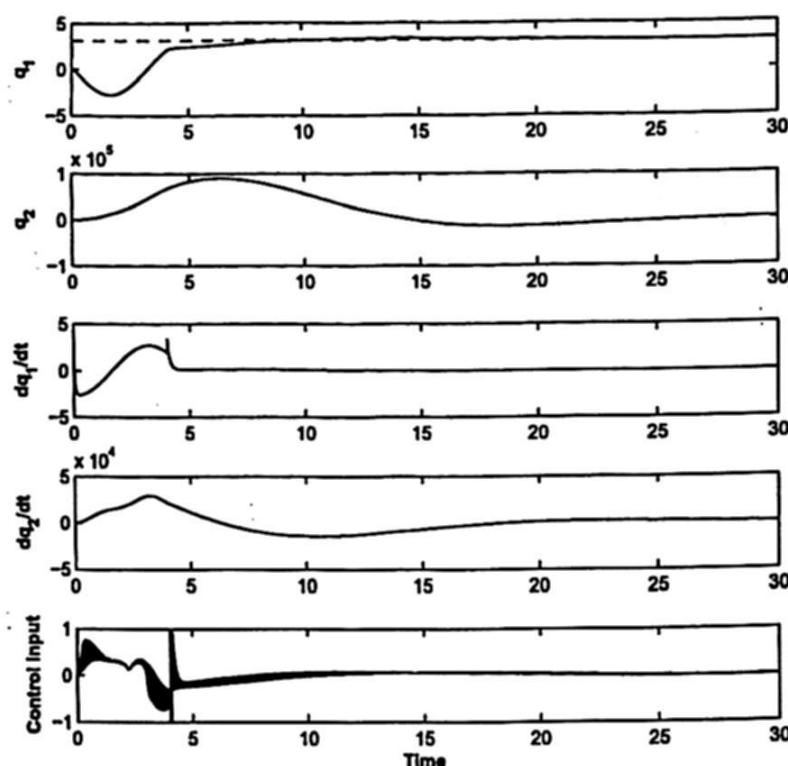


Fig. 3. Time responses of the closed-loop system.

## 6 Conclusions

In this paper we propose a nonlinear  $\mathcal{H}_\infty$ -controller to solve the stabilization problem, at the upright position, of a inertia wheel pendulum operating under uncertain conditions assuming position feedback only. The proposed controller consists of the swing-up part and the stabilization part. For the swing-up control, we proposed a Van-der-Pol equation to generate a continuously differentiable desired periodic trajectory where the  $\mathcal{H}_\infty$ -control successfully drives the pendulum to the region of attraction of the desired equilibrium point. Finally, the nonlinear  $\mathcal{H}_\infty$  regulator stabilizes the pendulum at the

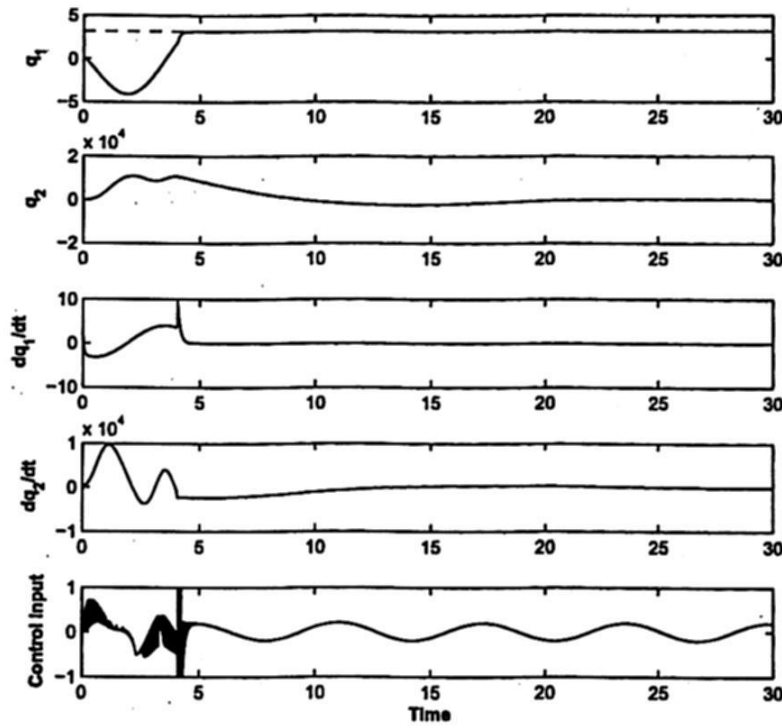


Fig. 4. Time responses of the closed-loop system.

desired position in spite of the presence of external disturbances, as was expected. The most important limitation lies in the swing-up control part where boundedness of the velocity of the wheel must be investigated.

## References

1. Doyle, J., Glover, K., Khargonekar, P., Francis, B.: State-space solutions to standard  $\mathcal{H}_2$  and  $\mathcal{H}_\infty$  control problems. *IEEE Transactions on Automatic Control* **34**(8) (1989) 831–847
2. Isidori, A., Astolfi, A.: Disturbance attenuation and  $\mathcal{H}_\infty$ -control via measurement feedback in nonlinear systems. *IEEE Trans. Autom. Control* **37**(9) (1992) 1283–1293
3. Acho, L., Orlov, Y., Solis, V.: Non-linear measurement feedback  $\mathcal{H}_\infty$ -control of time-periodic systems with application to tracking control of robot manipulators. *International Journal of Control* **74**(2) (2010) 190–198
4. Choi, Y., Chung, W.: *PID Trajectory Tracking Control for Mechanical Systems. Lecture Notes in Control and Information Sciences.* Springer, Berlin (2004)
5. Westerberg, S., Mettin, U., Shiriaev, A., Freidovich, L., Orlov, Y.: Motion planning and control of a simplified helicopter model based on virtual holonomic constraints. In: *Proc. 14th International Conference on Advanced Robotics, Munich, Germany* (2009) 1–6
6. Bentsman, J., Orlov, Y., Aguilar, L.: Nonsmooth  $\mathcal{H}_\infty$  output regulation with application to a coal-fired boiler/turbine unit with actuator deadzone. In: *American Control Conference ACC'13, Washington DC, USA* (2013) 3900–3905
7. Thum, C., Du, C., Chen, B., Ong, E.H., Tan, K.P.: A unified control scheme for track seeking and following of a hard disk drive servo system. *IEEE Transactions on Control Systems Technology* **18**(2) (2010) 294–306

8. Sun, W., Gao, H., Yao, B.: Adaptive robust vibration control of full-car active suspensions with electrohydraulic actuators. *IEEE Transactions on Control Systems Technology* **21**(6) (2013) 2417–2422
9. Ye, H., Gui, W., Jiang, Z.: Backstepping design for cascade systems with relaxed assumption on Lyapunov functions. *IET Control Theory and Applications* **5**(5) (2011) 700–712
10. Lopez-Martinez, M., Acosta, J., Cano, J.: Nonlinear sliding mode surfaces for a class of underactuated mechanical systems. *IET Control Theory and Applications* **4**(10) (2010) 2195–2204
11. Turker, T., Gorgun, H., Cansever, G.: Stabilisation of a class of 2-dof underactuated mechanical systems via direct Lyapunov approach. *International Journal of Control* **86**(6) (2013) 1137–1148
12. Block, D.J., Åström, K.J., Spong, M.W.: The Reaction Wheel Pendulum. *Synthesis Lectures on Control and Mechatronics* no. 1. Morgan & Claypool (2007)
13. Aguilar, L., Orlov, Y., Aho, L.: Nonlinear  $\mathcal{H}_\infty$ -control of nonsmooth time varying systems with application to friction mechanical manipulators. *Automatica* **39** (2003) 1531–1542
14. Brogliato, B., Lozano, R., Maschke, B., Ekeland, O.: *Dissipative Systems Analysis and Control*. 2nd edn. Springer, London (2006)
15. Orlov, Y., Aguilar, L., Aho, L., Ortiz, A.: Asymptotic harmonic generator and its application to finite time orbital stabilization of a friction pendulum with experimental verification. *International Journal of Control* **81**(2) (2008) 227–234

# Type-2 Fuzzy Control Lyapunov Approach for Position Trajectory Tracking

R. Farfan-Martinez<sup>1</sup>, J.A. Ruz-Hernandez<sup>2</sup>, J.L. Rullan-Lara<sup>2</sup>, W. Torres-Hernandez<sup>1</sup>, and L.A. Cambrano-Bravata<sup>1</sup>

<sup>1</sup>Universidad Tecnológica de Campeche. Carretera Federal 180 s/n,  
San Antonio Cardenas, C.P. 24381, Carmen, Campeche. Mexico.  
Tel: 01 (938) 3816700, Ext. 120

{farfan678,williantorreshernandez,luzbra}@hotmail.com

<sup>2</sup>Universidad Autonoma del Carmen. Calle 56, No. 4 Esq. Avenida Concordia,  
Col. Benito Juarez, C.P. 24180. Ciudad del Carmen, Campeche, Mexico.

{jruez,jrullan}@pampano.unacar.mx@hotmail.com

Paper received on 12/10/13, Accepted on 01/19/14.

**Abstract.** The paper presents the design of type-2 fuzzy controller using the *fuzzy Lyapunov synthesis approach* in order to systematically generate the rule base. To construct the rule base, the error signal and the derivative of the error signal are considered. It also presents the performance analysis to determine the value of the separation interval  $\xi$  between the upper and lower membership functions of the type-2 fuzzy set used. The controller is implemented via simulation to solve trajectory tracking problem for angular position of a servo trainer equipment in presence of backlash. Simulation results are successful and show better performance than a classic controller.

**Keywords:** Fuzzy Control, Lyapunov Approach, Nonlinearities, Servo Trainer.

## 1 Introduction

The fuzzy sets were introduced by L. A. Zadeh in the mid-sixties in order to process data affected by non-probabilistic uncertainty [1]. The type-1 fuzzy systems can handle the linguistic variables and experts reasoning and also reproduce the knowledge of systems to control, however, it can not handle uncertainties such as dispersions in linguistic distortion measurements and expert knowledge [2]. On the other hand, type-2 fuzzy systems can handle such kinds of uncertainties and also have the ability to model complex nonlinear systems. In addition, controllers designed using type-2 fuzzy systems achieve better performance than those of type-1. The type-2 fuzzy sets were also originally proposed by Zadeh in 1975 [3].

In [4] a fuzzy logic type-2 based controller using genetic algorithms is performed to control the shaft speed of a DC motor. Genetic algorithms are used to optimize triangular and trapezoidal membership functions. The controller is

implemented in a FPGA and its performance is compared with fuzzy logic type-1 and PID controllers.

A type-2 fuzzy controller (T2FC) is designed for an automatic guided vehicle for wall-following in [5]. In this case, T2FC has more robustness to sensor noise and better guidance performance than one of type-1. Another application of T2FC to mobile robots is presented in [6]. Trajectory tracking is applied first at simulation level and then on a Digital Signal Controller (DSC) of a experimental platform. The reported results show that performance of type-1 controller is poor comparing to type-2 controller.

Some applications of type-2 fuzzy controller in real-time can be also found in literature. For example, the classical inverted pendulum and the magnetic levitation system which are both highly non-linear. In [7], a low-cost microcontroller is used to validate the performance of T2FC for the inverted pendulum. For magnetic levitation system, [8] compared performance of type-1 and type-2 fuzzy controllers and a PID controller. Given that the system is unstable and non-linear, T2FC is showed better performance. Finally, position and velocity type-1 controller are designed in [9]. In this case, stability of both controllers are assured by means of Fuzzy Lyapunov Approach [10]. Results are presented in real time and are compared with classic controllers.

The paper is organized as follows. In section 2 we describe the servo trainer equipment and the characteristics of backlash. Then, section 3 presents the control design methodology using the fuzzy Lyapunov approach. Simulation results are presented in section 4. Finally, concluding remarks are presented in section 5.

## 2 Servo Trainer Equipment

The equipment used as plant to control in this paper is the *CE110 Servo Trainer* from *TQ Education and Training Ltd* [11]. This apparatus is used to help in teaching linear control theory and to implement validate some control algorithms (classical and non classical) in real-time.

The equipment have a variable load which is set using a current direct generator, by changes of different inertial load and using the engage a gearbox or by set all of them together. Besides, the apparatus have three modules to introduce some typical nonlinearities.

The mathematical model of servotrainer is set by equations [9]:

$$\begin{aligned}\dot{x}_1 &= x_2 \\ \dot{x}_2 &= -\frac{1}{T}x_2 + \frac{G_1 G_2}{T}u\end{aligned}\tag{1}$$

where  $x_1 = \theta$  and  $x_2 = \omega$  are the angular position and angular velocity, respectively. The gains  $G_1$  and  $G_2$  are defined by  $G_1 = k_i k_\omega$  and  $G_2 = k_\theta / 30 k_\omega$  where  $k_i = 3.229$  (rev/sec-Volts) is the motor constant,  $k_\omega = 0.3$  (Volts/(rev/sec)) is the velocity sensor constant and  $k_\theta = 20$  (Volts/rev) is the angle sensor constant. The time constant  $T$  change according to size of load:  $T = 1.5$  (sec) for small load (one inertial disc);  $T = 1$  (sec) for medium load



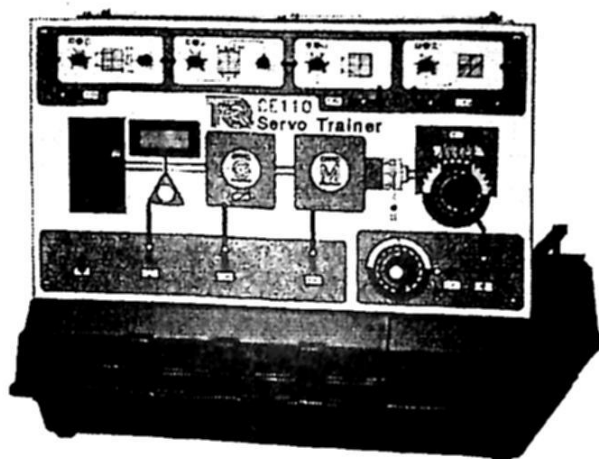


Fig. 1. Servo Trainer

(two inertial discs);  $T = 0.5$  (sec) for large load (three inertial discs). (see Fig. 1).

The equipment provides a hysteresis block, to simulate and study the important feature of backlash in the use of the gearbox and the effect on Servo Trainer.

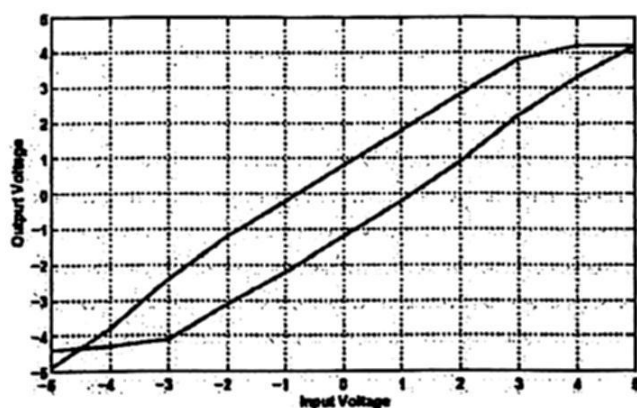


Fig. 2. Hysteresis Characteristics

The characteristics of the Servo Trainer hysteresis block are calculated experimentally (see Fig. 2) with width of 1 volt. Input and output voltage are taken in the block to determine its characteristics.

### 3 Controller Design

#### 3.1 Fuzzy Lyapunov Approach

The goal is to design a control law  $u$  such that the velocity and position of servo trainer follows a reference signal  $y_{ref}$ . One way of achieving this goal is to choose

a Lyapunov function candidate  $V(x)$ . Then, this Lyapunov function must meet the following requirements [10]:

$$V(0) = 0, \quad (2)$$

$$V(x) > 0, \quad x \in N \setminus \{0\}, \quad (3)$$

$$\dot{V}(x) = \sum_{i=1}^n \frac{\partial V}{\partial x_i} \dot{x}_i < 0, \quad x \in N \setminus \{0\}. \quad (4)$$

where  $N \setminus \{0\} \in R^n$  is some neighborhood of  $\{0\}$  excluding the origin  $\{0\}$  itself, and  $\dot{x}_i$  ( $i = 1, 2, \dots, n$ ). If  $\{0\}$  is an equilibrium point of (1) and such  $V(x)$  exist, then  $\{0\}$  is locally asymptotically stable.

The conditions (2) and (3) are satisfied by taking such Lyapunov function candidate  $V = \frac{1}{2}(e^2 + \dot{e}^2)$  where  $e$  is the tracking error. Differentiating  $V$  we have  $\dot{V} = e\dot{e} + \dot{e}\ddot{e}$ . Substituting  $w = \ddot{e}$ , is required then:

$$\dot{V} = e\dot{e} + \dot{e}w < 0 \quad (5)$$

Analyzing the equation (5), we can establish four basic fuzzy rules for  $w$  such that conditions (4) is satisfied:

- IF  $e$  is *positive* AND  $\dot{e}$  is *positive* THEN  $w$  is *negative big*
- IF  $e$  is *negative* AND  $\dot{e}$  is *negative* THEN  $w$  is *positive big*
- IF  $e$  is *positive* AND  $\dot{e}$  is *negative* THEN  $w$  is *zero*
- IF  $e$  is *negative* AND  $\dot{e}$  is *positive* THEN  $w$  is *zero*

### 3.2 Type-2 Fuzzy Systems

A fuzzy type-2 system denoted by  $\approx A$ , is characterized by a membership function type-2  $\mu_{\approx A} = (x, u)$ , where  $x \in X$ ,  $u \in J_x^u \subseteq [0, 1]$  and  $0 < \mu_{\approx A} = (x, u) < 1$ . It is defined as follows [12]

$$\approx A = \{(x, \mu_A(x) \mid x \in X)\} = \left[ \int_{x \in X} \left[ \int_{u \in J_x^u \subseteq [0, 1]} f_x(u)/u \right] / x \right] \quad (6)$$

If  $f_x(u) = 1$ ,  $\forall u \in [J_x^u, \bar{J}_x^u] \subseteq [0, 1]$ , membership function type-2  $\mu_{\approx A}$  is expressed by a lower membership function type-1  $\underline{J}_x^u = \underline{\mu}_A(x)$  and upper membership function type-1  $\bar{J}_x^u = \bar{\mu}_A(x)$ . Then,  $\mu_{\approx A}$  is called an fuzzy type-2 interval, denoted by equation (7)

$$\approx A = \left[ \int_{x \in X} \left[ \int_{u \in [\underline{\mu}_A(x), \bar{\mu}_A(x)] \subseteq [0, 1]} 1/u \right] / x \right] \quad (7)$$

If  $\approx A$  is a fuzzy type-2 singleton, then the membership function is defined by equation (8)

$$\mu_{\approx A}(x) = \begin{cases} 1/1, & \text{if } x = x' \\ 1/0, & \text{if } x \neq x' \end{cases} \quad (8)$$

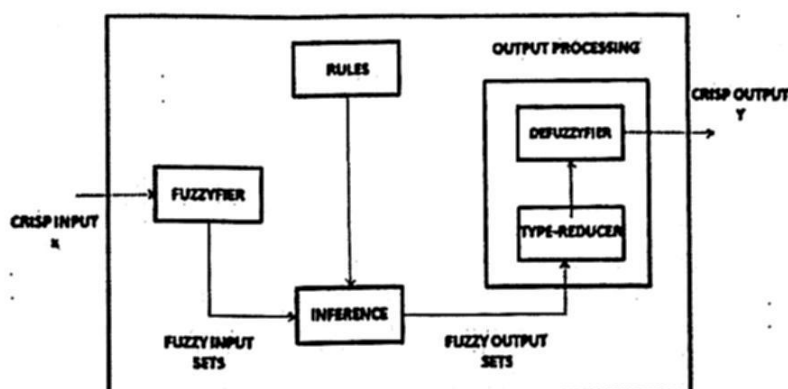


Fig. 3. Components of a type-2 fuzzy system

The type-2 fuzzy systems consist of a fuzzyfier which converts a value from real world into a fuzzy value, a fuzzy inference engine that applies a fuzzy reasoning to obtain a fuzzy output, an output processor comprising a reducer that transforms a fuzzy set type-2 into a fuzzy set type-1 and defuzzifier which converts a fuzzy value into a precise value (see Fig. 3).

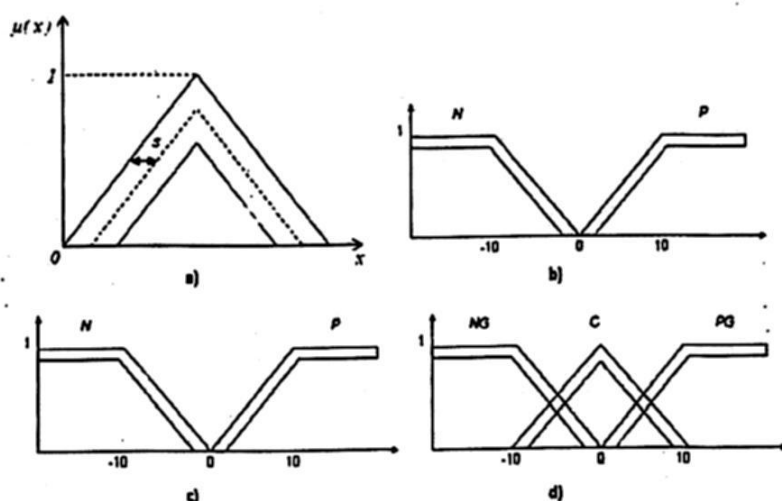


Fig. 4. Type-2 fuzzy sets: (a) Definition of type-2 fuzzy set; (b) Fuzzy set for  $e$ ; (c) Fuzzy set for  $\dot{e}$ ; (d) Fuzzy set for  $w$

As mentioned above, membership functions in type-2 fuzzy systems are characterized by having two membership functions of type-1; an upper and a lower membership function. The interval  $\xi$  between these two functions can be varied in order to obtain optimal performance [13]. Figure 4a shows such type-2 membership function.

In this paper we have used the Matlab Toolbox developed and described in [12] to implement the type-2 fuzzy system in order to generate values of  $w$ . Figures 4b-c shows fuzzy sets for error  $e$ , for the derivative of error  $\dot{e}$  and for variable  $w$ , respectively.

### 3.3 Mamdani Position Controller

The goal is to design a control signal  $u$  such that the angular position  $x_1$  follows a desired reference signal  $y_\theta$ . That is,  $e_\theta \rightarrow 0$  as  $t \rightarrow \infty$  where  $e_\theta = x_1 - y_\theta$ . In this case,  $\ddot{e}_\theta$  is related to  $w$  by  $\ddot{e}_\theta = w = \ddot{x}_1 - \ddot{y}_\theta$ . From equation (1), we have that  $\ddot{x}_1 = \dot{x}_2$  and the expression for  $w$  is  $w_\theta = -\frac{1}{T}x_2 + \frac{G_1 G_2}{T}u - \ddot{y}_\theta$ . Then, the control signal  $u$  for position tracking is

$$u = \frac{T}{G_1 G_2} (w_\theta + \ddot{y}_\theta) + \frac{1}{G_1 G_2} x_2 \quad (9)$$

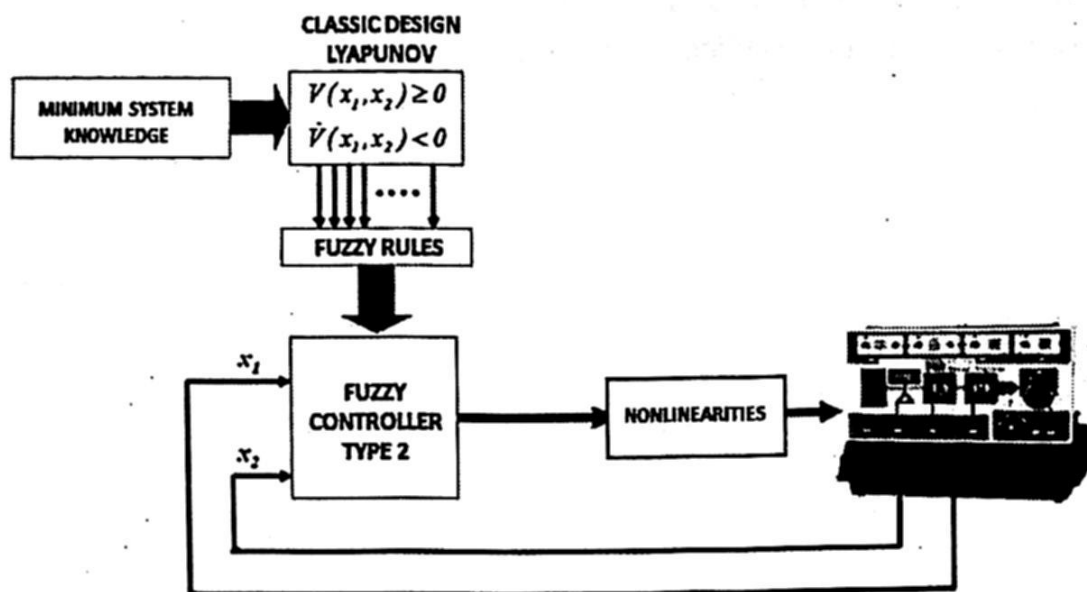


Fig. 5. Control scheme

Figure 5 shows the control scheme used for the simulations. The mathematical model of Servo Trainer and hysteresis block characteristics are determined experimentally in the equipment.

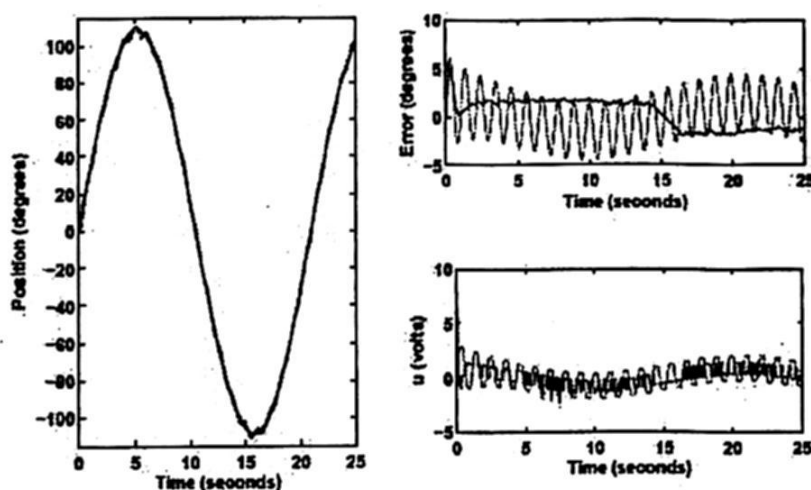
## 4 Simulation Results

In this section the *integral of the absolute value of the error (IAE)* and the *integral square error (ISE)* are used as performance criteria of proposed controller.

#### 4.1 Position controller

The reference signal is the sine signal  $y_\theta = 108 \sin 0.3t$  degrees and small load conditions are considered. Parameter  $\xi$  for type-2 fuzzy sets is set to 0.1. Performance of our controller is compared to those of a classical controller with a proportional controller for  $x_1$  ( $k_p = 10$ ) combined with a velocity feedback loop gain with  $k_v = 0.01$  [11].

In Figure 6(Left) we can observe the trajectory tracking. Both controllers have acceptable performance in tracking the position trajectory, but the classic controller has oscillations in the output from the start and kept until the end of the simulation. Looking the tracking errors (Up-Right side of Fig. 6) shows that the error T2FC is smaller in magnitude than the classic controller error. It is also observed that both control signals have oscillations during the simulation, but the control signal T2FC presents smaller variations. Both signals are within the limits of the actuator of the equipment.



**Fig. 6.** Position Tracking  $y_\theta$  (blue line), T2FC (green line), classical controller (red line). Left: Trajectory tracking; Up-Right: error signal; Down-Right: control signal

**Table 1.** Performance of trajectory tracking

	Controller	IAE	ISE
Position Tracking	Type-2 ( $\xi = 0.1$ )	2.06	0.1818
	PI ( $k_p = 10, k_i = 0.01$ )	2.919	0.4669

Finally, the Table 1 shows the performance in terms of IAE and ISE error criterions. Our type-2 fuzzy controllers had proved a good performance of the proposed approach and surpass performance of classical controller.



## 5 Conclusions

In this paper we have design type-2 fuzzy controller using the *fuzzy Lyapunov synthesis approach* in order to systematically generate the rule base. Controller is designed to solve the position trajectory tracking problem in a servo trainer system. To tuning the type-2 fuzzy controller, the separation between upper and lower membership functions is commanded by parameter  $\xi$  in steps of 0.1 units. The best tuning was obtained with  $\xi = 0.1$ .

The performance of our proposed controller is compared to classical controller under same simulations conditions for the servo trainer. The *IAE* and *ISE* are used as performance criterions. Simulation results had proved good performances of our proposed approach in position tracking applications and surpass performances of classical controller.

Actual research is conducted to test our controllers in medium and full load conditions in the servo trainer equipment. In order to demonstrate the effectiveness of our approach, authors are motivated to compare performances of type-2 fuzzy controller with performance of type-1 fuzzy controller.

## 6 Acknowledgments

The first and the fourth author thank to Universidad Tecnologica de Campeche for financial support and to Faculty of Engineering of the Universidad Autonoma del Carmen for providing the facilities for the use of equipment used in this work.

## References

1. Zadeh, L., A. Fuzzy Sets. Information and control 8, 338–353 (1965)
2. Kwak, H.J., Kim, D.W., Park, G.T., A New Fuzzy Inference Technique for Singleton Type-2 Fuzzy Logic Systems. International Journal of Advanced Robotic Systems 9, 1–7 (2012)
3. Zadeh, L., A. The concept of a linguistic variable and its application to approximate reasoning-1, Informat. Sci. 8, 199–249 (1975)
4. Maldonado, Y., Castillo, O. Genetic Design of an Interval Type-2 Fuzzy Controller for Velocity Regulation in a DC Motor. International Journal of Advanced Robotic Systems 9, 1–8 (2012)
5. Yao, L., Chen, Y.S., Type-2 Fuzzy Control of an Automatic Guided Vehicle for Wall-Following. In: Lucian Grigorie (ed.), Theory and Applications, pp. 243–252, ISBN: 978-953-307-543-3 (2011)
6. Leottau, L., Melgarejo, M. An Embedded Type-2 Fuzzy Controller for a Mobile Robot Application. In: Andon Topalov (Ed.), Recent Advances in Mobile Robotics, pp. 365–384, ISBN: 978-953-307-909-7. (2011)
7. Sierra, G.K., Bulla, J.O., Melgarejo, M.A., An Embedded Type-2 Fuzzy Processor For The Inverted Pendulum Control Problem. IEEE Latin America Transactions 9(3), 263–269 (2011)
8. Kumar, A., Kumar, V., Design and Implementation of IT2FLC for Magnetic Levitation System. Advances in Electrical Engineering Systems 1 (2), 116–123 (2012)

9. Ruz, J.A., Rullan, J.L., Garcia, R., Reyes, E.A., Sanchez, E. Trajectory Tracking Using Fuzzy-Lyapunov Approach: Application to a Servo Trainer. In: Castillo, O., Melin, P., Montiel Ross, O., Sepulveda Cruz, R., Pedrycz, W., Kacprzyk, J. (Eds.), *Theoretical Advances and Applications of Fuzzy Logic and Soft Computing*, Springer-Verlag, pp. 710-718 (2007)
10. Margaliot, M., Langholz, G., Fuzzy Lyapunov-based approach to the design of fuzzy Controllers. *Fuzzy Sets and Systems*, Elsevier 106, 49-59 (1999)
11. TecQuipment LTD: CE110 Servo Trainer, User's Manual, England (1993)
12. Castro, J.R., Castilo, O., Melin, P., Martinez, L. G., Escobar, S., Camacho, I., Building Fuzzy Inference Systems with the Interval Type-2 Fuzzy Logic Toolbox. In Melin, P. et al. (Eds.), *Analysis and Design of Intelligent Systems using Soft Computing Techniques*, Springer-Verlag, pp. 53-62 (2007)
13. Cazarez N. R., Castillo O, Aguilar L. and Cardenas S.: Lyapunov Stability on Type-2 Fuzzy Logic Control, *Proceedings of International Seminar on Computational Intelligence*, Mexico D. F., pp. 32-41, (2005)

# High-order Sliding-Mode Based Sub-optimal Linear Quadratic Regulator with Application to Roll Autopilot Design

Jorge Dávila

National Polytechnic Institute. Section of Graduate Studies and Research,  
ESIME-UPT, Av. Ticoman 600, Col. San Jose Ticoman, Gustavo A. Madero, Mexico  
D.F. (Tel: +52-55-57296000, Ext. 56100; e-mail: jadavila@ipn.mx)

*Paper received on 12/11/13, Accepted on 01/19/14.*

**Abstract.** A sub-optimal robust controller is designed for linear systems affected by external disturbances or bounded uncertainties. The sub-optimal controller is composed by two terms: first, a linear quadratic regulator, that provides optimal stabilization, is designed for the linear system in the absence of perturbations; second, a nonlinear compensation term, designed using the high-order sliding-modes techniques, is used to compensate the perturbations that affect the linear system. The proposed sub-optimal controller is applied to the design of a roll autopilot for a missile.

**Keywords:** Roll control, high-order sliding-modes, sub-optimal control

## 1 Introduction

### 1.1 State of the art

The design of controllers for perturbed systems is one of the most challenging problems in control theory. The control of systems under uncertain conditions has been successfully addressed using adaptive and robust techniques. However, for the aerospace applications, the appearance of other important phenomena during the flight significantly complicate the application of the adaptive control techniques. The fast changes in the plant's parameters caused by a fault must be identified and the control law reconfigured online. For example, both, the control derivatives and the stability derivatives, undergo significant changes due to a control surface fault, and control surface failure causes a trim disturbance that needs to be rejected by the flight control system [1].

Sliding mode control is known as an effective technique to deal with perturbed or uncertain systems, the application of sliding mode control techniques is restricted by the appearance of the chattering effect [2] (an undesirable high-frequency oscillation appearing on the system variables).

The High-Order Sliding-Modes techniques (see, for example, [3, 4]) mitigates the application problems related to standard sliding-modes by reducing the

switching frequency needed to maintain the sliding motion. This High-Order Sliding-Mode control techniques have been, in general, designed to stabilize systems with a relative degree greater than one preserving the robustness and accuracy of the standard sliding-modes.

The High-order sliding-modes techniques have already been applied to flight control problems with satisfactory results. In [6] a sliding modes based control is designed using two control loops, the proposed control ensures asymptotic tracking of the command deflections. The smooth second-order sliding-modes is applied for missile guidance in [7]. The dynamic sliding-manifolds technique is applied in combination with a transformation for stabilization of nonminimum phase aircrafts in the work by [8]. A fault detection algorithm and a fault tolerant control for a large aircraft with specific application in a B747 simulation model is presented in [9]. Recently [10] developed a 2nd order sliding-modes based black-box control for signal tracking, the proposed controller is tested in simulations with a 6-DOF UAV model.

An alternative to obtain the characteristic robustness of the sliding-modes, without applying directly the discontinuous control signals on the system, is the use of these techniques for the design of estimation algorithms. Disturbance identification algorithms are usually applied for fault tolerant and robust control design and the High-Order Sliding-Modes have been successfully applied for the design of observers and algorithms for estimation of disturbances (see, for example, [11,12]). The main advantages of the observers basing on high-order sliding-modes is their robustness against external perturbations [13–16], and that they bring the possibility of exploit the equivalent output injection for the designing of the identification algorithms for the disturbances.

The combination of sliding-mode control techniques with conventional control has allowed the development of robust control algorithms that are capable to solve the stabilization problem under uncertain conditions. Some of these works are briefly described below. In [17] a Linear Quadratic Regulator is applied to stabilize a nonlinear affine system using a compensation term generated by the use of integral sliding modes. In [18] the high-order sliding mode based hierarchical observer is applied to identify disturbances acting on a perturbed system, the identified signal is used to add robustness to the smooth control signal generated by standard feedback control. A backstepping design that combines the high-order sliding modes differentiator and the feedback linearization is proposed in [19].

In [20] the general model of an autopilot for tactical missiles is proposed. In this article, the dynamic of the missile is spliced into two decoupled dynamics. In one side the first order rigid body effect is considered; while by the other side, the dynamic corresponding to the flexible body dynamics is considered. Using this model, in [21] a robust autopilot is proposed using a Linear Quadratic Regulator with a compensation term designed using the methodology proposed in [22].



## 1.2 Main contribution

In this paper, a sub-optimal robust controller is designed for linear systems affected by external disturbances. The proposed controller is composed by two terms:

- A linear quadratic regulator capable to perform optimal stabilization of the linear system in the absence of perturbations.
- A nonlinear compensation term, designed using the high-order sliding-mode observer, that is used to compensate the perturbations that affect the linear system.

The controller is used to design a roll autopilot for a missile. The performance of the proposed observer is illustrated by simulations in the model presented in [20] and [21].

## 1.3 Paper structure

In Section 2, the class of systems under study is defined. The robust Linear Quadratic Regulator is designed in section 3, in particular, the linear part of the controller is presented in Subsection 3.1, and the high-order sliding-modes based compensation term is designed in Section 3.2. The application of the proposed technique to the roll autopilot design is given in Section 4. Section 5 provides conclusions to this study.

## 2 Problem statement

Let consider the following perturbed linear system

$$\dot{x} = Ax + Bu + Df \quad (1)$$

$$y = Cx \quad (2)$$

where  $x \in \mathbb{R}^n$  is the state vector,  $y \in \mathbb{R}^p$  is the system output,  $u \in \mathbb{R}^m$  and  $f \in \mathbb{R}^q$  are the control signal and disturbances, respectively. The matrices  $A$ ,  $B$ ,  $C$ , and  $D$  are all conformable matrices.

The Rosenbrock matrix of the triplet  $\{A, C, D\}$  is defined as:

$$R(s) = \begin{bmatrix} sI_n - A & -D \\ C & 0 \end{bmatrix} \quad (3)$$

The invariant zeros of the triplet  $\{A, C, D\}$  are given by the points  $s_0$  for which the Rosenbrock matrix  $R(s_0)$  loses rank.

It is considered that the system (1)-(2) satisfies the following assumptions:

**Assumption 1** *The triplet  $\{A, C, D\}$  does not have invariant zeros.*



**Assumption 2** The perturbation signal  $f$  satisfies

$$\|f\|_{\infty} \leq f^+$$

for a known scalar  $f^+ > 0$ , here  $\|\cdot\|_{\infty}$  denotes the infinite norm.

**Assumption 3** The control distribution matrix  $B$  and the disturbance term  $Df$  satisfy:

$$Df \in \text{span} B$$

Differential equations are understood in the Filippov sense [23] in order to provide for the possibility to use discontinuous signals in controls. Filippov solutions coincide with the usual solutions, when the right-hand sides are Lipschitzian. It is assumed also that all considered inputs allow the existence of solutions and their extension to the whole semi-axis  $t \geq 0$ .

The aim of this paper is designing a Linear Quadratic Regulator algorithm that can stabilize the state of the system (1)-(2) even in the presence of the disturbances vector  $f$ .

### 3 Robust Linear Quadratic Regulator Design

The proposed controller is composed by two signals. The first one, is used to provide for optimal stabilization the nominal system, while the second one is designed to guarantee robustness against the perturbation  $f$ . The control takes the following form.

$$u = u_1 + u_2 \quad (4)$$

where  $u_1$  and  $u_2$  will be designed below.

#### 3.1 Linear Quadratic Regulator

The Linear Quadratic Regulator is designed to minimize a quadratic performance index of the form

$$J = \int_0^{\infty} (x^T Q x + u_1^T R u_1) dt$$

where  $Q \geq 0$  and  $R > 0$  are weights to be chosen. The resulting is an optimal control law given by:

$$u_1 = -Kx \quad (5)$$

where the gain  $K$  is computed as  $K = R^{-1}B^T P$ , where  $P$  is computed as the solution of the matrix algebraic Riccati equation

$$PA + A^T P - PBR^{-1}B^T P + Q = 0$$

By an appropriate selection of matrixes  $Q$  and  $R$ , one can obtain the desired performance. This optimal controller is usually applied to solve a wide variety of problems, as for example, this controller is an usual tool for the design of autopilots.

### 3.2 High Order Sliding Modes based compensator

Under Assumption 1 the system (1),(2) is strongly observable (to see a deeper study about strong detectability, the reader can refer, for example, to the tutorial book [24]). This assumption allows us to reconstruct exactly and in a finite-time the state, even in the presence of the disturbance  $f$  [14].

With this aim, an observer which is based on the high-order sliding modes is proposed as:

$$\begin{aligned}\dot{z} &= Az + Bu + L(y - y_z) \\ e_y &= y - y_z \\ \hat{x} &= z + U^{-1}v(e_y)\end{aligned}\tag{6}$$

where the matrix  $U$  takes the form

$$U = \begin{bmatrix} C \\ C(A - LC) \\ \vdots \\ C(A - LC)^{n-1} \end{bmatrix}$$

the compensation term  $v(e_y)$  is composed by the variables

$$v(e_y) = [v_1 \ v_2 \ \cdots \ v_n]$$

where the components of the vector  $v_i$   $i = 1, \dots, n$ , and the additional variable  $v_{n+1}$  are taken from the high order sliding mode differentiator [25] given by:

$$\begin{aligned}\dot{v}_1 &= w_1 \\ w_1 &= -\alpha_{n+1}M^{1/(n+1)}|v_1 - e_y|^{n/(n+1)}\text{sign}(v_1 - e_y) + v_2 \\ \dot{v}_2 &= w_2 \\ w_2 &= -\alpha_n M^{1/n}|v_2 - w_1|^{(n-1)/n}\text{sign}(v_2 - w_1) + v_3 \\ &\vdots \\ \dot{v}_n &= w_n \\ w_n &= -\alpha_2 M^{1/2}|v_n - w_{n-1}|^{1/2}\text{sign}(v_n - w_{n-1}) + v_{n+1} \\ \dot{v}_{n+1} &= -\alpha_1 M \text{sign}(v_{n+1} - w_n)\end{aligned}\tag{7}$$

where the parameter  $M$  is chosen sufficiently large, in particular  $M > |d|f^+$ , where  $d = C(A - LC)^{n-1}D$ . The constants  $\alpha_i$  are chosen recursively sufficiently large as in [25]. In particular, one of the possible choices is  $\alpha_1 = 1.1$ ,  $\alpha_2 = 1.5$ ,  $\alpha_3 = 2$ ,  $\alpha_4 = 3$ ,  $\alpha_5 = 5$ ,  $\alpha_6 = 8$ , which is sufficient for  $n \leq 6$ . Note that (7) has a recursive form, useful for the parameter adjustment. In any computer realization one has to calculate the internal auxiliary variables  $v_j$  and  $w_j$ ,  $j = 1, \dots, n$ , using only the simultaneously-sampled current values of  $e_y$  and  $v_j$ .

The auxiliary output estimation error  $e_y$  and its first  $n$  derivatives take the following form

$$\begin{aligned}e_y &= y - Cz = C(x - z) \\ \dot{e}_y &= C(A - LC)(x - z) \\ &\vdots \\ e_y^{(n)} &= C(A - LC)^n(x - z) + C(A - LC)^{n-1}Df\end{aligned}$$

On the other hand, the high order sliding mode differentiator (7) brings an estimation of the derivatives up to order  $n - 1$ . Hence, after the convergence of the differentiator, the derivative of order  $n$  satisfies:

$$-\alpha_2 M^{1/2} |v_n - w_{n-1}|^{1/2} \text{sign}(v_n - w_{n-1}) + v_{n+1} = C(A - LC)^n(x - z) + C(A - LC)^{n-1}Df$$

Thus, the following equality holds after a finite time transient

$$v_{n+1} = C(A - LC)^n(x - z) + C(A - LC)^{n-1}Df. \quad (8)$$

The equation (8) is called the equivalent output injection. Given the properties of the differentiator (7),  $v_{n+1}$  is a continuous term.

The perturbation  $f$  can be identified through the equivalent output injection as:

$$\hat{f} = (C(A - LC)^{n-1}D)^{-1} (v_{n+1} - C(A - LC)^n U^{-1} v(e_y))$$

Notice that  $C(A - LC)^{n-1}D \neq 0$  otherwise no perturbations affects the system.

The control term that provides robustness against the disturbance  $f$  is proposed as

$$u_2 = H\hat{f} \quad (9)$$

where the matrix  $H$  is computed as  $H = B^+D$ , the matrix  $B^+$  is the Moore-Penrose left pseudoinverse of  $B$ , i.e.,  $B^+ = (B^T B)^{-1} B^T$ .

**Theorem 1.** *Be the system (1)-(2). Under Assumptions 1 - 3 the controller (4) provides suboptimal exact regulation under the presence of external perturbations.*

## 4 Application to the Robust Roll Autopilot Design

A wide variety of missiles possesses a cruciform configuration which brings to them a high accuracy and quick manoeuvring in any direction. However, the inherent instability of the roll yields in undesirable rolling motions that degrades the performance. To overcome this problem, the roll autopilots are proposed (see [21, 20]). The main objective of the above mentioned controllers is to maintain the attitude of the missile under system variations and external disturbances.

The block diagram of the missile roll dynamic is shown in Figure 1. The airframe flexibility is considered in the flexible body dynamics block. External disturbances  $d_{ext}$  are used to describe external perturbations.

The control is designed disregarding the flexible body effects and the external disturbances. In this sense, for analysis proposes, the flexible body dynamics, external perturbations and any other coupling effect derived from the pitch and yaw motions are concentrated in a single term,  $f$ . The state equations of the system can be written as:

$$\begin{aligned} \dot{x} &= Ax + Bu + Df \\ y &= Cx \end{aligned}$$

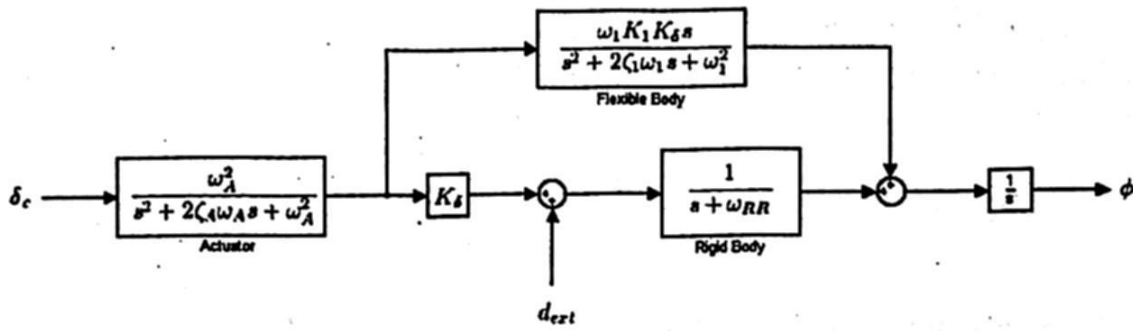


Fig. 1. System block diagram.

where the matrices  $A$ ,  $B$  and  $C$  take the following form

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -a_3 & -a_2 & -a_1 & -a_0 \end{bmatrix}, \quad B = \begin{bmatrix} 0 \\ 0 \\ 0 \\ b \end{bmatrix}, \quad D = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}$$

$$C = [1 \ 0 \ 0 \ 0]$$

where  $a_0 = 2\zeta_A\omega_A + \omega_{RR}$ ,  $a_1 = 2\zeta_A\omega_A\omega_{RR} + \omega_A^2$ ,  $a_2 = \omega_A^2\omega_{RR}$ ,  $a_3 = 0$ ,  $b = \omega_A^2 K_\delta$ .

The parameters are given in the following table:

Symbol	Variable	Value
$\omega_{RR}$	Roll rate bandwidth	2 rad/s
$K_\delta$	Fin Effectiveness	9000 1/s <sup>2</sup>
$\omega_A$	Actuator bandwidth	100 rad/s
$\zeta_A$	Actuator damping	0.65
$\omega_1$	Torsional mode frequency	250 rad/s
$\zeta_1$	Torsional mode damping	0.01
$K_1$	Torsional mode gain	-0.0000129
$\phi_{max}$	Maximum desired roll angle	10
$\dot{\phi}_{max}$	Maximum desired roll rate	300 deg/s
$\delta_{c(max)}$	Maximum desired fin deflection	30 deg

Table 1. System parameters.

For simulation purposes, the perturbation is given by:

$$d_{ext} = 54600 + 50000 \sin(\sin(t) \sin(0.1t) + 0.2)$$

The weighting matrices  $Q$  and  $R$  are the same as in [20]:

$$Q = \begin{bmatrix} \frac{1}{\phi_{max}^2} & 0 & 0 & 0 \\ 0 & \frac{1}{\dot{\phi}_{max}^2} & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \quad R = \begin{bmatrix} \frac{1}{\delta_{c(max)}^2} \end{bmatrix}$$

where  $\phi_{max}$ ,  $\dot{\phi}_{max}$ ,  $\delta_{c(max)}$  are the maximal permissible values of the respective variables.

Using the solution of the algebraic Matrix Riccati equation, the gain  $K$  for the controller (5) is given by

$$K = [3 \quad 0.1286 \quad 0.001 \quad 0]$$

The eigenvalues of matrix  $A$  are 0, -2,  $-65+75.9934i$ ,  $-65-75.9934i$ , notice that the system is marginally stable. The Luenberger gain of the observer is designed to obtain a stable estimation error. The gain  $L$  is chosen to place the roots of the estimation error dynamics matrix  $(A-LC)$  in -40, -41, -42, -43, as

$$L = \begin{bmatrix} 34 \\ -4417 \\ 499890 \\ -18385220 \end{bmatrix}$$

Using the compensated system matrix  $(A-LC)$ , the matrix  $U$  is given by:

$$U = \begin{bmatrix} 1 & 0 & 0 & 0 \\ -34 & 1 & 0 & 0 \\ 5573 & -34 & 1 & 0 \\ -839550 & 5573 & -34 & 1 \end{bmatrix}$$

The high-order sliding mode differentiator takes the form:

$$\begin{aligned} \dot{v}_1 &= w_1 = -\alpha_5 M^{1/5} |v_1 - e_y|^{4/5} \text{sign}(v_1 - e_y) + v_2 \\ \dot{v}_2 &= w_2 = -\alpha_4 M^{1/4} |v_2 - w_1|^{3/4} \text{sign}(v_2 - w_1) + v_3 \\ \dot{v}_3 &= w_3 = -\alpha_3 M^{1/3} |v_3 - w_2|^{2/3} \text{sign}(v_3 - w_2) + v_4 \\ \dot{v}_4 &= w_4 = -\alpha_2 M^{1/2} |v_4 - w_3|^{1/2} \text{sign}(v_4 - w_3) + v_5 \\ \dot{v}_5 &= -\alpha_1 M \text{sign}(v_5 - w_4) \end{aligned}$$

where the gains are chosen as  $\alpha_1 = 1.1$ ,  $\alpha_2 = 1.5$ ,  $\alpha_3 = 2$ ,  $\alpha_4 = 3$ ,  $\alpha_5 = 5$  and  $M = 2000$ .

The control signal (4) is given by  $u = -K\hat{x} + H\hat{f}$ , where

$$H = [0 \quad 0 \quad 0 \quad 0.1111 \times 10^{-7}]$$

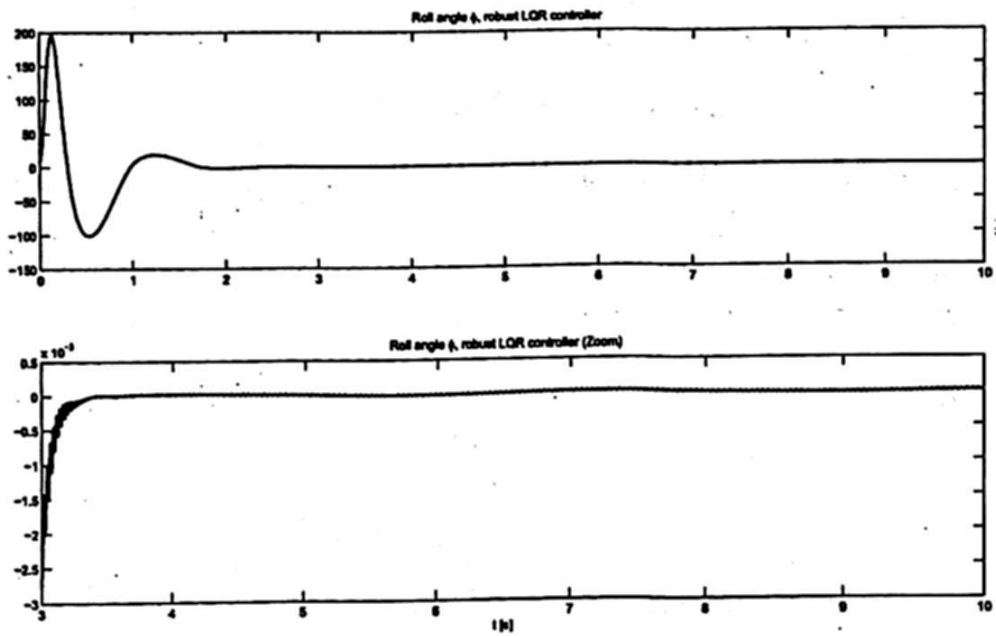
The convergence of the roll angle  $\phi$  to zero after a finite time transient is shown in the Figure 2. Deflection of the ailerons  $\delta$  and its ratio are presented in the Figure 3. The perturbation identification  $\hat{w}$  is shown in Figure 4.

The results obtained with the proposed methodology are compared with a standard Linear Quadratic Regulator. With this aim, the control signal takes the form:

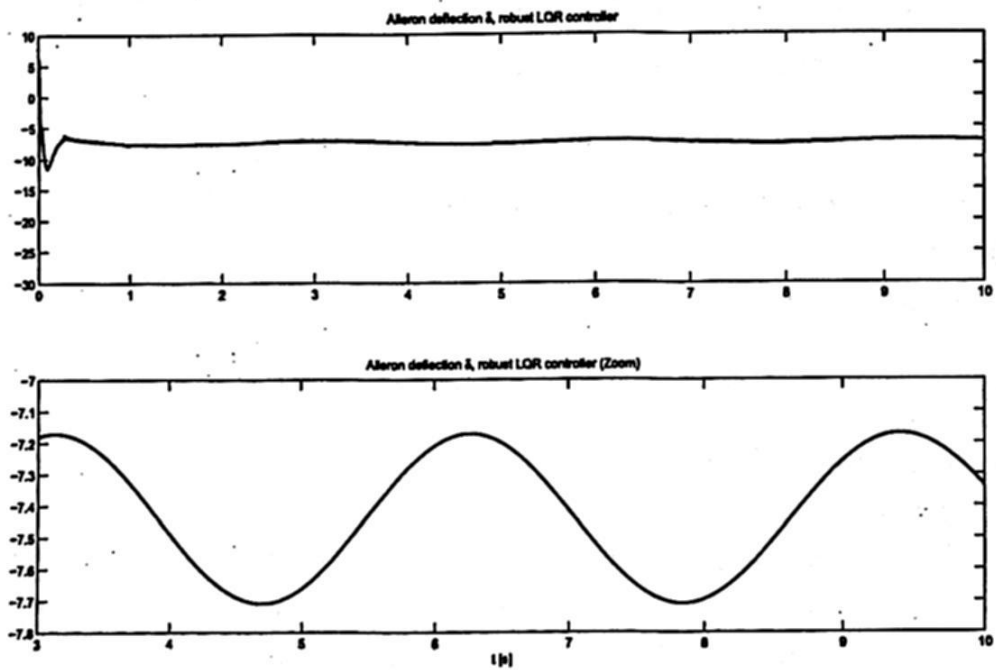
$$u = -Kx$$

The roll angle and a zoom on the graphic using the standard Linear Quadratic Regulator are shown in the Figure 5. In comparison with the standard Linear Quadratic Regulator, the robust Linear Quadratic Regulator is exact with





**Fig. 2.** Roll angle  $\phi$  (above) and a zoom on the image (below) for the robust LQR controller.



**Fig. 3.** Aileron Deflections  $\delta$  for the robust LQR controller.

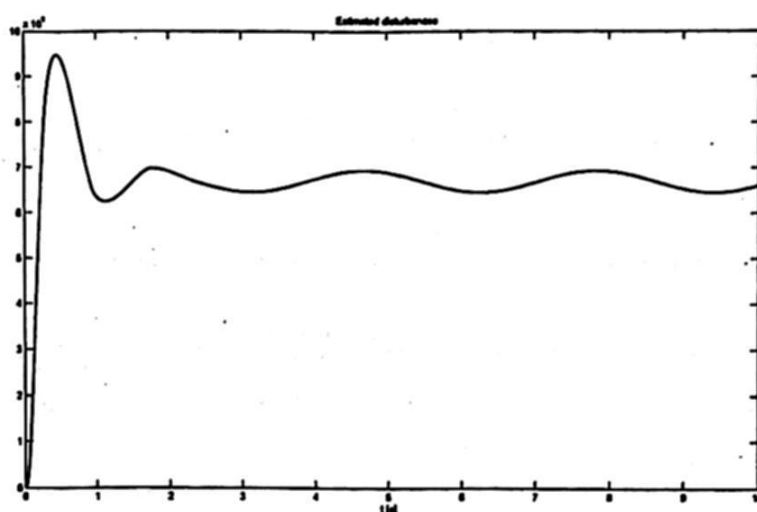


Fig. 4. Estimated disturbances.

respect to the coordinate  $\phi$ . The aileron deflection  $\delta$  for the standard Linear Quadratic Regulator is shown in the Figure 6. Notice that after 2.5 seconds, the deflections obtained for both controllers are very similar, then the most important contribution of the nonlinear compensation term takes place during the transient.

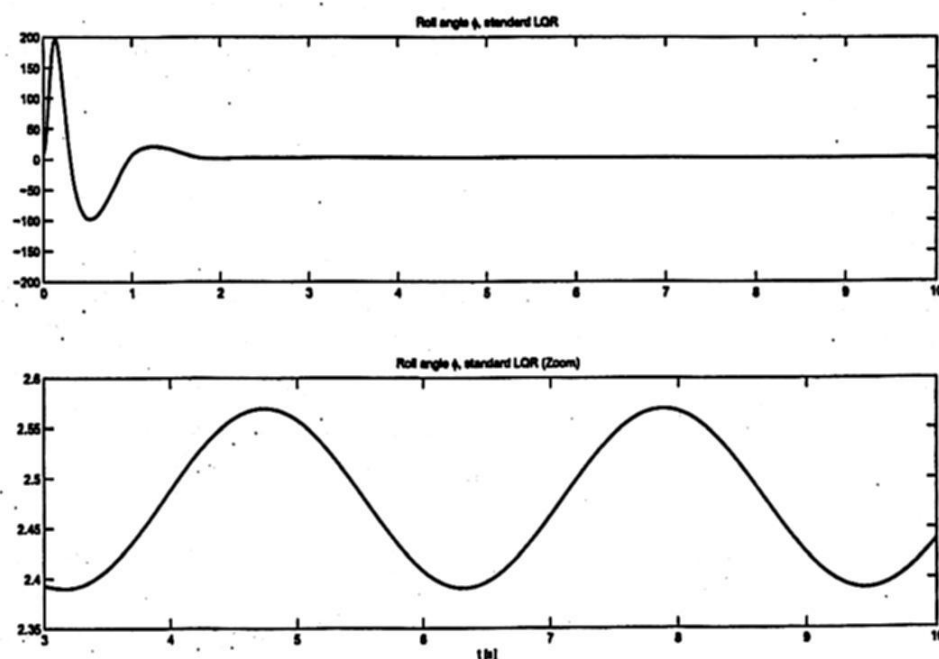
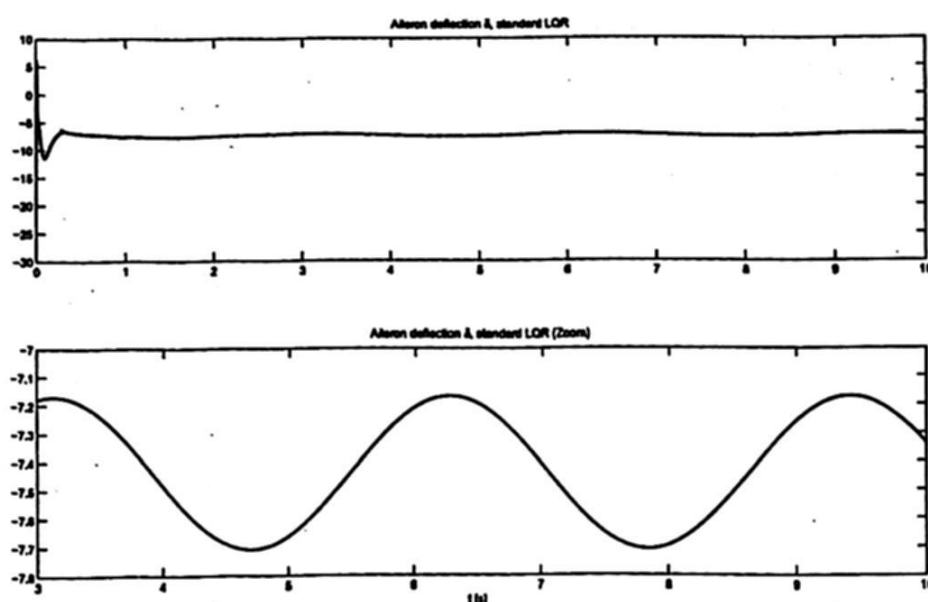


Fig. 5. Roll angle  $\phi$  (above) and a zoom on the image (below) for the standard LQR controller.



**Fig. 6.** Aileron Deflections  $\delta$  (above) and a zoom on the image (below) for the standard LQR controller.

## 5 Conclusions

In this paper a sub-optimal robust controller is proposed for linear systems. The controller is composed of two terms. The linear quadratic regulator term allowed the optimal stabilization of the linear system in the absence of perturbations, while the high-order sliding-mode based compensation term compensates the effect of perturbations disrupting the system. The proposed method is used to design a roll autopilot for a missile model. The robustness of the proposed controller is illustrated by simulations.

## References

1. Pachter, M., Chandler, P., Mears, M.: Reconfigurable tracking control with saturation. *AIAA Journal of Guidance, Control, and Dynamics* 18(5) (1995) 1016–1022
2. Fridman, L.: The problem of chattering: an averaging approach. In Young, K., Ozguner, U., eds.: *Variable Structure, Sliding Mode and Nonlinear Control*. Number 247 in *Lecture Notes in Control and Information Science*. Springer Verlag, London (1999) 363–386
3. Levant, A.: Sliding order and sliding accuracy in sliding mode control. *International Journal of Control* 58 (1993) 1247–1263
4. Bartolini, G., Pisano, A., Punta, E., Usai, E.: A survey of applications of second-order sliding mode control to mechanical systems. *International Journal of Control* 76 (2003) 875–892
5. Levant, A.: Quasi-continuous high-order sliding-mode controllers. *IEEE Trans. Automat. Contr.* 50(11) (2005) 1812–1816
6. Shtessel, Y., Buffington, J., Banda, S.: Multiple timescale flight control using reconfigurable sliding modes. *AIAA Journal of Guidance, Control, and Dynamics* 22(6) (1999) 873–883

7. Shtessel, Y., Shkolnikov, I., Levant, A.: Smooth second-order sliding modes: Missile guidance application. *Automatica* **43**(8) (2007) 1470–1476
8. Shkolnikov, I., Shtessel, Y.: Aircraft nonminimum phase control in dynamic sliding manifolds. *AIAA Journal of Guidance, Control, and Dynamics* **24**(3) (2001) 566–572
9. Alwi, H., Edwards, C.: Fault detection and fault-tolerant control of a civil aircraft using a sliding-mode-based scheme. *IEEE Trans. Contr. Syst. Technol.* **16**(3) (2008) 3675–3683
10. Bartolini, G., Pisano, A.: Black-box position and attitude tracking for underwater vehicles by second-order sliding-mode technique. *International Journal of Robust and Nonlinear Control* **20**(14) (2010) 1594–1609
11. Bartolini, G., Ferrara, A., Levant, A., Usai, E.: Sliding mode observers. In X.Yu, J.-X.Xu, eds.: *Variable Structure Systems: Towards the 21st Century. Lecture Notes in Control and Information Science*. Springer Verlag, Berlin (2002) 391–415
12. Poznyak, A.: Stochastic output noise effects in sliding mode estimations. *International Journal of Control* **76** (2003) 986–999
13. Davila, J., Fridman, L., Levant, A.: Second-order sliding-mode observer for mechanical systems. *IEEE Trans. Automat. Contr.* **50**(11) (November 2005) 1785–1789
14. Fridman, L., Levant, A., Davila, J.: Observation of linear systems with unknown inputs via high-order sliding-modes. *Int. J. System Science* **38**(10) (2007) 773–791
15. Bejarano, F., Fridman, L., Poznyak, A.: Exact state estimation for linear systems with unknown inputs based on hierarchical super-twisting algorithm. *Int. J. Robust Nonlinear Control* **17**(18) (2007) 1734–1753
16. Davila, J., Fridman, L., Pisano, A., Usai, A.: Finite-time state observation for nonlinear uncertain systems via higher order sliding modes. *International Journal of Control* **82**(8) (2009) 1564–1574
17. Pang, H., Chen, X.: Global robust optimal sliding mode control for uncertain affine nonlinear systems. *Journal of Systems Engineering and Electronics* **20**(4) (2009) 838–843
18. Ferreira, A., Bejarano, F., Fridman, L.: Robust control with exact uncertainties compensation: With or without chattering? *IEEE Trans. Contr. Syst. Technol.* **19**(5) (2011) 969–975
19. Davila, J.: Exact tracking using backstepping control design and high-order sliding modes. *IEEE Trans. Automat. Contr.* **58**(8) (2013) 2077–2081
20. Nesline, F., Wells, B., Zarchan, P.: Combined optimal/classical approach to robust missile autopilot design. *AIAA Journal of Guidance, Control, and Dynamics* **4**(3) (1981) 316–322
21. Talole, S., Godbole, A., Kolhe, J.: Robust roll autopilot design for tactical missiles. *AIAA Journal of Guidance, Control, and Dynamics* **34**(1) (2011) 107–1017
22. Wang, W., Gao, Z.: A comparison study of advanced state observer design techniques. In: *Proc. of the 2003 American Control Conference, Cleveland State Univ., OH, USA* (2003) 4754 – 4759
23. Filippov, A.: *Differential Equations with Discontinuous Right-hand Sides*. Kluwer Academic Publishers, Dordrecht, The Netherlands (1988)
24. Trentelman, H.L., Stoorvogel, A.A., Hautus, M.: *Control theory for linear systems*. Springer-Verlag, London, Great Britain (2001)
25. Levant, A.: High-order sliding modes: differentiation and output-feedback control. *International Journal of Control* **76**(9-10) (2003) 924–941

# Limit cycles in second order systems through sliding surface design

Raúl Rascón<sup>†</sup>, Andrés Calvillo<sup>‡</sup> and Luis Moreno<sup>†</sup>

<sup>†</sup> UABC Engineering faculty, Blvd. Benito Juárez y Calle de la Normal S/N, 21280 Mexicali, México.

<sup>‡</sup> Citedi-IPN, avenida del parque 1310 Mesa de Otay 22510 Tijuana B.C., México.

*Paper received on 12/11/13, Accepted on 01/19/14.*

**Abstract.** In this paper it is designed a sliding mode controller to generate a limit cycle in a nonlinear second order system using only one control input, it is considered that the systems is affected by a discontinuous function and a periodic disturbance. The proposed controller does not need an exact knowledge of the discontinuous function and disturbances; it only needs an upper bound of their magnitudes. It is proved that the limit cycle is reached in finite time. It is worth mention that the designed controller application is straightforward to a second order mass-spring-damper system. The performance of the proposed controller is illustrated in a numerical simulation.

**Keywords:** Limit cycles, sliding surfaces, second-order systems

## 1 Introduction

A recent work of generation of stable limit cycles with prescribed frequency and amplitude via polynomial feedback is presented in [Knoll and Robenack(2012)], also another previous work of generation of limit cycles in linear systems without perturbations is given by [Bacciotti et al.(1996)Bacciotti, Mazzi, and Sabatini].

This paper is about the generation of limit cycles through sliding mode technique, the main feature of this class of controllers is to allow the sliding mode to occur on a prescribed switching surface, so that the system is governed by the sliding equation only, and remains insensitive to a class of disturbances and parameter variations [Utkin(1978)]. This control method has been successfully tested for motion control of robotic manipulators, see [Sabanovic(2008)] and references therein. Besides, a previous work of sliding-mode control can be found in [Rascon et al.(2012)Rascon, Alvarez, and Aguilar].

The problem addressed in the present paper is the generation of limit cycles in nonlinear second order systems, the limit cycle could have prescribed frequency and amplitude as mentioned later, moreover the trajectories reach the limit cycle in finite time in spite of perturbations and nonlinear discontinuous phenomena, in order to achieve the control objective is necessary to know the upper bounds of the perturbations and the nonlinear term affecting the system.



The rest of the paper is outlined as follows: In Section II we describe the second order nonlinear system with perturbations. The control design and sliding surface proposed are presented in Section III. Section IV presents stability analysis. Section V presents an academic example performed with MATLAB®. Finally Section V includes some final comments.

## 2 Problem statement

Basically, the proposed problem it is about to design a discontinuous controller for the dynamic system

$$\begin{aligned}\dot{x}_1 &= x_2 \\ \dot{x}_2 &= f(x) + g(x)u + w(t)\end{aligned}\tag{1}$$

capable of guaranteeing that the trajectories of (1) will converge to a limit cycle in finite time. Consider that  $x = [x_1, x_2]^T$  is the state vector,  $u \in \mathbb{R}$  is the control input, the nominal dynamics of the system are giving by  $f(x) = -Ax_1 - Bx_2 - \alpha \text{sign}(x_2)$ ,  $g(x) \in \mathbb{R}$  is a well known function and  $w(t)$  is a matched uncertainty/perturbation with an upper bound  $M$  that is assumed known a priori, so it satisfies

$$\sup_t |w(t)| \leq M, \quad M > 0\tag{2}$$

for all  $t$  and some constant  $M > 0$ . Also, the term  $\alpha$  is considered with a level denoted as  $\alpha > 0$ . The  $\text{sign}(x_2)$  denotes the signum function defined as

$$\text{sign}(x_2) = \begin{cases} 1 & x_2 > 0 \\ [-1, 1] & x_2 = 0 \\ -1 & x_2 < 0. \end{cases}\tag{3}$$

Then, for a constant force input  $u = \bar{u}$  and zero disturbance ( $w(t) = 0$ ), the system (1) has the equilibrium point  $\bar{x}_2 = 0$  and  $\bar{x}_1 \in [(\bar{u} - \alpha)/A, (\bar{u} + \alpha)/A]$ .

## 3 Control design

Let us suppose that the disturbance  $w(t)$  affecting system (1) satisfies (2), and the discontinuous term is such that  $0 < \alpha \leq \alpha_c$ , for some known bound  $\alpha_c$ . The control objective is to find a control  $u$ , depending on  $x_1$  and  $x_2$ , such that the closed-loop response of system (1)-(2) satisfies

$$x_1^2 + x_2^2 = r^2\tag{4}$$

where  $r$  is the amplitude of the oscillation signal that we propose. Based on (4) two sliding surfaces are designed to have a sliding surface as in Figure 1

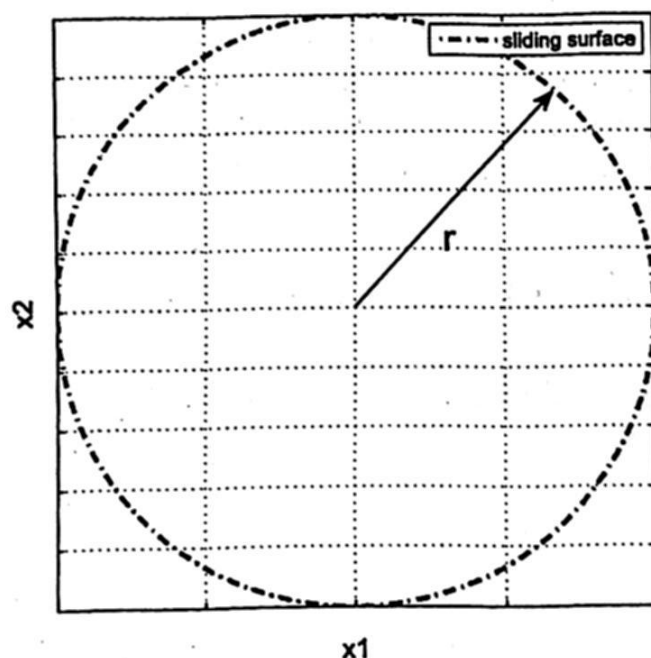


Fig. 1. Phase portrait of the proposed sliding surface.

$$s = \begin{cases} \sqrt{-x_1^2 + r} + x_2, & \text{if } x_2 \leq 0, \\ -\sqrt{-x_1^2 + r} + x_2, & \text{if } x_2 > 0. \end{cases} \quad (5)$$

The switching surface (5) can be represented as one under the following expression

$$s = -\frac{x_2}{|x_2|} \sqrt{-x_1^2 + r} + x_2. \quad (6)$$

A control law designed from (6), which can ensure us that trajectories  $(x_1, x_2)$  are going to converge to the sliding surface is given by

$$u = g(x)^{-1} \left( Ax_1 + Bx_2 + \frac{-x_1|x_2|}{\sqrt{-x_1^2 + r}} - \lambda s - \beta \text{sign}(s) \right). \quad (7)$$

where the parameters  $\lambda$  and  $\beta$  are positive; they will be tuned to ensure the motion of the trajectories be directed towards the sliding surface.

#### 4 Stability analysis

We analyze in this section the stability of the closed-loop system (1), controlled by (7), and conclude about the overall stability.

By substituting (7) into (1), the closed-loop system takes the form

$$\begin{aligned} \dot{x}_1 &= x_2, \\ \dot{x}_2 &= \frac{-x_1|x_2|}{\sqrt{-x_1^2 + r}} - \lambda s - \beta \text{sign}(s) - \alpha \text{sign}(x_2) + w(t). \end{aligned} \quad (8)$$

Now, we ensure the existence of sliding modes by verifying  $s\dot{s} < 0$ . To this end, note that, from (2) and the fact that  $\alpha \leq \alpha_c$ , then

$$\begin{aligned} s\dot{s} &= s \left( \frac{-|x_2| + x_2 \operatorname{sign}(x_2) + |x_2|^2}{|x_2|^2} \dot{x}_2 + \frac{x_1 x_2}{\sqrt{-x_1^2 + r|x_2|}} \dot{x}_1 \right) \\ &= s (-\lambda s - \beta \operatorname{sign}(s) + w(t) - \alpha \operatorname{sign}(x_2)) \\ &\leq -\lambda s^2 - \beta |s| + (M + \alpha_c) |s| \\ &\leq -\lambda s^2 - (\beta - (M + \alpha_c)) |s|. \end{aligned}$$

We conclude the existence of sliding modes on the surface (6) while the condition  $\beta > M + \alpha_c$  be satisfied. This gives a guide to tune the parameter  $\beta$  of the controller (7). In fact, we can demonstrate that the trajectories reach the surface  $s = 0$ , in finite time, using the quadratic function

$$V(s) = s^2, \quad (9)$$

and compute its time derivative along the solutions of (8),

$$\begin{aligned} \dot{V}(s(t)) &\leq -2\lambda s^2 - 2(\beta - (M + \alpha_c)) |s| \leq -2(\beta - (M + \alpha_c)) |s| \\ &= -2(\beta - (M + \alpha_c)) \sqrt{V(s(t))}. \end{aligned} \quad (10)$$

From (10) it follows that

$$V(t) = 0 \quad \text{for} \quad t \geq t_0 + \frac{\sqrt{V(t_0)}}{(\beta - (M + \alpha_c))} = t_f. \quad (11)$$

Hence,  $V(t)$  converges to zero in finite time and, in consequence, a motion along the manifold  $s = 0$  occurs in the discontinuous system (8). Notice that the reaching time can be reduced by increasing the value of parameter  $\beta$ .

## 5 Academic example

Performance issues and robustness properties of the proposed sliding mode controller have been tested with some numerical experiments under the following parameters as shown in Table 1

**Table 1.** Plant parameters, controller gains, amplitude of the discontinuous term, disturbances and initial conditions.

A	B	$\lambda$	$\beta$	r	$\alpha$	w	$x_1(0)$	$x_2(0)$
65	7	340	5	1	1.5	$0.5 \sin(10t)$	$\pi/4$	$\pi$

According to (1) the academic example takes the form

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} x_2 \\ -65x_1 - 7x_2 - 1.5 \operatorname{sign}(x_2) \end{bmatrix} + \begin{bmatrix} 0 \\ u \end{bmatrix} + \begin{bmatrix} 0 \\ 0.5 \sin(10t) \end{bmatrix} \quad (12)$$

The Figure 2 displays the system trajectories  $x_1$  and  $x_2$  which must converge to the proposed sliding surface as is shown in Figure 3. The control signal is shown in Figure 4, the chattering effect can be reduce by adjusting the  $\beta$  parameter, this can occur because  $\beta$  denotes the amplitude of the discontinuous control term, just keeping in mind that  $\beta > M + \alpha_c$  must be satisfied. The sliding motion which converge to  $s = 0$  in finite time approximately in  $t = 0.2$  seconds is presented in Figure 5.

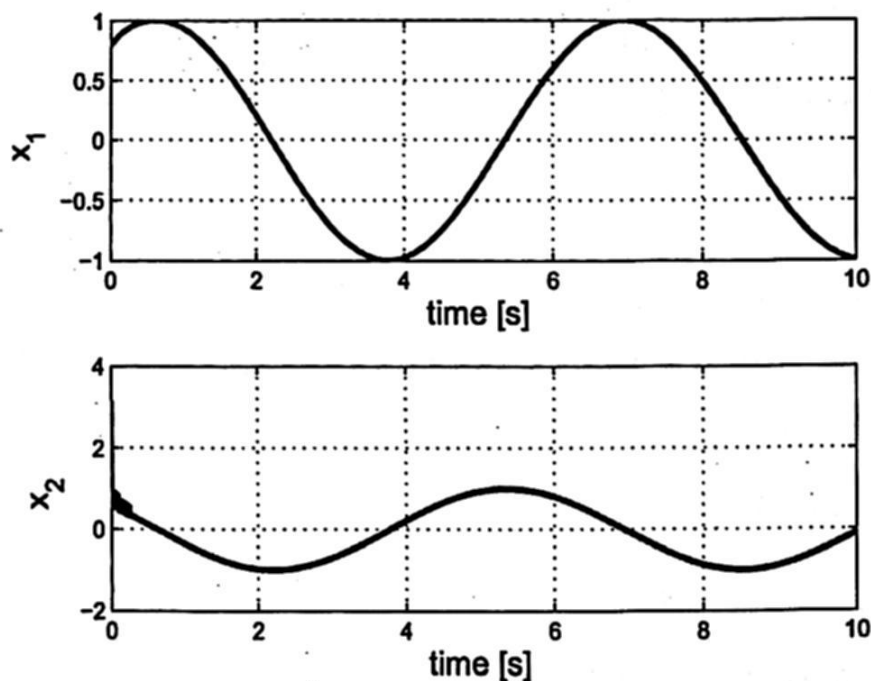
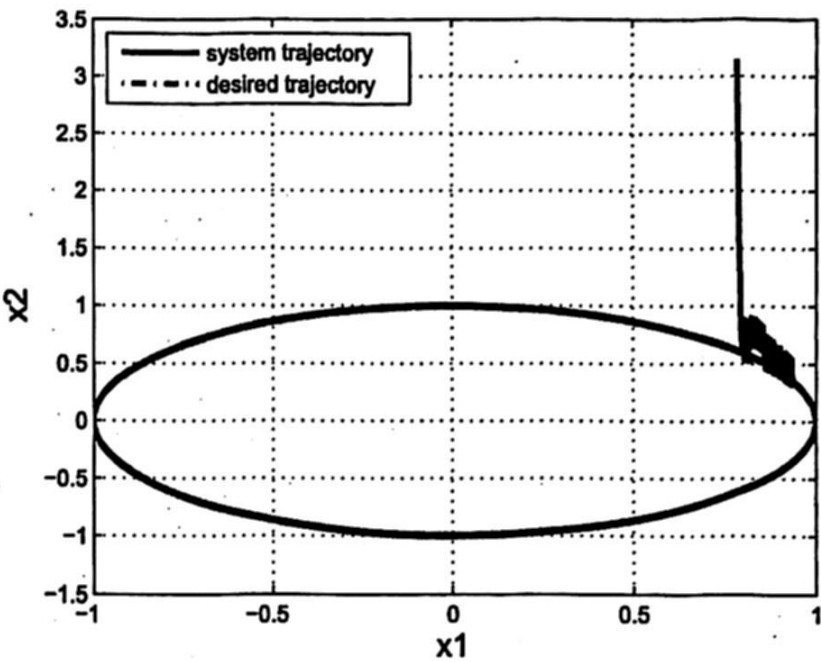


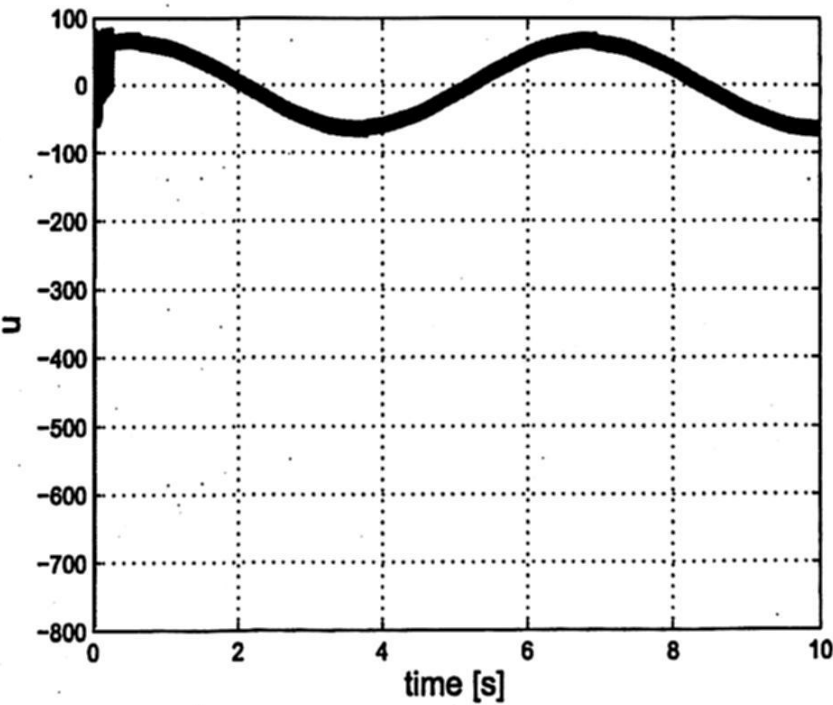
Fig. 2. Trajectories  $x_1$  and  $x_2$ .

## 6 Conclusions

Has been introduced a way to generate a limit cycle in a nonlinear second order system through a sliding surface design, beside of circular phase portraits it can be propose ellipsoidal sliding surfaces by using the methodology mentioned afore, moreover it can be change the frequency of oscillation by adding a constant gain  $\theta$  to the sliding surface as follows  $s = -\frac{x_2}{|x_2|} \sqrt{-x_1^2 + r} + \theta x_2$ . It is worth to mention that the proposed controller does not need an exact knowledge of the discontinuous function and disturbances; it only needs an upper bound of their magnitudes. It is proved that the limit cycle is reached in finite time in spite of the aforementioned uncertainties. This control technique can be directly applied to second order mechanical systems, like a mass-spring-damper system. As a future work the authors are interested in analyze the dynamical behavior of the trajectories once reached the sliding surface, also it would be of our interest to generate limit cycles with different geometrical shapes. It is important to point out that it has not been proved that the trajectories will remain in the sliding

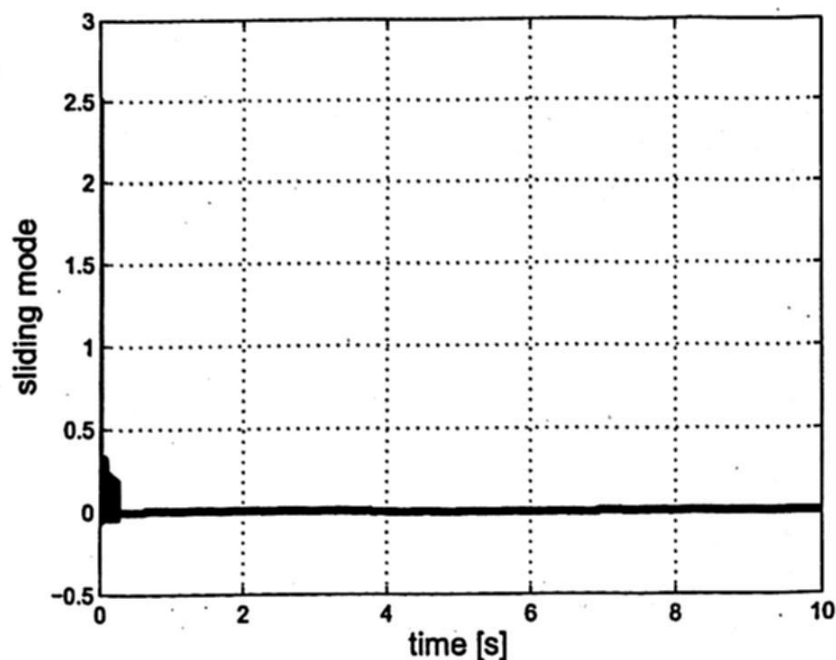


**Fig. 3.** Phase portrait, where can be seen the trajectories convergence to the proposed sliding surface.



**Fig. 4.** Control signal, where the chattering phenomenon appears in steady state.





**Fig. 5.** Sliding motion, observe that the chattering phenomenon has a relatively small amplitude.

surface for all time, so it is likely that trajectories can scape from the sliding surface at a time, further analysis should be done about this concern.

## References

- [Bacciotti et al.(1996)Bacciotti, Mazzi, and Sabatini] A. Bacciotti, L. Mazzi, and M. Sabatini. Generation of stable limit cycles in controllable linear systems. *Systems & Control Letters*, 28(1):43 – 48, 1996. ISSN 0167-6911. doi: [http://dx.doi.org/10.1016/0167-6911\(96\)00005-9](http://dx.doi.org/10.1016/0167-6911(96)00005-9). URL <http://www.sciencedirect.com/science/article/pii/0167691196000059>.
- [Knoll and Robenack(2012)] C. Knoll and K. Robenack. Generation of stable limit cycles with prescribed frequency and amplitude via polynomial feedback. In *9th International Multi-Conference on Systems, Signals and Devices (SSD)*, 2012, pages 1–6, 2012. doi: 10.1109/SSD.2012.6197994.
- [Rascon et al.(2012)Rascon, Alvarez, and Aguilar] R. Rascon, J. Alvarez, and L.T. Aguilar. Regulation and force control using sliding modes to reduce rebounds in a mechanical system subject to a unilateral constraint. *Control Theory Applications, IET*, 6(18):2785–2792, 2012. ISSN 1751-8644. doi: 10.1049/iet-cta.2011.0314.
- [Sabanovic(2008)] A. Sabanovic. Sliding modes in constrained systems control. *IEEE Transactions on Industrial Electronics*, 55:3332–3339, 2008.
- [Utkin(1978)] V. Utkin. *Sliding Modes and Their Applications*. Mir, Moscow, 1978.

## Part II

# Intelligent systems



# Water Distribution Systems Optimization

Avi Ostfeld

Civil and Environmental Engineering,  
Technion Israel Institute of Technology,  
Technion City, Haifa 32000, ISRAEL  
ostfeld@tx.technion.ac.il

*Invited paper*

**Abstract.** A water distribution system is a complex assembly of hydraulic control elements connected together to convey quantities of water from sources to consumers. The common high number of constraints and decision variables, the nonlinearity, and the non-smoothness of the head flow water quality governing equations are inherent to water distribution systems planning and management problems. This paper provides a brief overview on some of the more traditional and new water distribution systems problem algorithms and solution methodologies. The manuscript concludes with challenges and a look into the future for water supply systems optimization.

**Keywords:** water distribution systems, optimization, review, water quality, robust optimization, genetic algorithms.

## 1 Introduction

A water distribution system is an interconnected collection of sources, pipes and hydraulic control elements (e.g., pumps, valves, regulators, tanks) delivering consumers prescribed water quantities at desired pressures and water qualities. Such systems are often described as a graph with the links representing the pipes and the nodes defining connections between pipes, hydraulic control elements, consumers, and sources.

The typical high number of constraints and decision variables, the nonlinearity, and the non-smoothness of the head flow water quality governing equations are inherent to water supply systems planning and management problems. An example of this is the least cost design problem of a water distribution system defined as finding the water distribution system's component characteristics (e.g., pipe diameters, pump heads and maximum power, reservoir storage volumes, etc.), which minimize the system capital and operational costs, such that the system hydraulic laws are maintained (i.e., Kirchhoffs Laws No. 1 and 2 for continuity of flow and energy, respectively), and constraints on quantities and pressures at the consumer nodes are fulfilled.

In addition, problems related to aggregation, maintenance, reliability, unsteady flow and security can be identified for gravity, and/or pumping, and/or

storage branched/looped water distribution systems. Flow and head, or flow, head, and water quality can be considered for one or multiple loading scenarios, taking into consideration inputs/outputs as deterministic or stochastic variables. Figure 1 provides a schematic map of water distribution systems related problems.

Traditional methods for solving water distribution systems management problems used linear or nonlinear optimization schemes which were limited by the system size, the number of constraints, and the number of loading conditions. More recent methodologies employ heuristic optimization techniques, such as genetic algorithms or ant colony optimization as stand alone or hybrid data driven heuristic schemes. Other recent methods employ robust optimization methodologies for incorporating uncertainty.

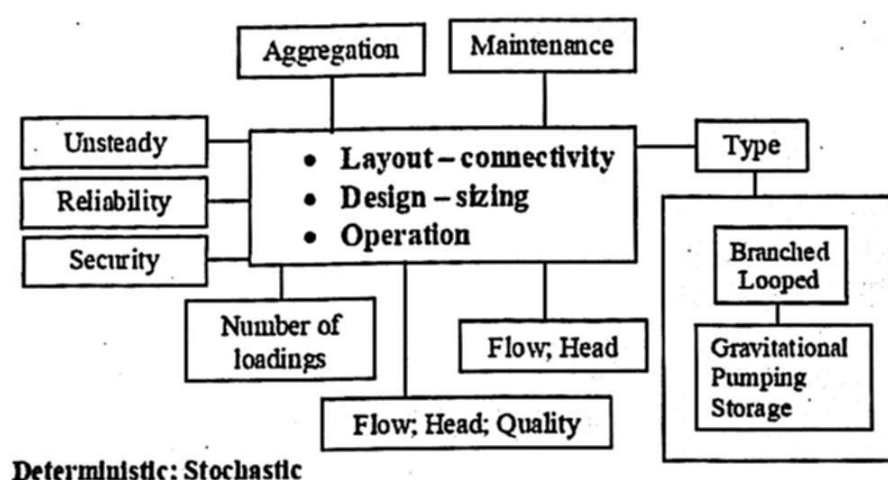


Fig. 1. A map of water distribution networks related problems  $z(t)$

This paper reviews part of the topics presented in Figure 1. It includes sections on least cost optimal design of water networks, reliability incorporation in water supply systems analysis, optimal operation of water networks, water quality considerations inclusion in distribution systems, water networks security, robust optimization employment for water distribution systems management, and a look into the future.

## 2 Least cost design of water networks

Numerous models for least cost design of water distribution systems have been published in the research literature during the last four decades. A possible classification for those might be: (1) decomposition: methods based on decomposing the problem into an "inner" linear programming problem which is solved for a fixed set of flows (heads), while the flows (heads) are altered at an "outer" problem using a gradient or a subgradient optimization technique [1, 38, 39, 22, 13, 34]; (2) linking simulation with nonlinear programming: methods based on



linking a network simulation program with a general nonlinear optimization code [29, 23, 50]; (3) nonlinear programming: methods utilizing a straightforward nonlinear programming formulation [55, 44]; (4) methods which employ evolutionary or metaheuristic techniques: genetic algorithms [46, 42, 40, 53, 56], simulated annealing [25], the shuffled frog leaping algorithm [14], ant colony optimization [27]; and (5) other methods: dynamic programming [48], integer programming [41].

The capabilities of solving water distribution systems optimization problems have improved dramatically since the employment of genetic algorithms [17]. Genetic algorithms are domain heuristic independent global search techniques that imitate the mechanics of natural selection and natural genetics of Darwins evolution principle. The premise is to simulate the natural evolution mechanisms of chromosomes, represented by string structures, involving selection, crossover, and mutation. Strings may have binary, integer, or real values. Simpson et al. [46] were the first to use genetic algorithms for water distribution systems least cost design. They applied and compared a genetic algorithm solution to the network of Gessler [16], to enumeration and to nonlinear optimization. Savic and Walters [42] used genetic algorithms to solve and compare optimal results of the oneloading gravity systems of the Two Loop Network [1], the Hanoi network [15], and the New York Tunnels system [43]. Salomons [40] used a genetic algorithm for solving the least cost design problem incorporating extended period loading conditions, tanks, and pumping stations.

### 3 Reliability of water supply

Reliability of water distribution systems gained considerable research attention over the last three decades. Research has concentrated on methodologies for reliability assessment and for reliability inclusion in least cost design and operation of water supply systems.

Shamir and Howard [45] were the first to propose analytical methods for water supply system reliability. Their methodology took into consideration flow capacity, water main breaks, and maintenance for quantifying the probabilities of annual shortages in water delivery volumes.

Reliability measures such as the probability of shortfall (i.e., total unmet demand), the probability of the number of failure events in a simulation period, and the probability of interfailure times and repair durations were used in various studies as reliability criteria. Bao and Mays [4] suggested stochastic simulation by imposing uncertainty in future water demands for computing the probability that the water distribution system will meet these needs at minimum pressures. Duan and Mays [12] used a continuous-time Markov process for reliability assessment of water supply pumping stations. They took into consideration both mechanical and hydraulic failure (i.e., capacity shortages) scenarios, all cast in a conditional probability frequency and duration analysis framework. Jacobs and Goulter [18] used historical pipe failure data to derive the probabilities that a particular number of simultaneous pipe failures will cause the entire system to fail.

Recently, Tanyimboh et al. [51] compared the surrogate measures of statistical entropy, network resilience, resilience index, and the modified resilience index for quantifying the reliability of water networks. Torii and Lopez [52] utilized first order reliability methods in conjunction with an adaptive response surface approach for analyzing the reliability of water distribution systems.

## 4 Water networks optimal operation

Following the well known least cost design problem of water distribution systems [21, 19, 1], optimal operation is the most explored topic in water distribution systems management. Since 1970 a variety of methods were developed to address this problem, including the utilizations of dynamic programming, linear programming, predictive control, mixedinteger, nonlinear programming, meta-modeling, heuristics, and evolutionary computation. Ormsbee and Lansey [30] classified to that time optimal water distribution systems control models through systems type, hydraulics, and solution methods. Examples for optimal operation of water distribution systems are described below.

Dreizin [11] was the first to suggest an optimization model for water distribution systems operation through a dynamic programming (DP) scheme coupled with hydraulic simulations for optimizing pumps scheduling of a regional water supply system supplied by three pumping units. Sterling and Coulbeck [49] used a dynamic modeling approach to minimize the costs of pumps operation of a simple water supply system. Olshansky and Gal [28] developed a two level linear programming methodology in which the distribution system is partitioned into subsystems for which hydraulic simulations are run and serve further as parameters in an LP model for pumps optimal scheduling. This approach was used also by Jowitt and Germanopoulos [20] who developed a linear programming model to optimize pumps scheduling in which the LP parameters are set through offline extended period hydraulic simulation runs. Biscos et al. [7] used a predictive control framework coupled with mixed integer nonlinear programming (MINLP) for minimizing the costs of pump operation. Biscos et al. [8] extended Biscos et al. [7] to include the minimization of chlorine dosage. Pulido-Calvo and Gutiérrez-Estrada [37] presented a model for both sizing storage and optimizing pumps operation utilizing a framework based on a mixed integer non linear programming (MINLP) algorithm and a data driven (neural networks) scheme. Ostfeld and Salomons [31] minimized the total cost of pumping and water quality treatment of a water distribution system through linking a genetic algorithm with EPANET ([www.epa.gov/nrmrl/wswrd/dw/epanet.html](http://www.epa.gov/nrmrl/wswrd/dw/epanet.html)).

Van Zyl et al. [54] utilized a genetic algorithm (GA) linked to a hillclimber search algorithm for improving the local GA search once closed to an optimal solution. LópezIbáñez et al. [26] proposed an ant colony optimization (ACO) [10] framework for optimal pumps scheduling. Boulos et al. [9] developed the H2ONET tool based on genetic algorithms for scheduling pump operation to minimize operation costs.

## 5 Inclusion of water queality

Research in modeling water quality in distribution systems started in the context of agricultural usage [24, 47] primarily in arid regions where good water quality is limited. In 1990 the United States Environmental Protection Agency (USEPA) promulgated rules requiring that water quality standards must be satisfied at the consumer taps rather than at the treatment plants. This initiated the need for water quality modeling, the development of the USEPA simulation water quantity and quality model EPANET (EPANET 2.0@2002), and raised other problems and research needs that commenced considerable research in this area to assist utilities.

Optimization models of water distribution systems can be classified according to their consideration of time and of the physical laws which are included explicitly [32, 33]. In time the distinction is between policy and real time models. Policy models are run off line, in advance, and generate the operating plans for several typical and/or critical operating conditions. Real time (online) models are run continuously in real time, and generate an operating plan for the immediate coming period. The classification with respect to the physical laws which are considered explicitly as constraints are: (1) QH (discharge - head) models: quality is not considered, and the network is described only by its hydraulic behavior; (2) QC (discharge quality) models: the physics of the system are included only as continuity of water and of pollutant mass at nodes. Quality is described essentially as a transportation problem in which pollutants are carried in the pipes, and mass conservation is maintained at nodes. Such a model can account for decay of pollutants within the pipes and even chemical reactions, but does not satisfy the continuity of energy law (i.e., Kirchoffs Law no. 2), and thus there is no guarantee of hydraulic feasibility and of maintaining head constraints at nodes; and (3) QCH (discharge - quality - head) models: quality constraints, and the hydraulic laws, which govern the system behavior, are all considered. The QH and QC problems are relatively easier to solve than the full QCH.

Since the events of 9/11 in the US the security of water distribution systems became a foremost concern. Threats on a water distribution system can be partitioned into three major groups according to their resulted enhanced security: (1) a direct attack on the main infrastructure: dams, treatment plants, storage reservoirs, pipelines, etc.; (2) a cyber attack disabling the functionality of the water supervisory control and data acquisition (SCADA) system, taking over control of key components which might result water outages or insufficiently treated water, changing or overriding protocol codes, etc.; and (3) a deliberate chemical or biological contaminant injection at one of the system's nodes.

The threat of a direct attack can be minimized by improving the system's physical security (e.g., additional alarms, locks, fencing, surveillance cameras, guarding, etc.), while a cyber attack by implementing computerized hardware and software (e.g., an optical isolator between communication networks, routers to restrict data transfer, etc.).

Of the above threats, a deliberate chemical or biological contaminant injection is the most difficult to address. This is because of the uncertainty of the



type of the injected contaminant and its effects, and the uncertainty of the location and injection time. Principally a contaminant can be injected at any water distribution system connection (node) using a pump or a mobile pressurized tank. Although backflow preventers provide an obstacle, they do not exist at all connections, and at some might not be functional.

The main course to enhance the security of a water distribution system against a deliberated contamination intrusion is through a sensor system [2, 3].

## 6 Robust Optimization

The approach presented in most previous studies is to treat the problem as deterministic assuming perfectly known parameters. Consequently, deterministic models are likely to perform poorly when implemented in reality when the actual problem parameters are revealed, hence the need to find more "robust" solution approaches. Perelman et al. [35, 36] proposed formulating a deterministic equivalent of the stochastic problem of optimal design/rehabilitation of water distribution systems using the non probabilistic robust counterpart (RC) approach [5, 6] for uncertainty inclusion into optimization modeling. The uncertainty of the information is quantified through a deterministic userdefined ellipsoidal uncertainty set, which can be probabilistically justified, with the decision maker seeking a solution that is optimal for all possible realizations in the uncertainty set. The robust counterpart makes no assumptions about the probability density function of the uncertain variables and their dependencies, does not require the construction of a representative sample of scenarios, and has the same size as the original deterministic model.

## 7 Conclusions

Traditionally, water distribution networks were designed, operated, and maintained through utilizing offline small discrete datasets. Those were the governing and limiting constraints imposed on modeling challenges and capabilities. This situation is dramatically changing: from a distinct framework of data collection to a continuous transparent structure. With multiple types of sensor data at multiple scales, from embedded real time hydraulic and water quality sensors to airborne and satellite based remote sensing, how can those be efficiently integrated into new tools for decision support for water distribution networks is a major challenge.

This new reality is expected to limit all current modeling efforts capabilities and require new thinking on approaches for managing water distribution networks: from a state of lack of data to a situation of overflowing big data information. New tools for data screening, algorithms and data driven modeling constructions, as well as computational efficiency are anticipated to govern all future developments for water distribution networks analysis and optimization.

## 8 Acknowledgments

This research was supported by the Fund for the Promotion of Research at the Technion, and by the Technion Grand Water Research Institute (GWRI).

## References

1. Alperovits E. and Shamir U. (1977). "Design of optiwater distribution systems." *Water Resources Research*, Vol. 13, No. 6, pp. 885-900
2. American Society of Civil Engineers (ASCE) (2004). "Guidelines for designing an online contaminant monitoring system." Available on line at: <http://www.water-simulation.com/wsp/2005/01/12/guidelines-for-designing-an-online-contaminant-monitoring-system/>
3. American Water Works Association (AWWA) (2004). "Security guidance for water utilities" Available on line at: [http://www.asce.org/uploadedFiles/ewri/Codes\\_and\\_Standards/4.pdf](http://www.asce.org/uploadedFiles/ewri/Codes_and_Standards/4.pdf)
4. Bao Y. and Mays L. W. (1990). "Model for water distribution system reliability." *Journal of Hydraulic Engineering*, Vol. 116, No. 9, pp. 1119-1137.
5. Ben-Tal A. and Nemirovski A. (1998). "Robust convex optimization." *Mathematics of Operational Research*, Vol. 23, pp. 769-805.
6. Ben-Tal A. and Nemirovski A. (1999). "Robust solutions of uncertain linear programs." *Operational Research Letters*, Vol. 25, pp. 113.
7. Biscos C., Mulholland M., Le Lann M. V., Brouckaert C. J., Bailey R., and Roustan M. (2002). "Optimal operation of a potable water distribution network." *Water Science and Technology*, Vol. 46, No. 9, pp. 155-162.
8. Biscos C., Mulholland M., Le Lann M.-V., Buckley C. A., and Brouckaert C. J. (2003). "Optimal operation of water distribution networks by predictive control using MINLP." *Water SA*, Vol. 29, No. 4, pp. 393-404.
9. Boulou P. F., Wu Z., Orr C. H., Moore M., Hsiung P., and Thomas D. (2011). "Optimal pump operation of water distribution systems using genetic algorithms." Available online at: <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.114.1935>
10. Dorigo M. (1992). "Optimization, learning and natural algorithms." Ph.D. thesis, Politecnico di Milano, Milan, Italy.
11. Dreizin Y. (1970). "Examination of possibilities of energy saving in regional water supply systems." M.Sc. Thesis, Technion Israel Institute of Technology, 85p.
12. Duan N. and Mays L. W. (1990). "Reliability analysis of pumping systems." *Journal of Hydraulic Engineering*, Vol. 116, No. 2, pp. 230-248.
13. Eiger G., Shamir U., and Ben-Tal A. (1994). "Optimal design of water distribution networks." *Water Resources Research*, Vol. 30, No. 9, pp. 2637-2646.
14. Eusuff M. M. and Lansey K. E. (2003). "Optimization of water distribution network design using the shuffled frog leaping algorithm." *Journal of Water Resources Planning and Management Division, ASCE*, Vol. 129, No. 3, pp. 210-225.
15. Fujiwara O. and Khang D. B. (1990). "A two-phase decomposition method for optimal design of looped water distribution networks." *Water Resources Research*, Vol. 26, No. 4, pp. 539-549.
16. Gessler J. (1985). "Pipe network optimization by enumeration." *Proceedings, Computer Applications for Water Resources, ASCE*, New York, N. Y., pp. 572-581.
17. Holland J. H. (1975). *Adaptation in natural and artificial systems*. Ann Arbor: The University of Michigan Press, Ann Arbor.



18. Jacobs P. and Goulter I. (1991). "Estimation of maximum cut-set size for water network failure." *Journal of Water Resources Planning and Management Division* ASCE, Vol. 117, No. 5, pp. 588-605.
19. Jacoby S. (1968). "Design of optimal hydraulic networks." *Journal of Hydraulic Division*, ASCE, Vol. 94, No. HY3, pp. 641-661.
20. Jowitt P. W. and Germanopoulos G. (1992). "Optimal pump scheduling in water-supply networks." *Journal of Water Resources Planning and Management Division* ASCE, Vol. 118, No. 4, pp. 406-422.
21. Karmeli D., Gadish Y., and Meyers S. (1968). "Design of optimal water distribution networks." *Journal Pipeline Division*, ASCE, Vol. 94, No. 1, pp. 1-9.
22. Kessler A. and Shamir U. (1989). "Analysis of the linear programming gradient method for optimal design of water supply networks." *Water Resources Research*, Vol. 25, No. 7, pp. 1469-1480.
23. Lansey K. E. and Mays L. W. (1989). "Optimization models for design of water distribution systems." In, *Reliability Analysis of Water Distribution Systems*, Mays L. W. Ed., pp. 37-84.
24. Liang T. and Nahaji S. (1983). "Managing water quality by mixing water from different sources." *Journal of Water Resources Planning and Management Division*, ASCE, Vol. 109, pp. 48 - 57.
25. Loganathan G. V., Greene J. J., and Ahn T. J. (1995). "Design Heuristic for Globally Minimum Cost Water-Distribution Systems." *Journal of Water Resources Planning and Management Division*, ASCE, Vol. 121, No. 2, pp. 182-192.
26. López-Ibáñez M., Prasad T. D., and Paechter B. (2008). "Ant colony optimization for optimal control of pumps in water distribution networks." *Journal of Water Resources Planning and Management Division*, ASCE, Vol. 134, No. 4, pp. 337-346.
27. Maier H. R., Simpson, A. R., Zecchin A. C., Foong W. K., Phang K. Y., Seah H. Y., and Tan C. L. (2003). "Ant colony optimization for design of water distribution systems." *Journal of Water Resources Planning and Management Division*, ASCE, Vol. 129, No. 3, pp. 200-209.
28. Olshansky M. and Gal S. (1988). "Optimal operation of a water distribution system." IBM - Israel, Technical Report 88.239, 52p.
29. Ormsbee L. E. and Contractor D. N. (1981). "Optimization of hydraulic networks." *Proceedings, International Symposium on Urban Hydrology, Hydraulics, and Sediment Control*, Kentucky, Lexington KY, pp. 255-261.
30. Ormsbee L. E. and Lansey K. E. (1994). "Optimal control of water supply pumping systems." *Journal of Water Resources Planning and Management Division*, ASCE, Vol. 120, No. 2, pp. 237-252.
31. Ostfeld A. and Salomons E. (2004). "Optimal operation of multiquality water distribution systems: unsteady conditions." *Engineering Optimization*, Vol. 36, No. 3, pp. 337-359.
32. Ostfeld A. and Shamir U. (1993a). "Optimal Operation of Multiquality Distribution Systems: Steady State Conditions." *Journal of Water Resources Planning and Management Division*, ASCE, Vol. 119, No. 6, pp. 645 - 662.
33. Ostfeld A. and Shamir U. (1993b). "Optimal Operation of Multiquality Distribution Systems: Unsteady Conditions." *Journal of Water Resources Planning and Management Division*, ASCE, Vol. 119, No. 6, pp. 663 - 684.
34. Ostfeld A. and Shamir U. (1996). "Design of Optimal Reliable Multiquality Water Supply Systems." *Journal of Water Resources Planning and Management Division*, ASCE, Vol. 122, No. 5, pp. 322 - 333.

35. Perelman L., Housh M. and Ostfeld A. (2013a). "Least cost design of water distribution systems under demand uncertainty: the robust counterpart approach", *Journal of Hydroinformatics*, Vol. 15, No. 3, pp. 737-750.
36. Perelman L., Housh M., and Ostfeld A. (2013b). "Robust optimization for water distribution systems least cost design." *Water Resources Research*, Vol. 49, Issue 10, pp. 6795-6809.
37. Pulido-Calvo I. and Gutierrez-Estrada J. C. (2011). "Selection and operation of pumping stations of water distribution systems." *Environmental Research Journal*, Vol. 5, No. 3 pp. 1-20.
38. Quindry G. E., Brill E. D., Liebman J. C., and Robinson A. R. (1979). Comment on "Design of optimal water distribution systems" by Alperovits E. and Shamir U., *Water Resources Research*, Vol. 15, No. 6, pp. 1651-1654.
39. Quindry G. E., Brill E. D., and Liebman J. C. (1981). "Optimization of looped water distribution systems." *Journal of Environmental Engineering, ASCE*, Vol. 107, No. EE4, pp. 665-679.
40. Salomons E. (2001). "Optimal design of water distribution systems facilities and operation." MS Thesis, Technion, Haifa, Israel (In Hebrew).
41. Samani M. V. and Mottaghi A. (2006). "Optimization of Water Distribution Networks Using Integer Linear Programming." *Journal of Hydraulic Engineering, ASCE*, Vol. 132, No. 5, pp. 501-509.
42. Savic D. and Walters G. (1997). "Genetic algorithms for least cost design of water distribution networks." *Journal of Water Resources Planning and Management Division, ASCE*, Vol. 123, No. 2, pp. 67-77.
43. Schaake J. C. and Lai D. (1969). "Linear programming and dynamic programming application to water distribution network design." Report No. 116, Department of Civil Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts.
44. Shamir U. (1974). "Optimal design and operation of water distribution systems." *Water Resources Research*, Vol. 10, No. 1, pp. 27-36.
45. Shamir U. and Howard C. D. (1981). "Water supply reliability theory." *Journal of the American Water Works Association*, Vol. 37, No. 7, pp. 379-384.
46. Simpson A. R., Dandy G. C., and Murphy L. J. (1994). "Genetic algorithms compared to other techniques for pipe optimization." *Journal of Water Resources Planning and Management Division, ASCE*, Vol. 120, No. 4, pp. 423-443.
47. Sinai G., Koch E., and Farbman M. (1985). "Dilution of brackish waters in irrigation networks - an analytic approach." *Irrigation Science*, Vol. 6, pp. 191-200.
48. Singh R. P. and Mahar P. S. (2003). "Optimal Design of Multidiameter, Multioutlet Pipelines." *Journal of Water Resources Planning and Management Division, ASCE*, Vol. 129, No. 3, pp. 226-233.
49. Sterling M. J. H. and Coulbeck B. (1975). "A dynamic programming solution to optimization of pumping costs." *Proceedings Institution of Civil Engineers*, Vol. 59, No. 4, pp. 813-818.
50. Taher S. A. and Labadie J. W. (1996). "Optimal Design of Water-Distribution Networks with GIS." *Journal of Water Resources Planning and Management Division, ASCE*, Vol. 122, No. 4, pp. 301-311.
51. Tanyimboh T. T., Tietavainen M. T., Saleh S. (2011). "Reliability assessment of water distribution systems with statistical entropy and other surrogate measures." *Water Science and Technology: Water Supply*, Vol. 11, No. 4, pp. 437-443.
52. Torii A. J. and Lopez R. H. (2012). "Reliability analysis of water distribution networks using the adaptive response surface approach." *Journal of Hydraulic Engineering*, Vol. 138, No. 3, pp. 237-246.

53. Vairavamoorthy K. and Ali M. (2005). "Pipe Index Vector: A Method to Improve Genetic-Algorithm-Based Pipe Optimization." *Journal of Hydraulic Engineering*, ASCE, Vol. 131, No. 12, pp. 1117-1125.
54. van Zyl J. E., Savic D. A., and Walters G. A. (2004). "Operational optimization of water distribution systems using a hybrid genetic algorithm." *Journal of Water Resources Planning and Management Division*, ASCE, Vol. 130, No. 2, pp. 160-170.
55. Watanatada T. (1973). "Least-cost design of water distribution systems." *Journal of Hydraulic Division*, ASCE, Vol. 99, No. HY9, pp. 1497-1513.
56. Wu Z. Y. and Walski T. (2005). "Self-Adaptive Penalty Approach Compared with Other Constraint-Handling Techniques for Pipeline Optimization." *Journal of Water Resources Planning and Management Division*, ASCE, Vol. 131, No. 3, pp. 181-192.

# Spectrum resource optimization for future cellular networks

Anabel Martínez-Vargas<sup>1</sup>, Ángel G. Andrade<sup>2</sup>, Roberto Sepúlveda<sup>1</sup>, Oscar Montiel-Ross<sup>1</sup>

<sup>1</sup>Instituto Politécnico Nacional, Centro de Investigación y Desarrollo de Tecnología Digital (CITEDI-IPN), Tijuana, México

amartinez@citedi.mx, {rsepulve, oross}@ipn.mx

<sup>2</sup>Universidad Autónoma de Baja California (UABC), Mexicali, México

aandrade@uabc.edu.mx

Paper received on 11/20/13, Accepted on 01/19/14.

**Abstract.** During the last few years cellular networks have increased the use of spectrum resources due to the success of mobile broadband services. Making new exclusive spectrum available to meet traffic demand is challenging since spectrum resources are finite therefore costly. Cognitive radio (CR) technology along with spectrum sharing strategies is proposed as a solution that can reuse the limited spectrum resource. In order to meet mobile traffic demand is expected that future cellular networks overlaid femto-cells (small cells) on the existing macrocell network. To extend and share a common spectrum, femto-base station (femto-BS) is empowered with CR technology. In this paper, we present evaluation of a cognitive cellular network using a spectrum resource strategy based on binary particle swarm optimization (BPSO).

**Keywords:** cognitive radio, spectrum sharing, binary particle swarm optimization

## 1 Introduction

The continued growth and demand satisfaction of future cellular networks depends on the availability of spectrum resource. Currently, spectrum resource is underutilized due to Static Spectrum Allocation Policy as some studies pointed out [1] underling the need for a more flexible and efficient spectrum management. In this context, spectrum sharing techniques are proposed as a solution to reuse available spectrum through Cognitive Radio (CR) technology. It will enable the coexistence of primary (user with higher priority to access the spectrum) and secondary (users with lower priority to access to spectrum) radio nodes on the same spectrum band to improve the spectral efficiency. In order to access to a channel, a secondary user could perform one of the following spectrum sharing strategies: transmit simultaneously with the primary user



as long as the resulting interference is constrained (spectrum underlay), or exploit an unused channel of primary user (spectrum overlay) [2].

In order to meet mobile traffic demand is expected that future cellular networks overlaid femto-cells (small cells) on the existing macrocell network. To extend and share a common spectrum, femto-base station (femto-BS) is empowered with CR technology. A femto-BS with CR technology is able to adapt optimally their operating parameters according to interactions with the surrounding radio environment [3]. Either if a femto-BS performs overlay or underlay spectrum sharing strategies an admission and interference control approach should be taken into consideration to assure protection to primary user, that is, to guarantee that its communication cannot be disrupted due to share the channel. However, certain Quality of service (QoS) should also be taken into account in the secondary user side to provide significant benefits to both primary user and secondary user from spectrum sharing. Therefore, it is necessary to quantify the effect of the femtocells networks (secondary users) interference on the macrocell network (primary user) performance. A potential application of femtocells is envisioned when a large number of users congregate at the same time such as in case of game stadiums. Under this situation, the macrocell network is likely to be overloaded due to the large amount of data generated. If some of this data can be offloaded to additional spectrum, such as femtocells, the users can be served [4].

In this paper we show performance system in terms of number of admitted secondary links coexisting with primary links, and maximum throughput. We consider throughput as a metric of spectral efficiency and spectrum underlay as the spectrum sharing technique. We focus on the downlink analysis since it is more critical in terms of femto-macro interference [5]. The solution procedure is based on an improved version of Binary Particle Swarm Optimization (BPSO) algorithm [6], known as Socio-Cognitive Particle Swarm Optimization (SCPSO) [7].

In some works performance results about admission and interference control in cognitive cellular networks are presented. In [8] is evaluated the performance of different sharing schemes (interweave, underlay, controlled underlay) in terms of transmission capacity. Numerical results conclude that controlled underlay scheme provides improved spatial reuse. On the other hand, in [9] an adaptive resource management strategy based on game-theory is proposed. It employs power control to mitigate interference among femtocells and macrocell. It concludes that when femto-BSs recognize the interference sources, interference management in cellular networks is enhanced therefore a higher throughput is achieved for femtocells. Its main drawback is that it only considers maximizing throughput of femtocells, furthermore, its shutdown process in femto-BSs can introduce additional computational time. The downlink spectrum sharing on overlay mode is addressed in [10] to improve network capacity, and mixed primal and dual decomposition methods are applied to solve it. However the time complexity is an issue. Another downlink spectrum sharing on overlay mode is presented in [11], a game theory is used to mitigate cross-tier and intra-tier interference, the spectral efficiency is in terms of the concept of effective capacity which is defined as the maximum constant arrival rate that can be supported by the system while satisfying the given QoS requirement. Channel sensing introduces an overhead since data transmission and reception cannot be performed within a sensing frame. In [12] an admission control algorithm to manage interference in two-tier femtocell network is proposed and QoS is provided for macro-cell and femto-cells. However, when the network



becomes congested, the admission control algorithm converges slowly down. It also presents unfairness since higher QoS for macro-users can be improved at the cost of degrading the QoS of the femto-users.

The remainder of this paper is organized as follows: In section 2, we present the system model and we introduce the solution procedure based on SCPSO. Section 3 shows simulation results. Section 4 concludes this paper.

## 2 Macro-Femto spectrum sharing approach

The macrocell consists of multiple macro-users and a macro-Base Station (macro-BS) located at the center of a coverage area  $A$ , then a number of femtocells are randomly distributed on  $A$ . A secondary link is represented by the union of a transmitter (femto-BS) and a receiver (femto-user) and it is identified by a number beside the link. Similarly, the union of a transmitter (macro-BS) and a receiver (macro-user) is referred as a primary link. A primary link has a primary channel to share (the numbers in braces in Fig. 1) and it can be assigned to several secondary links (the number in brackets in Fig. 1), as long as they, together, do not disrupt communication in the primary link.  $P$  and  $S$  represent the set of primary and secondary links respectively. To assure successful communications for those secondary links that attempt to exploit concurrently a channel with a primary link, a certain QoS is also guaranteed for them (denoted by  $\alpha$ ). Macro-BS and femto-BSs transmit at any given channel at full power; therefore, transmission power is maintained constant.

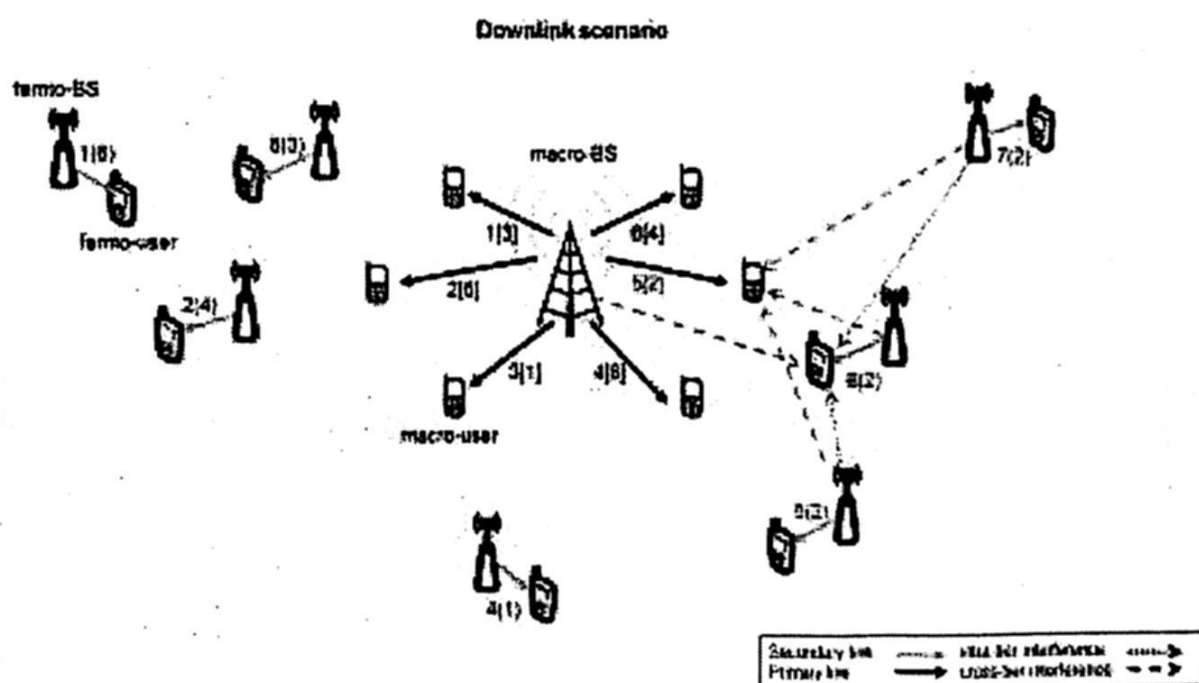


Fig. 1. Downlink interference scenario in macro-femtocellular networks

A successful reception of a transmission at a primary link depends on whether the signal-to-interference-plus-noise ratio (SINR) observed by macro-user is larger than an SINR threshold (denoted by  $\beta$ ). The SINR at the receiver of primary link  $v$  is given by:

$$SINR_v = \frac{P_v / ldp(v)^n}{\sum_{k \in \Phi} P_k / dps(k, v)^n}, 1 \leq v \leq Pl \quad (1)$$

where  $P_v$  is the transmit power of primary link  $v$ ,  $ldp(v)$  is the link distance of primary link  $v$ ,  $n$  is the path loss exponent (a value between 2 and 4). Those parameters characterized the desired signal. On the other hand  $P_k$  is the transmit power of secondary link  $k$ ,  $dps(k, v)$  is the distance from transmitter in secondary link  $k$  to receiver in primary link  $v$ .  $k$  is the index of active secondary transmitters.  $\Phi$  is the set of active secondary transmitters. The aforementioned refers to the aggregated cross-tier interference, that is, the total interference from those secondary links using the same channel that primary link being analyzed as show in Fig. 1 in which  $SINR_v$  is computed in primary link 5.

In contrast, the SINR at the receiver of secondary link  $u$  is given by:

$$SINR_u = \frac{P_u / lds(u)^n}{\sum_{k \in \Phi} P_k / dss(k, u)^n + P_v / dps(v, u)^n}, 1 \leq u \leq Sl \quad (2)$$

where  $P_u$  is the transmit power of secondary link  $u$  and  $lds(u)$  is the link distance of secondary link  $u$ . Meanwhile,  $P_k$  is the transmit power of transmitter of secondary link  $k$ ,  $dss(k, u)$  is the distance from transmitter of secondary link  $k$  to receiver of secondary link  $u$ . The above represents the aggregate intra-tier interference. In contrast,  $P_v$  is the transmit power of primary link  $v$ ,  $dps(v, u)$  is the distance from transmitter of primary link  $v$  to receiver of secondary link  $u$ . That refers the cross-tier interference perceived by a receiver of secondary link  $u$ .  $\alpha$  represents SINR threshold for secondary links. Fig. 1 shows  $SINR_u$  computed in secondary link 6.

Data rate contributions of the secondary and primary links are derived from equations (3) and (4) respectively. The data rate depends on channel bandwidth  $B$  that secondary and primary links can share and the conditions of the propagation environment (attenuation and interference).

$$c_u' = B \log_2(1 + SINR_u) \quad (3)$$

$$c_v'' = B \log_2(1 + SINR_v) \quad (4)$$

The metric considered as a measure of spectral efficiency on the cognitive cellular network is data rate, therefore the objective of resource allocation is to find the maximum data rate of system (5) subject to the SINR requirements of the secondary links (6) and primary links (7), that is:

$$\text{Max} \sum_{u=1}^{Sl} c_u' x_u + \sum_{v=1}^{Pl} c_v'' \quad (5)$$

$$SINR_u \geq \alpha \quad (6)$$

$$SINR_v \geq \beta \quad (7)$$

$$c_u' > 0, u=1, 2, \dots, Sl \quad (8)$$

$$c_v'' > 0, v=1, 2, \dots, Pl \quad (9)$$

$$c_u', c_v'' \in R^+ \quad (10)$$

$$x_u = \begin{cases} 1, & \text{if } SINR_u \geq \alpha \text{ and } SINR_v \geq \beta \end{cases} \quad (11)$$

0, otherwise

where  $x_u = 1$  if secondary link  $u$  is included in the solution and  $x_u = 0$  if it remains out as indicated in (11).

To find the set of secondary links that can maximize the data rate of cognitive cellular network without degrading the QoS of both the macrocell and the femtocells, a systematic procedure based on BPSO is used, in particular, the SCPSO. In BPSO methods, a swarm is composed as a number of particles  $S$  and a particle (vector  $X_i$ ) represents a candidate solution of the problem. Each particle  $X_i$  has its own velocity (vector  $V_i$ ) and memory (vector  $P_i$ ) in which the best solution found by the particle so far is recorded ( $pbest$ ). On the other hand, the best solution found by the whole swarm is called  $gbest$  (vector  $P_g$ ). At each iteration, the particle evolves taking into account the best solution found in its path,  $pbest$ , and the leader,  $gbest$ , until a stop condition is met. The algorithm to address the spectrum sharing problem in the cognitive cellular network is as follows:

**Input:** The number of secondary links  $Sl$ , the number of primary links  $Pl$ , SINR thresholds  $\alpha=\beta$ , the number of particles  $S$ , and the number of iterations and  $T_{max}$ , the number of runs.

**Output:** Maximum data rate in the system  $f(P_g)$ , the set of selected secondary links  $P_g$ , channel allocation for primary links *vector Spectrum Status*, the best channel allocation for secondary links  $P'_g$ , SINR level at primary links, and SINR level at secondary links.

**Step 1:** It is the initialization stage, it includes:

- 1.1: Locate randomly  $Sl$  and  $Pl$  over the coverage area  $A$
- 1.2: Initialize randomly candidate solution vector  $X_i$ , where  $x_{id} \in \{0,1\}$
- 1.3: Initialize randomly velocity vector  $V_i$ , where  $v_{id} \in [-V_{max}, V_{max}]$
- 1.4: Set  $P_i = X_i$
- 1.5: Let coincide the personal best channel allocation vector  $P'_i$  and candidate channel allocation vector  $X'_i$
- 1.6: Initialize randomly vector *Spectrum Status* with values from  $Pl$ .

**Step 2:** The update  $P_i$  stage. Particle compares  $f(X_i) > f(P_i)$  according to objective function in equation (5) and restrictions in (6)-(11), and overwrites  $P_i$  if  $f(X_i)$  is higher than  $f(P_i)$ .

**Step 3:** The update  $P_g$  stage.  $P_i$  values will be compared with current  $P_g$  value, so if there is a  $P_i$  which is higher than current  $P_g$ , it will be overwritten.

**Step 4:** Update elements in  $X_i$  and  $V_i$  according to the following equations:

$$v_{id} = w \times v_{id} + c_1 r_1 (p_{id} - x_{id}) + c_2 r_2 (p_{gd} - x_{id}) \quad (12)$$

$$v_{id} = w^l \times v_{id} + c_3 (gbest - pbest) \quad (13)$$

$$x_{id} = x_{id} + v_{id} \quad (14)$$

$$x_{id} = x_{id} \bmod (2) \quad (15)$$

where  $c_1$  and  $c_2$  are the learning factors,  $c_3$  is the socio-cognitive scaling parameter,  $r_1$  and  $r_2$  are random numbers uniformly distributed in  $[0,1]$ ,  $w$  and  $w^l$  are the inertia weights.

**Step 5:** If  $x_{id} = 1$  then allocate randomly a new channel to  $x'_{id}$  from the set of primary links  $Pl$ .

**Step 6:** For each particle in the swarm, perform Step 2 – Step 5.

**Step 7:** Repeat Step 2- Step 6 until stopping criterion met.

### 3 Performance evaluations

We deploy the downlink of a CDMA (Code Division Multiple Access) in two-tier heterogeneous network following the model described in section 2 and shown in Fig. 1. A number of secondary links  $Sl$  are randomly spread over an area of 5000 m x 5000 m (macrocell range). Also, there are six primary links,  $Pl$ , which have fixed locations. Each secondary transmitter (femto-BS) is assumed to have an assigned receiver at a random limited distance (30 m),  $r$ , away. Moreover, each transmitter is assumed to employ unit transmission power and the channel strength to be determined by path loss. We treat interference as noise, assume that the ambient/thermal noise is negligible, and assert transmission success to be determined by the SINR lying above a specific threshold where  $\alpha=\beta$ .

The snapshot of the location of  $Pl$  and  $Sl$  present in the area is called a scenario. An experiment is defined by a set of scenarios (with same  $Sl$  and  $Pl$ ) at given SINR threshold. By imposing different SINR thresholds, we simulate different requirements for mobile applications to guarantee a good service. At each experiment, 500 independent runs are taken. Each run represents a different placement of the  $Sl$  in a scenario. The stopping criterion for a run is defined by the maximum number of iterations  $T_{max}$ . The simulation parameters are given in Table 1.

Table 1. Simulation parameters

Parameters	Value
Number of secondary links $Sl$ =	20
Number of primary links $Pl$ =	6
Channels to share =	1, 2, 3, 4, 5, 6
Runs =	500
SINR thresholds $\alpha=\beta$	4, 6, 8, 10, 12, 14 dB
Channel bandwidth =	20 MHz
Swarm size $S$ =	40
Maximum number of iterations $T_{max}$ =	100
Cognitive, social and socio-cognitive factors $c_1, c_2, c_3$ =	2, 2, 12
Inertia weight $w$ =	0.721
Maximum velocity $V_{max}$ =	[-6,6]

From the set of 500 runs that are evaluated for a given experiment at a SINR threshold ( $\alpha=\beta$ ), the run containing the maximum data rate of the system is taken and that information is analyzed and reported in Table 2. Those results suggest that the higher the SINR threshold, the data rate decreases since the requested QoS is higher and it restricts the number of admitted secondary links coexisting with the primary links. This last observation is consistent with results reported in [13] which concludes that increasing the SINR threshold decreases the permissible number of secondary links. Also from Table 2, it is shown that for higher SINR thresholds is more challenging to share a channel, in this case, from SINR values at 10 dB. Then the



network capacity directly depends on the interference limit established in the cognitive cellular network.

**Table 2.** The best found at each experiment

$\alpha=\beta$ (dB)	Maximum system throughput (Mbps)	Number of selected secondary links allocated to primary channels						Total number of se- lected second- ary links
		Ch 1	Ch 2	Ch 3	Ch 4	Ch 5	Ch 6	
4	9828.3486	4	2	2	2	2	4	16
6	9376.8802	2	5	1	3	2	3	16
8	9585.0736	1	2	3	4	5	2	17
10	8877.6855	3	3	3	5	2	0	16
12	9295.4222	0	2	5	2	2	5	16
14	8648.4162	4	2	1	3	3	2	15

Reusing spectrum bands represents benefits in terms of spectral efficiency as long as band-specific conditions are imposed to those which are allowed to access the same range of frequencies. If not conditions are imposed, it can lead to a “tragedy of the commons” in which many users try to access the same spectral resource and neither is able to communicate reliably given the amount of interference. Those conditions are regulatory policies which represent specifications of network deployment and operation to avoid harmful interference among coexisting systems. Some examples of specifications of network deployment are number of selected secondary users, exclusion zones (radius of protection of a primary user), transmission power, and SINR thresholds.

## 4 Conclusion

In two-tier heterogeneous network, we study the spectrum resource optimization problem. We aim at maximize the throughput of the system as a metric of spectral efficiency under QoS constraints. Then an adaptive resource management framework



based on an improved version of binary particle swarm optimization is applied to solve the problem.

The numerical results have shown that, network performance depends directly on the value of requested QoS in the system i.e., taking into consideration the requirements of primary and secondary users. They also suggest that is possible that a set of secondary users can share simultaneously with the primary user a channel, as long as, certain conditions are imposed to secondary users.

Regulation is a key role to support spectrum sharing, in this context, identify design requirements for deploying future cellular networks based on CR technology is helpful for developing regulatory policies that assure a peaceful coexistence among heterogeneous systems.

## References

1. Roberson, D.A., Hood, C.S., LoCicero, J.L., MacDonald, J.T.: Spectral Occupancy and Interference Studies in support of Cognitive Radio Technology Deployment. 1st IEEE Workshop on Networking Technologies for Software Defined Radio Networks, 2006. SDR '06. pp. 26–35. IEEE (2006).
2. Lu, L., Zhou, X., Onunkwo, U., Li, G.Y.: Ten years of research in spectrum sensing and sharing in cognitive radio. *EURASIP J. Wirel. Commun. Netw.* 2012, 28 (2012).
3. Ahokangas, P., Matinmikko, M., Yrjola, S., Okkonen, H., Casey, T.: “Simple rules” for mobile network operators’ strategic choices in future cognitive spectrum sharing networks. *IEEE Wirel. Commun.* 20, 20–26 (2013).
4. Wang, J., Ghosh, M., Challapali, K.: Emerging cognitive radio applications: A survey. *Commun. Mag. IEEE.* 49, 74–81 (2011).
5. Palanisamy, P., Nirmala, S.: Downlink interference management in femtocell networks - a comprehensive study and survey. 2013 International Conference on Information Communication and Embedded Systems (ICICES). pp. 747–754 (2013).
6. Kennedy, J., Eberhart, R.C.: A discrete binary version of the particle swarm algorithm. *Systems, Man, and Cybernetics*, 1997. “Computational Cybernetics and Simulation”. 1997 IEEE International Conference on. pp. 4104–4108 vol.5 (1997).
7. Deep, K., Bansal, J.C.: A Socio-Cognitive Particle Swarm Optimization for Multi-Dimensional Knapsack Problem. International Conference on Emerging Trends in Engineering and Technology (ICETET). pp. 355–360 (2008).
8. Cheng, S.-M., Ao, W.-C., Tseng, F.-M., Chen, K.-C.: Design and Analysis of Downlink Spectrum Sharing in Two-Tier Cognitive Femto Networks. *IEEE Trans. Veh. Technol.* 61, 2194–2207 (2012).
9. Liu, Y., Cai, L.X., Shen, X., Luo, H.: Deploying cognitive cellular networks under dynamic resource management. *IEEE Wirel. Commun.* 20, 82–88 (2013).
10. Xiang, J., Zhang, Y., Skeie, T., Xie, L.: Downlink Spectrum Sharing for Cognitive Radio Femtocell Networks. *IEEE Syst. J.* 4, 524–534 (2010).

11. Lien, S.-Y., Lin, Y.-Y., Chen, K.-C.: Cognitive and Game-Theoretical Radio Resource Management for Autonomous Femtocells with QoS Guarantees. *IEEE Trans. Wirel. Commun.* 10, 2196–2206 (2011).
12. Ngo, D.T., Le, L.B., Le-Ngoc, T., Hossain, E., Kim, D.I.: Distributed Interference Management in Two-Tier CDMA Femtocell Networks. *IEEE Trans. Wirel. Commun.* 11, 979–989 (2012).
13. Roy, S.D., Mondal, S., Kundu, S.: A new algorithm for admission control of secondary users in CDMA based Cognitive Radio Network. 2010 International Conference on Computer and Communication Technology (ICCCT). pp. 35–39 (2010).

# Segmenting supervised activities in a video sequence based on handling of artifacts towards intelligent systems

Francisco E. Martínez-Pérez<sup>1</sup>, Héctor G. Pérez-González<sup>1</sup>, José A. González-Fraga<sup>2</sup>, Juan Carlos Cuevas-Tello<sup>1</sup> and Sandra Edith Nava-Muñoz<sup>1</sup>

<sup>1</sup> Facultad de Ingeniería, Universidad Autónoma de San Luis Potosí, Av. Manuel Nava No. 8, Zona Universitaria, San Luis Potosí, S.L.P. 78290, México  
eduardo.perez@uaslp.mx, hectorgerardo@acm.org, cuevas@uaslp.mx, senavam@uaslp.mx,

<sup>2</sup> Facultad de Ciencias, Universidad Autónoma de Baja California, Carretera Transpeninsular Ensenada-Tijuana No. 3917, Colonia Playitas, Ensenada, B.C. 022860, México

angel.fraga@uabc.edu.mx

*Paper received on 11/22/13, Accepted on 01/19/14.*

**Abstract.** Nowadays intelligent systems community has conducted research on human activity recognition. For example, in a healthcare environment, we would like to know what activities (feeding, blood pressure, hygiene and medication) are performed by a caregiver given a video sequence (recorded by a surveillance system). Specifically, it is complicated to infer those activities that are performed using one or several artifacts at different times, so the activity inference and video segmentation are complex tasks. Additionally, it is desirable to perform the video segmentation in an automatic fashion. Therefore, in this paper we present an intelligent system for video segmentation. We present an example in a realistic environment in which the analysis of video sequences per day was reduced by using video segmentation.

**Keywords:** video segmentation, activity recognition, roaming beat.

## 1 Introduction

Generally, an intelligent system is composed by several modules with the aim to support robustness and computational complexity. The system relies on signal processing and machine learning techniques [1]. It first applies preprocessing to remove noise from the signals and segment them. Next, the system extracts signal features to enhance the characteristics unique to each activity and to reduce data dimensionality, and then it uses classifiers to map these features to discrete activity or context classes. The goal of an intelligent system for human activity recognition consists of automatically analyzing and classifying activities with information gathered from different capture sources like video cameras or other sensors. However, human activity recognition is complicated. It is due to

there exists a lot of way to perform an activity. For example, humans perform several physical activities such as walking, running and so on. In this kind of activity to get signal processing, some authors have reported physical recognition results using a set of templates or people silhouettes corresponding to each activity [2, 3]. A related work in which physical activities are analyzed, can be found in [4]. Other works of different kind are focused on activity classification, for example some works involve physical artifacts as shown in [5]. These authors show activity classification using several kinds of sensors. Moreover, the authors classify activities, that human performs, on three classes such as sequential, interleaved and concurrent activities. For example, when someone brushes his teeth and continues to another activity like arranging his hair, this corresponds to a sequential activity. Interleaved activities correspond, for instance, to the case in which someone is eating dinner and then answers the phone, at the end both are performed at the same time. Finally, two activities are concurrent when these activities start at the same time, i.e. eating and watching TV, both activities carried out always at the same time.

For example, a comparison of inferring activities considering sequential, interleaved and concurrent is done in [6]. They propose some methods to perform inference, comparing four techniques: the hidden Markov model (HMM), the conditional random field (CRF), a variant of CRF and emerging patterns (EP). They reported acceptable results using EP on inferring physical activities. Other work is presented in [5] in which they show an effectiveness rate of 66.13 percent in the inference of sequential, interleaved and concurrent activities. Its main disadvantage is that authors take into account sequential process and their validations require a controlled space. Therefore it is necessary to take into account an understanding such as presented in [7], which comprises three elements that correspond to a person, the use of devices or tools, and the purpose of the activity. The activities are never performed alone, these are interleaved with other activities with the same goals, or instruments used in producing the intended activity. In the development of a specific activity, the motion artifact may be viewed as a change of activity or a change in the purpose of the actions, or a grouping of actions. Therefore, it is established that the actions of the activities should never be labeled as unknown, according to [8], they create a list of ordered actions, the first one being the most likely.

For aforementioned reasons, activities must be seen as multiple flows of actions that are performed on different times, so these activities do not have a specific behavior in order to recognize them. Therefore, it is necessary to take into account all variants of actions when activity is started, performed and finished. In this sense, to solve this problematic, we use a concept and its methodology called Roaming Beat (RB) [9] (see section 2.1).

On the other hand, video sequence segmentation is a problem that activity recognition has tackled. For example, [10–13] show the work in which a video sequence segmentation is implemented using image processing. Background technique is used in [10]. Other authors treat video segmentation as non-local spatiotemporal structure using regions technique [11]. In the same direction, in [12] spatiotem-



poral technique is used; these authors show an algorithm for video segmentation using motion cues from past and future frames. Another way to make a video segmentation is presented by [14]. The process performed by these authors is based on a set of labels in the video sequence. These allow them to automatically discover a set of relevant tags and extract a set of key words correlated with an activity. However the proposed technique depends on the activities being related to video labels. Such labels consist of contextual information describing the activity of the video, i.e. the video was previously analyzed and labeled. Additionally activity recognition is not obtained in an automatic way. Unfortunately, all the video sequences in real life are not automatically labeled. Furthermore, these sequences involve a multitude of people both static and in movement. This allows us to give meaning to an activity as these movements are signals for the occurrence of an event, forming a set of activity events.

The main contribution of this paper is the development of an intelligent system for video segmentation. We describe the process to store an activity in a knowledge base and the process to recover an activity from the knowledge base. This paper is organized as follows: Section 2 introduces our methodology for activity segmentation in a video sequences. In Section 3, the results showing an example implemented in a healthcare environment. Finally section 4 provides our conclusions and future work.

## **2 Labeling a video sequence and reducing video analysis**

The artifacts play an important role for activity inference, because the artifacts are a trigger within an event in the setting. It is caused through activity performance because an activity is mediated by one or more instruments and is directed toward a certain artifact[7].

### **2.1 Labeling a video sequence through Roaming Beat concept**

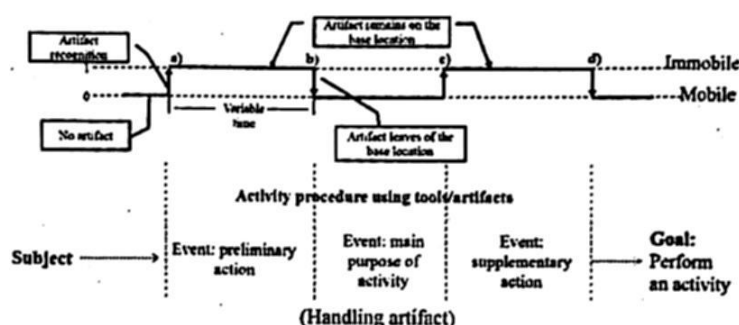
In a previous work, for inferring human activities in a Healthcare Environment the Roaming Beat (RB) concept was used. A RB is defined as:

"The ability of an artifact to give a time stamp (hour and date) to change from motionless to a mobile state and change from the base location to any other" [15].

This concept is useful for describing the artifact behavior when an activity is performed. The behavior is obtained as a result of data conversion when the artifacts are recognized in a base location through a sensor [9, 15]. Also using RBs methodology it is possible to label video sequences through identifying artifacts.

Each artifact produces its own roaming beat when it is handled. The change of state is called "beat", so several changes of state can be observed as behavior of the artifact (see a, b, c, d in Fig. 1) as a result of the artifact recognition in a base location in a time span.





**Fig. 1.** Set of events related to an activity using an artifact. Artifact's behavior representation (or roaming beat) (a) first beat; (b) second beat; (c) third beat; (d) fourth beat

This concept is useful for describing the artifact behavior when an activity is performed. The behavior is obtained as a result of data conversion when the artifacts are recognized in a base location through a sensor 43. Also using RBs methodology it is possible to label video sequences through identifying artifacts. Each artifact produces its own roaming beat (RB) when it is handled. The change of state is called "beat", so several changes of state can be observed as behavior of the artifact (see a, b, c, d in Fig. 1) as a result of the artifact recognition in a base location in a time span. Each signal produced by the object recognition method, is converted into a train of pulses, as shown in Fig. 1.

Moreover, a camera is involved at the same setting. Each time that a beat is produced, an index of video sequence is related to it, so it is possible to get a specific video sequence segment related to the activity, i.e. in Fig. 1(a) starts the activity and Fig. 1(d) ends the activity. Additionally it is possible to get several images related to the activity using the indexes. It is important to note that an activity can be related to one or more artifacts, and each artifact produces its own beats when it is handled. Therefore, by each beat produced, there is a record in a database that it is composed by a specific index related to a video sequence (the first beat is equal to start index related to the activity and the last one beat is equal to the end index). Hence, it is possible to get a video segment related to the activity performed.

## 2.2 Identifying when activities or events happen

As mentioned above, each artifact produces its own beats when they are handled, this manipulation is detected because the artifact begins to have movement between the base location and any coordinate location in space (roaming), and can be a significant movement in the scenario when someone performed an activity. For this reason, a beat is interpreted as an event when the activity is performed. Therefore, each event is related to an index in a video sequence.

For activity recognition is necessary to define two terms: a) recognized activity, it is a set of arranged events produced by an artifact in a specific activity; and b) an event is when someone takes one or some artifacts and places them on the base location. The event can be an isolate movement.

An activity is recognized when the activity is completed by events produced by each artifact. The activity behavior is composed by three steps: 1) preliminary actions; 2) the main purpose of the activity; and 3) supplementary actions. These actions are accomplished when the all beats are obtained in the activity. The whole number of beats must be greater than or equal to 4 as shown in Fig. 1. Moreover, the beats or events number must complete a specific time  $t$  based on the activity 4. However, there are some events that may occur, so the artifact has not accomplished its established behavior in an activity. In this sense, there exist some events that can be postponed in an activity. Despite, the activity is not affected. Those events are called: preliminary and supplementary actions. These actions can be interrupted, but can be performed later in a time  $t+t_1$ , and then perform the main purpose of the activity. Where  $t_1$  is the long time to take it into account in a whole activity. These kinds of events may be important; it depends of both the scenario and the user requirements. In this kind of event, the artifact can leave or remain on the base; therefore the main reason is that an activity is not finished so far.

In summary, the activity recognition and events are very important for the activity inference process. Because an activity is composed by a starting event; a behavior related to an activity is composed of a set of events; and an end activity. The end activity is performed within the activity in a time  $t$ . Those events that are started but the artifact has not accomplished, may be taken into account as important events and considered in future analysis for other intelligent systems.

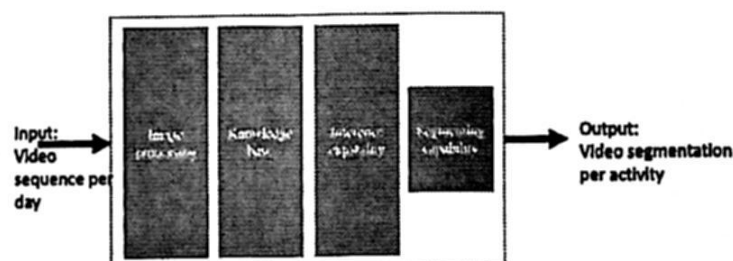


Fig. 2. General architecture of the intelligent system

### 2.3 Describing the activity recognition process

In Fig. 2, we present the architecture of the proposed intelligent system. The input is a video sequence and the output is a segmented video that represents an activity. The Image Processing module employs correlation filters for artifact recognition [16]; this process gives us the roaming beats. The knowledge base stores the RB of each artifact (see Fig. 3). The module of Inference Capability contains the process for recovering an activity from a knowledge base (see Fig. 4).

As aforementioned, the artifacts behavior produce a set of events within a time span, so an activity is recognized. The system shown in Fig. 3 is composed by two objects recognized, our example is based on video cameras (two video cameras for artifact recognition). The object recognition is just implemented in

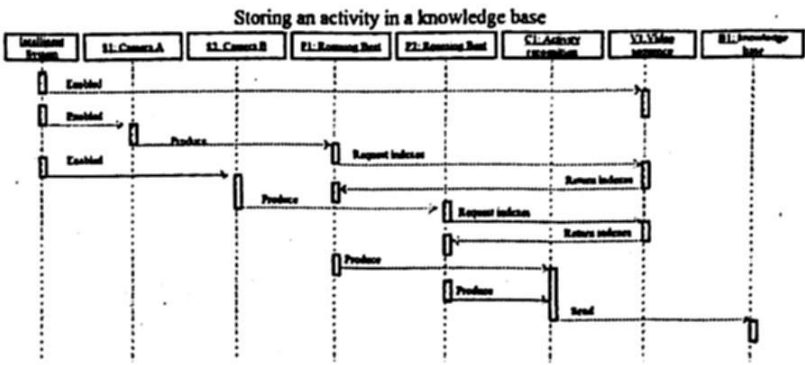


Fig. 3. Process for storing an activity in a knowledge base

a camera 1 but each beat produced is related to each index of both Camera A (S1) and Camera B (S2). Also the system includes a knowledge base in which all activities are recorded. To clarify the whole process, we divided it in two steps showed in Figs. 3 and 4. The first one is the process for storing each activity in a knowledge base and the second one is for recovering and showing an activity from the knowledge base.

Fig. 3 shows the activity recognition process 4, in which the system starts when it enables the cameras (S1 and S2) and the video sequence (V1). Each time a user handles one or several artifacts an event is produced (P1 and P2). Each event obtains an index related to image that it is requested to the video sequence and the index is returned (P1 and P2 through V1). Once all artifacts have performed an activity, the activity is sent and recorded into the knowledge base (P1 and P2 toward B1).

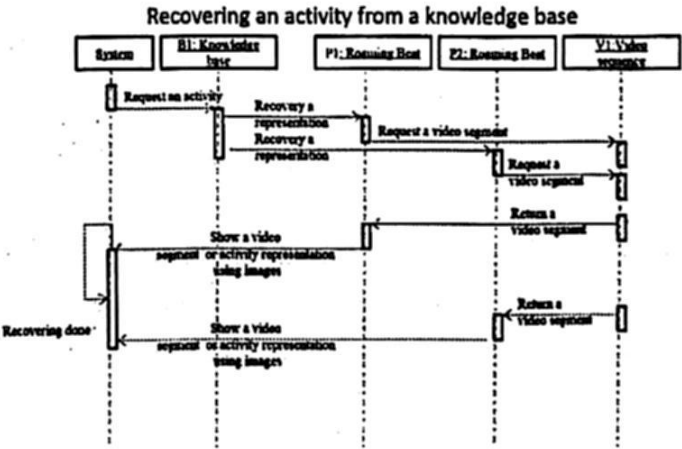


Fig. 4. Process for recovering an activity from a knowledge base

Fig. 4 depicts the process for recovering an activity from a knowledge base through a query. First, the users need to obtain information related to the performed activities and they request it to the system. It recovers an activity from the knowledge base (B1), but before that, the knowledge base joins the RBs involved in the activity requested. Each RB requests its own video segment (P1 and P2) from the whole video sequence (i.e. from current day). The video seg-

ment is returned to each RB. Once that all RB are obtained is possible to show an activity and can be in two ways: the first one is an activity representation using images obtained from each index related to an event or beat; and another is showed in a video segment. The system recoveries the indexes related to the start and end of and an activity.

In this section, we described the whole process related to activity recognition in which it was divided in two steps: the purpose of the first one is activity recognition through to obtain the artifacts' behavior; and the second one is to show the activity representation to the user in two ways: using images or video segments. However, it is necessary to evaluate our applications, so we present the evaluation protocol in the next section.

3 Results in situ

We proposed to evaluate our application in a healthcare environment. It is important to infer the activities that the caregivers perform during their workday with the aim to provide a proper care to the elders. The following sections describe the context of the environment and the main results of the evaluation.

3.1 Environment

The evaluation of the application was conducted in a private nursing home for a period of ten days during 12 hours per day. Two video cameras were installed in a room where an elderly patient with restricted mobility (ERM) was living. One camera was installed in the ceiling of the room and another was installed close to the base location (2 meters approximately). The cameras used were a wvc53gca Linksys model. The video was captured using MPGe-4 format having a resolution of 320 240 pixels, the frame rate was 2.4 per second. Storage and processing were done in a model T3500 Dell Precision workstation.

Table 1. Record of events and activities performed in a healthcare environment, Act: Activities, Ev: Events, D1 : Day 1, D2: Day 2... D10: Day 10, Tot: Total

Descrtiption	D1	D1	D2	D2	D3	D3	D4	D4	D5	D5	D6	D6	D7	D7	D8	D8	D9	D9	D10	D10	Tot	Tot
Activities/Events	Act	Ev	Act	Ev	Act	Ev	Act	Ev	Act	Ev	Act	Ev	Act	Ev	Act	Ev	Act	Ev	Act	Ev	Act	Ev
Hygiene	1	1	1	0	2	0	2	0	2	1	3	0	3	0	2	1	2	1	3	0	21	4
Blood Pre	1	0	2	0	2	0	1	0	1	0	1	0	1	0	3	0	1	1	1	1	14	2
Feeding	2	0	3	0	2	0	3	0	2	0	3	0	2	0	3	0	1	0	2	0	23	0
Medications	2	1	2	0	2	0	1	0	1	0	2	0	1	0	2	0	1	0	2	0	16	1
Total by day	6	2	8	0	8	0	7	0	6	1	9	0	7	0	10	1	5	2	8	0	74	7
Total																						81

There were a total of 109 hours recorded by camera which 941,760 images were analyzed in real time. The analysis consisted of applying 6 filters for recognition using image correlation methods for video sequence generated by the



camera that was on the base location 3. There were a total of 81 video segments corresponding to 74 completed activities and 7 events are shown in Table 1.

### 3.2 Describing activities and events results

The application that was developed for this evaluation was able to infer a total of 74 activities and 7 events based on our classification as described (see Section 2.2). This assessment generates information of activities of the three types: sequential, concurrent and interleaved. The following describes the considerations taken into account with respect to the four activities as shown in Table 1:

1. The activity of taking blood pressure, was always performed independently, i.e. there was no additional activity or event when this event was performed. It was considered as a simple activity. The total number of simple activities correspond to 14 blood pressure activities, 10 of feeding activity and 3 medications, whose results were 27 activities and 2 events recorded.
2. The feeding and medication activities are sequential. These activities are sometimes performed in a concurrent way. The result was a total of 26 inferred activities and 1 registered event. These activities were considered concurrent because in healthcare environments, medicines are supplied during two feeding events corresponding to breakfast and dinner. Therefore, Table 1 shows a difference between these two activities (23 food and 16 drugs). These two activities are a total of 13 feeding activities performed concurrently and 10 performed in a sequential way. With respect to medication activity, 13 were performed concurrently and 3 were performed sequentially.
3. Hygiene activity is an activity with complex actions, because it includes a set of artifacts. Each artifact has its own behavior within the activity, so this activity was seen as interleaved activity, given the behavior of the artifacts and the people who performed the activity. We had 21 interleaved activities.

These three types of activities were successfully completed. Each behavior per artifact and per activity was established previously. However there were some events recorded that did not showed the expected behaviors. The results of these events were: two sequential activities were related to events (blood pressure event), an event associated with a concurrent activity (medicine) and four interleaved activities were related to hygiene events.

Subsequently, video segments generated by events were analyzed in order to determine the possible causes of the occurrence. The main reasons were identified: in the case of events of the blood pressure activity, it was observed that the artifact was removed from the base location for placement elsewhere; because the previous time of this activity had been placed on the base location. Another reason was the occlusion of the artifacts in its recognition, when reviewing the video related to medication; it was observed that this activity could be considered as a complete activity.

Finally, hygiene activity was the only one with a higher number of events. In



the analysis of the video segments generated by these events (4 events), it was observed that the completed behavior was not accomplished in this activity. However, it is possible to include these four events within another activity that correspond to the healing activity. This is because the video segments were very clear in this activity. Examples of movement indexes are presented below.

### 3.3 Showing activities based on roaming beat in a video segment

Following the activities in Table 1 and the results, let us show you the way how the events allow to recover from RB video clips or images that represent an activity.

Fig. 5 depicts how a complex activity is performed. This activity corresponds to the hygiene activity (interleaved activity). This figure shows the representation of video sequences captured by two cameras. Furthermore, this representation shows the behavior of the use of two artifacts that correspond to the paper towel and the physiological solution. Remember that, the image processing is performed on the video stream from the Camera A described in section 2.2.

Fig. 5 shows 16 beats produced by the paper towel (from 1 to 16). Each of these has an associated beat or event related to labeled picture with its number. There are also five beats produced by physiological solution (from 17 to 21) and displayed the images associated with their number.

The activity in the video segment is inferred from knowing the first and last beats corresponding to the relationship that is created when you registered the first beat with the index of the video stream into the knowledge base, so that the first beat was linked to the index 156 and the latter with the index 2160. These indexes are benchmarks that allow us to move to one side or the other. That is, index 156 tags the beginning of the sequence corresponding to the interaction between the caregiver and the patient. The caregiver enters to the room at the index number = 35 as shown in the image labeled "Initialized" in Fig. 5. It was possible to determine it by looking at the video stream from the Camera B starting at index 35. Specifically in this activity the preliminary actions of the activity start at index 35. But our application labeled the activity initialization at the index 156, also it was considered an activation of the activity from index 156 to 2160 and labeled the termination of the activity at the index 2160. The image labeled "Suspended" in Fig. 5 shows the time at which the caregiver leaves the room as a signal of completion of the activity, which corresponds to index 2220 of the sequence of video.

Another example from our application is shown in Fig. 6, which illustrates the execution of concurrent activities. Fig. 6a shows the feeding activity. This activity is related to the recognition of the tray. Figure 6b shows the medication activity. These activities were considered concurrent because when the tray arrives at the base location then the caregiver immediately performs preliminary actions to execute the activity of medication. It can be seen in Fig. 6 a small difference (in seconds) from the first beat corresponding to the activity of feeding and the first beat of medication activity. However, the activity of medication is the first one to be completed.

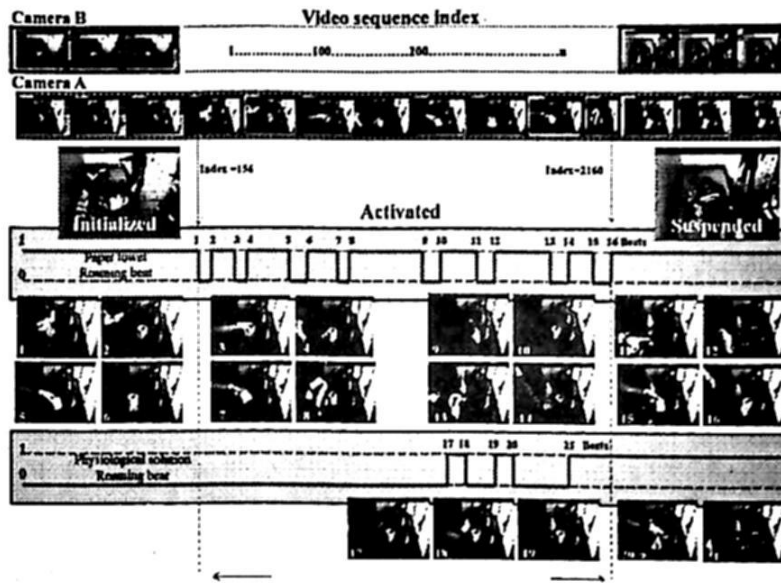


Fig. 5. Video segment related to hygiene activity using two artifacts.

As mentioned earlier, the indexes allow to move into the video stream backwards. As shown by the image obtained before the tray arrived to the base location, which was labeled as the image that initializes the activity and the beginning of the video segment. Fig. 6 shows images of the video stream from the Camera B. This is because each beat is related to the rate of each video sequence, i.e. for each camera.

On average 19 hours with 49 minutes were recorded per day. Using the activity inference application based on the manipulation of artifacts was possible to reduce the analysis of video sequences to only 2 hours 19 minutes which represents 11 percentage of the total hours in a day captured for analysis.

#### 4 Conclusions and future work

We conclude that the activity inference is a complicated process, so it is necessary to take into account several details related to human activities. The role of the knowledge base is important to store and recover activities for obtaining a video segmentation in an automatic way. Therefore, performing activities using physical artifacts should be seen as multiple flows of actions at different times. For this reason, the technique of Roaming beat is used as a solution for the inference of sequential, interleaved and concurrent activities, in which multiple events take place at the same time. These events are part of the inferred activities from motion cues generated by artifacts while the activities are performed. Hence, it was possible to obtain video segmentation and an intelligent system in an easy way.

Each event provides a meaning of the activity based on artifacts handling. The meaning is related with three phases of the activity: start, center (main development), and end; we consider these phases as preliminary actions, actions in the activity performed, and complementary actions, respectively. In our case of

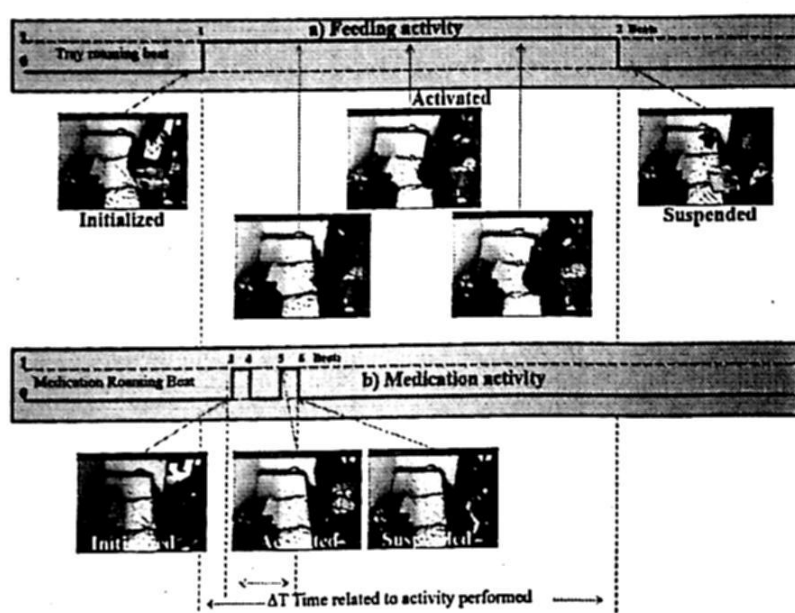


Fig. 6. Video segments related to two activities, feeding and medications activities.

study (healthcare environment), the meaning is related to the video sequence indexes; hence the indexes were used for video segmentation in an automatic way.

Moreover, each beat produced by artifacts allows us to recover video segment related to simple events or activities. These video segments are transformed to activity representation or in a means for decision making of the user.

We detailed our results for the purpose of linking multiple streams of actions. In this way it was possible to infer 27 sequential activities, 21 interleaved activities, and 26 concurrent activities. This produced 74 activity related, video segments, reducing the video sequence analysis to about 11 percentage of total daily video. Thus, the review that a person would have to make of a 12 hours video is reduced to only 2 hours.

In this work, the result obtained in a healthcare environment based on prior work [9], is presented in detail. As a future work, we pretend to implement the RB technique in other kind of scenario, the knowledge base is used only to store and recover activities and the knowledge base is static. We wish to apply data mining and artificial intelligence techniques in order to infer new activities based on the behavior of humans.

## 5 Acknowledgments

This work was founded by PROMEP under the contract PROMEP-103.5-13-6575 (UASLP-PTC-452) provided to the first author.

## References

1. Cook, Diane J. and Augusto, Juan Carlos and Jakkula, Vikramaditya R.: Ambient intelligence: Technologies, applications, and opportunities. Pervasive and Mobile

- Computing, pages 1–38,(2009)
2. Hu, Jinhui and Boulgouris, Nikolaos V.: Fast human activity recognition based on structure and motion. *Pattern Recognition Letters*, pages 1814–1821, 32, (2011)
  3. Qian, Huimin and Mao, Yaobin and Xiang, Wenbo and Wang, Zhiqun: Recognition of human activities using SVM multi-class classifier. *Pattern Recognition Letters*, 31, 100–111,(2010)
  4. Aggarwal, J.K. and Ryoo, M.S.: Human activity analysis: A Review *ACM Computing Surveys*, 3, 1–43,(2011)
  5. Gu, Tao and Wang, Liang and Wu, Zhanqing and Tao, X.: A pattern mining approach to sensor-based human activity recognition. *IEEE Transactions on Knowledge and Data Engineering*, 23, 9, 1359–1372 (2011)
  6. Kim, Eunju and Helal, Sumi and Cook, Diane: Human activity recognition and pattern discovery. *Pervasive Computing, IEEE*, 9,1, 48–53, (2010)
  7. Bødker, Susanne: Context and consciousness activity theory and human-computer interaction, *Applying Activity Theory to Video Analysis: How to Make Sense of Video Data in Human- Computer Interaction*. 7, 147–174, (1996)
  8. Rincón, J., Santofimia, M. J., and Nebel, J.: Common-Sense Reasoning for Human Action Recognition. *Pattern Recognition Letters*, (2012)
  9. Martínez-Pérez, Francisco E. and González-Fraga, Jose Ángel and Cuevas-Tello, Juan C. and Rodríguez, Marcela D.: Activity Inference for Ambient Intelligence Through Handling Artifacts in a Healthcare Environment. *Sensors*, 12, 1, 1072–1099, (2012)
  10. Lu, Guoliang and Kudo, Mineichi and Toyama, Jun: Temporal segmentation and assignment of successive actions in a long-term video. *Pattern Recognition Letters*, (2012)
  11. Ahuja, N.: Exploiting nonlocal spatiotemporal structure for video segmentation. In: *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 741–748. IEEE (2012)
  12. Lezama, J., Alahari, K., Sivic, J., Laptev, I.: Track to the future: Spatio-temporal video segmentation with long-range motion cues. In: *Cvpr 2011*, pp. 3369–3376, IEEE (2011)
  13. Trichet, R., Nevatia, R.: Video Segmentation with Spatio - Temporal Tubes. In: *International Conference on Advanced Video and Signal Based Surveillance Video*, pp. 330–335, IEEE (2013)
  14. Cho, S., Kwak, S., Byun, H.: Recognizing humanhuman interaction activities using visual and textual information. *Pattern Recognition Letters*, 1–9, (2012)
  15. Martinez-Perez, F. E., Gonzalez-Fraga, J. A., Tentori, M.: Artifacts' Roaming Beats Recognition for Estimating Care Activities in a Nursing Home. In: *4th International Conference on Pervasive Computing Technologies for Healthcare 2010*, (2010)
  16. Martinez-Perez, F. E., González-Fraga, J.A., Tentori, M.: Automatic activity estimation based on object behaviour signature. In: *Proceedings of SPIE*, 1, pp 77980E, (2010)



# A Hybrid ECJ+BOINC Tool for Distributed Evolutionary Algorithms

Francisco Fernández de Vega<sup>1</sup>, Leonardo Trujillo<sup>2</sup>, Francisco Chávez<sup>1</sup>, Enrique Mediero<sup>1</sup>, and Luis Muñoz<sup>2</sup>

<sup>1</sup> Universidad de Extremadura, C/. Santa Teresa de Jornet, 38, 06800, Mérida, España

fcofdez@unex.es, fcchavez@unex.es, enmediero@gmail.com

<sup>2</sup> TREE-LAB, Posgrado en Ciencias de la Ingeniería, Instituto Tecnológico de Tijuana, Av. Tecnológico S/N, Fracc. Tomás Aquino, Tijuana, B.C., México

www.tree-lab.org

leonardo.trujillo@tectijuana.edu.mx, lmunoz@tectijuana.edu.mx

*Paper received on 11/21/13, Accepted on 01/19/14.*

**Abstract.** This paper presents an improvement to ECJ (Evolutionary Computation in Java), the popular evolutionary computation tool, which allows users to exploit distributed computational resources through the use of volunteer computing. In particular, the BOINC (Berkeley Open Infrastructure for Network Computing) middleware is used to distribute ECJ client software on top of a virtualization layer, this allows researchers to parallelize their experiments without having to port their software to make it compatible with BOINC. In this way, an interested researcher can use the ECJ+BOINC software in the same way that they use the standard ECJ tool. Moreover, the system allows the user to choose between different distribution models based on their preferences. Finally, the ECJ+BOINC system is developed in a modular manner, allowing for easy updates and modifications based on the possible requirements of future ECJ versions.

**Keywords:** Evolutionary Computation, Distributed Computation, ECJ, BOINC

## 1 Introduction

Every year, the importance of distributed and parallel computing models becomes more apparent, with potential impacts in all of computer science. Recently, the president of the IEEE Computational Intelligence Society, in the May 2013 issue of the Computational Intelligence Magazine, listed a series of challenges that must be addressed by researchers in the field of Computational Intelligence. In particular, two problems stand out, commonly referred to as *Big Data* and *Real Time* [1]. For most researchers, these terms suggest the need to process large amounts of data on the one hand, and the need for fast and efficient processing on the other. It is also relevant to mention that from the perspective



of Evolutionary Computation (EC), a sub-field of Computational Intelligence, both concepts are not part of the canonical formulation of evolutionary algorithms (EA) for search and optimization [2].

Over recent years many proposals have attempted to integrate notions of parallel and distributed computation within traditional EA, such as parallel EA models or traditional algorithms that are simply parallelized at the implementation level. However, despite such work, there is still a tendency for researchers to use sequential models and implementations, what is commonly referred to as immobilization within the software industry. In general, researchers prefer to use well-known computational tools before investing time and resources in modifying existing tools or developing new ones.

Therefore, some works have focused on developing computational tools that facilitate the use of distributed resources while keeping the amount of time and effort at a minimum. For instance, in [8] and [6] the authors proposed a model that emphasizes the cloud computing model and allowed users to deploy virtual machines that internally executed an evolutionary algorithm within a volunteer computing setting. While the proposal met the stated goals, it suffered from two possible shortcomings: (a) the size of the distributed virtual machines was quite large (several gigabytes); and (b) researchers needed to prepare and deploy the necessary infrastructure to run the Berkeley Open Infrastructure for Network Computing (BOINC) middleware [4].

The current work goes one step further, by considering one of the most well-known and widely used EC tools called Evolutionary Computation in Java (ECJ) [5], and integrating within it the BOINC technology. The goal is to allow any researcher that is familiar with ECJ to launch experimental distributed experiments based on BOINC. Moreover, another goal is to make the required learning curve of the new technology less severe; this is accomplished by fully integrating a BOINC server within ECJ, and the necessary virtual machine within the BOINC client. In this way, an ECJ user only has to perform two simple tasks: (1) run the virtual machine that acts as server, by compiling the problem specific EA within this server just as it is normally done in ECJ; and (2) run the client virtual machines. When the EA is executed, the ECJ+BOINC system automatically distributes its execution across all available distributed resources or nodes.

This paper presents a proof-of-concept implementation of the ECJ+BOINC system, and also outlines specific areas for future work and improvements. The remainder of this paper is organized as follows: Section 2 reviews related work in volunteer and distributed computing models for EAs. Then, Section 3 describes the proposed methodology. Afterwards, experiments and results are summarized in Section 4. Finally, concluding comments and future work are discussed in Section 5.

## 2 Volunteer Computing and Distributed EAs

Volunteer computing, a specific distributed computing model, has gained large acceptance over recent years. It is based on users voluntarily cooperating with research projects, by offering their computing resources (personal computers or other devices) that are connected to the Internet to act as data processing units. The model uses specialized software that exploits computing and storage resources that normally would be wasted, and assigns these resources to collaborative research projects [7]. The most widely used and successful middleware for this task is BOINC [4]. Currently there are more than one million volunteers collaborating worldwide on BOINC projects, donating their resources and CPU cycles<sup>3</sup>.

In particular, the BOINC model is now also used with some EAs, achieving interesting results when evaluated based on system robustness and fault tolerance [6]. Therefore, there is a current interest in further exploiting such models in EC projects and applications, that normally incur high computational costs and need to process and store large amounts of data.

BOINC was developed two decades ago with the *seti@home* project, providing all of the basic software tools required to deploy a volunteer computing system or project. It includes a BOINC server, that distributes the computational load among a set of volunteer client machines that are connected to a specific project. Moreover, it also includes a BOINC client, that is in charge of communicating with the server and executing the computational tasks assigned to the client machine by the server. While the general model is simple and straightforward, particularly compared to other GRID-based models, to use the basic BOINC tools on a specific project, the source code must first be ported to operate within the BOINC environment through the provided API. This can be a big limitation in some instances, since not all programming languages are supported by the BOINC API, sometimes requiring a large scale re-coding project that will require some IT management experience in distributed computing environments. Moreover, in some cases it simply is not possible to port a particular system that has strong dependencies with specialized software or simulators; for instance, one can imagine such a case for an application that requires Matlab or Mathematica.

To overcome these and other possible problems, some researchers have developed partial solutions based on virtualization, in a sense encapsulating projects before distributing them over a BOINC network of clients [6]. However, this model imposes limits over the distributed computing paradigm: it is only possible to execute a single application multiple times. For instance, it is not possible within a traditional EA, to distribute individual candidate solutions for fitness evaluation over BOINC or perform load balancing, when fitness evaluations are not homogeneous, a common problem in genetic programming (GP) systems for example.

<sup>3</sup> <http://boinc.berkeley.edu/>

In summary, while the technology is currently widely available, and some partial simplified solutions have also been developed, there is still a lack of a complete system capable of integrating the BOINC middleware within an evolutionary search. This paper presents such an alternative, the first attempt at integrating the BOINC advantages with an EC toolkit.

## 2.1 ECJ & BOINC

ECJ is a very popular and well-known tool used by researchers within the Computational Intelligence community, with many works publishing results that either use ECJ or that are compared with the tools provide by ECJ. However, there is no off-the-shelf way of distributing an ECJ application through BOINC.

While ECJ does offer some parallel computing models to exploit several processors or cores, it is not possible to launch an ECJ algorithm within a distributed BOINC environment. This work focuses on one of the simplest possible models, where individuals from the evolving population are distributed by BOINC to be evaluated by a set of connected clients. To achieve this, ECJ must be modified accordingly, which is the contribution presented in this work.

Another goal is to develop software tools that simplify the use of such a distributed model for an EA. Therefore, a BOINC server is integrated within ECJ, configured to distribute work units composed of sub-sets of individuals from an EA developed with ECJ, to clients that are configured and connected with the server. Moreover, to further simplify the process for the user, all of this was encapsulated within dedicated virtual machines, allowing researchers to simply download the virtual machine server and run it on the server machine. Then, the user must define its fitness function and other problem-specific details of their ECJ project, and include these functions in the sever before it is compiled, like any other ECJ applications. When the application is finally executed, the work units and all of the BOINC middleware are configured and launched automatically. At the other end, client machines must download the BOINC clients, and connect to the BOINC server to participate with a project. The client-side virtual machine is configured to automatically connect with the server-side virtual machine, without requiring any additional administrative or configuration tasks by the user. In what follows, the proposed system is described in detail, referred to as the ECJ+BOINC system.

## 3 Proposal

As stated before, the goal of this work is to develop the necessary software tools that allows a researcher to exploit the BOINC paradigm of distributed volunteer computing with EAs written with the ECJ toolkit. In what follows, the underlying technical implementation of the proposed system is described, focusing on each of the technologies that were employed.



### 3.1 Virtualization

To start, let's consider the basic requirements to be able to incorporate BOINC within ECJ. As stated before, there are two main pieces of software that need to be included: (a) a BOINC server that distributes the work units; and (b) the BOINC clients that perform the requested computations. Both software modules, the server and clients, are distributed over a virtualization layer, in this case VirtualBox<sup>4</sup>. This virtualization layer can be used by clients running on different operating systems, it is simple to setup and install.

The virtual machines are configured with Linux, along with all of the necessary software tools to administrate the server and client tasks independently. Moreover, these virtual machines also have all the required software to run ECJ applications, both on the server and the client.

### 3.2 ECJ

ECJ is probably the most popular EC tool, developed by Sean Luke et al. ECJ is written in Java and includes a large set of features and state-of-the-art methods [5]. It was designed to be flexible and highly configurable, with almost all of its parameters determined at run time from a specified set of configuration text files.

The structure of ECJ includes a large set of specialized classes, among them the most relevant to the current work is the *Evolve* class. Among others, the *Evolve* class includes the *main()* method, which is executed to start all evolutionary algorithms implemented in ECJ. Another noteworthy class is *Evaluator*, to which the task of evaluating an evolving population is assigned, of particular importance for a system that distributes population evaluation through BOINC. As stated above, the client machines will evaluate subsets of the evolving population, hence each client needs to employ an *Evaluator* object, that returns the necessary output to the server, to guide the search process being executed by an *Evolve* object on the server.

### 3.3 BOINC

BOINC is written and developed in C/C++, thus it most easily integrates with systems written in the same programming language. Since not all projects that are deployed with BOINC are written in C/C++, it is necessary to port the application or employ specialized *Wrapper* objects. Through a wrapper object or function, it is possible to run non C/C++ code within a BOINC client. On the other hand, the BOINC architecture is based on the client-server model that allows for multiple *work units* to be deployed simultaneously by the server, which are then processed on the BOINC clients, while the server waits for the results returned from each client. The server can validate the results returned by the clients, and determines if the results are accepted or rejected, in which case the corresponding work unit can be sent to another client.

<sup>4</sup> <https://www.virtualbox.org/>

### 3.4 ECJ+BOINC

To integrate BOINC with ECJ, it is necessary to modify the normal execution sequence of a BOINC project, since part of the ECJ application is kept on the server side, and only population evaluation is distributed over BOINC. Normally, in a BOINC project the server only generates and distributes work units, it then waits idle while the clients perform all of the data processing associated with a project. On the other hand, in this work the server is in charge of executing an EA, and only during fitness evaluation the EA is paused on the server side. At this moment, the server generates a series of work units that are sent to the BOINC clients, each work unit contains a subset of individuals from the population that must be evaluated on the client side. Once the clients compute the corresponding fitness values, these are returned to the BOINC server, which then continues with the normal EA execution.

This program flow also requires new code for the ECJ classes used during evolution. First, a new class has been derived from the ECJ *SimpleEvaluator.java* class, by adding a Shell Script that generates a single work unit, intended for a single BOINC client; this is the simplest possible case. The server then waits idle, until the results are uploaded onto the *upload* directory of the corresponding project.

Once program execution is halted on the server, the current state of the project run is saved using the *checkpoint.java* class. This class is used to be able to send the current state of the run to the client, along with the population that will be evaluated. The client evaluates the population, and updates the state of the current run, which is returned to the server so the search can continue. Figure 1 presents the basic program flow of the ECJ+BOINC system.

Based on this simple single client version, the next step is to extend the ECJ+BOINC system to run with several clients, each assigned the responsibility of evaluating a subset of individuals from the population.

Similarly to the simpler case, once ECJ execution is paused on the server, the requested number of work units are generated for each client; note that not all clients will necessarily evaluate the same number of individuals. Once each client terminates its evaluation process and the results are returned to the server, the partial results of each client must be integrated into a single state file that updates the state of the EA running on the server. This requires the following modifications to the ECJ library:

- Class *EvaluatePopulation*: before the population is evaluated, the state of the EA is saved.
- Class *evalPopChunk*: calls a process that creates the requested work units and distributes them to the clients.

Once the clients have returned their results normal ECJ execution continues on the server, as depicted graphically in Figure 2, where ECJ+BOINC is running with  $x$  number of clients.



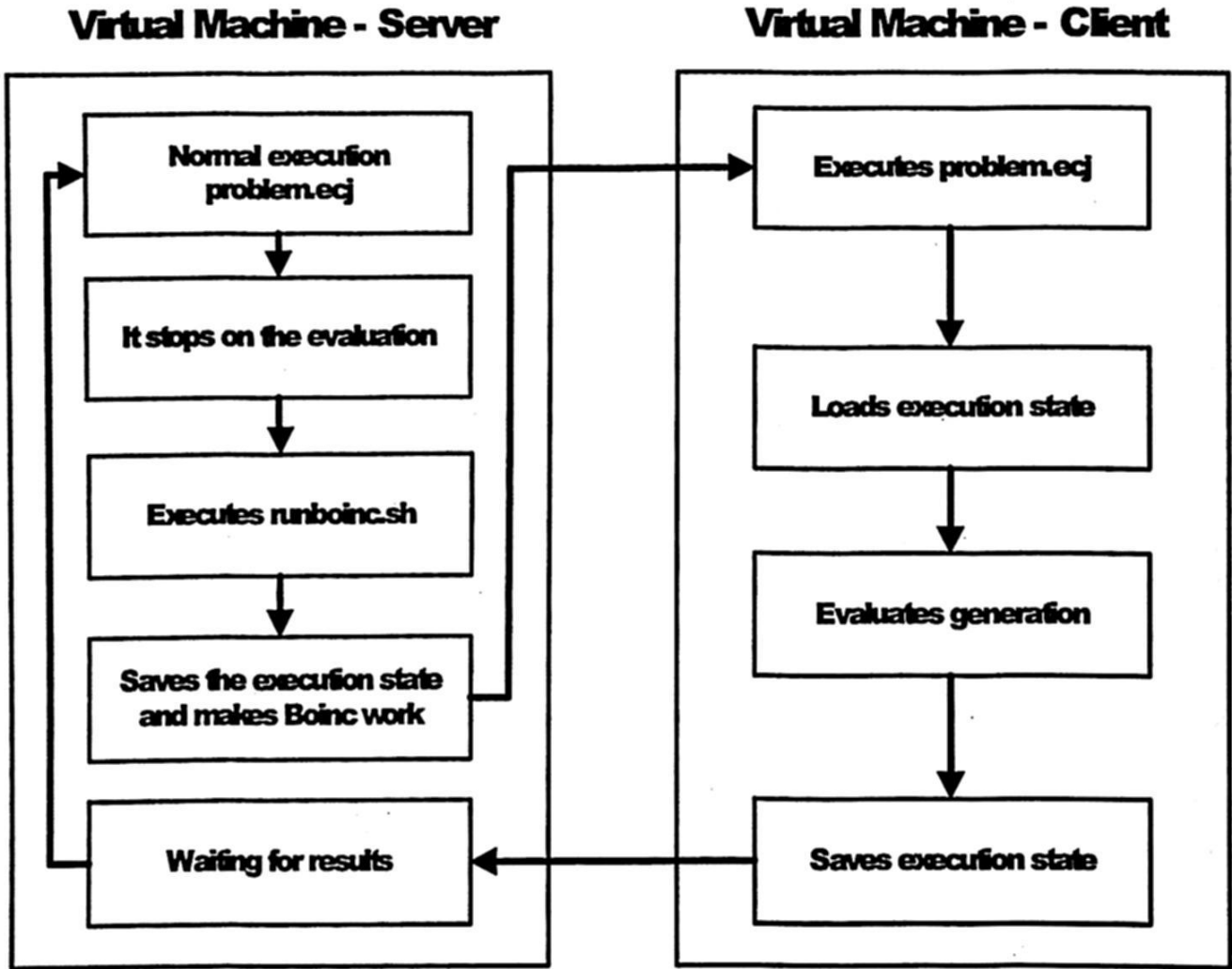


Fig. 1. Program flow of an ECJ+BOINC system running with a single client.

## 4 Experiments and Results

To test the proposed ECJ+BOINC system, this work uses a typical GP benchmark problem with a high computational cost, the Even Parity problem [3]. This problem is widely used for benchmarking purposes, however instead of using the common 5-bit problem, the number of bits was set to 30, increasing the difficulty of the problem and forcing the system to incur higher computational costs during fitness evaluation. In this scenario, a normal ECJ GP run required approximately 20 minutes to evaluate the fitness of 10 individuals on a standard PC.

Besides the small number of individuals used during the run, 10, and the relatively small number of generations set to 25, all other parameters were set at the default values provided by ECJ. Indeed, the goal of the tests are not to evaluate if the GP algorithm provided by ECJ can solve this problem, this is irrelevant to the tests, the goal is to illustrate the performance gains provided by BOINC-based distribution of fitness evaluations. To test ECJ+BOINC, five clients were connected to the server machine, on which the virtual machine client and BOINC client were installed. During fitness evaluation, a single individual

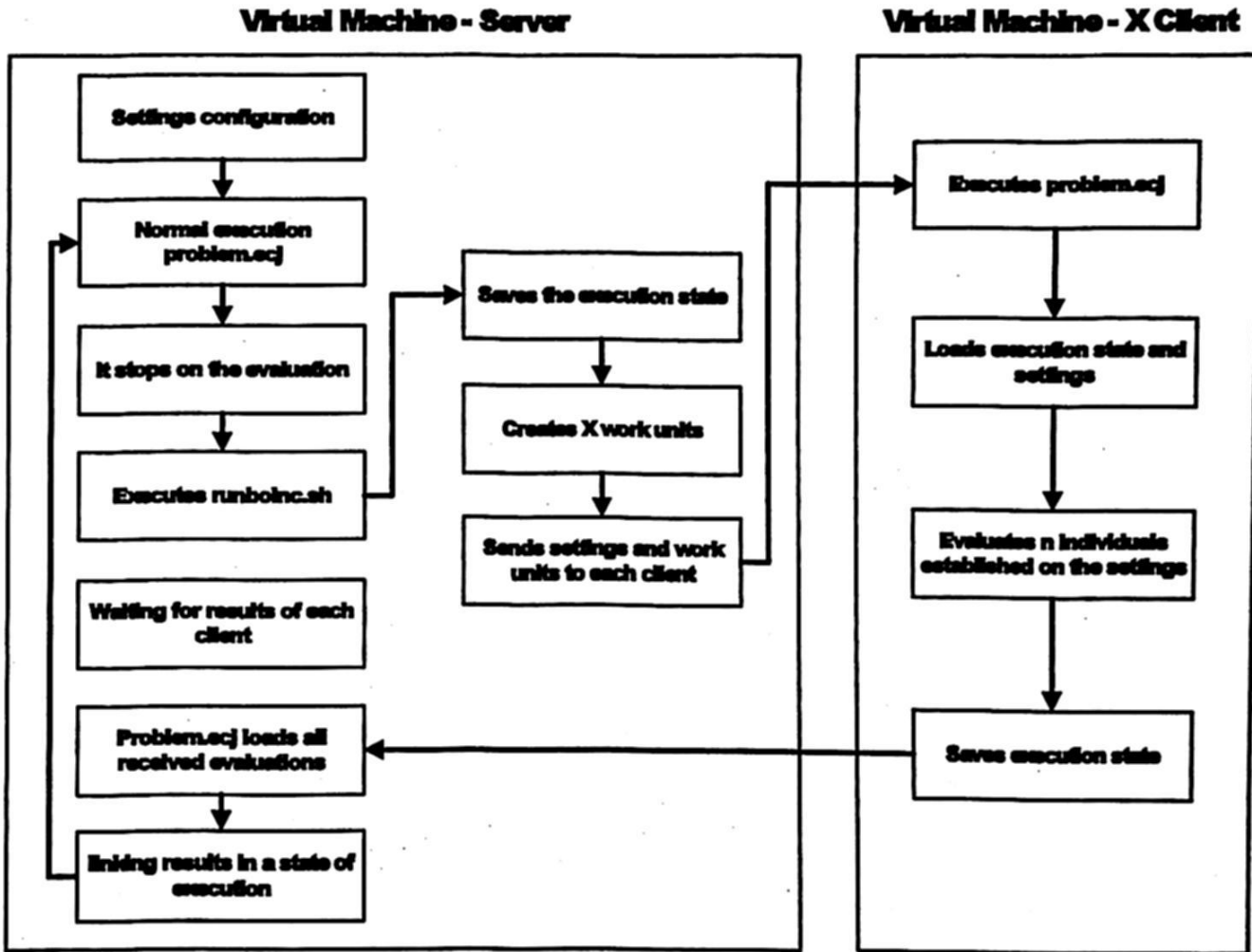


Fig. 2. Program flow of an ECJ+BOINC project running with  $x$  number of clients.

is distributed to each client at each generation. Ten runs were executed, using standard ECJ and the ECJ+BOINC system.

Figure 3 summarizes the results of these experiments in boxplots comparing the ECJ system with the ECJ+BOINC configuration. First, Figure 3a shows that there is no difference between both systems based on the quality of the results found. This is desired, the underlying BOINC distributed model is not designed to modify or alter the search dynamics, its sole purpose is to reduce computation time by distributing costly fitness evaluations. Similarly, the fact that ECJ+BOINC does not effect search dynamics is confirmed by Figure 3b, that shows the average program size evolved in each GP run. Again, ECJ and ECJ+BOINC basically produce the same results, which means that the BOINC-related modifications did not inadvertently modify the search process. Finally, Figure 3c summarizes the results related to the run time of each system. As expected, ECJ+BOINC clearly reduces the computation time required to evolve solutions for the benchmark problem, basically reducing the median run time in half.

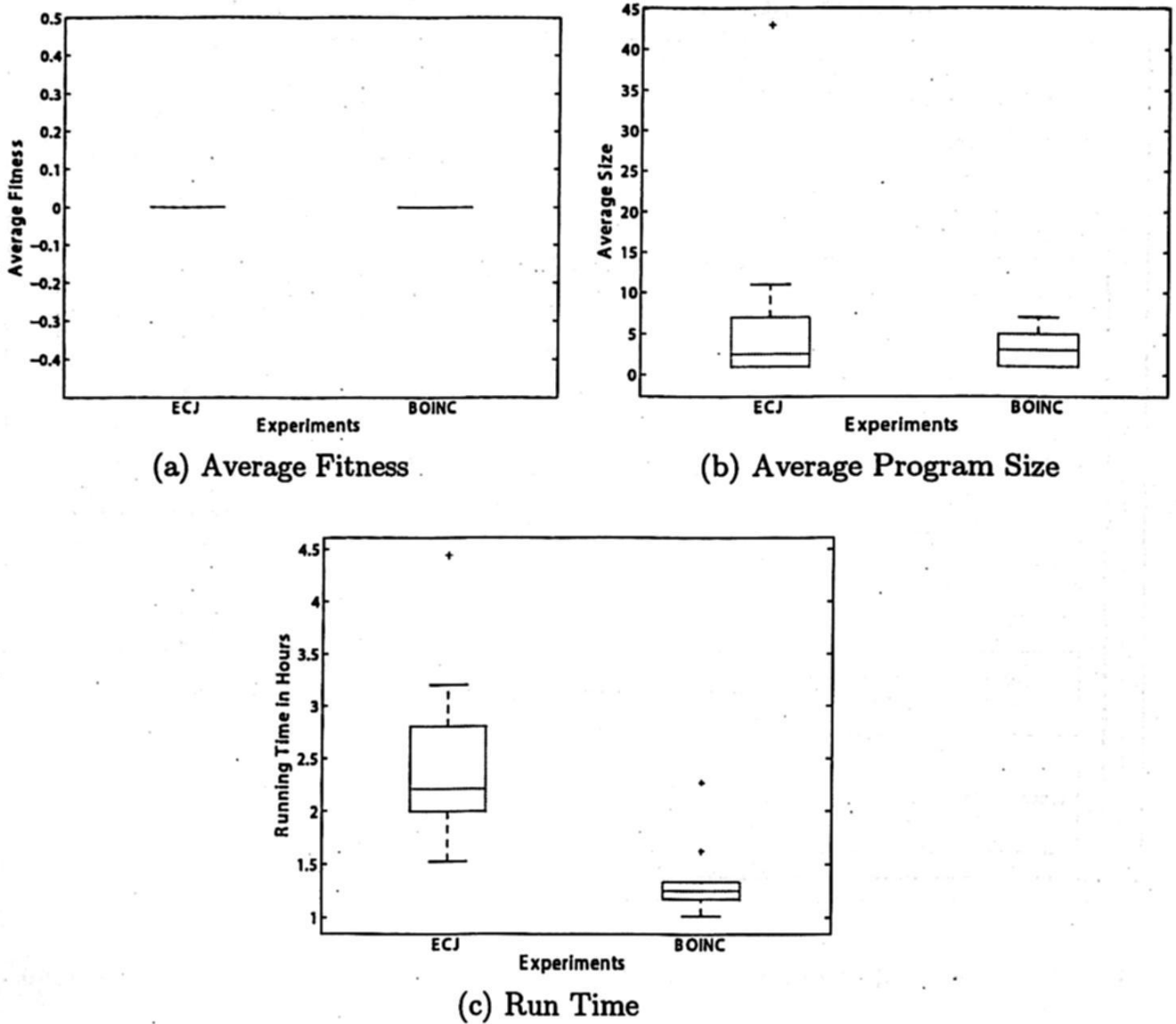


Fig. 3. Summary of the experimental comparison between ECJ and ECJ+BOINC.

## 5 Summary, Conclusions and Future Work

Overall, distributed and parallel computing has a large impact every research area where large amounts of data need to be quickly processed. One way to obtain the computational power and storage resources required to solve such problems, is to exploit the volunteer computing model popularized by the BOINC middleware in many real-world projects. In general, however, BOINC has still not been fully integrated within EC systems, that almost always need to improve efficiency and reduce run times.

This work represents the first attempt to integrate BOINC into a popular and widely used EC tool, the well-known ECJ library. In this work, BOINC is used to distribute the population over a set of connected clients, basically parallelizing fitness evaluation during an ECJ run, normally the most severe bottleneck in an EA. Moreover, this was done in such a way so as to make the entire process transparent for an ECJ user, who has to do little additional configurations or setup compared to a basic ECJ run. This is achieved by deploying

the BOINC clients and server within specialized virtual machines, something that is automatically accomplished by the ECJ+BOINC system.

Results show that the proposed solution integrates nicely with ECJ, without affecting the search process in any way, only producing a substantial improvement in run time on a benchmark problem for GP. While these initial results are encouraging, future work is still required. For instance, future research should focus on employing the ECJ+BOINC system on a real-world problem, where fitness evaluation is more costly, for instance with evolutionary robotics, where fitness depends on a costly simulation process [9]. Moreover, in such a case the use of a virtual machine will be even more beneficial, easily deploying a large amount of fitness-specific software libraries over a large set of distributed clients. Finally, it will be of interest to explore how a distributed system, such the one presented in this paper, can be used to enhance the search dynamics of an EA [10, 11]. These algorithms could produce interesting epiphenomenons, such as reducing bloat in a GP-based search [12].

## 6 Acknowledgments

Funding for this work provided by Ministerio de Ciencia e Innovación, through the ANYSELF project (TIN2011-28627-C04), Universidad de Extremadura and Gobierno de Extremadura Consejería de Economía Comercio e Innovación and FEDER project No. GRU10029. Additionally support provided by CONACYT (Mexico) Basic Science Research Project No. 178323, DGEST (Mexico) Research Projects No.5149.13-P and TIJ-ING-2012-110, and IRSES project ACoBSEC from the European Commission. Fifth author supported by CONACYT scholarship No. 302526.

## References

1. Polycarpus, M.: Computational Intelligence in the Undergraduate Curriculum. *Computational Intelligence Magazine*, 8:2 3 (2013).
2. De Jong, K.: *Evolutionary Computation: A Unified Approach*. The MIT Press. (2001)
3. Koza, J.R.: *Genetic Programming*. MIT Press. 1992.
4. Anderson, David P.: BOINC: A System for Public-Resource Computing and Storage, *Proceedings of the 5th IEEE/ACM International Workshop on Grid Computing. GRID '04* 4–10 (2004).
5. David R. White: Software review: the ECJ toolkit, *Genetic Programming and Evolvable Machines* 13:1 65–67 (2012)
6. F. Fernández de Vega, G. Olague, L. Trujillo, D. Lombría: Customizable execution environments for evolutionary computation using BOINC + virtualization, *Natural Computing* 12:2 163–177 (2013)
7. David P. Anderson: Volunteer computing: the ultimate cloud, *ACM Crossroads* 16:3 7–10 (2010).
8. Chávez, Francisco and Guisado, Jose Luís and Lombrana, D and Fernández, Francisco: Una herramienta de programación genética paralela que aprovecha recursos públicos de computación, *MAEB* 167–173 (2007)



9. L. Trujillo, G. Olague, E. Lutton, F. Fernández de Vega, L. Dozal and E. Clemente: Speciation in Behavioral Space for Evolutionary Robotics. *Journal of Intelligent & Robotic Systems* 64:34, 323–351 (2011)
10. M. García-Valdez, L. Trujillo, F. Fernández de Vega, J.J. Merelo Guervás and G. Olague. EvoSpace: a distributed evolutionary platform based on the tuple space model. In *Proceedings of the 16th European conference on Applications of Evolutionary Computation (EvoApplications'13)*, Anna I. Esparcia-Alcázar (Ed.). Springer-Verlag, Berlin, Heidelberg, 499–508 (2013).
11. M. Garcia-Valdez, J.J. Merelo, L. Trujillo, A. Mancilla and F. Fernandez-de-Vega: Is there a free lunch for cloud-based evolutionary algorithms? *IEEE Congress on Evolutionary Computation 2013*, Cancun, Mexico, 20 - 23 June, 2013. IEEE Press, 2871–2879 (2013).
12. Robin Harper: Spatial co-evolution: quicker, fitter and less bloated. In *Proceedings of the fourteenth international conference on Genetic and evolutionary computation conference (GECCO '12)*, Terence Soule (Ed.). ACM, New York, NY, USA, 759–766 (2012).

# GPU implementation of nonlinear anisotropic diffusion for medical image enhancement

Fernando Villalbazo, Juan J. Tapia, and Julio C. Rolón

CITEDI Research Center, Instituto Politécnico Nacional, Avenida del Parque 1310,  
Mesa de Otay, Tijuana, Baja California, México, 22510

`fvillabazo@citedi.mx`

`{jtapiaa,jcrolon}@ipn.mx`

*Paper received on 11/29/13, Accepted on 01/19/14.*

**Abstract.** In this work, the implementation of the diffusion process in a graphics processing unit (GPU) with application to medical image enhancement is presented. The method is implemented in the many-core architecture of the GPU and uses its resources of texture and shared memory. The diffusion equation is used to perform the enhancement of medical images. Due to the characteristics of the diffusion algorithm, it is possible to take advantage of the resources of the processor and reduce the image processing time.

A gain of up to 20 times in the execution time of the GPU implementation of anisotropic diffusion algorithm has been achieved, with respect to the implementation in a general purpose processor. An important noise reduction has also been accomplished by using the diffusion parameters found with the proposed fast search algorithm. A gain of up to three times in the GPU implementation of diffusion filtering execution time has been achieved using the fast search method instead of exhaustive search.

**Keywords:** GPU, Image enhancement, Nonlinear diffusion, Parallel algorithm

## 1 Introduction

In recent years, applications that combine general purpose processors with graphics processors have increased significantly, due to the high processing power of the GPU, their ability to perform massively parallel processing and their scaling performance that is achieved by increasing the number of graphics processors. For these reasons, a GPU is a suitable cost-effective solution to achieve high performance computing for general purpose applications. An important field of application is the medical image processing. The medical image applications are time consuming, because of the amount of data to be processed and the requirement of high spatial resolution [1].

Medical images contain certain amount of noise that is inherent to the acquisition process and affects the image quality. Noise reduction without loss of

desired information in the image is an important challenge in the area of digital image processing. The diffusion process is a method that applies a selective smoothing filter and performs noise reduction inside the homogeneous regions. The noise is removed while the sharpness of the edges is preserved. The numerical algorithm is implemented with an explicit finite difference method, therefore there are not data dependencies and properly fits the architecture of GPUs, enabling the parallelization down to the pixel level.

The diffusion process has been widely used in image denoising and applied to several types of medical images [2–7]. In [8], a form to estimate the stopping time  $T$  for the diffusion is proposed.  $T$  is calculated by correlating the original and the enhanced images. In [9], a multigrid solver is proposed to solve the anisotropic diffusion equation and to estimate the diffusion parameters. However, the parameter localization through these methods is based on a estimation, not on the evaluation of the diffusion equation. The main objective of this work is the effective utilization of a GPU to reduce the processing time of the medical image enhancement through the diffusion process and the estimation of the diffusion parameters that ensure a minimal noise in the image, using a rapid search optimization method.

Section 2 summarizes the nonlinear diffusion method; section 3 describes the parallelization strategy we have followed; section 4 presents the experimental results for five synthetic MRI (Magnetic Resonance Imaging) images; finally, some conclusions and comments about future work are outlined.

## 2 Diffusion filtering

The diffusion process may be considered as a dispersion phenomena in which concentration of either mass, heat, or any other physical variable of interest moves from an area of high concentration to an area of low concentration, until the equilibrium is reached. The diffusion equation is

$$\frac{\partial u}{\partial t} = \nabla \cdot (\alpha \nabla u), \quad (1)$$

where  $\nabla \cdot$  is the divergence operator and  $\alpha$  is a diffusion parameter which defines the diffusion intensity. If  $\alpha$  is constant in the medium, a linear diffusion is produced. If  $\alpha$  is a function of some parameter of the medium, a nonlinear diffusion is produced.

The result of the diffusion when applied to images is a family  $u(x, y, t)$  at different scales  $t$ , where

$$u(x, y, t + 1) = G_\sigma(x, y) * u(x, y, t), \quad (2)$$

here  $G_\sigma(x, y)$  is a Gaussian filter with variance  $\sigma$  [11].

In image processing, the diffusion is the process through which clusters of high-energy pixels within an image are dispersed, which results in the softening of the image. Since the process evolves in time, at  $t = 0$ , the original image

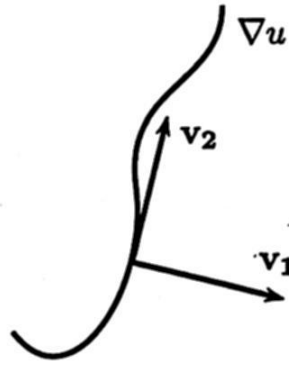


Fig. 1. Orientation of eigenvectors  $\mathbf{v}_1$  and  $\mathbf{v}_2$

$u(x, y)$  is  $u(x, y, 0) = u^0$ . The Gaussian scale is  $0 \leq t \leq T$ , where  $T$  is the total number of scales of the process. Each new scale only depends on the image produced at the previous scale. The image at  $u(x, y, t+1)$  is a smoother version of the image at  $u(x, y, t)$ . If  $T \gg 0$ , all the image pixels will have the same gray level.

## 2.1 Nonlinear Diffusion

Introduced in [12], the nonlinear isotropic diffusion process makes use of Eq. (1) where the diffusion parameter  $\alpha$  is defined as a function of the contour intensity. The contour estimator used is the image gradient  $\nabla u$ .

In the nonlinear isotropic diffusion, the parameter  $\alpha$  is a scalar value

$$\alpha(x, y, t) = g(\|\nabla u(x, y, t)\|), \quad (3)$$

where  $g$  is known as the diffusivity function. Conditions are imposed on  $g$  such that whenever the gradient is high, e.g. a contour in the image, diffusion is not applied; conversely, when gradient is low, e.g. an homogeneous region, the diffusion is maximized:

$$\lim_{\|\nabla u\| \rightarrow \infty} g(\|\nabla u\|) \rightarrow 0, \quad \lim_{\|\nabla u\| \rightarrow 0} g(\|\nabla u\|) \rightarrow 1. \quad (4)$$

In the nonlinear anisotropic diffusion, the smoothing depends on the gradient intensity and its direction [13]. To obtain information about the pixel neighbourhood, a structure tensor  $\mathbf{J}$  is used [14], and it is calculated from the image gradient as

$$\mathbf{J}(\nabla u) = \nabla u \nabla u^T = \begin{bmatrix} J_{11} & J_{12} \\ J_{12} & J_{22} \end{bmatrix}. \quad (5)$$

The direction of the diffusion propagation is indicated by the eigenvectors  $\mathbf{v}_1$  and  $\mathbf{v}_2$  of the structure tensor  $\mathbf{J}$ . The orientation of eigenvectors  $\mathbf{v}_1 \parallel \nabla u$  and  $\mathbf{v}_2 \perp \nabla u$  with respect to the image gradient is illustrated in Fig. 1. The diffusion parameter  $\alpha$  from the Eq. (1) is defined as the diffusion tensor

$$\mathbf{D} = \begin{bmatrix} a & b \\ b & c \end{bmatrix} = [\mathbf{v}_1 \ \mathbf{v}_2] \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \begin{bmatrix} \mathbf{v}_1 \\ \mathbf{v}_2 \end{bmatrix}, \quad (6)$$



where the parameters  $\lambda_1$  and  $\lambda_2$  define the diffusion intensity along the direction of the eigenvectors  $\mathbf{v}_1$  and  $\mathbf{v}_2$ , respectively.

The approach used in this work is the edge enhancing diffusion, which performs a controlled diffusion along the direction corresponding to the contour gradient, and a maximum diffusion in the direction normal to the contour gradient. The  $\lambda_1$  parameter in the edge enhancing approach, is defined by a diffusivity function  $g$ . In the edge enhancing diffusion, the values of  $\lambda_1$  and  $\lambda_2$  are

$$\begin{aligned}\lambda_1 &= g(\|\nabla u\|) \\ \lambda_2 &= 1\end{aligned}\quad (7)$$

The vector  $\mathbf{v}_1$  can be represented as  $\mathbf{v}_1 = [\cos \theta, \sin \theta]^T$ , where  $\theta$  is defined as the angle of inclination of the image gradient [15]. The edge orientation can be computed as

$$\theta = \arctan\left(\frac{2J_{12}}{J_{22} - J_{11}}\right) + \frac{\pi}{2}. \quad (8)$$

The vectors  $\mathbf{v}_1$  and  $\mathbf{v}_2$  are orthogonal, so the vector  $\mathbf{v}_2$  can be computed as

$$\mathbf{v}_2 = [-\sin \theta, \cos \theta]^T. \quad (9)$$

## 2.2 Discrete solution of the diffusion equation

The numerical implementation for the solution of the diffusion equation in Eq. (1) is obtained by finite differences centered in the space, and forward differences in the time discretization, which leads to an explicit scheme, where the image  $u^{t+1}$  is obtained from  $u^t$  with

$$u^{t+1} = u^t + \tau(\nabla \cdot (\alpha \nabla u)), \quad (10)$$

where  $\tau$  is a parameter that keeps the numerical solution stable. In image processing, the parameter  $\tau$  is chosen in the range of  $0 < \tau < 1/4$ . Each scale  $t$  represents one iteration of the numerical algorithm.

For the nonlinear anisotropic diffusion, the spatial derivative  $\nabla \cdot (\mathbf{D} \nabla u)$  is approximated by the discretized expression  $\mathbf{A}u^t$

$$\nabla \cdot (\mathbf{D} \nabla u) \approx \mathbf{A}u^t \quad (11)$$

The expression  $\mathbf{A}u^t$  is represented by a mask that is function of the diffusion tensor  $\mathbf{D}$  [14, 16]

$$\begin{aligned}A_{i,j}u_{i,j} &= \frac{c_{i,j+1} + c_{i,j}}{2}u_{i,j+1} - \frac{b_{i-1,j} + b_{i,j+1}}{4}u_{i-1,j+1} + \frac{b_{i+1,j} + b_{i,j+1}}{4}u_{i+1,j+1} \\ &\quad - \frac{a_{i-1,j} + 2a_{i,j} + a_{i+1,j} + c_{i-1,j} + 2c_{i,j} + c_{i+1,j}}{2}u_{i,j} \\ &\quad + \frac{a_{i-1,j} + a_{i,j}}{2}u_{i-1,j} + \frac{a_{i+1,j} + a_{i,j}}{2}u_{i+1,j} + \frac{c_{i,j-1} + c_{i,j}}{2}u_{i,j-1} \\ &\quad - \frac{b_{i+1,j} + b_{i,j-1}}{4}u_{i+1,j-1}.\end{aligned}\quad (12)$$

**Algorithm 1** Pseudocode for the anisotropic process

---

```

u = noisy image
for t = 1 to T do
    Diff = 0
    compute J
    compute  $\lambda_1$ , set  $\lambda_2$ 
    compute  $\theta$ 
    compute  $v_1, v_2$ 
    compute D
    for i = 1 to M do
        for j = 1 to N do
            compute  $A_{i,j}$ 
            compute  $Diff_{i,j} = Diff_{i,j} + A_{i,j}$ 
        end for
    end for
     $u = u + \tau \cdot Diff$ 
end for

```

---

The sequential iterative process to obtain the enhanced image  $u$  from the noisy image, is shown in the Algorithm 1. The image in the scale  $t$  is computed from the data of the previous scale  $t - 1$  and then overwritten over the same memory space, therefore only the images in  $t$  and  $t - 1$  have to be stored in the entire process. The images are updated in each scale.

### 3 Parallel implementation of the nonlinear diffusion algorithm

The execution time of the anisotropic diffusion algorithm can be reduced by through a parallel implementation on the GPU and the efficient handling of the different types of memories. Parallelization of the complete diffusion process algorithm is performed at the pixel level, exploiting the fine grain level parallelism allowed by the GPU architecture. The CUDA programming model has been used instead of other models to obtain the best performance in GPU applications [17].

A graphics processor consists of a set of multiprocessors, where each of these has several cores that compute the same operation on different data, according to the SPMD (Single Program, Multiple Data) model [18, 19]. The work is distributed equally between the multiprocessors, in order to avoid overload or lack of work on some multiprocessors.

The management of different types of GPU memories enables the possibility to improve the algorithm performance. The global memory may be addressed by all the active threads of the GPU, with the disadvantage that when many threads need to access memory at the same time, some latency may arise that limits the performance of the GPU. The texture, constant and shared memories are on-chip memories, consequently, the access time for these memories is lower than the global memory.

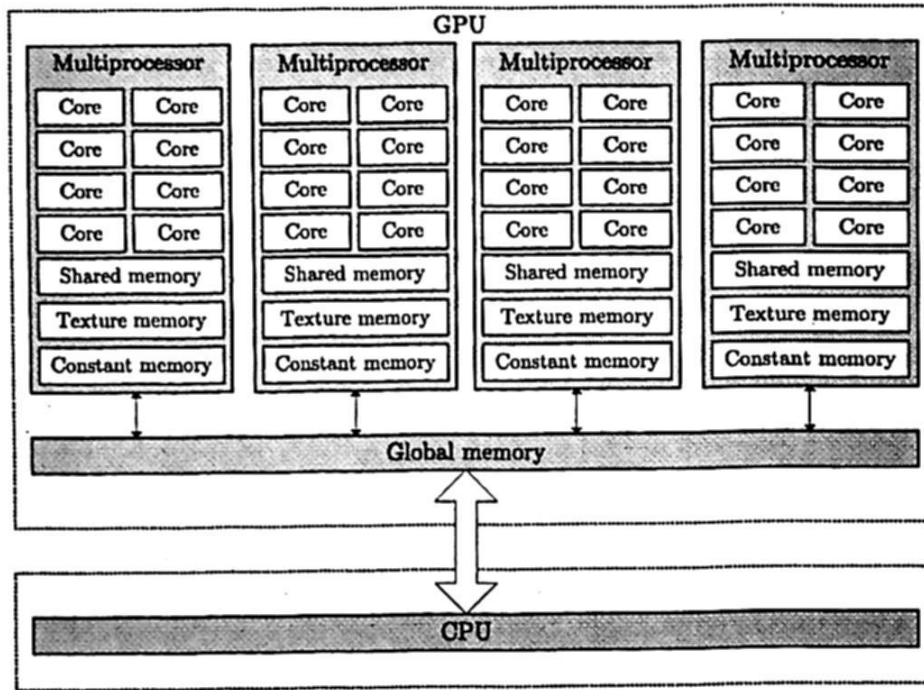


Fig. 2. Functional subdivision of the memory of a GPU device.

To transfer data from the CPU to the GPU on-chip memories, is necessary to transfer the data first to the GPU global memory and next to the corresponding memory locations. Texture memory is used to accelerate data reading from global memory, taking advantage of the two-dimensional image structure. Once written, the contents of these memories may only be transferred to the corresponding memory location of the actual processing unit, it is not possible for the processing unit to update them directly from the multi-thread cores. Only the host may instruct the GPU to transfer the contents of the global memory to the texture memory of the GPU. The GPU architecture and the distribution of the GPU memories into global, texture, constant and shared memories is shown in Fig. 2.

Synchronization points are added in the algorithm to prevent errors at the thread level. According to the explicit scheme, the image  $u_{i,j}^{t+1}$  is obtained directly from  $u_{i,j}^t$  and overwrites the image data on the global memory. In this process, image data are always kept inside the GPU and texture memory is used to store the images at time  $u_{i,j}^{t+1}$ . The gradient  $\nabla u$ , the eigenvectors  $\mathbf{v}_1$  and  $\mathbf{v}_2$ , the diffusion parameters  $\lambda_1$ ,  $\lambda_2$  and the tensors  $\mathbf{J}$ ,  $\mathbf{D}$  are calculated at the pixel level using the local memory of each thread.

The texture memory is read-only, consequently, the result of the nonlinear diffusion for each pixel,  $u_{i,j}^{t+1}$  overwrites the value  $u_{i,j}^t$  in the global memory. To begin a new iteration, the CPU instructs to the GPU to update the texture memory with the image data of the global memory.

Each function that is executed in parallel by the graphics processor is called kernel. The main program is executed in the CPU, and calls two kernel functions. Both kernels are two-dimensional, of size  $[N/blocks, N/blocks]$ , where *blocks* is the number of threads that can be executed by a block. Each thread computes

**Algorithm 2** Pseudocode for the parallel anisotropic process

---

```

CPU memory allocation
GPU memory allocation
u = noisy image
MxN = image size
copy u from CPU to GPU
for  $t = 1$  to  $T$  do
    kernel1 <<< (N/blocks, N/blocks), (blocks, blocks) >>> (u, D)
    update D in texture memory
    kernel2 <<< (N/blocks, N/blocks), (blocks, blocks) >>> (u, D)
    update u in texture memory
    copy u from GPU to CPU
    free CPU memory
    free GPU memory
end for
__global__ kernel1 (u, D)
{
     $x = blockIdx.x * blockDim.x + threadIdx.x$ 
     $y = blockIdx.y * blockDim.y + threadIdx.y$ 
     $id = y * M + x$ 
    compute J
    compute  $\lambda_1$ , set  $\lambda_2$ 
    compute  $\theta$ 
    compute  $v_1, v_1$ 
    compute D[id]
}
__global__ kernel2 (u, D)
{
     $x = blockIdx.x * blockDim.x + threadIdx.x$ 
     $y = blockIdx.y * blockDim.y + threadIdx.y$ 
     $id = y * M + x$ 
    compute Diff
    compute new u[id]
}

```

---

its identifier  $id$ , from their position in the kernel. The first kernel computes  $J$ ,  $\lambda_1$ ,  $\lambda_2$ ,  $\theta$ ,  $v_1$  and  $v_2$ .

Each of these terms are stored in the local memory of the GPU. The diffusion tensor  $D$  is calculated and stored in the global memory of the GPU, that is linked to the texture memory. The CPU must indicate to the GPU to update the diffusion tensors in texture memory from the global memory. The second kernel function calculates the amount of diffusion  $Diff$  by reading the diffusion tensors and the image in the previous iteration from the texture memory. The new image is stored in global memory and then updated by the CPU for the next iteration. The parallel implementation of the anisotropic diffusion process is summarized in the Algorithm 2.

## 4 Experimental Results

The parallel algorithm for the nonlinear anisotropic diffusion was tested in a GPU with 336 cores (NVIDIA GeForce 460). All the calculations were made with single precision floating point arithmetic. Two implementations of the nonlinear anisotropic diffusion have been compared, through the computation of Eq. (10). The first implementation makes use of the global memory of the GPU, and the second uses the texture memory. In both cases, the execution time was measured using the GPU timer. The time measured includes the execution of the CPU and the GPU functions for the diffusion computation until the diffusion stopping time  $T$  is reached, it also includes the time spent in data transfers between both architectures, the texture memory allocation and memory updates. The experiment was conducted over 181 synthetic MRI images of  $181 \times 217$  pixels with 8 bits per pixel [20]. Noise added into the process of image capture is modelled by a Gaussian distribution with zero mean and variance  $\sigma = 0.002$ .

The diffusivity function  $g$  used in the experiment is [12]

$$g(\nabla u) = \frac{1}{1 + \left(\frac{\|\nabla u\|}{K}\right)^2}, \quad (13)$$

where  $K$  is a contrast parameter which controls the amount of diffusion the function exerts.

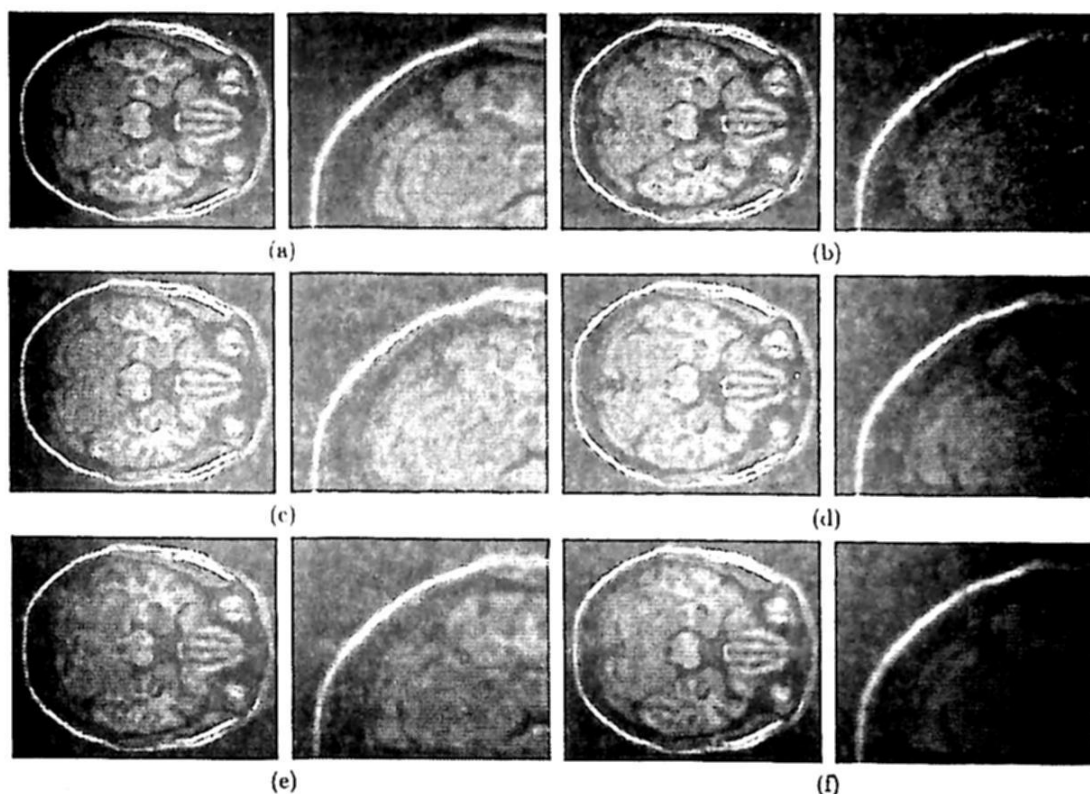
**Table 1.** Average execution time in msec for the nonlinear anisotropic diffusion algorithm.

$t$	1	2	3	4	5
$T_{CPU}$	29.20	42.20	55.80	64.90	77.70
$T_{global}$	0.84	0.95	1.15	1.38	1.59
$T_{tex}$	0.36	0.53	0.71	0.89	1.07
$T_{global}/T_{tex}$	2.35	1.77	1.63	1.56	1.49
$T_{CPU}/T_{tex}$	20.33	19.79	19.73	18.33	18.24

The average execution time of the algorithm implemented in the CPU and in the GPU is shown in Table 1. The term  $T_{CPU}$  identifies the execution time in a CPU,  $T_{global}$  denotes the global memory approach in the GPU, and  $T_{tex}$  identifies the case of texture memory in the GPU. A total of  $t = 5$  Gaussian scales were used. If more scales are used the error grows rapidly and the image becomes indistinguishable. A gain of up to 20 times in the algorithm execution time was achieved using the GPU architecture with the memory optimization, compared against the CPU implementation. In the GPU, the implementation with the memory optimization is 1.6 times faster than the implementation using global memory. A more detailed comparison between the CPU and the GPU is presented in [10], they use as metrics: execution time, occupancy and FLOPS.

The evolution of the nonlinear anisotropic diffusion process on an image for different scales is shown in Fig. 3, using the edge enhancing approach. It





**Fig. 3.** Nonlinear anisotropic diffusion process evolution. (a) Noiseless image, (b) Noisy image. Diffusion (c)  $t = 2$ , (d)  $t = 5$ , (e)  $t = 10$ , (f)  $t = 20$

is possible to see the smoothing effect of the diffusion filter and the gradual decreasing of noise in the images. If  $T \gg 0$ , an undesirable effect is generated in the image edges, because of the extreme diffusion that is applied to those areas. The diffusion parameters are  $\tau = 0.25$  and  $K = 20$ .

The quality of the enhanced image is measured by the mean squared error ( $MSE$ ), which is given by

$$MSE = \frac{1}{M \times N} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} (u_{i,j}^0 - u_{i,j}^t)^2, \quad (14)$$

where  $M \times N$  is the image size,  $u^0$  and  $u^t$  are the original and enhanced images. The  $MSE$  error behavior of the anisotropic diffusion for values of the contrast parameter  $K$ , from  $K = 0$  to 100 was calculated for the image of Fig. 3. The lowest  $MSE$  obtained is achieved with  $K = 43$  and  $t = 4$ , which is  $MSE = 27.33$ . The error is reduced in the edges, improving the visualization of the image. The skull area keeps details through the diffusion scales, but the area inside the brain is smoothed after several diffusion scales. The test image is shown in Fig. 3b. We have found experimentally that the  $MSE$  error always has only one minimum. The error measurement is performed to optimize the diffusion parameters that produce the lowest error. The  $MSE$  calculation is performed by reading the pixel values from the texture memory and saving partial summations on GPU shared memory, combined with a reduction algorithm [19, 21]. The operation  $u^0 - u^t$  is computed in parallel by the GPU and the summations of the difference are

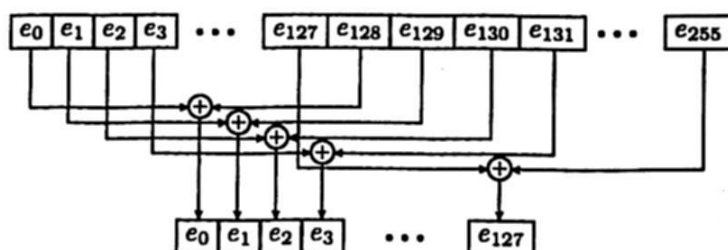


Fig. 4. First step of the reduction algorithm

stored in the shared memory of the corresponding multiprocessor. When all the summations are stored, these are used to compute the total error, which finally is transferred to the CPU. The summation is performed in parallel through the reduction algorithm, which allows to calculate the sum of the error with a reduced number of iterations. The error array is split in 2, and the elements  $i$  and  $i + N/2$  are summed, where  $i$  is the element position, and  $N$  is the size array. When the calculation of a partial sum is completed, the array is split in 2 again and the summation is performed again. The algorithm continues until there is just one element, with the value of the sum of all elements. The first step of the reduction algorithm, with an array of 256 elements is shown in Fig. 4.

#### 4.1 Fast search method to optimize the diffusion parameters

To optimize the diffusion parameters, an exhaustive search of the whole space should provide the answer. However, this method is computationally intensive. In order to reduce the computational load generated by the exhaustive search, a fast search method is implemented, which is a variant of the three-step search method [22]. The algorithm consists in dividing the search space in 4 regions of the same size. The parameters that correspond to the locations of the centroids of each region are evaluated. The region with the lower error is chosen to split in the next iteration. The algorithm continues until a certain number of iterations is reached. This method has the advantage of reducing the computational cost of the search method, when the search space is very large. However, as with any optimization method that approximates the solution, the parameters not always match the real minimum with the approximate one.

Table 2. Minimum  $MSE$  found by the exhaustive and fast search algorithm.

Image	$MSE_{ES}$	$MSE_{FS}$	$MSE_{dif}$	$(K, t)_{ES}$	$(K, t)_{FS}$
10	25.9229	25.9364	0.0134	(43,4)	(45,5)
40	27.3342	27.3367	0.0025	(43,4)	(41,5)
70	28.1007	28.1033	0.0026	(43,4)	(41,5)
100	27.0664	27.1540	0.0198	(43,4)	(41,5)
130	26.8103	26.9351	0.1247	(41,4)	(39,5)

The fast search method was applied to five MRI images, T1 weighted. The image modality was chosen due to its high contrast. The minima  $MSE$  calculated using the exhaustive and fast search method are defined by the  $MSE_{ES}$  and the  $MSE_{FS}$ , respectively. The difference of the  $MSE$  between both methods is  $MSE_{dif} < 1$ , which does not represent a significant difference in the resulting images.

Table 2 shows the diffusion parameters and the error ( $MSE$ ) that results from applying the exhaustive and fast search method on a search space  $1 \leq K \leq 100$  and  $1 \leq t \leq 30$ . Furthermore, the diffusion parameters ( $K, t$ ) localized with both methods, for comparison.

Average execution time of exhaustive search algorithm  $T_{ES}$  and fast search algorithm  $T_{FS}$  is:  $T_{ES} = 113.13$  msec. and  $T_{FS} = 31.19$  msec, for  $1 \leq K \leq 100$  and  $1 \leq t \leq 30$  for 181 MRI images. Processing the complete set of 181 images and calculating the error by using the fast search algorithm, a gain  $G = 3.62$  of the execution time was achieved in the algorithm. For the search interval proposed in the experiment, using the exhaustive search needs to perform 3000 iterations of the algorithm, whereas with the fast search algorithm with 140 iterations the minimum is reached.

## 5 Conclusions

It has been demonstrated that the use of the GPU in-chip memories significantly decreases the processing time of anisotropic diffusion algorithm and error computation, because the algorithm is naturally adapted to the GPU architecture. Texture memory aim to fast read pixel values and shared memory is useful to fast error calculation, through the parallel reduction algorithm. In the case of medical image datasets, this reduction may lead to an improvement in the global performance of the system. The error computation is performed totally in the GPU, in a way to minimize the data transference between both architectures and to improve the performance of the implementation.

A fast search of the parameters that minimize the  $MSE$  in the enhancement process was implemented. The results of the  $MSE$  obtained with the fast search algorithm are very close to those obtained with the exhaustive search algorithm. The differences are negligible.

## 6 Acknowledgments

This work has been partially supported by COFAA-IPN, and by grants IPN-SIP-20120606 and IPN-SIP-20130489.

## References

1. Pratz, G., Xing, L.: GPU computing in medical physics: A review. *Med. Phys.* **38**(5) (2011) 2685–2697

2. Fernandez, J.J., Li, S.: Anisotropic Nonlinear Filtering of Cellular Structures in Cryoelectron Tomography. *Comput. Sci. Eng.* **7**(5) (2005) 54–61
3. Gerig, G., Kubler, O., Kikinis, R., Jolesz, F.A.: Nonlinear Anisotropic Filtering of MRI Data. *IEEE Trans. Med. Imag.* **11**(2) (June 1992) 221–232
4. Harry, M.S., Fernández, D.C.: Comparison of PDE-Based Nonlinear Diffusion Approaches for Image Enhancement and Denoising in Optical Coherence Tomography. *IEEE Trans. Med. Imag.* **26**(6) (June 2007) 761–771
5. Guo, S., Xiaoming, W., Renjing, C., Jing, Z.: An approach to suppress speckle noise and enhance edge. *J. Electron.* **23**(2) (2006) 225–230
6. Xuejun, S., Land, W., Samala, R.: Deblurring of Tomosynthesis Images Using 3D Anisotropic Diffusion Filters. In: *Proc. SPIE. Volume 6512.*, San Diego, USA (2007)
7. Sun, Q., Hossack, J.A., Tang, J., Acton, S.T.: Speckle reducing anisotropic diffusion for 3D ultrasound images. *Comput. Med. Imaging Graphics* **28** (2004) 461–470
8. Mrazek, P., Navara, M.: Selection of optimal stopping time for nonlinear diffusion filtering. In: *Int. J. Comp. Vision, Netherlands* (2003) 189–203
9. Chen, D., MacLachlan, S., Kilmer, M.: Iterative parameter-choice and multigrid methods for anisotropic diffusion denoising. *SIAM J. Sci. Comput.* **33**(5) (October 2011) 2972–2994
10. Alvarado, R., Tapia, J.J., Rolón, J.C.: Medical image segmentation with deformable models on graphics processing units. *J. Supercomput.* DOI 10.1007/s11227-013-1042-4 (Published online: 17 december 2013)
11. Black, M.J., Sapiro, G., Marimont, D., Heeger, D.: Robust Anisotropic Diffusion. *IEEE Trans. Image Proc.* **7**(3) (March 1998) 421–424
12. Perona, P., Malik, J.: Scale-Space and Edge Detection Using Anisotropic Diffusion. *IEEE Trans. Pattern Anal. Mach. Intell.* **12**(7) (July 1990) 629–639
13. Weickert, J., Hagen, H.: *Visualization and Processing of Tensor Fields.* Springer (2006)
14. Weickert, J.: *Anisotropic Diffusion in Image Processing.* Teubner-Verlag (1998)
15. Terebes, R., Borda, M., Germain, C., Lavialle, O., Pop, S.: Asymmetric Directional Diffusion Based Image Filtering and Enhancement. In: *Proc. of the IEEE Intl. Conf. on Automation, Quality and Testing, Robotics. Volume 3.*, Washington, DC, USA (2008) 413–418
16. Weickert, J., Zuiderveld, K., ter Haar Romeny, B., Niessen, W.: Parallel Implementations of AOS Schemes: A Fast Way of Nonlinear Diffusion Filtering. In: *Proc. of IEEE International Conference on Image Processing, Santa Barbara, CA* (Oct 1997) 396–399
17. Malik, M., et al.: Productivity of GPUs under different programming paradigms. *Concurrency and Computation: Practice and Experience* **24**(2) (2012) 179–191
18. Kirk, D.B., Hwu, W.m.W.: *Programming Massively Parallel Processors: A Hands-on Approach.* Morgan Kaufmann (February 2010)
19. Sanders, J. and Kandrot, E.: *CUDA by Example: An Introduction to General-Purpose GPU Programming.* Addison-Wesley (July 2010)
20. Cocosco, C.A., Kollokian, V., Kwan, R.K.S., Pike, G.B., Evans, A.C.: BrainWeb: Online Interface to a 3D MRI Simulated Brain Database. In: *Third Int. Conf. on Functional Mapping of the Human Brain, Copenhagen, Denmark* (May 1997) 425
21. Sanderson, A.R., Meyer, M.D., Kirby, R.M., Johnson, C.R.: A framework for exploring numerical solutions of advection-reaction-diffusion equations using a GPU-based approach. *Comput. Vis. Sci.* **12**(4) (March 2009) 155–170
22. Mitchell, J.L., Pennebaker, W.B., Fogg, C.E., Legall, D.J.: *MPEG Video Compression Standard.* Chapman & Hall (1996)



# Pattern Identification by an Artificial Neural Network implemented in a DSP using a touchscreen

M. Méndez<sup>1</sup>, J. Buendía<sup>2</sup>, J. Galindo<sup>2</sup>, M. Rodríguez<sup>2</sup>, J. L. Mata-Machuca<sup>1</sup>, L. Fonseca<sup>1</sup>

<sup>1</sup> UPIITA-IPN Academia de Mecatrónica, Av. Instituto Politécnico Nacional 2580, Barrio La Laguna Ticomán, Gustavo A. Madero, 07340, México .D.F.  
mmendezma{ lfonseca, jmatam}@ipn.mx

<sup>2</sup> UPIITA-IPN Ingeniería Mecatrónica, alumnos PIFI, Av. Instituto Politécnico Nacional 2580, Barrio La Laguna Ticomán, Gustavo A. Madero, 07340, México .D.F.  
jbuendiar0900@alumno.ipn.mx,  
jgalindoo0900(mrodriguez0809}@ipn.mx

*Paper received on 11/29/13, Accepted on 01/19/14.*

**Abstract.** This paper describes an application of pattern recognition by using an artificial neural network (ANN) employing a resistive touch screen as input method. The processing related to the ANN was implemented in a TMS320C6713 DSP Starter Kit (DSK). Some results obtained in a DSK6713 shown that the proposed methodology is able to recognize among three different patterns drawn in a resistive touch panel.

**Keywords:** artificial neural network, patterns, Bayesian network, back propagation, digital signal processor

## 1 Introduction

An artificial neural network is a system based on the operation of biological neural networks, in other words, is an emulation of a biological neural system (Hilera-González, 2000; Isasi and Galván, 2004). A neural network has at least two physical components, namely, the processing elements and the connections between them. The processing elements are called neurons, and the connections between the neurons are known as links. Every link has a weight parameter associated with it. Each neuron receives stimulus from the neighboring neurons connected to it, processes the information, and produces an output. Neurons that receive stimuli from outside the network



(i.e., not from neurons of the network) are called input neurons. Neurons whose outputs are used externally are called output neurons. Neurons that receive stimuli from other neurons and whose output is a stimulus for other neurons in the neural network are known as hidden neurons. There are different ways in which the information can be processed by a neuron, and different forms of connecting neurons to another one. Different neural network structures can be constructed by using different processing elements and by the specific manner in which they are connected (Bishop, 2006; Chen, 2010).

One of the main applications for artificial neural networks is to recognize among different patterns (Goltsev, 2012; Jeong and Lee, 2012). In order to carry out this task it is needed a processing tool according to the computational load.

Nowadays microcontrollers have a limited mathematical processing capacity (Ibrahim, 2008; Tremberger et al., 2012), however, DSP's were made specifically for signal processing as its name means "Digital Signal Processor". These were developed by the U.S. semiconductor company Texas Instruments (TI), at present there are different companies that have developed their own versions of DSP as Motorola, Analog Devices and others.

The main contribution of this work is to tackle the pattern recognition problem applying an artificial neural network where the implementation is carried out in a TMS320C6713 DSP Starter Kit (DSK) employing a resistive touch screen as input method. For simplicity, the proposed technique implemented in a DSK6713 is able to recognize among three different patterns drawn in a resistive touch panel. The paper presents an implementation of an artificial neural network in a digital signal processor chip, for identification of the alphabet letters: A, M, and W. Although the processing level of the DSK6713 far exceeds the requirements of this application, the process of identifying patterns may contain, as future work, more challenging letters, for instance E and F or M and N, and others symbols

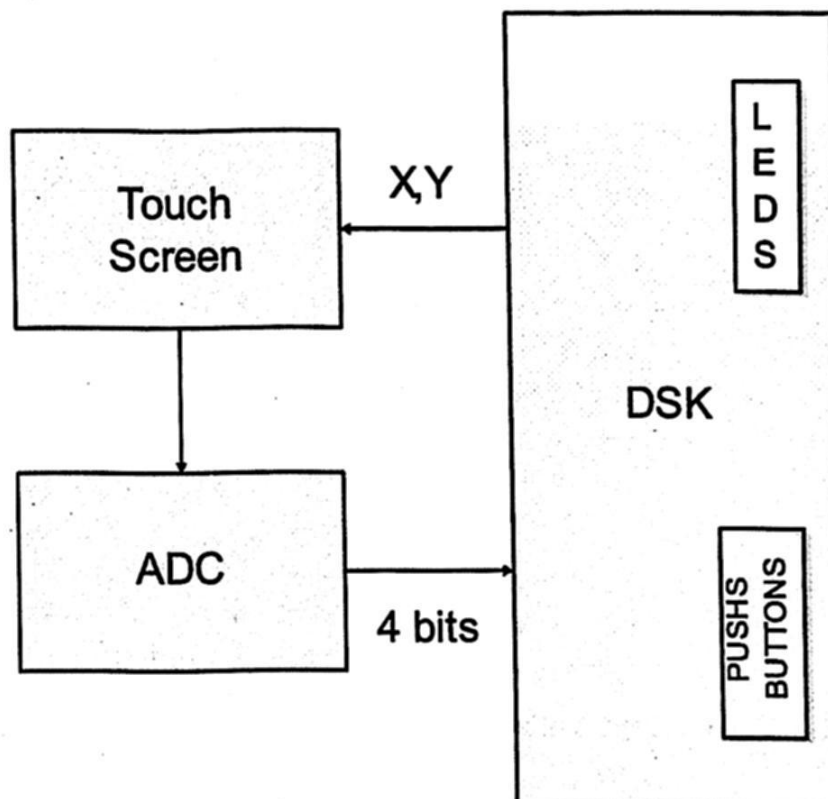
Some DSPs have the advantage of execute different processes in parallel that is why they can process much more information than other devices, so the applications of DSP are very diverse as communications, motor control, image processing, voice recognition, digital cameras or camcorders, MP3 players, high definition televisions (HDTV), etc. (Lee et al., 2010; Madiseti, 2010; Yolacan, 2011).

## 2 Problem statement and solution

The present paper describes the way we trained and applied an artificial neural network in a DSK6713 in order to be able to recognize among three different patterns drawn in a resistive touch panel.

A resistive touch screen were used for the acquisition of patterns, this screen has five terminals, two for power supply, two for x-axis and y-axis selection, and an analog output which is proportional to the pressed position. Subsequently, the analog output is converted by a 4-bit ADC and interpreted by the DSK (see Fig. 1).

The screen was discretized at a resolution of  $9 \times 9$ , which means, each pattern has a total of 81 elements.



**Fig. 1 Block diagram of the proposed prototype .**

Within the DSK, each element pressed from the screen in a matrix form (XY) is interpreted as 1 and introduced a vector of 81 elements, the rest of them are 0. This vector is used as input to the ANN recognition stage.

Input / output peripherals of the DSK were used, LED's to indicate that DSK is in acquisition mode or to indicate the result of the pattern recognition and DIP switches to select between the 2 stages of the process.

While pressing DIP switch 0 of DSK LED0 flashes, during this time the DSK is receiving data from the ADC and touch screen in order to fill the input vector. When DIP switch 3 is pressed ANN is implemented to this vector and the results are displayed by activating one of the remaining 3 LEDs, each of these refers to a pattern.

The implemented ANN is Bayesian type and it was trained by back-propagation method (Fig. 2). It consists of 3 layers; the first layer (the input layer) has 81 inputs and 81 neurons, which output feeds the second layer (the hidden layer) which consists of 30 neurons, which output is the input for the last layer of 3 neurons (the output layer) which gives us the result of the ANN as is shown in Fig. 3.

The training of the ANN was performed in MATLAB ®. For these three training patterns (letters M, A and W), we have used three variants for each one, giving a total of 9 training vectors. These training vectors are shown in a graphical representation in Fig. 4.

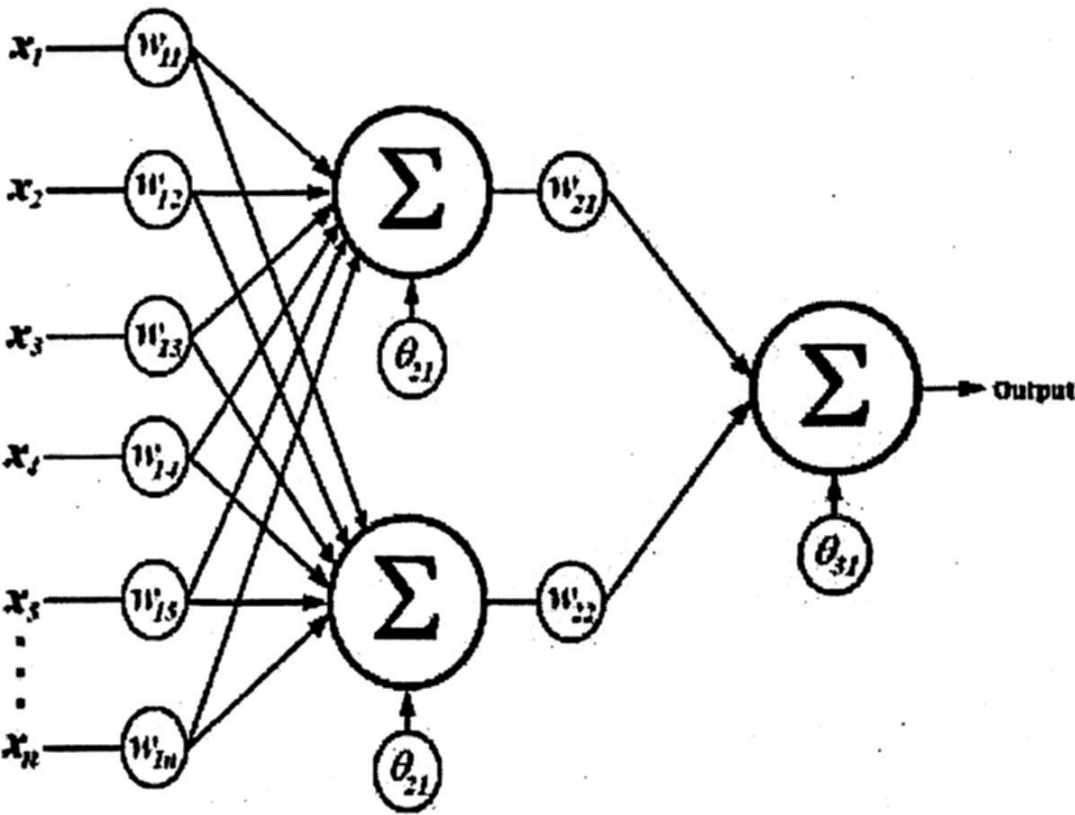


Fig. 2. Graphic representation of an artificial neural network.

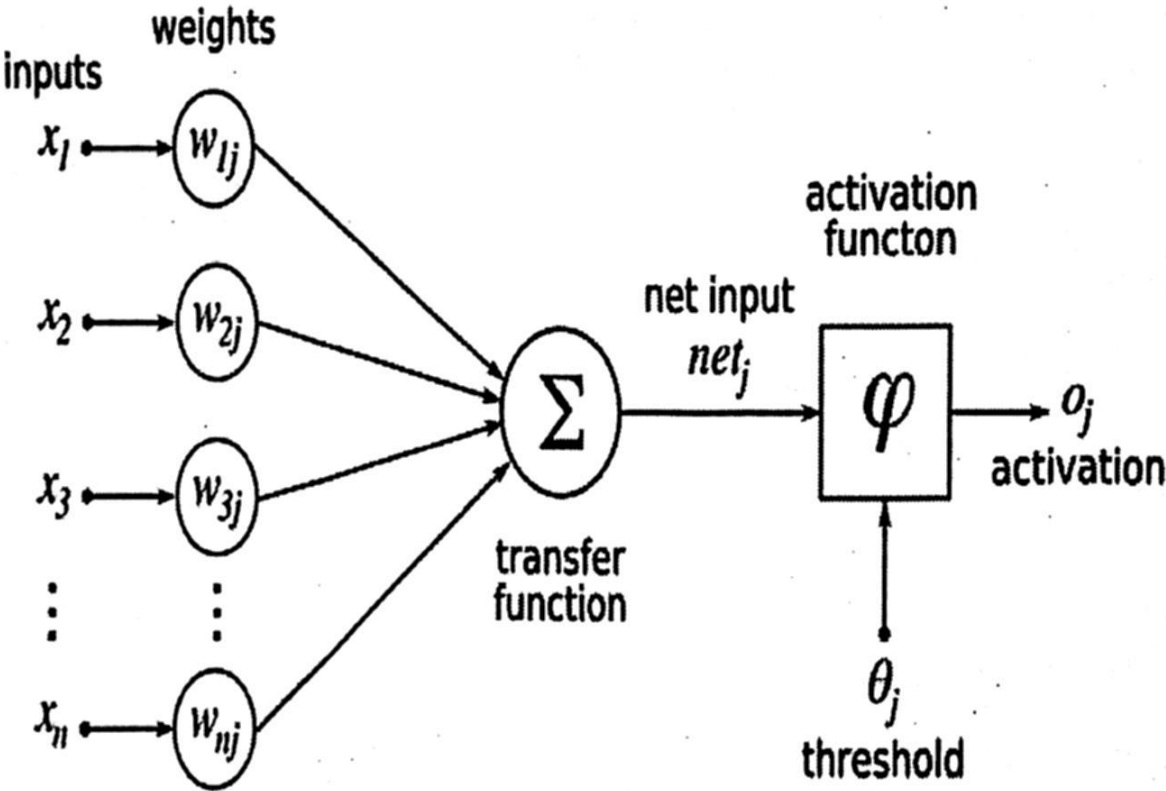


Fig. 3. Graphic representation of a neuron.

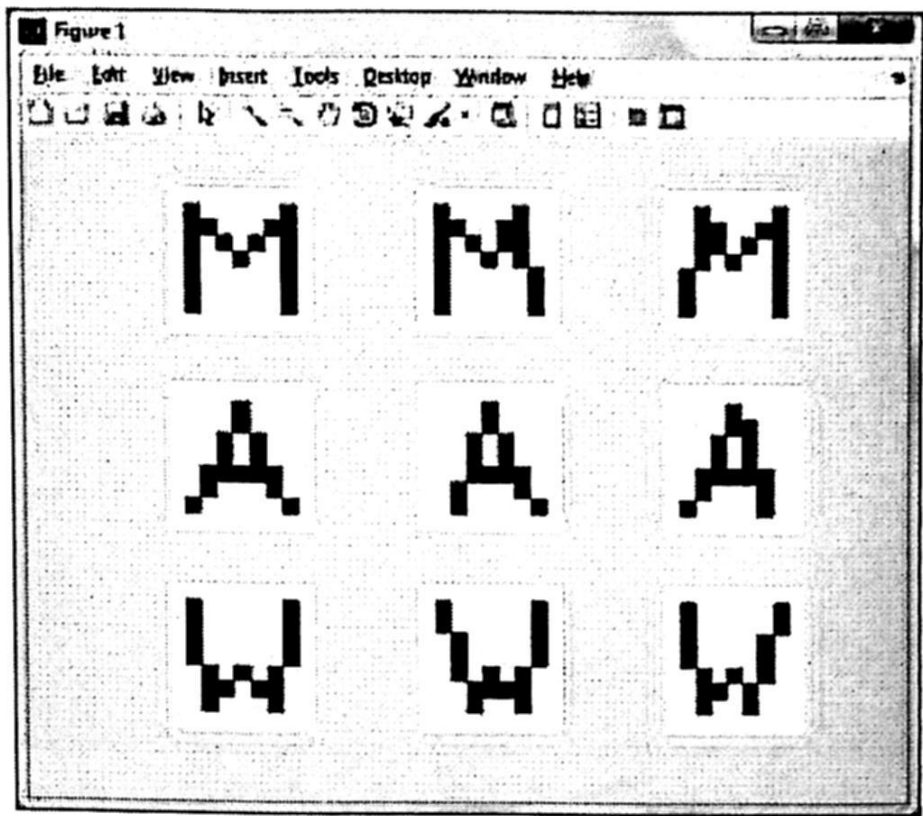


Fig. 4. Target vectors used for training.

For each pattern a target vector was defined, whose length is the number of neurons in the output layer, see Table 1.

Table 1. Patterns and output layers

Output layer	Pattern		
	'M'	'A'	'W'
Neuron 1	1	0	0
Neuron 2	0	1	0
Neuron 3	0	0	1

As a validation method for the ANN, it was introduced into the ANN the same training vectors previously described in Fig. 4. The results obtained from the validation are the expected one in each case.

Pattern 'M' outputs

1.0000  
0.0000



0.0000

Pattern 'A' outputs

0.0000

1.0000

0.0000

Pattern 'W' outputs

0.0000

0.0000

1.0000

In this way we tested that the ANN responds appropriately to the training vectors.

### 3 Results

In order to demonstrate that the current project is working in a proper manner, around 100 different testing patterns were tested; these patterns and their results were stored in the DSK in order to be analyzed later. Figure 5 depicts the experimental platform.

The results were satisfactory in 96% of cases, since the patterns drawn on the touch screen were properly detected. Twelve of these patterns are shown in Table 2.

The results in Table 2 show the values of the output layer of the ANN, which means that the closer to 1, the output of neuron 1 pattern is identified by the network as an 'M' and so on. This is evident if compared with the target vector.

According to the tests the best identified pattern is the 'W' followed by 'A' and finally the pattern 'M'. Although some trials have values in the other two neurons, these are insignificant because they do not affect the result.

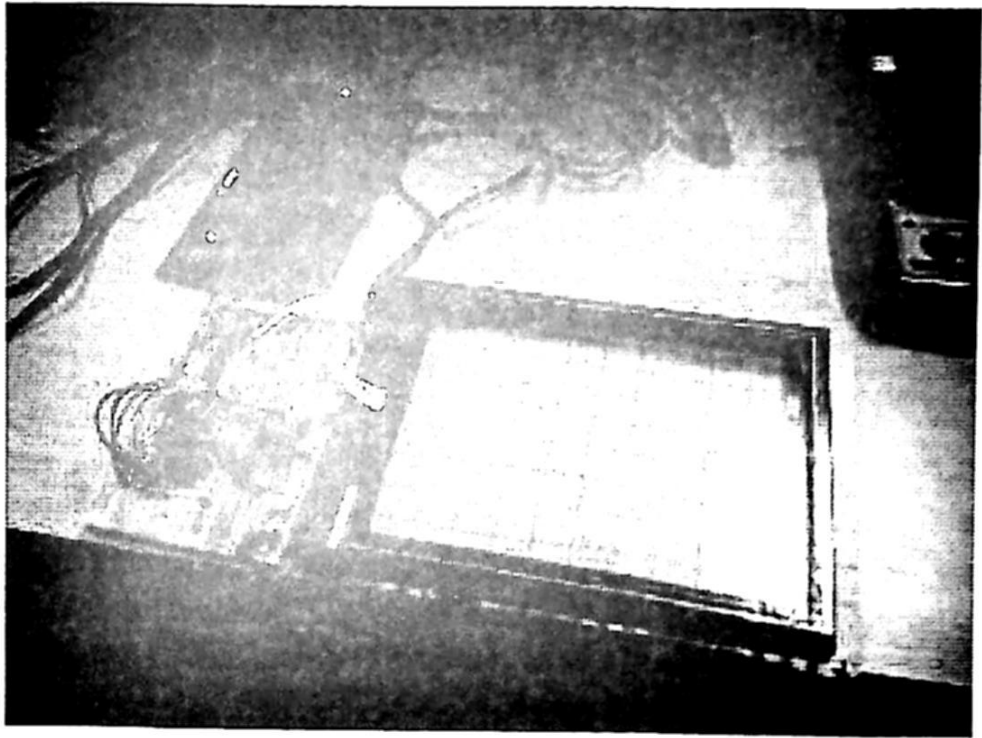
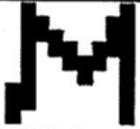





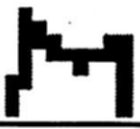







Fig. 5. Experimental implementation of the proposed prototype.

Table 2. Test Patterns.

	Pattern 1		Pattern 2		Pattern 3	
Neuron 1	0.9999		0.0116		0.0000	
Neuron 2	0.0000		0.9357		0.0000	
Neuron 3	0.0000		0.0000		1.0000	

	Pattern 4		Pattern 5		Pattern 6	
Neuron 1	1.0000		0.0000		0.0002	
Neuron 2	0.0000		0.9960		0.0000	
Neuron 3	0.0000		0.0000		0.9986	

	Pattern 7		Pattern 8		Pattern 9	
Neuron 1	0.9951		0.0000		0.0000	
Neuron 2	0.0000		1.0000		0.0000	
Neuron 3	0.0000		0.0000		1.0000	

	Pattern 10		Pattern 11		Pattern 12	
Neuron 1	0.9982		0.0000		0.0000	
Neuron 2	0.0001		0.9999		0.0000	
Neuron 3	0.0000		0.0000		1.0000	

## 4 Conclusion

Based on the performed tests, it can be determined that the system is able to successfully distinguish among the three patterns, even if the pattern has significant changes. Considering that the processing level of the DSK used far exceeds the requirements of this application, the process of identifying patterns may contain, as future work, all letters of the alphabet and others symbols, it also would be possible to increase the screen resolution for more detailed patterns.

## 5 Acknowledgements

M. Méndez, J. Mata-Machuca, L. Fonseca thank to the Secretaría de Investigación y Posgrado, Instituto Politécnico Nacional. J. Buendía, J. Galindo and M. Rodríguez are grateful for the support provided by the Instituto Politécnico Nacional through the PIFI scholarship program with the project SIP 20131714.

## References

1. Bishop C (2006) Pattern recognition and machine learning. Springer, New York, USA.
2. Chen C(ed) (2010) Handbook of pattern recognition and computer vision, 4<sup>th</sup> Edition, World Scientific Publishing Co, Danvers, MA, USA
3. Goltsev A (2012) Investigation of efficient features for image recognition by neural networks, Neural Networks 28: 15-23
4. Hilera-González J (2000) Redes neuronales artificiales: fundamentos, modelos y aplicaciones, Alfaomega, México
5. Ibrahim D (2008) Advanced PIC microcontroller projects in C: from USB to Zigbee with the PIC 18F Series, Newnes, Elsevier, MA, USA
6. Isasi P, Galván I (2004) Redes neuronales artificiales: un enfoque práctico, Pearson-Prentice Hall, Madrid, España
7. Jeong S, Lee M (2012) Adaptive object recognition model using incremental feature representation and hierarchical classification, Neural Networks 25: 130-140
8. Lee H, Chakrabarti C, Mudge T (2010) A low-power DSP for wireless communications, IEEE Transactions on Very Large Scale Integration (VLSI) Systems 18(9): 1310-1322
9. Madisetti V (2010) The digital signal processing handbook, second edition, Digital signal processing fundamentals, CRC Press, Taylor & Francis Group, FL, USA
10. Ruge I, Alvarado D (2013) FPGA-based neural network evaluation system for image recognition, Tecnura 17(36): 87-95
11. Tremberger G, Armendariz R, Takai H, Holden T, Austin S, Johnson L P, Cheung T (2012). Applications of Arduino microcontroller in student projects in a community college. In American Society for Engineering Education. American Society for Engineering Education
12. Yolacan E., Aydin S, Ertunc H (2011) Real time DSP based PID and state feedback control of a brushed DC motor, In XXIII International Symposium, Information, Communication and Automation Technologies (ICAT), pp. 1-6

# Muscle Pain and Blink Classification using a Brain Computer Interface

Oscar Montiel, Roberto Sepúlveda, Gerardo Díaz, Daniel Gutierrez, and Oscar Castillo

Instituto Politécnico Nacional, CITEDI. Tijuana, B.C., México

{oross,rsepulvedac}@ipn.mx

{gdiaz,dgutierrez}@citedi.mx

Instituto Tecnológico de Tijuana-ITT. Tijuana, B.C., México

{ocastillo}@tectijuana.mx

*Paper received on 11/30/13, Accepted on 01/19/14.*

**Abstract.** In this paper, two different architectures of artificial neural network (ANN) for the classification of blinking and arm pain caused by an external agent are used. The electroencephalographic (EEG) dataset is obtained from 20 people in the range of 23 to 30 years of age using a Brain Computer Interface (BCI), it is divided into a batch of necessary patterns to train and test the ANN. Experimental results using different training algorithms are shown.

**Key words:** EEG, BCI, ANN, FFT, Arm Pain, Blink, Adaptation Algorithm, MLP

## 1 Introduction

The human brain is a complex network of synaptic connections between neurons, which generate the electric impulses necessary to develop human functions like movements, communication, language, feelings, memory, reasoning, etc.; these functions are represented by EEG signals [1]. Since Human Computer Interface (HCI) technology has allowed to read EEG signals in humans, it was thought for interpreting and using them as communication channels with auxiliary devices that can help people with mental and physical problems [7].

EEG signals are read and interpreted by a BCI system; these electrical signals are produced by the different stimulus as physical action (motion) or mental status as feeling, imagination, memory, etc. Using BCI devices have many applications in robotic prostheses, pattern recognition, studies of pathologies such as epilepsy, Alzheimer, Parkinson, etc.

This paper presents a methodology to classify EEG signals using a multilayer perceptron (MLP) trained with the backpropagation algorithm, to find patterns produced by different external stimulus, specifically muscle pain and eye blinking. It is organized as follows: In Section II, the Backpropagation algorithm is explained as a method of ANN training as well as an EEG overview for understanding how signals are interpreted and manipulated. Section III deals with the

problem formulation, here the process used for implementing EEG Signal Processing and Classification is shown. In Section IV the methodology to present training patterns to the ANN is explained. In Section V and Section VI, the analysis of results and conclusions are shown, respectively.

## 2 Artificial Neural Network Overview

An Artificial Neural Network (ANN) is an assembly of interconnected and hierarchical organized simple processing elements; its functionality is inspired by the biological nervous system. The processing ability of the network is contained in the strength of the interconnection (weights) of its units, which is obtained through a process of adaptation of its parameters; the idea is to learn a set of patterns, which are the training examples [18] [20]. An ANN is a machine learning method inspired in how the brain works to solve any kind of problems by the association of neural information.

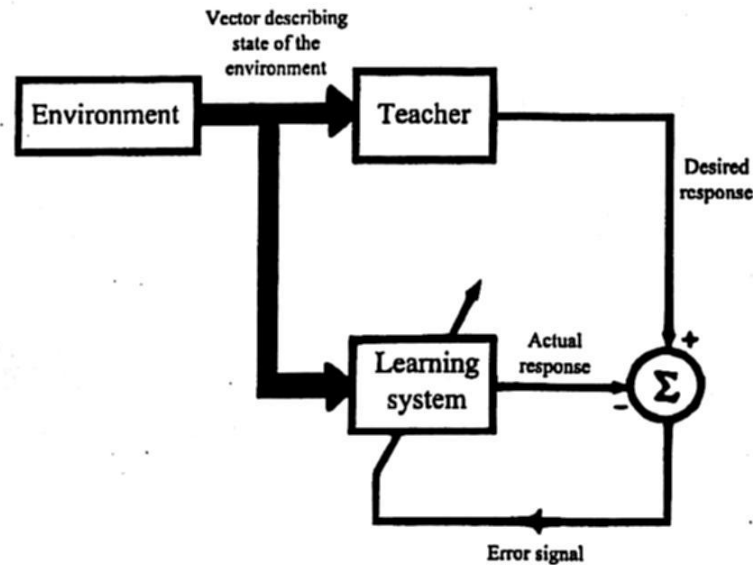


Fig. 1. Model representation of supervised mode training. This figure shows the different elements that conform this learning process.

### 2.1 Backpropagation Algorithm

This algorithm is a method of ANN training; it basically consists in using an error signal calculated by the difference between the actual  $y_j(n)$  and desired output  $d_j(n)$  of the network to correct the weights, in this algorithm the error is retropropagated from the output to the input then an optimization algorithm modifies the weights with the aim to reduce the error  $e_j(n)$  at the individual neurons, hence the global output error called Mean Squared Error (MSE), the training finished when the MSE is less than the convergence error  $\varepsilon_0$ , see



Algorithm 1. Once trained the network, it has the ability to focus on the characteristics of an arbitrary input that resembles other previously seen, regardless any noise signals affecting the patterns [17].

The Backpropagation Algorithm could be used in two modes of training; supervised and non-supervised mode, respectively. The first one is defined as a type of network learning with a teacher, which is the expert problem solver, the knowledge is provided by tagged pairs of inputs and outputs to achieve the training, see Fig. 1. In the non-supervised mode, the ANN learns with no teacher, in this mode, there are no tagged examples of a function to be learn [18] [17].

The experiments shown in this paper were achieved using the supervised training mode; so, the teacher presents to the network the training dataset. In Algorithm 1, the pseudocode for implementing the Backpropagation Algorithm is shown [17].

---

#### Algorithm 1 Backpropagation Algorithm

---

```

1: procedure BACKPROPAGATION(pattern,maxepoch, $\epsilon_0$ ,Target) ▷
    $\epsilon_0 :=$ convergency error
2:   while  $|MSE| < \epsilon_0$  do
3:     repeat
4:       Calculate for each node
5:        $w_{ji} := d_j(n) + y_j(n)$ 
6:        $MSE(n) := \frac{1}{2} \sum_j |e_j(n)|^2$ 
7:        $w_{ji}(n+1) := d_j(n+1) + y_j(n+1)$ 
8:        $MSE(n+1) := \frac{1}{2} \sum_j |e_j(n+1)|^2$ 
9:        $epoch := epoch + 1$ 
10:    until maxepoch ||  $\epsilon_0$  achieved
11:   end while
12:   return y
13: end procedure

```

---

## 2.2 EEG Overview

The technique of electroencephalography (*EEG*) is used to analyze the brain activity, which is manifested in electric waves. To accomplish EEG technique, an array of electrodes are placed on the scalp over multiple areas of the brain to detect and record the patterns of electrical activity. The electrodes are placed on the scalp according to the international 10-20 system of electrode position [10].

It is important to know that the brain is divided by sections, two hemispheres (left and right) and four lobules (frontal, parietal, temporal and occipital), where each section is related to specific sections of the human body. For example, to analyze a stimulus on the left side of the body the right hemisphere has to be analyzed, and viceversa. The scheme of a BCI presented in the Fig. 2 is divided into three main sections: In the first section, the electric activity generated by the brain as well as the interface to acquire the activity is shown. In the second

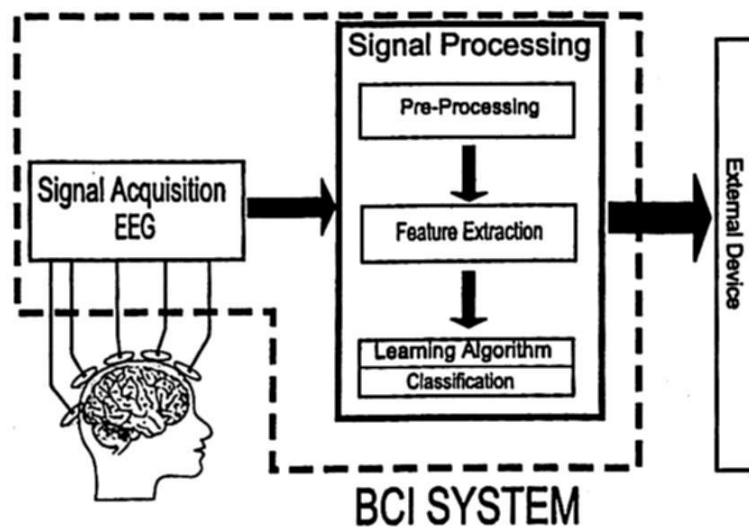


Fig. 2. Scheme of a Brain Computer Interface (BCI) system

section, the signal-processing block is in charge of conditioning and manipulating data using methods as correlation, Artificial Neural Networks, ANFIS, etc., to accomplish a task. The third section consists in the application, here a variety of algorithms can be implemented, for example, wheel chair control by mind, health assistant, rehabilitation, epilepsy detection, mind control of a prosthesis [7] [12] [8] [15].

The brain activity has specific characteristics like time, frequency, amplitude, magnitude and kurtosis; some of them can be analyzed in time or frequency domain.

### 3 Problem Formulation

The experimental platform of Fig. 3 shows the modules used to process EEG signals; it is a flexible system based on the described block diagram shown in Fig. 2 that supports the process of research and development.

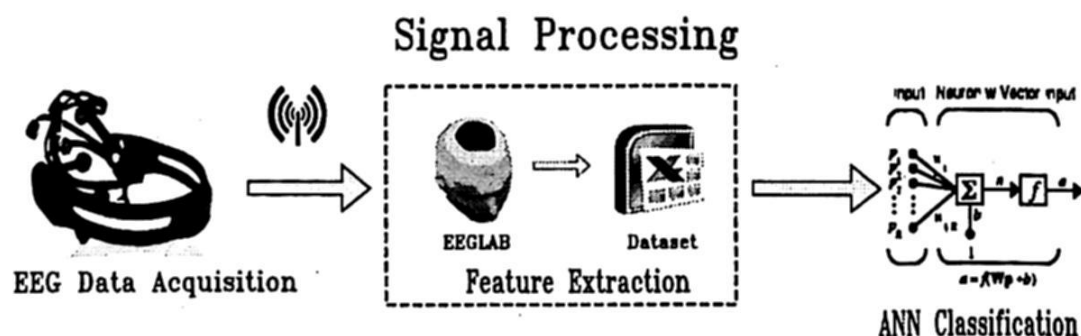


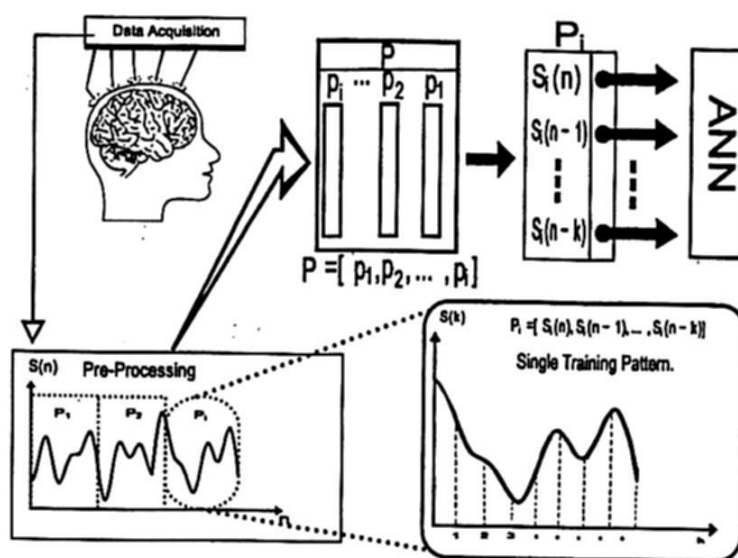
Fig. 3. EEG Signal Processing and Classification using a BCI system.

In the first section of Fig. 3, the EEG Data is acquired by the EPOC Neuroheadset and is transmitted using wireless communication. Once the data has

been received by the computer system, it is recorded for postprocessing, then using a Feature Extraction method some aspects of the signal like, amplitude, time duration, form, magnitude, frequency bands, etc., are determined, to accomplish this EEGLab and Excel are used. The EEGLab is an open-source toolbox for Matlab, which is used to study the offline EEG data already recorded; this software package has different properties for analyzing dipole sources, Independent Component Analysis (ICA), Fast Fourier Transform (FFT), Wavelet Transform, etc. The EEGLab software can read samples from EPOC Neuroheadser and save them in Excel format. Then, the dataset is divided in patterns that are used for training and testing the Neural Network [21].

## 4 Proposed Methodology

Once the data has been processed, the next step is to select patterns based on a teacher heuristic. All EEG signals are divided in patterns to build the batch of patterns used as input of the ANN. Fig. 4 illustrates how the batch of patterns feed the ANN.



**Fig. 4.** The EEG signals are sampled using the EPOC neuroheadset; a time series is divided into lots of training patterns  $P$ , each pattern consists of multiple data streams that will serve either to train the network, or to verify the training.

Each pattern contains a number of samples to train and test the ANN; the number of samples for each pattern depends on the stimulus that required to be classified. In Algorithm 2, the pseudocode for EEG signal processing is shown.

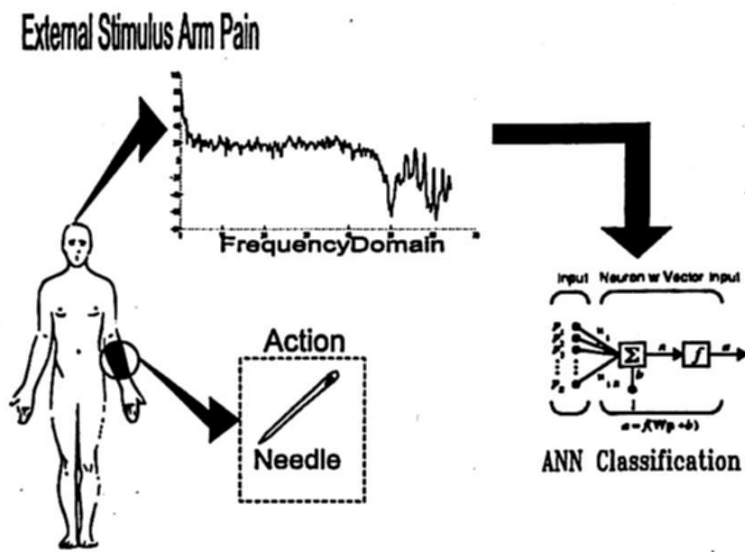
### 4.1 Muscle Pain Classification

The first stimulus to classify is the pain induced by an external agent [4] [14], this stimulus is induced by a prick in the right arm; the subject must be in a relaxed status for two minutes approximately, and then, the pain is induced [6].

**Algorithm 2** EEG Signal Processing general method proposal

```
1: Initialize variables
2: Select electrodes to be analyzed
3: Create Inputelectrode matrix from EEG readings ▷ If it is necessary, apply FFT
   to Inputelectrode matrix
4: Normalize Inputelectrode matrix already changed
5: Select TRAINDATA, TARGET and TESTDATA to train ANN
6: while Any data be minor than size of train dataset do           ▷ Training Data
7:   TRAINDATA accumulate pattern for each iteration
8: end while
9: while Any data be minor than size of test dataset do           ▷ Testing Data
10:  TRAINDATA accumulate pattern for each iteration
11: end while
12: Define ANN TARGET           ▷ Depending activation function
13: Define ERROR, MAXEPOCH      ▷ ANN training parameters
14: Train ANN
15: Test ANN
```

The EEG signal obtained after the prick is the most important information used for training and testing the ANN [5]. Fig. 5 explains the experimental process of pain activity classification.



**Fig. 5.** Scheme for implementing EEG signal processing and classification of muscle pain.

It is worthwhile to mention that the pain cannot be appreciated in the time domain; therefore, the signal has to be converted to frequency domain using the Fast Fourier Transform, then a filter is applied to eliminate noise in the signal. All the experiments were saved in an Excel Table (dataset); it was divided into two sets; one set is used to train the ANN, and the second set for testing it, which is important for proving the network knowledge generalization capability.

## 4.2 Eye Blinking Classification

Fig. 6 shows the experimental process to achieve the eye blinking classification; although, it is considered an artifact, it is important to take care of it because this artifact is present in the whole EEG encephalographic readings. Blinking is a natural body movement that helps to maintain the eyes wet and protected from external elements [16] [11].

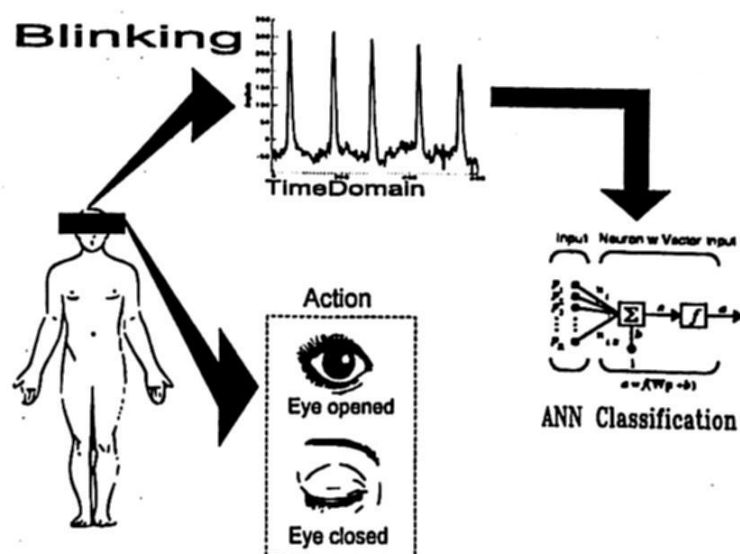


Fig. 6. Implementing eye blinking signal processing and classification.

As it is seen in Fig. 6 each time the eyelid changes from open to close, an increment of amplitude in microvolts appears in the readings, this data can be used to generate the dataset for the ANN [9].

It is important to achieve a proficient identification and classification of eye blinking for a good management of EEG signals, because it may affect the analysis of other EEG stimulus classifications [16].

## 5 Experimental Results

Two cases of study for the classification of EEG signals are presented using ANN multilayer perceptron type with hyperbolic tangent activation functions, it was trained in supervised mode with the backpropagation algorithm. For training, a +1 value is assigned to a valid stimulus, and -1 for a non-valid stimulus. In the operation mode, the ANN returns values between 1 and -1, to make the classification possible a threshold of +0.9 is applied, hence any value equals or over the threshold is considered a valid stimulus, otherwise it is a non-valid stimulus.

The first case shown in Fig. 5 consists in identifying muscle pain induced by an external agent. The second case is to identify the blink of both eyes as it is shown in Fig. 6; although the blink signals are not true encephalographic signals, they are artifacts and their study is considered important, because they interfere



with the EEG results and their interpretation. In all the conducted experiments in this work, the scheme shown in Fig. 3 was used.

The adaptive algorithms have the purpose to accelerate the convergence of the error for neural network by optimizing the weights in the learning process. In this research, the adaptive algorithms shown in Table 1 were used to investigate which one have the best characteristics in this kind of classification.

**Table 1.** Adaptive algorithms used for experiments.

Acronym	Adaptive algorithm
LM	Levenberg Marquardt
OSS	One Step Secant
BFGS	BFGS Quasi-Newton
RP	Resilient Backpropagation
GD	Gradient Descent

The ANN architectures used for both experiments are summarized in Table 2. To establish a benchmark to compare different adaptive algorithms, a convergence error of  $1e^{-3}$  was established. Fig. 4 illustrates in general terms, how the different training patterns were generated and used in batch mode. To perform the experiments, a computer with a processor I7 2.67Ghz 920A, with 6GB of RAM and OS Windows 7 of 64 bits was used.

**Table 2.** ANN architectures for external stimulus classification.

Stimulus	ANN Architectures
Blinking (both eyes)	120:20:10:5:1
Pain (right arm)	1280:20:10:5:1

Table 3 shows some parameters about training ANN for the pain arm stimulus. The LM algorithm has the best qualities for training and testing (i.e. knowledge generalization), 99.7% and 98.8%, respectively; moreover, it is the fastest algorithm in the training mode, it last a meantime of 1167 seconds.

Table 4 shows statistical values of time about training ANN for blinking. The RP last a meantime of 0.50 seconds, which means that RP was the fastest algorithm to train the blinking patterns. Also, the best rate of classification was obtained with this algorithm, for the training patterns a 99.5% of successful was achieved, and 96.4% for the test pattern.

**Table 3.** Statistical results of ANN training to classify pain in the right arm. Different adaptive learning algorithms were used.

Adaptive learning algorithm	Mean time(sec)	Min time(sec)	Max time(sec)	Desv. Std(sec)	Mean epochs
LM	1167.00	420.00	2700.00	468.53	33344
OSS	1864.43	840.00	5687.00	1012.52	36462
BFGS	1403.60	600.00	4140.00	706.33	33677
RP	1568.00	840.00	4800.00	771.39	48982
GD	1290.00	720.00	4140.00	728.87	39677

**Table 4.** Statistical results of ANN training to classify eye blinking. Different adaptive learning algorithms were used.

Adaptive learning algorithm	Mean time(sec)	Min time(sec)	Max time(sec)	Desv. Std(sec)	Mean epochs
LM	9.43	2.00	28.00	5.89	11
OSS	0.83	0.49	1.70	0.26	27
BFGS	145.97	59.00	269.00	60.32	18
RP	0.50	0.36	2.79	0.43	27
GD	22.43	10.00	55.00	10.20	3825

**Table 5.** Classification results for the blinking activity using the training and testing data. The values express the percentage of true classification.

Learning algorithm	Train data	Test data
LM	98.7	93.5
OSS	99.3	94.1
BFGS	99.3	93.0
RP	99.5	96.4
GD	99.7	88.8

**Table 6.** Classification results for the right arm pain using the training and testing data. The values express the percentage of true classification.

Learning algorithm	Train data	Test data
LM	99.7	96.9
OSS	96.6	95.4
BFGS	97.4	93.1
RP	95.3	93.0
GD	98.7	96.8

## 6 Conclusions and Future Work

It is known that the use of ANNs is a powerful and efficient tool to classify EEG signals; however, there is not enough research works focused to pain detection caused by external agents; this work provides valuable information in this field. In addition, eye blinking signals classification has been included because at the present, it is well known its precise classification importance to improve EEG interpretation.

Several experiments with different learning algorithms for a multilayer perceptron type ANN were achieved. In all the experiments, we used a dataset consisting of 60 patterns of each person; 30 of them were used for training, and the rest for testing the network; a total of 20 people in the range of 23 to 30 years of age were used to complete the whole dataset.

The muscle pain classification experiments were performed by inducing pain with a needle in the right arm of a person. Statistical results demonstrate that the LM algorithm performs better than the others shown in Table 3, with this algorithm once trained the network, we obtain a 99.7% of reliability classifying the training dataset, and 96.9% using the test dataset which is fine because the patterns in the test set are unknown for the ANN. It is worth mentionion that the LM proved to be the fasted algorithm for training.

For eye blinking classification, the GD was the better recognizing the training patterns but the less reliable with the test pattern. The RP is almost as good as the GD recognizing training patterns, and it is the most reliable classifying the test patterns; with the RP, a reliability percentage of 99.5% with the training patterns and 96.4% of reliability with the test patterns was obtained.

With respect to the statistical results, hence the reliability percentages, it is important to mention that a 0.9 recognition umbral was used; therefore, these results might be improved by reducing the umbral value.

## 7 Acknowledgments

The authors thank to the "Instituto Politécnico Nacional (IPN)", to the "Instituto Tecnológico de Tijuana (ITT)", and the "Consejo Nacional de Ciencia y Tecnología (CONACYT)" for their support.

## References

1. Arias, G. , Felipe, H.: Detección y clasificación de artefactos en señales eeg. In Memorias de STSIVA 09. Universidad Tecnológica De Pereira. (2009).
2. Balakrishnan, D.; Puthusserypady, S.: Multilayer perceptrons for the classification of brain computer interface data. Bioengineering Conference, 2005. Proceedings of the IEEE 31st Annual Northeast , pp.118,119, 2-3 (2005).
3. Chambayil, B., Singla, R., Jha, R.: Virtual keyboard bci using eye blinks in eeg. In Wireless and Mobile Computing, Networking and Communications (WiMob), 2010 IEEE 6th International Conference on. 466-470. (2010).

4. Chang, P., Arendt-Nielsen, L., Graven-Nielsen, T., Svensson, P., Chen, A.: Different eeg topographic effects of painful and non-painful intramuscular stimulation in man. *Experimental Brain Research*, 141, 195-203. (2001).
5. Chang, P. F., Arendt-Nielsen, L., Graven-Nielsen, T., Svensson, P., Chen, A. C.: Comparative eeg activation to skin pain and muscle pain induced by capsaicin injection. *International journal of psychophysiology*, 51(2), 117-126. (2004).
6. Chen, A. Rappelsberger, P.: Brain and human pain: Topographic eeg amplitude and coherence mapping. *Brain Topography*, 7(2), 129-140. (1994).
7. De la O Chavez, J. R.: BCI para el control de un cursor basada en ondas cerebrales. Masters thesis, Universidad Autónoma Metropolitana. (2008).
8. Erfanian, A.; Gerivany, M. : EEG signals can be used to detect the voluntary hand movements by using an enhanced resource-allocating neural network. *Engineering in Medicine and Biology Society*, 2001. Proceedings of the 23rd Annual International Conference of the IEEE , vol.1, no., pp.721,724 vol.1 (2001).
9. Gabor AJ, Seyal M.: Automated interictal EEG spike detection using artificial neural networks. *Electroencephalogr Clin Neurophysiol* 1992; 83: 271-280 (1992).
10. Hirsch, L. Richard, B.: Atlas of EEG in Critical Care, chapter EEG basics. Wiley, 1-7. (2010).
11. Kutlu, Y.; Isler, Y.; Kuntalp, D.: Detection of Spikes with Multiple Layer Perceptron Network Structures. *Signal Processing and Communications Applications*, 2006 IEEE 14th, vol., no., pp.1, 4, 17-19. (2006).
12. Lin, J.-S., Chen, K.-C., Yang, W.-C.: Eeg and eye-blinking signals through a brain-computer interface based control for electric wheelchairs with wireless scheme. In *New Trends in Information Science and Service Science (NISS)*, 2010 4th International Conference on. 731-734. (2010).
13. Papadourakis G, Vourkas M, Micheloyannis S, Jervis B.: Use of artificial neural networks for clinical diagnosis. *Math Comput Simulation* 1996; 40: 623-635 (1996).
14. Pera, D. L., Svensson, P., Valeriani, M., Watanabe, I., Arendt-Nielsen, L., Chen, A. C.: Long-lasting effect evoked by tonic muscle pain on parietal EEG activity in humans. *Clinical Neurophysiology*, 111(12), 2130-2137. (2000).
15. Pérez, M. Luis, J.: Comunicación con Computador mediante Señales Cerebrales. Aplicación a la Tecnología de la Rehabilitación. Ph.D. thesis, Universidad Politécnica de Madrid. (2009).
16. Sovierzoski, M., Argoud, F., De Azevedo, F.: Identifying eye blinks in eeg signal analysis. In *Information Technology and Applications in Biomedicine*, 2008. ITAB 2008. International Conference on. 406-409. (2008).
17. Mitchell, T.M.: *Machine Learning*, Chapter 4, McGraw-Hill. (1997).
18. Haykin, S.: *Neural Networks A Comprehensive Foundation*, Delhi, India: Pearson Prentice Hall. (1999).
19. Haselsteiner, E.: Using Time-Dependent Neural Networks for EEG Classification, *IEEE TRANSACTIONS ON REHABILITATION ENGINEERING*, vol. 8, no. 4, pp. 457-463, (2000).
20. Hagan, M. T.: *Neural Network Design*, Boston, MA: PWS Publishing Company, (1996).
21. Morchen, F.: Time series feacture extraction for data mining using DWT and DFT, Philipps University Marburg, Marburg, Germany, (2003).



# Emotion recognition and emotional incentive model

Adrian R. Aguiñaga\* and Miguel Ángel López Ramírez , Arnulfo Alanis Garza

Instituto Tecnológico de Tijuana

adrian.rodriquez

@tectijuana.edu.mx

Paper received on 11/30/13, Accepted on 01/19/14.

**Abstract.** Few years has passed since Rosalind W. Picard founded in 1997, the MIT Research Group Affective Computing. Since then a large number of developments aimed at improve the relationship between humans and computers has been done. Following Picard ideas, the outline of a emotional rules based model are presented in this paper, this model arises as the result of analyze theoretical emotion models, EEG pattern recognition and its possible application in computational models. The base idea are simple, develop a rule based model based on emotions, able to interact with human emotional states and take decisions based on them, in order to reduce the burden generated in the interaction between human-machine interfaces. As part of this research an emotion recognition based EEG analysis are presented, as well as the necessary rules to generate associated interpretations for the bio-signal analysis. A multidisciplinary effort of several research fields were required to create this model, such as neurology, digital signal processing, artificial intelligence, physiology, psychology and behavior analysis; Just to provide the references and the emotional interpretation to create a incentive model, capable to interpreted and execute a process modeled under the user perception.

**Keywords:** Affective computing, EEG analysis, Emotional ruled systems

## 1 Introduction

On the last decade brain computer interfaces (BCI) research has been increase dramatically, mainly due the ever-increasing development in the computational and sensors technologies [1][2]; Leading to a accelerated development in affective computing, in order to satisfy one of its principal objectives, "*create devices that allows a natural interaction between humans and machines*", to reduce the burden in the human-machine relationship. However each development in this area, involves as previously mentioned a multidisciplinary effort and the brain electrical signatures analysis, has been emerged mostly to analyze physiological disorders, such as epilepsy and sleep illness [3], however the implementation of this kind of analysis are wide diversified (i.e the analysis of cognitive processes and motor imaginary processes in humans[4][5]).

Another aspect that allows this kind of research, are the advances in technologies to analyze the brain activity; Magnetic resonance imaging (MRI), functional MRI magnetic resonance imaging (fMRI), electroencephalography (EEG) are just a some of the

---

\* Postgraduate Department



techniques developed up to date, however the portability and cost still are an important considerations to provide feasibility to a research and the viability of being applicable; The low cost of implementations and its relatively easy implementations makes the EEG, the best candidate to be implemented in the BCI's. However a drawback of this technique are the noisy information provide it by it, that requires robust techniques to process and analyze the signals. The rest of the paper are arranged as follows, in section 2, the problems related of analyze emotions and behavior over biological signals are discussed, section 3, specifies the proposed rules to create a feedback systems model, section 4, describes the proposed methodology and topology, section 5, presents the results and finally in section 6, the conclusion are presented.

## 2 Characterization

The EEG signals, used for this analysis were extracted from the DEAP data base [6], which is up to our best knowledge, the most complete database related to physiological signals recorded from emotional stimulus. DEAP contains the brain activity of thirty-two persons from England and Switzerland, using the 10/20 electrode placement model,<sup>1</sup> recording 40 different emotional states related to arousal-valence space model (AVS), provided by exposition to a audio/visual stimuli associated to single emotion over one minute for each trial, also the front video of the face for each person was recorded<sup>2</sup>. Each one of the stimulus, was selected by experts selection and a online ranking survey.

### 2.1 Wavelet analysis

Wavelets are a class of functions that are used to locate a particular function in both space and scaling. A family of wavelets can be constructed from a function  $\psi(x)$ , sometimes known as a *Mother wavelet*, which is confined in a finite interval. *Wavelets Daughter*  $\psi^{(a,b)}(x)$  are then formed by translation ( $b$ ) and contraction ( $a$ ), and wavelets are especially useful for the compression of image data and capable of handling the complex behavior of the EEG signals, by their properties that are superior to some conventional Fourier transformation aspects, it has also shown good bio-signals behavior characterization, since WT allows to obtain spatial temporal information being this fundamental on the biological signals processing. For this research, the feature extraction was performed by WT<sup>3</sup>. One of the most convenient advantage of WT, is that the WT kernel coefficients could be used as features to perform classification.

An individual wavelet can be defined by

$$\psi^{(a,b)}(x) = |a|^{(-1/2)}\psi((x-b)/a). \quad (1)$$

<sup>1</sup> Also contains the respiration rate, electrocardiogram, temperature, galvanic and myoelectric information as well as a relation of relevant information for each user

<sup>2</sup> Each user completes a survey to verify the relationship between each real test results and expected

<sup>3</sup> Many authors reported good performance of WT in EEG signals analysis [7][8][9][10][11][12]

In other way

$$W_{\psi}(f)(a, b) = 1/(\sqrt{a}) \int_{-\infty}^{\infty} f(t) \psi((t - b)/a) dt, \quad (2)$$

Unfortunately the selection of the best wavelet to implement, still are an exhaustive search process, to select the appropriate for this research a comprehensive search and trials has been performed, and the Daubechies 6, was selected as the best candidate for this analysis due to the orthogonality and asymmetry properties of this family of wavelets.

**Rhythms** In neurology the EEG bands of frequencies are known as rhythms <sup>4</sup>, shown in table 1, the analysis of behavior of this rhythms are fundamental for the emotional process analysis, due that previous analysis denotes that each rhythm are associated to a natural behavior or specific task <sup>5</sup>[14][13].

Table 1. Brain Rhythms

Brain Rhythms	
Delta	0.1 to 4 Hz
Theta	4 to 8 Hz
Alpha	8 to 12 Hz
Mu	8 to 13 Hz
Beta	12 to 30 Hz
Gamma	25 to 100 Hz

A four level WT decomposition were performed to obtain individual rhythms analysis.

**Filter and domain reduction** The wavelet analysis could be generalized as a band-pass filter, perform a wavelet decomposition of the signal on the desired coefficient of contraction and this feature can also be used as domain reduction, and the complementary filters could be suited for the rhythms previously defined.

## 2.2 Brain bounded areas

In order to reduce the computational burden, a boundary model based on the Broadmann areas was created, considering that each of these areas are related to a specific task, and discrete regions are provided by the 10/20 model, a delimited area could be generated, <sup>6</sup> following the considerations shown as follow [13][14][15]:

<sup>4</sup> This term are more associated to music than engineering

<sup>5</sup> I.e., the mu rhythms are recently related to the motor process and the theta are related to hippocampal process.

<sup>6</sup> With the intervention of an expert in neurology Dr. Carlos Francisco Romero Gaitán, emeritus member of the Mexican Society of Neurology

- Vision<sup>7</sup> Primary areas : 18,19 ; Secondary areas: 20, 21 and 37.
- Audition<sup>8</sup> Primary areas : 41 ; Secondary areas: 22, 42.
- Body sensations: Primary areas : 1,2,3 ; Secondary areas: 5,7 ; Tertiary areas: 22,37,39 y 40.
- Motor system: Primary areas : 4,6,8,44 ; Secondary areas: 9,10,11,45,46,47.

Also most of the literature refers to the limbic system as one the main brain regions related to the emotional process, taking all of this in considerations the occipital, temporal and parietal regions electrodes are associated to create a bounded model as shown in figure 1.

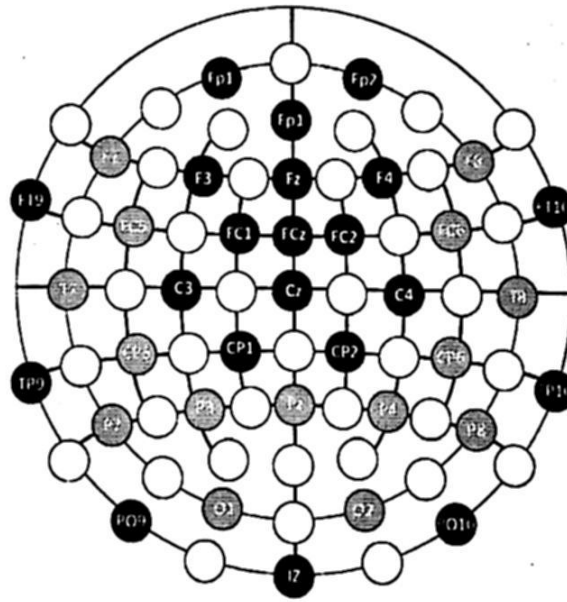


Fig. 1. Emotion regions bounded as an arc model for the 10/20 model.

### 3 Incentive rule based system

An emotional rule based model, denominated Emotional Incentive Model (EIM) is a proposal of this paper, the implementation of this model to interact with systems that recognize and classify an emotional process in a humans are the main goal of it. Also as a model capable of interfacing BCI and systems, providing feedback through a set of rules interacting with the extracted emotional status at real time<sup>9</sup>.

Six rules have been generated to this model shown in table 2, this rules are based on the six basic emotions associated to the human survival process[18][19] [16] [17]

<sup>7</sup> As a Audio -visual characterization the visual area activated have to be considered.

<sup>8</sup> As a Audio -visual characterization the audition areas activated have to be considered.

<sup>9</sup> I.e A prosthesis capable to explore new configurations to avoid or reduce the stress and frustration from it user by increasing or decreasing torque or tracing new trajectories, based on the information provided by the emotion recognition system, will take into account the comfort level of the person beyond the interaction models previously created with a general purpose

[20][21]. The fundamentals from this EIM are taken from traditional notions of emotions and most accepted theories, where emotions are discretized in several ways by different authors as the six basic emotions proposed by Ekman and Friesen [22] and tree structure of emotions proposed by Parrot [23], however hard tags as such are not adequate to define the strength of an emotion considering that emotions are a continuous phenomena rather than discrete, so a dimensional emotion scales models must be considered, such as Plutchiks emotion wheel[18] and the valence-arousal scale by Russell [24] .

Table 2. Emotion based rules

Emotions	
Joy	As incentive to continue a process.
Fear	As incentive to prioritize a process.
Surprise	As incentive for good results performing a new process.
Sadness	As incentive to explore new ways to perform a process.
Disgust	As incentive to stop actual process and explore new process.
Anger	Force to stop a process.

As previous mentioned one of the main goals of affective computing systems, are focused on reducing the burden that is generated in the interaction between humans and machines; And this model aims to increase the efficiency of a system, based on the mental states considerations directly from the user <sup>10</sup> perception, see figure 2. [25][26][27].

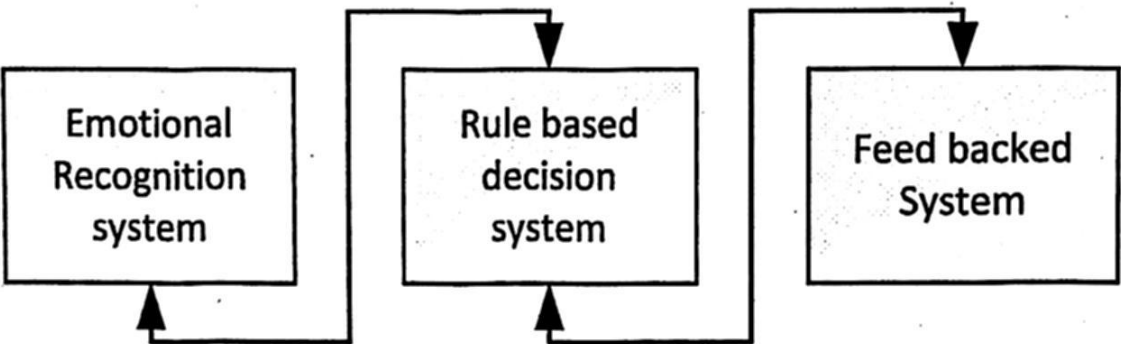


Fig.2. Emotion Incentive Model applied to feed backed models.

In figure 2, the basic steps of the proposed model are illustrated:

*EIM Basic procedures*

- Step 0: Standby, waiting for Actuator System (AS) start up.
- Step 1: AS process start.
- step 2: Rule based decision system start.

<sup>10</sup> The perception of the performance by the user, independently of the optimal model that could be determined

Step 3: Emotional status monitor start.  
 Step 4: Consistent emotion detected.  
 Step 4: Incentive updated.  
 Step 5: Feedback AS.  
 Step 6: Process updated.  
 Step 7: Go to Step 0.

Also the implementation of all the elements to sensing and acting <sup>11</sup>, are as substantiated as affective wearables.

## 4 Methodology

The implementation of this model are based on the actual performance on the emotional detection systems, which are between 70 and 85% based of our own results and several consulted literature, as shown in table 3.

**Pre-processing stage** A band pass filter between 0.5 and 47 Hz was applied to the raw signals and a Laplacean filter were applied to reduce the artifacts contained on the signals as on[7].

**Feature extraction** Daubechies 6, Discrete Wavelet transform(DWT) are applied to obtain the level 4 decomposition coefficients, that would be implemented on a neuronal network pattern recognition task.

**Inputs** Only 15 of the 21 electrodes where selected to perform the analysis as described on section 2, (OP3,OP4,PT8,PT7,PF7,TF8,T7,T8,FC5,FC6,Fp5,Fp6) and the relation ship of each emotions separation where made by a arousal and valence model as on figure 3, whit non negative values and uniformly distributed emotions<sup>12</sup> as high or low statements:

- HA/HV:High Arousal and High Valence.
- HA/LV:High Arousal and Low Valence.
- LA/HV:Low Arousal and High Valence.
- LA/LV:Low Arousal and Low Valence.

Each experiment consists of the sum of 15 electrodes model, three uncorrelated emotion from 32 users in order to avoid the trivial case, this means that independent samples are taken for training and not just the average of all users to explore the generalization of the results. Then the process are evaluated with other set of emotions, tagged to a different set of classes as in figure 3, and two architectures were tested for each of them.<sup>13</sup>

<sup>11</sup> Any sensor attached to a person could affect its normal behavior [2]

<sup>12</sup> Emotions selected from the Ekman model.

<sup>13</sup> Scaled conjugated gradient and back propagation network whit two layers and 10-fold cross validations.



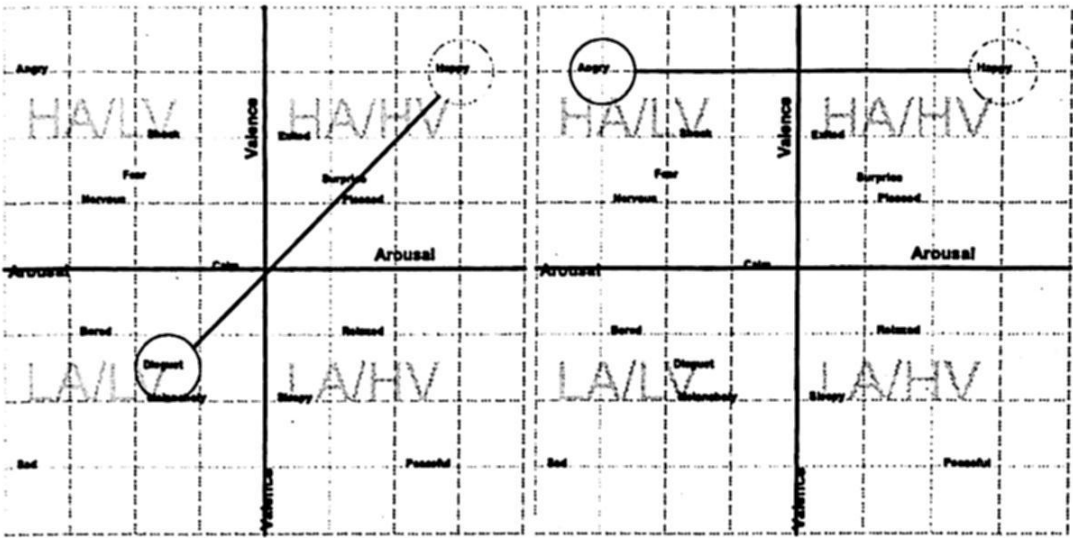


Fig. 3. Emotion selection by a AVS model and Ekman emotion distribution

10-fold cross-validation was performed to evaluate the mean performance of the analysis and identify the average behavior of classifiers test, shown in the figure 4 and in the figure 5, the average of each cross validations are presented. Other configurations were also monitored with similar results, as shown in Figure 4. A 70-30% configuration were implemented as training configurations <sup>14</sup>

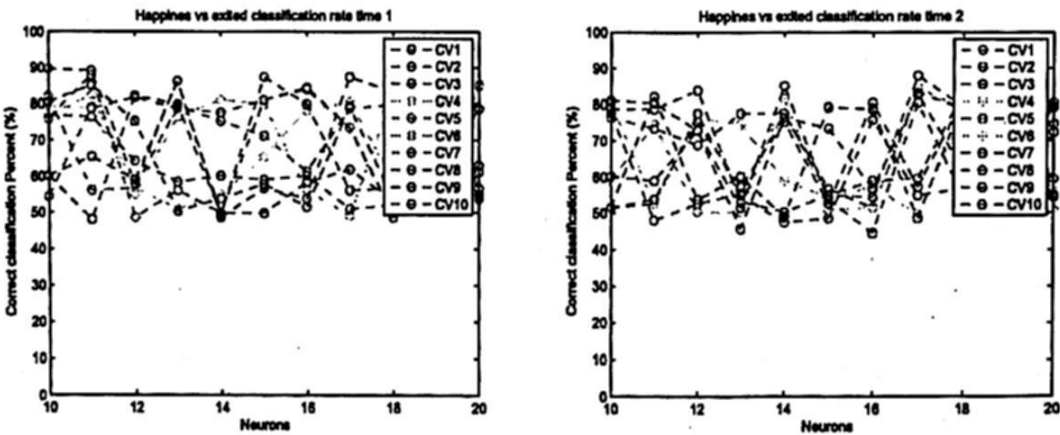


Fig. 4. 10 fold cross validations performance

The results of the recognition task provided by the EIM, are then carry out to the process described in section 3, and improve the user experience while they are interacting with an external system.

<sup>14</sup> Also the exploration of different configurations as 60-30 and 50-50 were made if, however 70/30 shows the better performance.

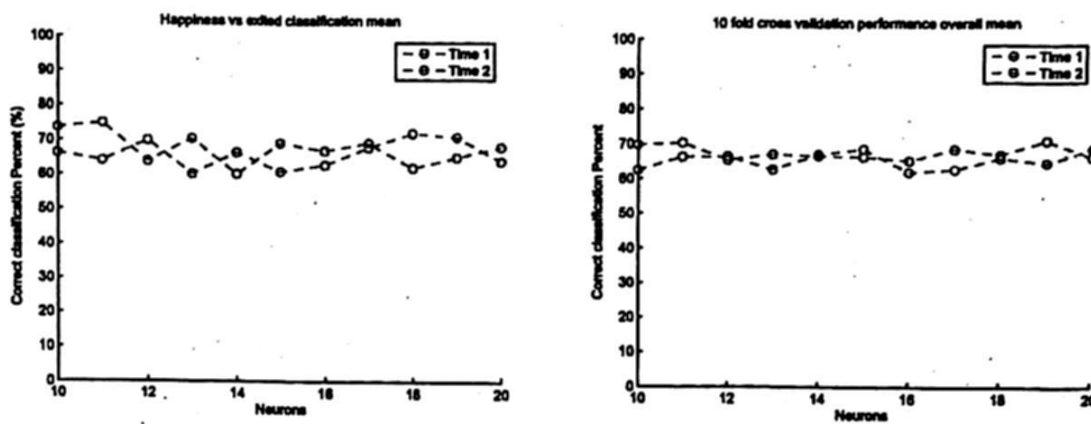


Fig. 5. Mean 10 fold cross validations performance for two temporal analysis (2 of 6).

## 5 Results

Our group <sup>15</sup> are currently developing the affective wearables, to implement this model <sup>16</sup>, however the task of emotion recognition are getting some good results as shown in the table 3, where the behavior of the application model from the defined area is presented.

Our initial data generate an overall behavior of 80% recognition rate, on four emotions (anger, sadness, happiness, disgust). This rate could be increased by using more robust pre-processing technique like DTC-WT (Dual tree Complex Wavelet transform) or MNF (maximum noise fraction) techniques, however as a first approach the proposed model here provides interesting results, because even that each of the trails includes 1440 EEG signals mixed from 32 persons, the average of classification are very competitive, see table 3.

A different trial, are shown in the figure 6, a total of 32 persons where involved on a single analysis with random selected characteristics, from the same four emotions; Then were evaluated all the by a 85% training/test and 15% for validations. The two combinations of temporal analysis, showing similar performance than most of the reported recognition task table 3.

Table 3. Reported works and models

Autor	Reported Classification rate (\%)
Sun [25]	70-76
Lin [26]	69
Murugappan [7]	81*
Narajan [8]	91*
Yaacob [9]	93*
Daimi [5]	67-83

\*Trivial cases (single emotion recognition)

<sup>15</sup> Laborarotio de Computación Afectiva (LabCAfe)

<sup>16</sup> Myoelectric devices and signal acquisition systems.

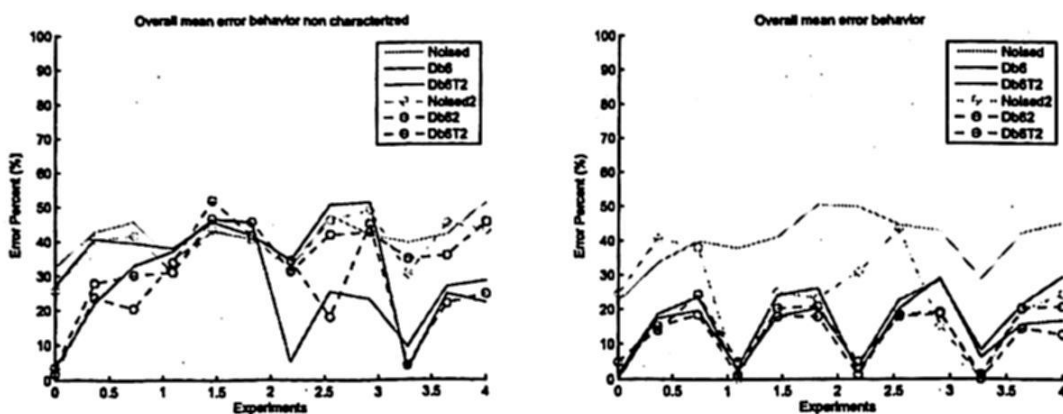


Fig. 6. Mean of 10 fold cross validations performance for two architectures

## 6 Conclusions

The implementation of systems that could interact with humans in a more natural way, are today a more feasible idea with technological advances, like powerful mobile devices and the reduction in the size of these, also the advances in smart computing enables to solve a complex task, in a more reliable and flexible way, this means that the need to explore new ways to apply the biological signals analysis are growing.

The emotion rule incentive introduced are a model that pretends interact with any device as translator between computer and human. Note that the main contribution lies in the EIM model, and the results that shows on its implementation versus the results provided several other models that reports emotions recognition systems [28] [29] [30] [31]. The proposed incentive rule based systems are one of the first approached models generated to be included as a computational model, not as virtual agent. Also one of the great challenges of these processes is that they lack generalizable data sources, but on the other hand research of emotion recognition are a increasing field, and the combination of disciplines generates less complex models, that can perform up to acceptable rates <sup>17</sup>, to sustain the acquisition of rules directly from the human, to provide a natural feedback which allows to a system adapt to its user and provide a better interaction.

## 7 Acknowledgments

To the ITT postgraduate department and Conacyt.

## References

1. Rosalind W. Picard: *Affective Computing*, MIT Press, (1997).
2. R. W. Picard and J. Healey: *Affective Wearables*, MIT Media Laboratory, Cambridge, MA, USA (1997).

<sup>17</sup> Near to acceptable rate of recognition, up to 90%

3. Jerald Yoo, LongYan, Dina El-Damak, Muhammad Awais Bin Altaf, Ali H. Shoeb, and Anantha P. Chandrakasan: An 8-Channel Scalable EEG Acquisition SoC With Patient-Specific Seizure Classification and Recording Processor, *IEEE Journal of solid state circuits*, vol. 48, no. 1, January (2013).
4. J. Wang and Y. Gong: Recognition of multiple drivers emotional state, *Proc. Int. Conference. Pattern Recognition* (2008).
5. Daimi Syed Naser, Goutam Saha: Recognition of Emotions Induced by Music Videos Using DT-CWPT, *Indian Conference on Medical Informatics and Telemedicine (ICMIT)* (2013).
6. S. Koelstra, C. Muehl, M. Soleymani, J.-S. Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt, I. Patras, : "DEAP: A Database for Emotion Analysis using Physiological Signals (PDF)" *IEEE Transaction on Affective Computing, Special Issue on Naturalistic Affect Resources for System Building and Evaluation*, in press (2011).
7. M. Murugapan, R. Narajan and S. Yaacob: Discrete wavelet transform based selection of EEG frequency for assessing human emotions, *journal of bio-medical Science and engineering*, vol 3 pp.390-936 (2010).
8. M. Murugapan, R. Narajan and S. Yaacob: Inferring of human emotional states using multi-channel EEG, *European Journal of Scientific research ISSN 1450-216X Vol. 48 No. 2 pp. 281-299* (2010).
9. M. Murugapan, R. Narajan and S. Yaacob: Classification of human emotion from EEG, using discrete wavelet transform, *Journal of bio-medical Science and engineering*, (2010).
10. Palle E.T. Jorgensen, K. A. Ribet *Analysis and Probability Wavelets, Signals, Fractals*, Springer Science, ISBN-10:0-387-29519-4, (2006).
11. Daniel Alpai *Wavelets, multiscale systems and hypercomplex analysis*, Verlag Basel, ISBN 3-7643-7587- 6, (2006).
12. Hans-Georg Stark: *Wavelets and Signal Processing*, ISBN 3-540-23433-0 Springer Berlin Heidelberg New York (2005).
13. Stuart Ira Fox *Human Physiology*, Sevent edition, Mc Graw Hill, US, NY, (2002).
14. Lauralee Sherwood *Human Physiology from cells to system*, Cengage Learning 161, ISBN-13: 978-1- 111-57743-8, (2013).
15. Sander, D., Grandjean, D., Pourtois, G., Schwartz, S., Seghier, M.L., Scherer, K.R., & Vuilleumier, P. Emotion and attention interactions in social cognition: Brain regions involved in processing anger prosody, *NeuroImage*, 28, 848858, (2005).
16. C. Peter, E. Ebert: *Affective Information Processing* Springer (2009).
17. Joseph E. LeDoux *Emotion circuits in the brain*, Annual Revision Neurocience, pages. 154184, (2000).
18. R. Plutchik: The nature of emotions *American Scientist*, vol. 89, p. 344, (2001).
19. K.R. Scherer, : "What are emotions? And how can they be measured", *Social Science Information*, vol. 44, no. 4, pp. 695-729, (2005).
20. Cristina M. Alberini y Joseph E. LeDoux, Memory reconsolidation, *Current Biology*, Volume 23, Issue 17, Pages R746-R750, ISSN 0960-9822, 9 September (2013).
21. Torfi Sigurdsson, Valerie Doyere, Christopher K. Cain, Joseph E. LeDoux, Long-term potentiation in the amygdala: A cellular mechanism of fear learning and memory, *Neuropharmacology*, Volume 52, Issue 1, Pages 215-227, ISSN 0028-3908, January (2007).
22. P. Ekman, W. V. Friesen, M. OSullivan, A. Chan, I. Diacoyanni- Tarlatzis, K. Heider, R. Krause, W. A. LeCompte, T. Pitcairn, and P. E. Ricci-Bitti: Universals and cultural differences in the judgments of facial expressions of emotion. *Journal of Personality and Social Psychology*, vol. 53, no. 4, pp. 712-717 (1987).
23. W. G. Parrott: *Emotions in Social Psychology: Essential Readings*. Psychology Press, (2001).
24. J. A. Russell: A circumplex model of affect *Journal of Personality and Social Psychology*, vol. 39, no. 6, pp. 1161-1178 (1980).

25. Sun. S and Zhang. C: Adaptive feature extraction for EEG signal classification, IEEE transactions on medical and biological engineering and computing, vol 44 (10), pp,931-935 (2006).
26. Lin. Y.P, Wang C.H, Wu. T.L, Jeng. S.K and Chen. J.H: Multilayer perceptron for EEG signal classification. In: Tencon 2007,pp 931-935 (2007).
27. Elliot M. Forney and Charles W. Anderson, Classification of EEG During Imagined Mental Task by Forecasting with Elman Recurrent Neural Networks, International Joint Conference on Neural Networks, San Jose, California, USA, July 31 August 5, (2011). Leon E, Clarke G, Callaghan V, Doctor F Affect-aware behaviour modelling and control inside an intelligent environment, Pervasive and Mobile Computing, Elsevier B.V, Volume 6, Issue 4, 2010.
28. Leon E, Clarke G, Callaghan, Sepulveda F, A user-independent real-time emotion recognition system for software agents in domestic environments, Engineering Applications of Artificial Intelligence, Volume 20, Issue 3, Pages 337-345, 2007.
29. E. Leon, with G. Clarke, V. Callaghan, F. Sepulveda Real-time detection of emotional changes for inhabited environments. Computers & Graphics Journal Special Issue on Pervasive Computing and Ambient Intelligence, 28(5): 635-642 2004
30. E. Leon, G. Clarke, F. Sepulveda, V. Callaghan Real-time Physiological Emotion Detection Mechanisms: Effects of Exercise and Affect Intensity 27th Conference of the IEEE Engineering in Medicine and Biology Society, Shanghai, pp. 4719-4722, September 1-4, 2005E.
31. E. Leon, G. Clarke, F. Sepulveda, V. Callaghan (2004) Optimised Attribute Selection for Emotion Classification Using Physiological Signals. Proceedings of the 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, San Francisco, p184 187, 1-4th Sept 2004



# Neurofuzzy Identification Applied to a Flow Control Equipment

William Torres Hernández<sup>1</sup>, Rosalío Farfán Martínez<sup>1</sup>, José A. Ruz Hernández<sup>2</sup>,  
Ramón García Hernández<sup>2</sup>, José L. Rullán Lara<sup>2</sup>

<sup>1</sup>Universidad Tecnológica de Campeche, Carretera Federal 180 s/n, San Antonio Cárdenas,  
Carmen, Campeche. México. C.P. 24381. Tel: 01 (938) 3816700, ext. 121  
{williantorreshernandez, farfan678}@hotmail.com

<sup>2</sup>Universidad Autónoma del Carmen, Calle 56, #4 esq. Avenida Concordia, Col. Benito Juárez,  
Carmen, Campeche, México. C.P. 24180. Tel: 01 (938) 3811018, ext. 1700  
{jruez, rghernandez, jrullan}@pampano.unacar.mx

*Paper received on 12/11/13, Accepted on 01/19/14.*

**Abstract.** In this paper a neurofuzzy identification scheme is designed for a flow control equipment. The identification procedure includes data collect, Adaptive Neuro Fuzzy Inference System (ANFIS) training and validation with data fresh. ANFIS training is performed online using a Pseudo-Binary Random Signal (PBRs) in order to obtain a neurofuzzy model. The feasibility of the proposed neurofuzzy identification scheme is validated in real time.

**Keywords:** Neurofuzzy Identification, ANFIS, Flow.

## 1 Introduction

Numerous advances in science have resulted in new research areas which are modeled on natural behavior of human beings, one of these fields is called Artificial Intelligence, which uses various techniques that mimic the processes of learning, reasoning and making decisions produced in the brain, but applied and directed to objects or systems and thereby provide intelligence. Although this is a relatively new field, since its inception with the contributions of scientists as Lotfi A. Zadeh in 1965 and J. J. Hopfield in 1982 among others, artificial intelligence techniques have been the subject of great interest and now smart devices or systems are in many cases replaces conventional (Ching Tai Lin & C. S., 1986).

One of the problems for the implementation of automatic control systems is to obtain a model that describes the system dynamics to be controlled. Usually this model is not available or is too complicated for design purposes. Therefore it is important to have a simple model to work with him, but that includes the essential features of the

process (Chiasson & Bodson, 1993). Models using neurofuzzy systems are useful to estimate from experimental data where the nonlinearities are included. The ANFIS model allows systems with high nonlinearity and time-invariant which combines the concepts of neural networks and fuzzy logic to form an intelligent system that highlights the ability for adaptation and automatic learning.

## 2 Description of experimental equipment

The flow measurements have a great importance in the processes and are commonly used for process control and accounting measures (turnover, import / export products), so selecting the best technology has great implication. For example, flowmeters are used to account products within the plant itself, with the outside. As for the process control, flow measurement is essential to perform automatic control and to optimize yields in production units applying material balances for this cause the flow to be measured and controlled carefully (Smith y Corripio, 2010).

In Fig. 1, a block diagram of the interconnected elements used in this work is shown, such as: A PWM voltage regulator module, centrifugal water pump, a flow sensor and data acquisition card DAQ PCI6071E.

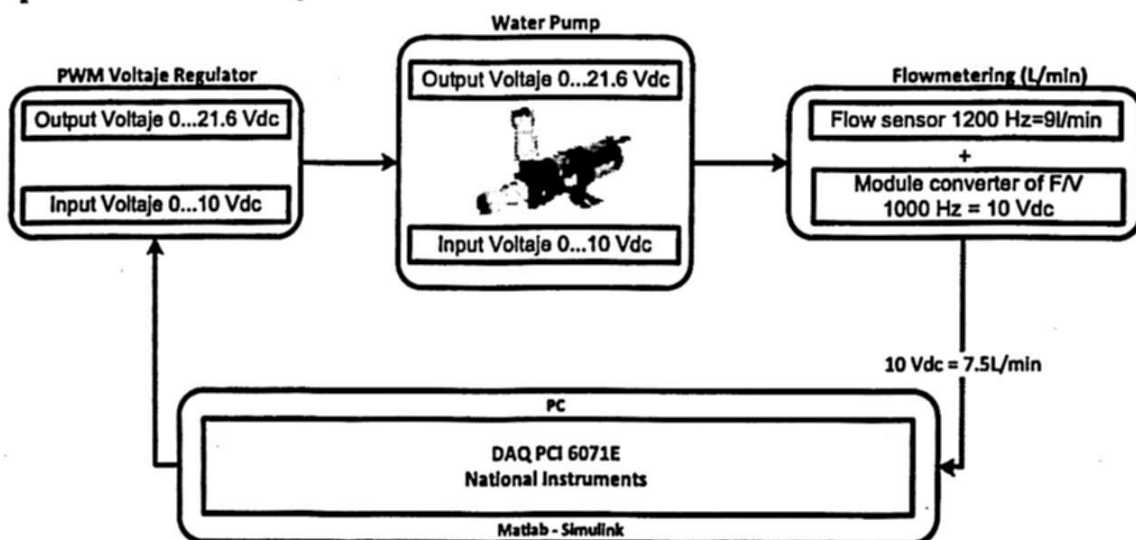


Fig. 1. Block diagram of the connection for flow regulating equipment.

This connection is made to obtain input-output data signals for process.

The pump used in this paper is a centrifugal pump brand Johnson CM30P7-1 model part number 10-24503-04 operates with a 24 Vdc and provides a maximum flow rate of 5 l/min., The module PWM controller brand Kaleja model D-73553 is a DC-DC used to control the voltage applied to the pump drive.

In Fig 2 the flow sensor (IR Opflow Type 2) is shown.

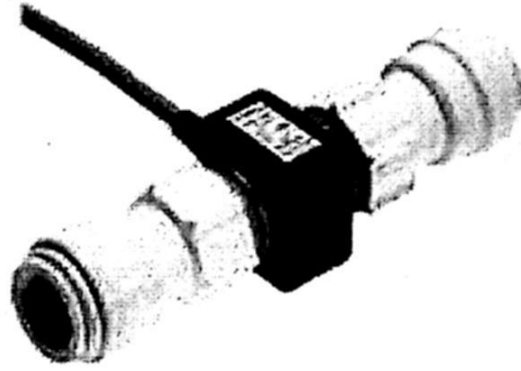


Fig 2. flow sensor (IR Opflow Type 2).

Its measurement range is 0.3-9.0 l / min and the frequency range of 40-1200 Hz output, and a K factor of 8000 pulse/dm<sup>3</sup>, Equation 1 is deduced.

$$1200 \text{ Hz} = 9 \text{ l/min} \quad (1)$$

The flow sensor has as output signal proportional to the amount of liters per minute making it necessary to use a converter frequency to voltage frequency. By design inverter v/f the maximum input frequency which can be applied to sensor is 1000 Hz by providing at its output 10 Vdc, Equation 2.

$$1000 \text{ Hz} = 10 \text{ Vdc} \quad (2)$$

Based on equation (1) and (2) the maximum flow that can be measured are:

$$1000 \text{ Hz} \frac{9 \text{ l/min}}{1200 \text{ Hz}} = 7.5 \text{ l/min} \quad (3)$$

So that the following relationship is obtained:

$$10 \text{ Vdc} = 7.5 \text{ l/min} \quad (4)$$

The maximum voltage that the inverter will provide with a flow rate of 5 l/min is:

$$(5 \text{ l/min}) \frac{10 \text{ Vdc}}{7.5 \text{ l/min}} = 6.66 \text{ V} \quad (5)$$

because of this there is no risk of saturation of the sensor and converter.

Fig 3 shows the electrical connection diagram.

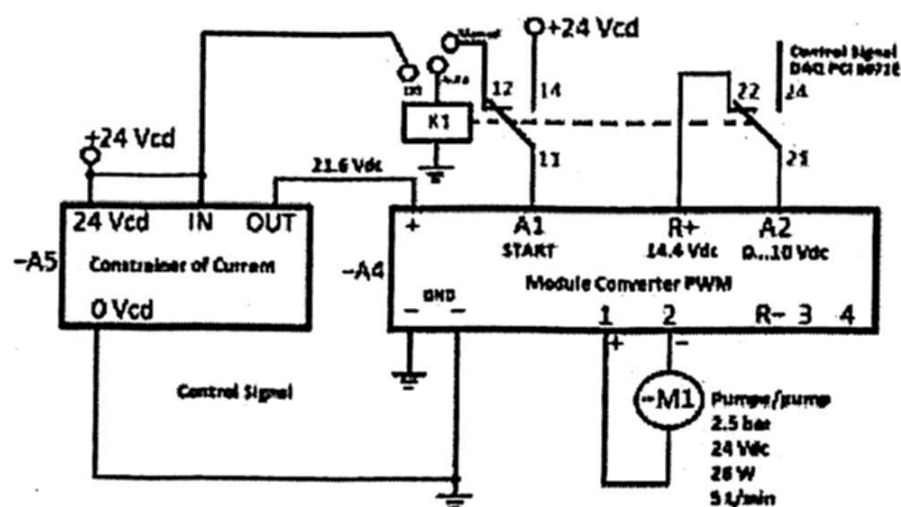


Fig 3. Electrical connection diagram PWM drive module.

### 3 Data Acquisition

The first step in the identification process is to perform a kind of experiment in the system studied, to collect input-output data are used to obtain the final model. To generate these data the experimental equipment is shown in Fig 4.

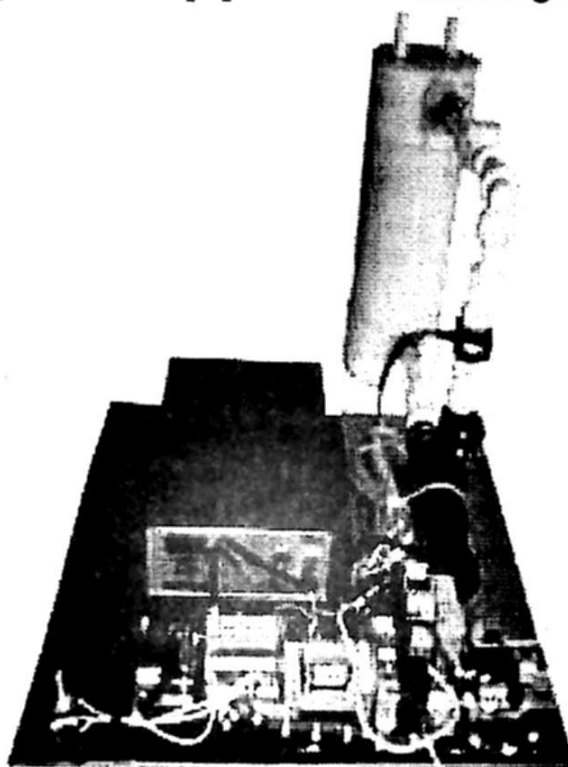


Fig 4. Experimental equipment.

When working with real-time systems, the sampling time also depends on some factors such as processing speed of data acquisition card and runtime model code, and has a decisive influence on the identification experiment.

The selected sampling period is  $T = 0.01$  seconds, the time of data acquisition for this study was 300 seconds obtaining a total of 30000 data which were divided in 15000 for training and the remaining 15000 for neurofuzzy model validation.

Identification scheme used is series-parallel, which is shown in Fig. 5. The error  $e(k)$  obtained by the difference between the response of the plant and the identifier is

used as a performance index in order to satisfy a set of input-output data in parameter estimation process.

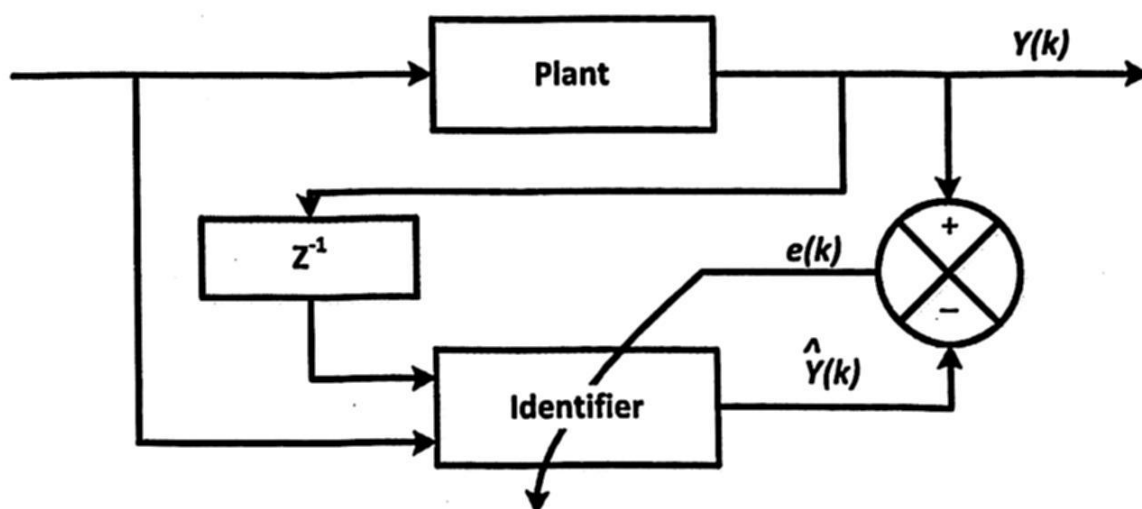


Fig 5. Series-Parallel Identification Scheme

In the data acquisition PRBS was used see Fig 6. For generation of the PRBS a 8 bits shift register with feedback to the first stage of the shift register is used by an exclusive OR operation in the registers 2, 3, 4 and 8, for a period of the sequence of 255.

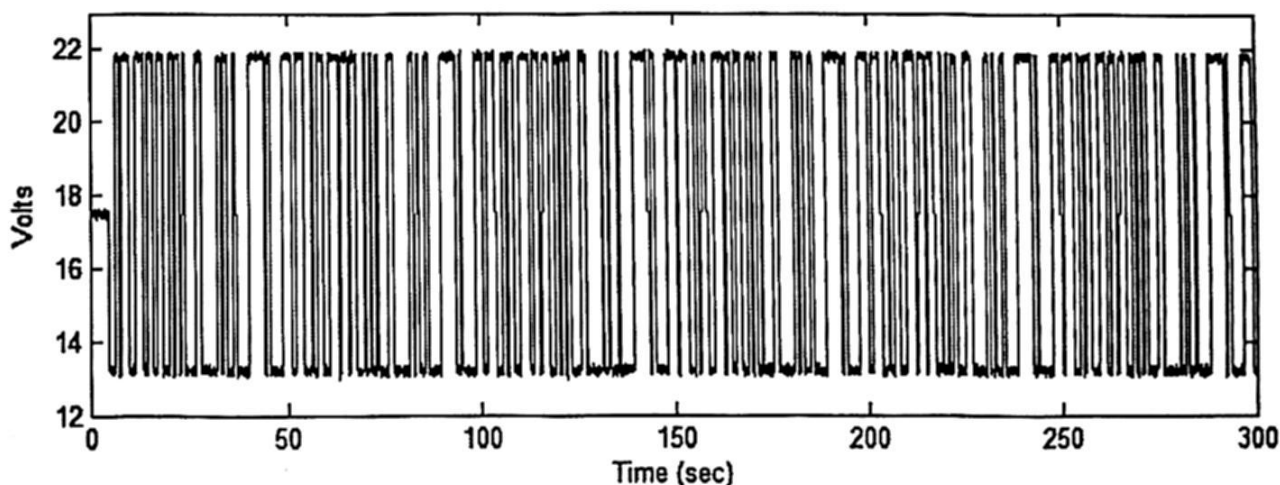


Fig. 6. Pseudorandom binary sequence, signal voltage applied to the pump.

ANFIS have not a recursive structure, past values of the inputs and outputs are used to capture the dynamics plant (Saludes Rodil & J. Fuente, 2007). Previously obtained data from voltage and flow can be used for training tool ANFIS EDIT using Matlab.

Fig. 7 shows the signal from the sensor (l/min.).



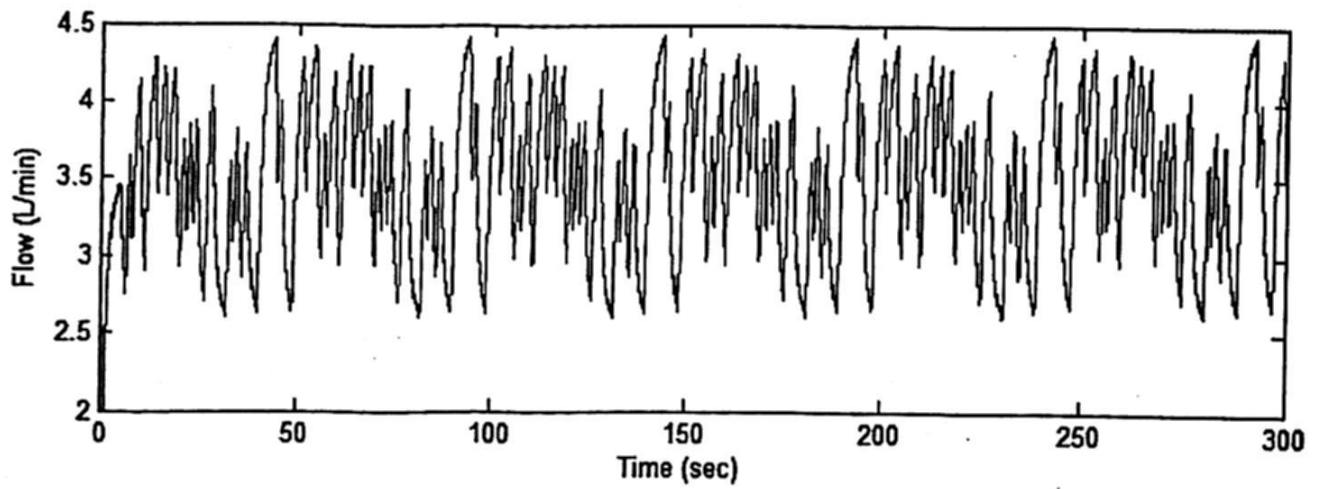


Fig. 7. Experimental data obtained for the sensor.

#### 4 Obtaining model neurofuzzy

The purpose of training is to obtain an elevational tuning neurofuzzy adaptive model adaptive advantage of artificial neural networks, seeking convergence near zero error, this convergence ensures that the model and validate the neurofuzzy be compared to the actual flow output team with the departure of the ANFIS network, the error between the two outputs is close to zero.

In complex systems, it is not always easy to determine the variables that can be used as inputs model, however, through the analysis of dynamic systems is known that for output system at any time  $t \geq t_0$ , it is necessary know the system at time  $t = t_0$  and the system input at time  $t \geq t_0$ . (Ogata, 1998). Based on this you can determine the inputs that are necessary to perform the routine training ANFIS network of an arbitrary system. In particular you can determine that the team dynamic flow control is completely characterized by the knowledge of the applied variables voltage (V) is the system input and flow rate (Q) is the system output. Thus the flow of equipment at any time  $t \geq t_0$  be determined by the knowledge of both the flow and the voltage applied at time  $t = t_0$ , as well as the flow for time  $t \geq t_0$ . In other words if you want to find the flow in an instant  $k+1$  is necessary to know the flow rate, applied voltage and  $k$ , (Ogata, Control Systems in Discrete Time, 1996). According to the above we can write:

$$Q(k+1) = f(V(k), Q(k)) \quad (6)$$

The flow rate  $Q$  at time  $k$  can be represented as:

$$Q(k) = f(V(k-1), Q(k-1)) \quad (7)$$

In the Toolbox ANFIS EDIT, the hybrid learning algorithm and fuzzy inference system (Takagi-Sugeno) T-S 5 membership functions for the Gaussian bell input into the layer 1, the hybrid learning algorithm and the consequent used T-S type fuzzy rules in layer 4. In Fig. 8 the structure for ANFIS network is shown.

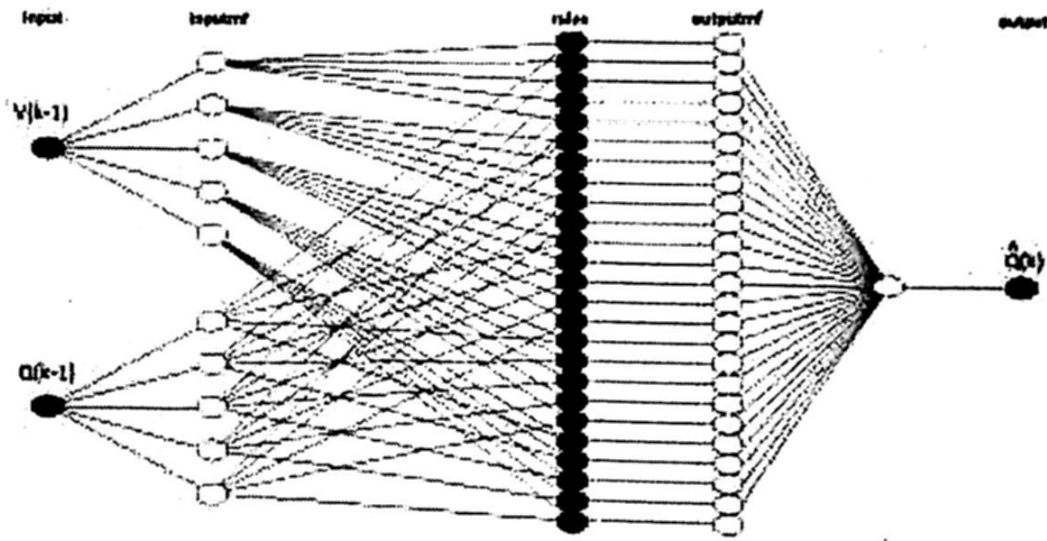


Fig. 8. Structure for neurofuzzy ANFIS Model.

In Fig. 9 neurofuzzy response model validation data is shown. It can be seen total convergence between actual output data of the plant and the neurofuzzy model.

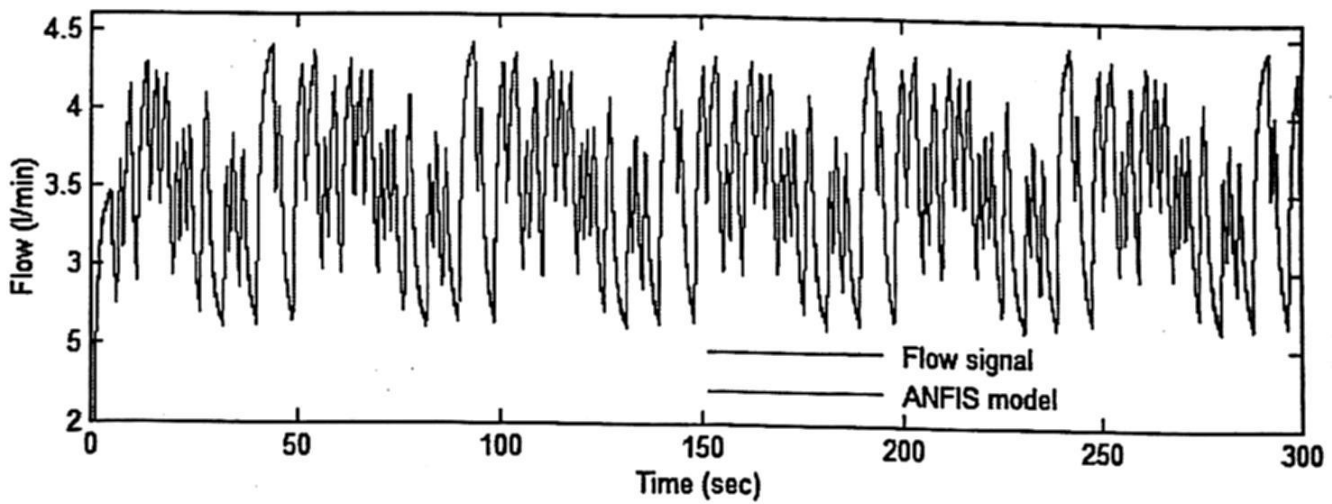


Fig. 9. Neurofuzzy model validation.

In Fig. 10 the prediction error between the neurofuzzy model and the plant is shown.

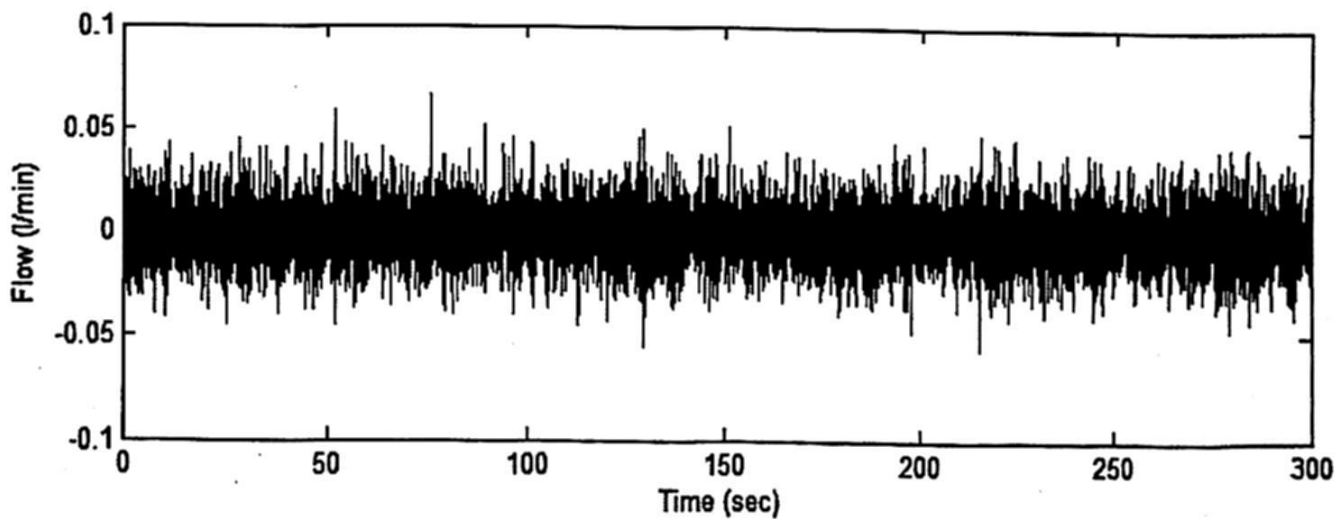


Fig. 10. Prediction error  $e(k)$ .

## 5 Real-time validation

For real-time validation a sinusoidal reference signal is applied. The flow signal and estimated flow are shown in Fig. 11.

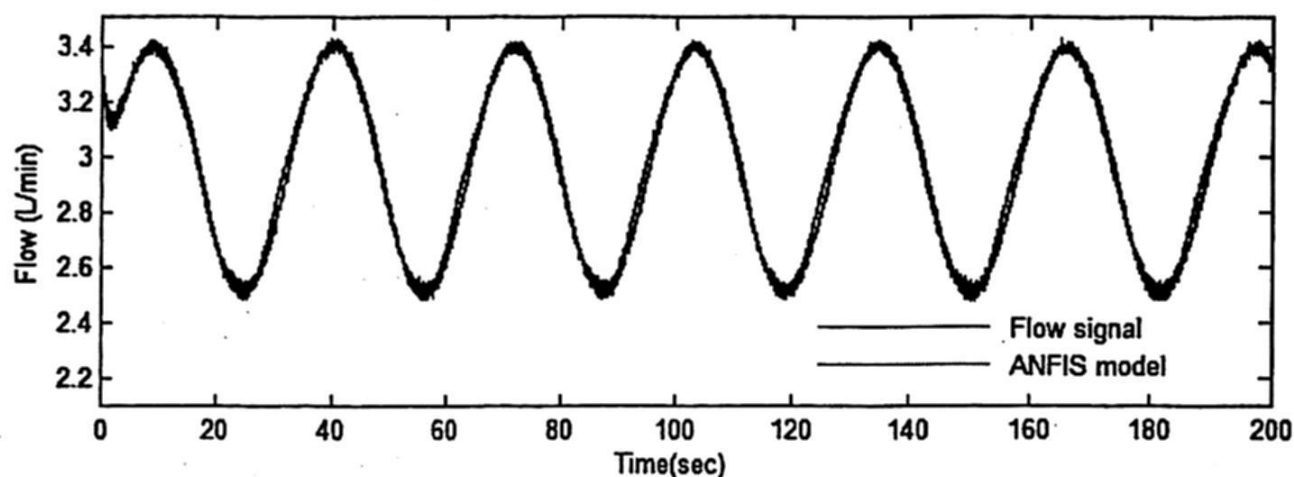


Fig. 11. Flow signal and estimated flow in real time.

## 6 Conclusion

The proposed neurofuzzy identification scheme was validated in real time using a flow control equipment. For process modeling is very useful to use the technique of Neurofuzzy identification due to it gives very good much information about the system. This technique is based on obtaining data directly from the computer through sensors.

## References

1. Ching Tai Lin, C., & C. S., G. (1986). *Neural Fuzzy Systems: A Neuro-fuzzy Synergism to Intelligent Systems*. Prentice Hall.
2. Chiasson, J., & Bodson, M. (1993). "Technical Notes and Correspondence Nonlinear Control of a Shunt DC". *IEEE Transactions on Automatic control*, 38(11), 935-942.
3. Saludes Rodil, S., & J. Fuente, M. (2007). Control IMC no Lineal Tolerante a Fallos. *Revista Iberoamericana de Automática e Informática Industrial*, 52-63.
4. Smith, C., & Corripio, A. (2010). *Control Automático de Porcesos Teoría y Práctica*. México: Limusa S.A. de C.V.
5. J. G. Pacheco Richard, J. A. Ruz-Hernandez, and E. Shelomov. Tecnicas neurodifusas aplicadas al control del equipo de la bola y la viga. In *Proceedings of Seminario Anual de Automatica, Electronica Industrial e Instrumentacion*, pages 1-6, Vigo, Spain, 2003.
6. B. Allaoua, A. Laoufi, B. Gasbaoui, and A. Abderrahmani. Neuro-fuzzy dc motor speed control using particle swarm optimization. *Leonardo Electronic Journal of Practices and Technologies*, (15):1-18, 2009.
7. W. Torres Hernandez, R. Farfan Martinez, E. Reyes-Pacheco, J. Rullan-Lara, and J. A. Ruz-Hernandez. Identificacion y control de velocidad de un motor de corriente directa. In *Proceedings of VIII Congreso Nacional de Ingenieria Electrica y Electronica del Mayab*, pages 217-227, Merida, Yucatan, Mexico, 2008.

8. J-S R. Jang and C-T Sun. Neuro-fuzzy modeling and control. In Proceedings of the IEEE, pages 378–406, 1995.
9. N. Pitalua, U. Castro Peñaloza, J. A. Ruz-Hernandez, and R. Lagunas Jimenez. Introduccion a los Sistemas Inteligentes. Departamento de Editorial Universitaria (UABC), Mexicali, Baja California, Mexico, 2008.
10. J-S R. Jang. ANFIS: Adaptive-network-based fuzzy inference system. IEEE Transactions on Systems, Man, and Cybernetics, 23:665–685, 1993.





## **Part III**

# **Information and communications technology**



# Remote Sensing Image Processing with Graph Cut of Binary Partition Trees

Philippe Salembier<sup>1</sup> and Samuel Foucher<sup>2</sup>

<sup>1</sup> Technical University of Catalonia, Barcelona, Spain

<sup>2</sup> Computer Research Institute of Montreal, Vision Team, Montreal, Canada  
*Invited Paper*

**Abstract.** This paper discusses the interest of hierarchical region-based representations of images such as Binary Partition Trees (BPTs) and the usefulness of graph cut to process them. BPTs can be considered as an initial abstraction from the signal in which raw pixels are grouped by similarity to form regions, which are hierarchically structured by inclusion in a tree. They provide multiple resolutions of description and easy access to subsets of regions. Their construction is often based on an iterative region-merging algorithm. Once constructed, BPTs can be used for many applications including filtering, segmentation, classification and object detection. Many processing strategies consist in populating the tree with features of interest for the application and in applying a specific graph cut called pruning. Different graph cut approaches are discussed and analyzed in the context of Polarimetric Synthetic Aperture Radar (PolSAR) images.

**Keywords:** Binary Partition Tree, graph cut, pruning, PolSAR images, super-pixel partition.

## 1 Introduction

Remote sensing technologies are currently undergoing an important evolution in terms of the quality and the quantity of information that is acquired. Sensors are able to capture data at an increasing resolution both in terms of spatial and spectral resolutions. This wealth of information generates real challenges with respect to signal processing tasks. One of the major issues is concerned with the handling of the signal correlation.

The traditional pixel-based image representation is not the most appropriate one to deal with the huge amount of information produced by high resolution remote sensing sensors while being able to deal with the signal correlation. A more appropriate representation should somehow group pixels of similar properties into elementary entities that should be easily handled and accessed. Moreover, the representation should be useful for different applications and therefore describe the data at multiple resolutions.

Recently, the interest of Binary Partition Trees (BPTs) [9] has been investigated for remote sensing including SAR [2] and hyperspectral images [10].

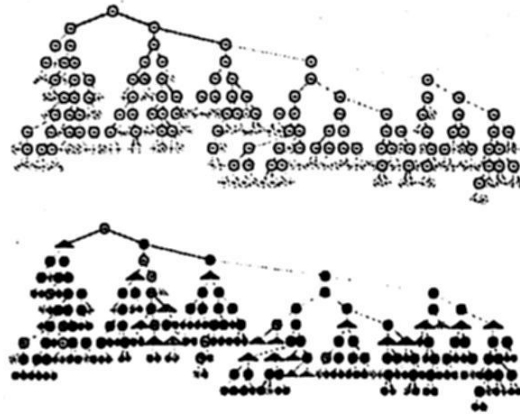


Fig. 1. Example of BPT (Top) and of pruning (Bottom).

BPTs are region-based representations in which pixels are grouped by similarity. They provide multiple resolutions of description and easy access to subsets of regions. Their construction is often based on an iterative region-merging algorithm: starting from an initial partition, the pair of most similar neighboring regions is iteratively merged until one region representing the entire image support is obtained. The BPT essentially stores the complete merging sequence in a tree structure. Once constructed, BPTs can be used for a large number of tasks including image filtering, object detection or classification [3].

The processing of BPTs often relies on a specific type of graph cut called *pruning*. In this paper, we will formally define this notion and show how it can be used to extract partitions from the tree. Then, several examples of useful pruning are presented and analyzed in the context of PolSAR images.

The paper is organized as follows: Sec. 2 discusses the principles of BPT creation and their processing by means of graph cut. A possible way to evaluate the quality of a BPT is presented in Sec. 3 and used in Sec. 4 to study the influence of the initial partition on the BPT construction. Sec. 5 presents and analyzes four pruning techniques for low-level processing of PolSAR images. Finally, conclusions are reported in Sec. 6.

## 2 Binary partition Tree creation and processing through graph cut

The BPT creation starts by the definition of an initial partition. The initial partition can be composed of individual pixels as in [2, 3]. While this strategy guarantees a high precision as starting point of the merging process, it also implies high computational and memory costs as a huge number of regions have to be handled. As an alternative, the initial partition may correspond to an over-segmentation as a superpixel partition [1]. Once the initial partition is defined, the BPT construction is done by iteratively merging the pair of most similar neighboring regions.

In the PolSAR case, the information carried by pixels of an image  $I$  corresponds to the covariance matrix  $Z_{ij}^I$  of the scattering vector:  $\mathbf{k} = [S_{hh}, \sqrt{2}S_{hv}, S_{vv}]^T$

measured on the resolution cell at location  $(i, j)$ . The subindices  $h$  and  $v$  indicate the horizontal and vertical polarization states and  $S_{pq \in \{h, v\}}$  represents the complex SAR data where the polarization states employed in reception and transmission are given by  $p$  and  $q$  respectively. To construct the BPT, regions  $R$  can be modeled as in [2] by their mean covariance matrix  $\mathbf{Z}_R = \frac{1}{|R|} \sum_{i,j \in R} \mathbf{Z}_{ij}^I$ , where  $|R|$  is the region number of pixels. The distance between neighboring regions defining the merging order can be measured as in [3] by the geodesic similarity adapted to the cone of positive definite Hermitian matrices [4]:

$$S(R_1, R_2) = \|\log(\mathbf{Z}_{R_1}^{-1/2} \mathbf{Z}_{R_2} \mathbf{Z}_{R_1}^{-1/2})\|_F \cdot \ln\left(\frac{2|R_1||R_2|}{|R_1| + |R_2|}\right) \quad (1)$$

where  $\log(\cdot)$  is the matrix logarithm and  $\ln(\cdot)$  the natural logarithm. Using this similarity measure, regions are iteratively merged until a unique region representing the entire image support is obtained. After each merging, a new region is created, its mean covariance matrix is computed and its similarity with its neighbors is updated.

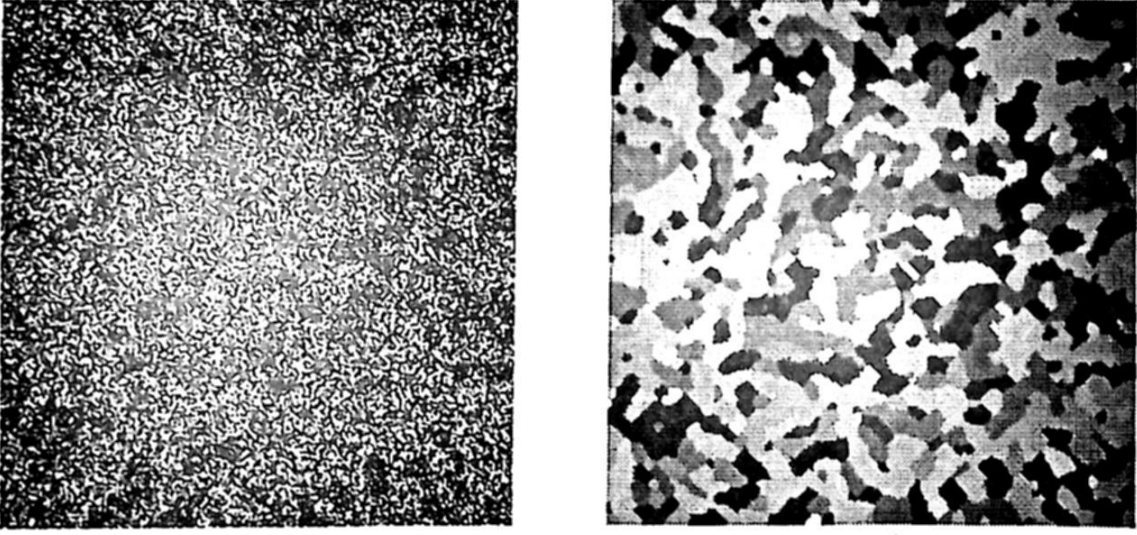
An example of BPT can be seen on the top part of Fig. 1. The regions belonging to the initial partition form the leaves of the BPT (shown with green circles in Fig. 1). In this example, only a small image portion has been used. The initial partition involves 108 regions and the BPT is therefore composed of 215 nodes. During the merging process, the BPT is constructed by creating a parent node for each pair of merged regions. In Fig. 1, the edge color represents the similarity value between the two merged regions: blue (red) indicates very similar (dissimilar) regions.

Once the BPT has been constructed, it can be used for a wide range of applications including filtering, segmentation or classification. A large number of applications relies on the extraction of a partition from the BPT. This process can be seen as a particular graph cut called *pruning*: Assume the tree root is connected to a *source* node and that all the tree leaves are connected to a *sink* node. A *pruning* is a graph cut that separates the tree into two connected components, one connected to the source and the other to the sink, in such a way that any pair of siblings falls in the same connected component. The connected component that includes the root node is itself a BPT and its leaves define a partition of the space. In the sequel, we discuss several examples of pruning in the context of PolSAR images. The first one will allow us to evaluate the quality of a BPT itself.

### 3 Graph cut to evaluate of the quality of a BPT

One of the major issues in PolSAR image is the speckle noise that results from the coherent integration of the scattered electromagnetic waves. Speckle filtering aims at reducing noise within homogeneous extended targets while preserving meaningful spatial details [5]. Insufficient noise reduction leads to important bias on derived polarimetric parameters therefore degrading the performance of





**Fig. 2.** Example of original PolSAR image (Left) and its corresponding ground-truth (Right). RGB-pauli color coding.

classification tasks or biophysical parameter retrieval. Most filters are based on adaptive strategies using a sliding square window with a fixed size [8, 7].

In this context of PolSAR images and speckle noise, we would like to be able to evaluate the quality of a BPT and see the influence on the constructed BPT of specific choices concerning the region model, the similarity function, the initial partition, etc. This is not a trivial task since a huge number of partitions can be extracted from a given BPT. However, here, we are concerned by low-level processing and by removing the speckle noise as much as possible to allow a precise estimation of the polarimetric parameters. As a result, we are going to rely on a dataset of PolSAR images on which the ground-truth polarimetric information is available. More precisely, we use the set of simulated PolSAR images [5] where the underlying ground-truth, i.e. the class regions, is modeled by Markov Random Fields. A set of typical polarimetric responses have been extracted from an AIRSAR image (L-band) so that they represent the 8 classes found in the  $H/\alpha$  plane and randomly assigned to each class. Finally, single look complex images have been generated from the polarimetric responses using a Cholesky decomposition [6]. An example of image and its corresponding ground-truth is presented in Fig. 2. Thanks to this dataset involving ground-truth information, we can measure the quality of a BPT.

Let us define the quality of a BPT as the quality of the best image, according to a given error measure  $E$ , that can be extracted from it. Extracting an image from the BPT consists in selecting a set of nodes forming a partition of the image and in assigning the mean covariance matrix of the region to its pixels. So the question is to identify the *ideal* partition that can be extracted from the tree.

The error measure between an image  $I(i, j)$  and the ground-truth image  $I_{GT}(i, j)$  we use is defined by [2]:

$$E(I, I_{GT}) = \frac{1}{N} \sum_{i,j} \frac{\|Z_{ij}^I - Z_{ij}^{I_{GT}}\|_F}{\|Z_{ij}^{I_{GT}}\|_F} \quad (2)$$

where  $N$  is the image number of pixels,  $\mathbf{Z}_{ij}^I$  ( $\mathbf{Z}_{ij}^{IGT}$ ) is the pixel value of image  $I$  ( $I_{GT}$ ) at location  $(i, j)$  and  $\|\cdot\|_F$  represents the Frobenius matrix norm. This measure is based on the average inverse signal to noise ratio.

As previously mentioned, the extraction of a partition from the BPT is defined by a pruning. To define the *ideal* pruning, let us use the following ideal criterion derived from Eq. 2 by noting that all pixels belonging to the same region  $R$  have the same covariance matrix  $\mathbf{Z}_R$ :

$$C_{\text{ideal}} = \sum_R \phi_R \text{ with } \phi_R = \sum_{i,j \in R} \frac{\|\mathbf{Z}_R - \mathbf{Z}_{ij}^{IGT}\|_F}{\|\mathbf{Z}_{ij}^{IGT}\|_F}, \text{ s.t. } \{R\} \text{ is a partition} \quad (3)$$

This criterion is ideal because it uses the ground-truth image  $\mathbf{Z}^{IGT}$  which will be unknown in practice. However, it is very useful to quantify the quality of a BPT and to define an upperbound on the performances of all possible pruning strategies.

This criterion can be efficiently minimized using an dynamic programming algorithm originally proposed in [9] for global optimization. The solution consists in propagating local decisions in a bottom-up fashion. To initialize the process, the leaves of the BPT are assumed to belong to the optimum partition. Then, one checks if it is better to represent the area covered by two sibling nodes as two independent regions  $\{R_1, R_2\}$  or as a single region  $R$  (the common parent node of  $R_1$  and  $R_2$ ). The selection of the best choice is done by comparing the criterion  $\phi_R$  evaluated on  $R$  with the sum of the costs  $\phi_{R_1}$  and  $\phi_{R_2}$ :

$$\text{If } \phi_R \leq \phi_{R_1} + \phi_{R_2} \begin{cases} \text{then } R \text{ belongs to the optimum partition} \\ \text{else } R_1 \text{ and } R_2 \text{ belong to the optimum partition} \end{cases} \quad (4)$$

The best choice (select either " $R$ " or " $R_1$  plus  $R_2$ ") is stored in the BPT node representing  $R$  together with the corresponding cost value (either  $\phi_R$  or  $\phi_{R_1} + \phi_{R_2}$ ). The procedure is iterated up to the root node and defines the best partition.

This algorithm guarantees to find the global optimum of the criterion on the tree because the criterion is additive with respect to regions. The bottom part of Fig. 1 shows the BPT after the nodes have been populated with the value  $\phi_R/|R|$  (as for edges, low (high) values are represented in blue (red)). The triangle-shaped nodes show where the pruning has been done. They form the leaves of the pruned BPT and equivalently the regions of the extracted partition. Then, each region is represented by its mean covariance matrix to create the filtered image. Finally, this image is used to compute the BPT quality with Eq. 2. Based on this strategy to evaluate the BPT quality, we can now investigate the impact of specific choices related to the BPT construction. As an example, we analyze the influence of the initial partitions in the following section.

Initial Partition	Ideal pruning $C_{ideal}$ (dB)	Number of regions of the initial partition	Acceleration factor w.r. to the pixel partition
Pixel	-16,12	65.536	1
Watershed	-13,53	7.720	46
SLIC (size=2)	-15,94	15.946	15
SLIC (size=3)	-15,38	7.257	56
SLIC (size=4)	-14,51	4.143	121

**Table 1.** Influence of the initial partition in terms of BPT quality and computational load. Results have been averaged over the entire dataset.

#### 4 Superpixel initial partition and its influence on the BPT quality

As previously mentioned, the initial partition used in [2, 3] was composed of individual pixels. The main drawback of this choice is the high number of initial regions and the corresponding cost in memory usage and computational load. To see whether the number of initial regions can be drastically reduced without losing too much in terms of quality, experiments with superpixel partitions have been conducted. Concretely, initial partitions have been generated either with a watershed applied on the vectorial image gradient and with the SLIC algorithm [1]. In both cases, only the diagonal elements of the covariance matrices have been used to generate the superpixel.

Thanks to the strategy presented in Sec. 3, we can compare the influence of various initial partitions on the BPT construction by extracting the ideal partition and measuring  $E(I, I_{GT})$ . The results are given in Table. 1. As can be seen, the best BPT is obtained with the pixel partition. However, the use of the SLIC superpixels almost preserves the BPT quality but can drastically reduce the complexity. It is therefore a very good alternative and, in the sequel, we use the SLIC (size 2) superpixels as initial partition.

#### 5 Low level processing through BPT pruning

This section discusses pruning techniques for low-level processing and grouping of PolSAR data. The main goal is to allow for a precise estimation of the polarimetric parameters. In the previous section, we have used an ideal pruning technique to assess the quality of the BPT. It defines the upperbound on the quality of the partitions that can be extracted from the BPT but it cannot be used in practice as it relies on the ground-truth data.

Here we study the interest of four pruning techniques useful in practical situations. The two first ones were proposed in [2]. The first one simply consists in following the merging sequence and in stopping the iterative merging process when a predefined number  $N_R$  of regions is obtained. Note that this can be viewed as a pruning of the BPT, but actually, there is no need to fully construct the BPT to compute the resulting partition.



The second pruning [2] consists in populating the tree nodes with a feature measuring the region homogeneity (difference between the pixel values and the region mean) given by:

$$\phi_R = \frac{1}{|R|} \sum_{i,j \in R} \frac{\|Z_{ij}^I - Z_R\|_F}{\|Z_R\|_F} \quad (5)$$

Once the tree has been populated, the feature value of each node is compared to a predefined threshold. Note that the feature value is expected to be rather high for large regions and low for small regions. In the extreme case of regions made of a single pixel,  $Z^R$  coincides with  $Z_{ij}^I$  and therefore  $\phi_R = 0$ . However, the feature of a parent node is not always larger or equal to the features of its siblings. To define the pruning, we have used the so-called *Max* rule [3] which consists in selecting on each branch the closest node to the root for which the homogeneity criterion is below the threshold.

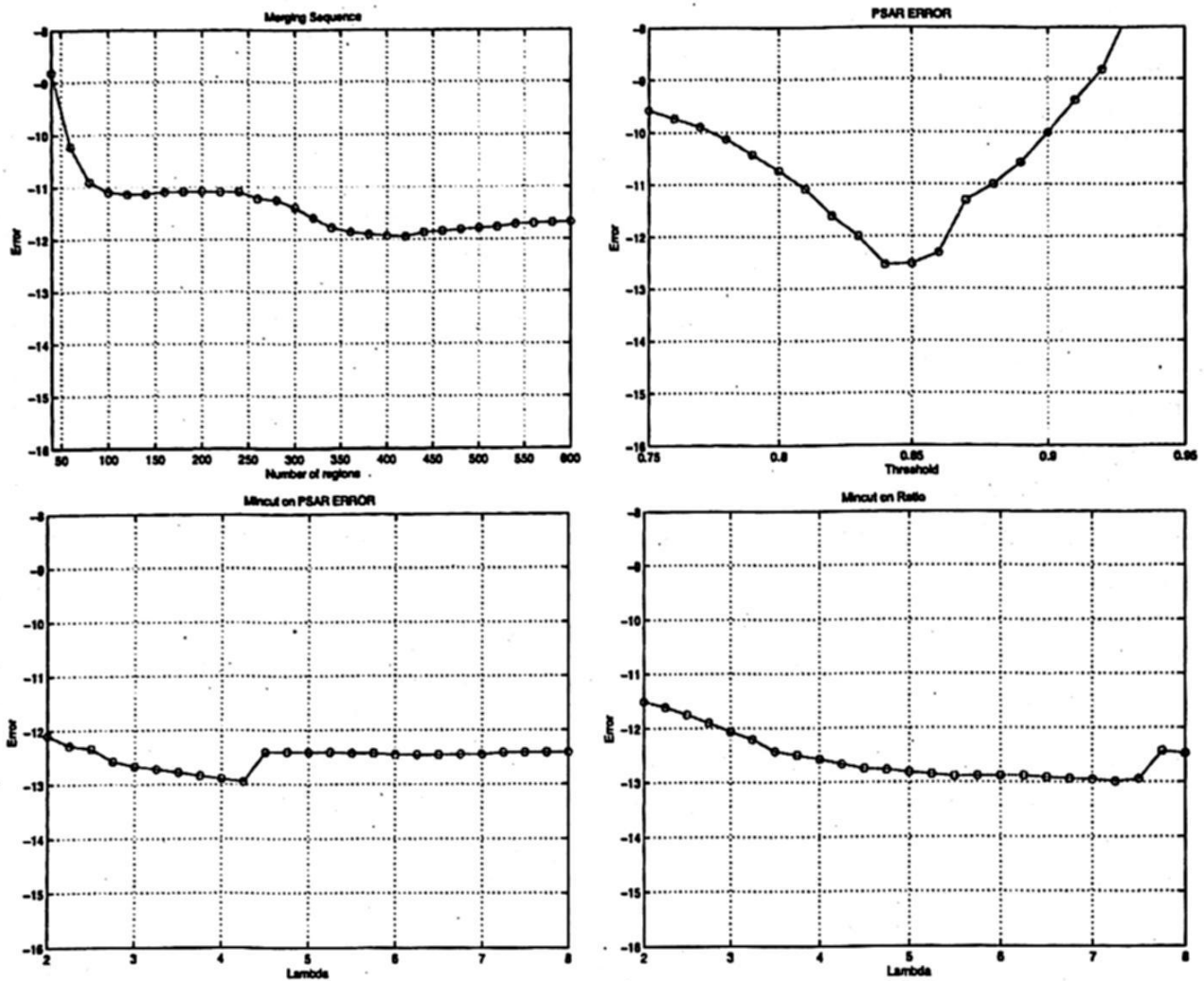
We introduce now two new pruning strategies based on the minimization of a global criterion as in Sec. 3. The initial idea is to use  $C = \sum_R \phi_R$  with  $\phi_R$  being the homogeneity criterion  $\phi_R = \sum_{i,j \in R} \|Z_{ij}^I - Z_R\|_F / \|Z_R\|_F$ . Note that this criterion is the same as the one defined by Eq. 5 without the averaging parameter  $|R|$ . However, on its own, it is not useful because a partition made of individual pixels sets the criterion to 0. Following classical approaches in functional optimization,  $\phi_R$  can be interpreted as a data fidelity term and combined with a data regularization term which encourages the optimization to find partitions with a reduced number of regions. As simple data regularization term, we use here a constant value  $\lambda$  that penalizes the region presence. Therefore the final homogeneity-based criterion to be minimized is:

$$C_{\text{Homog.}} = \sum_R \phi_R \text{ with } \phi_R = \sum_{i,j \in R} \frac{\|Z_{ij}^I - Z_R\|_F}{\|Z_R\|_F} + \lambda, \text{ s.t. } \{R\} \text{ is a partition} \quad (6)$$

Finally, the last pruning is also based on a graph cut minimizing a global criterion but here the idea relies on ratio filters: if the ideal image structure is known (here represented by  $Z_R$ ), then the ratio  $Z_{ij}^I / Z_R$  should only contain noise of variance 1 and no structure information. If the structure information is absent, the energy of the ratio should be minimum. This reasoning leads to the following minimization criterion involving as before a data fidelity term and a data regularization term:

$$C_{\text{Ratio}} = \sum_R \phi_R \text{ with } \phi_R = \sum_{i,j \in R} \left\| \frac{Z_{ij}^I}{Z_R} \right\|_F + \lambda, \text{ s.t. } \{R\} \text{ is a partition} \quad (7)$$

Fig. 3 shows, for the image of Fig. 2, the evolution of the error measure  $E(I_{\text{Processed}}, I_{\text{GT}})$  as a function of the pruning parameters: the number of regions for the first pruning, the threshold on the homogeneity value for the second pruning and the  $\lambda$  value for the remaining pruning. In terms of global error measure, we can see that the best pruning techniques are the two based on



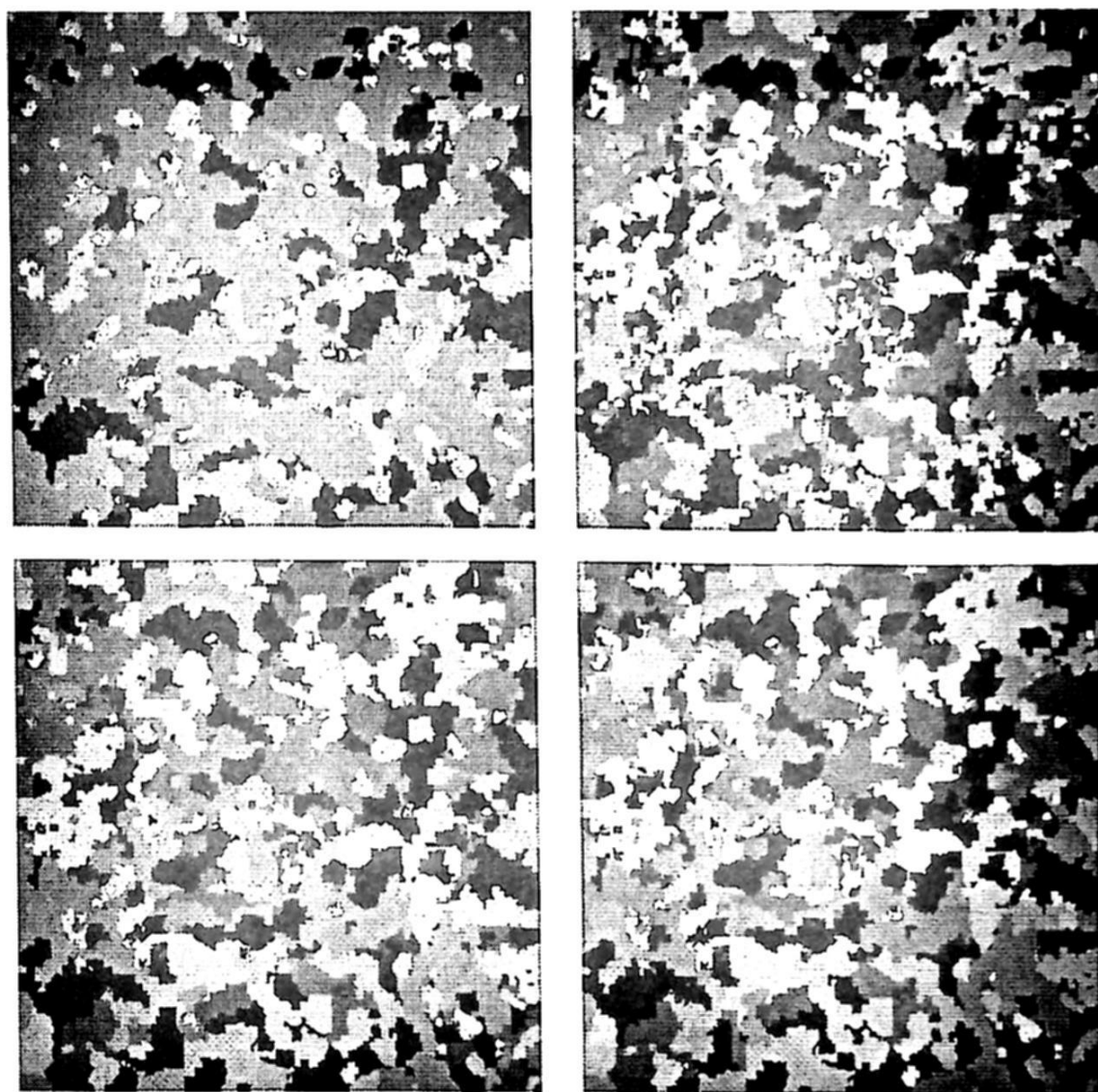
**Fig. 3.** Error measure  $E(I_{Processed}, I_{GT})$  in dB as a function of the pruning parameters. Top row: Merging sequence and threshold on homogeneity (Eq. 5). Bottom row: mincut on homogeneity criterion (Eq. 6) and on ratio image (Eq. 7).

global optimization (Eq. 6 and Eq. 7). They provide a lower value of the error and moreover their dependency on the  $\lambda$  value is smooth. The pruning following the merging sequence does not lead to the best estimation of the polarimetric parameters. This result highlights the interest of constructing the BPT to extract partitions that have not been observed during the merging process. Moreover, in practice, it is difficult to define a priori the appropriate number of regions. Finally, the pruning involving the thresholding on the homogeneity criterion provides intermediate results but we may also note that the value of the threshold has a very strong impact on the results.

Processing technique	Original image	Boxcar	Refined Lee	BPT: homog. mincut	BPT: ratio mincut
$E(I_{Processed}, I_{GT})$ in dB	-1,87	-9,11	-12,17	-14,57	-14,43

**Table 2.** Results of low-level processing (average over the entire dataset).

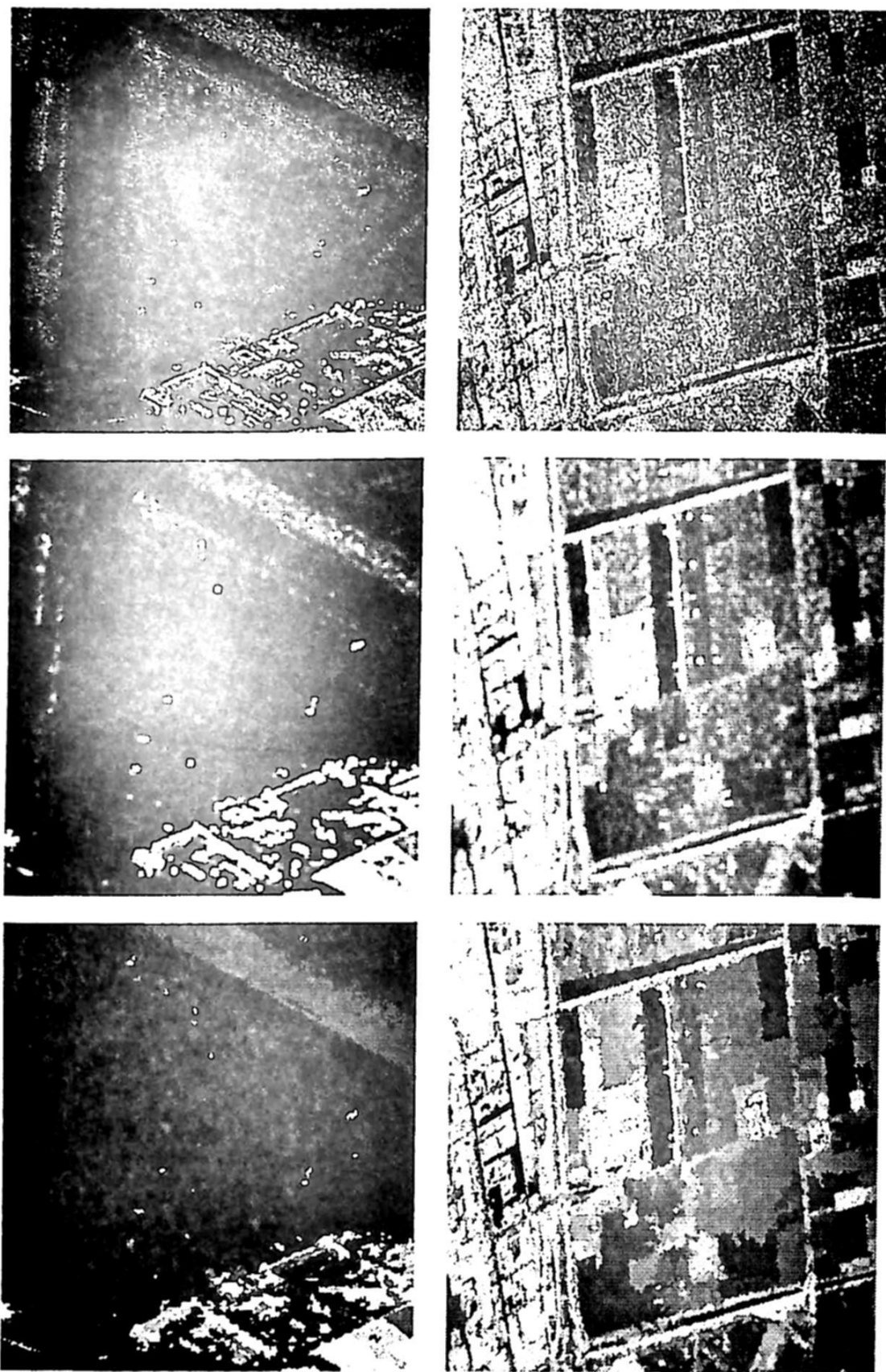




**Fig. 4.** Processed images with the optimum pruning parameters. Top row: Merging sequence and threshold on homogeneity (Eq. 5). Bottom row: mincut on homogeneity criterion (Eq. 6) and on ratio image (Eq. 7).

Fig. 4 shows the images resulting from the four pruning techniques using the optimum parameter in each case. The visual inspection of these images corroborates the analysis done on Fig. 3. The pruning following the merging sequence merges too many regions represented in dark and light green in the ground-truth image (Fig. 2). The image corresponding to the thresholding on the homogeneity criterion is still noisier than the one involving the global optimization.

The results obtained with best pruning techniques are given in Table 2 as well as results for classical filtering strategies. These results have been obtained by averaging  $E(I_{Processed}, I_{GT})$  over the entire database of 10 images. The table shows the interest of the pruning techniques involving global optimization. It also allows us to quantify the gap between these pruning techniques reaching about -14.5dB and the ideal pruning corresponding to almost -16dB (see Table 1). It can be concluded that there is still some room from improvement of the pruning.



**Fig. 5.** Results on real images. Top row: Original images (RGB Pauli composition). Middle row: Multilook filtering, Bottom row: Mincut on homogeneity criterion (Eq. 6).

Finally, results on applying the pruning with global optimization of the homogeneity (Eq. 6) are shown in Fig. 5 together with the original image and the result of the classical multilook filter. These results visually highlight the interest

of the BPT to perform a low-level processing of PolSAR images while preserving the spatial resolution of the content.

## 6 Conclusions

This paper has discussed the interest of Binary Partition Trees (BPTs) for remote sensing applications such as PolSAR. These hierarchical region-based representations of images are useful for many tasks. Here, we have mainly focussed on low-level processing of PolSAR covariance matrices. The paper has highlighted the usefulness of a specific type of graph cut called pruning that extracts partitions from the BPT. Specific pruning techniques have been defined to evaluate the quality of BPT and to perform low-level grouping allowing a precise estimation of the polarimetric information to be done without losing in terms of spatial resolution. In this context, the pruning techniques resulting from the global optimization of a criterion minimizing the region homogeneity or the energy of the ratio image have proved to be very efficient and robust.

## References

1. R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk. SLIC superpixels compared to state-of-the-art superpixel methods. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(11):2274 – 2282, May 2012.
2. A. Alonso-Gonzalez, C. Lopez-Martinez, and P. Salembier. Filtering and segmentation of polarimetric SAR data based on binary partition trees. *IEEE Transactions on Geoscience and Remote Sensing*, 50(2):593–605, 2012.
3. A. Alonso-Gonzalez, S. Valero, J. Chanussot, C. Lopez-Martinez, and P. Salembier. Processing multidimensional SAR and hyperspectral images with binary partition tree. *Proceedings of the IEEE*, 101(3):723–747, 2013.
4. F. Barbaresco. Interactions between symmetric cone and information geometries: Bruhat-tits and siegel spaces models for high resolution autoregressive doppler imagery. In *Emerging Trends in Visual Computing*, volume 5416, pages 124–163. Lecture Notes in Computer Science, F. Nielsen, Ed., Springer Berlin / Heidelberg, 2009.
5. S. Foucher and C. López-Martínez. An evaluation of PolSAR speckle filters. In *IEEE International Geoscience and Remote Sensing Symposium, IGARSS*, 2009.
6. J.-L. Lee, T.L. Ainsworth, J.P. Kelly, and C. López-Martínez. Evaluation and bias removal of multilook effect on entropy/alpha/anisotropy in polarimetric SAR decomposition. *IEEE Trans. Geoscience and Remote Sensing*, 46(10):3039–3051, 2008.
7. J.S. Lee, M.R. Grunes, and G. De Grandi. Polarimetric SAR speckle filtering and its implication for classification. *IEEE Trans. Geoscience and Remote Sensing*, 37(5):2363–2373, 1999.
8. J.S. Lee, M.R. Grunes, and S.A. Mango. Speckle reduction in multipolarization and multifrequency SAR imagery. *IEEE Trans. Geoscience and Remote Sensing*, 29(4):535–544, 1991.

9. P. Salembier and L. Garrido. Binary partition tree as an efficient representation for image processing, segmentation, and information retrieval. *IEEE Trans. on Image Processing*, 9(4):561 – 576, apr 2000.
10. S. Valero, P. Salembier, and J. Chanussot. Hyperspectral image representation and processing with binary partition trees. *IEEE Transactions on Image Processing*, 22(4):1430 – 1443, 2013.



# Homogeneous Quality Video in Multi-Sources Systems

Francisco de Asís López-Fuentes<sup>1</sup>

Networking and Distributed Systems Group

<sup>1</sup>Department of Information Technology

Universidad Autónoma Metropolitana - Cuajimalpa

Av. Vasco de Quiroga 4871, Santa Fe, Cuajimalpa de Morelos

053000 México, D. F.

{flopez}@correo.cua.uam.mx

*Paper received on 11/20/13, Accepted on 01/19/14.*

**Abstract.** Scalable video coding (SVC) has emerged as an important video standard to provide more functionality to video transmission and storage applications. This paper evaluates a strategy based on scalable video coding for peer-to-peer (P2P) video streaming services. This strategy uses SVC to reach a homogeneous video quality between different videos from different sources. Our scheme is implemented under the H.264/MPEG-4 SVC compression standard. Our evaluations were realized over a local area network (LAN). We evaluate our implementation in terms of overall throughput, delivery time (delay) and video quality. The obtained results show that our proposed solution strategies achieve good performance and introduce benefits in the peer-to-peer video streaming systems.

**Keywords:** scalable video coding, peer-to-peer systems, video streaming architectures.

## 1 Introduction

Video streaming over the Internet has gained significant popularity during the last years. This fact has generated a dramatic technological and social revolution in video distribution and consumption. People download videos from several online video portals or form community networks to share their common interest. Thus, video playback from on-line video or news site has become part of the daily life of most Internet users. Streaming video applications have a strong impact in different scenarios such as videoconferencing, Peer-to-peer (P2P) content distribution, or event broadcast (e.g. Internet Protocol Television (IPTV)). Different video streaming applications for live streaming or video on demand services have emerged as valuable tools to improve communication. Subsequently, many P2P media streaming systems such as ZigZag [1], CoolStreaming [2] or Mutualcast [3], have been developed. P2P paradigm has become a promising solution for video streaming, because it offers characteristics which cannot be provided by the client-server model. In contrast to client-server model, P2P networks do not have a single point of failure, the upload capacity is shared



among all peers, the bottlenecks are avoided, the contents can be shared by all participating peers, and they provide scalability. On the other hand, scalable Video Coding (SVC) can provide encoding of a high quality video bit stream, which contains some subset bit streams that can themselves be decoded with a complexity and reconstruction quality similar to that achieved using the existing H.264/AVC (Advanced Video Coding) design with the same quantity of data as in the subset bitstream [4], [5]. Using scalable video coding (SVC), parts of a video stream can be removed and the resulting sub-stream constitutes another valid video stream for some target decoder. In this way, we distribute video with different quality to requesting peers with different bandwidth characteristics or when the network characteristics are time-varying.

In this paper we use scalable video coding to adapt the flow rate from multiple sources in order to effectively use the available upload capacity from each source to deliver a homogeneous video quality for all streams. Our SVC scheme uses a peer-to-peer (P2P) structure described in previous work [9]. P2P structure used in this work is inspired from a scheme called Mutualcast [3], which is a scheme that reaches the maximum possible throughput. Our proposal scheme considers as participating peers to the source peers, the requesting peers and the helper peers. We assume that all requesting peers need to receive all videos. The helper peers are not interested in receiving the videos and just contribute their resources during distribution. We evaluate a particular case for a multi-source system using SVC for a reduced number of peers. Scalability of our scheme is limited, but it can be used in systems with a reduced number of participating nodes, such as video conferencing services or on-line games. The main contribution of this paper is to show how scalable video techniques can help to obtain a homogeneous video quality in LAN (local area network) applications with multi-sources and videos encoded with different bit-rates.

The rest of this paper is organized as follows. We first present the scalable video coding modes used to encode a video into layers in Section 2. Video streaming strategies based on scalable video coding (SVC) are discussed in Section 3. The first method aims to provide differentiated video quality according to the capacity of each node, while the second method aims to provide homogeneous quality to different resources of the network. Finally, Section 4 describes the manner in which these strategies are implemented and evaluated. Conclusions are given in Section 5 where we also suggest further work.

## 2 Scalable Video Coding Background

Scalable video coding (SVC) is a technique that encodes the video into layers. SVC is well established concept in the video area, and incorporates the following scalability modes [5], [6] and [7]:

- *Temporal scalability*: subset bitstream represents lower temporal resolution. With subset bitstream a part of frames in one group of pictures (GOP) can be decoded.
- *Spatial scalability*: lower subset bitstream can only playback a video with lower frames size.
- *SNR (Signal-to-noise ratio)/fidelity scalability*: the base layer bit stream can only playback a video of very low quality. And the more enhancement layers the client receives, the better quality the video has.

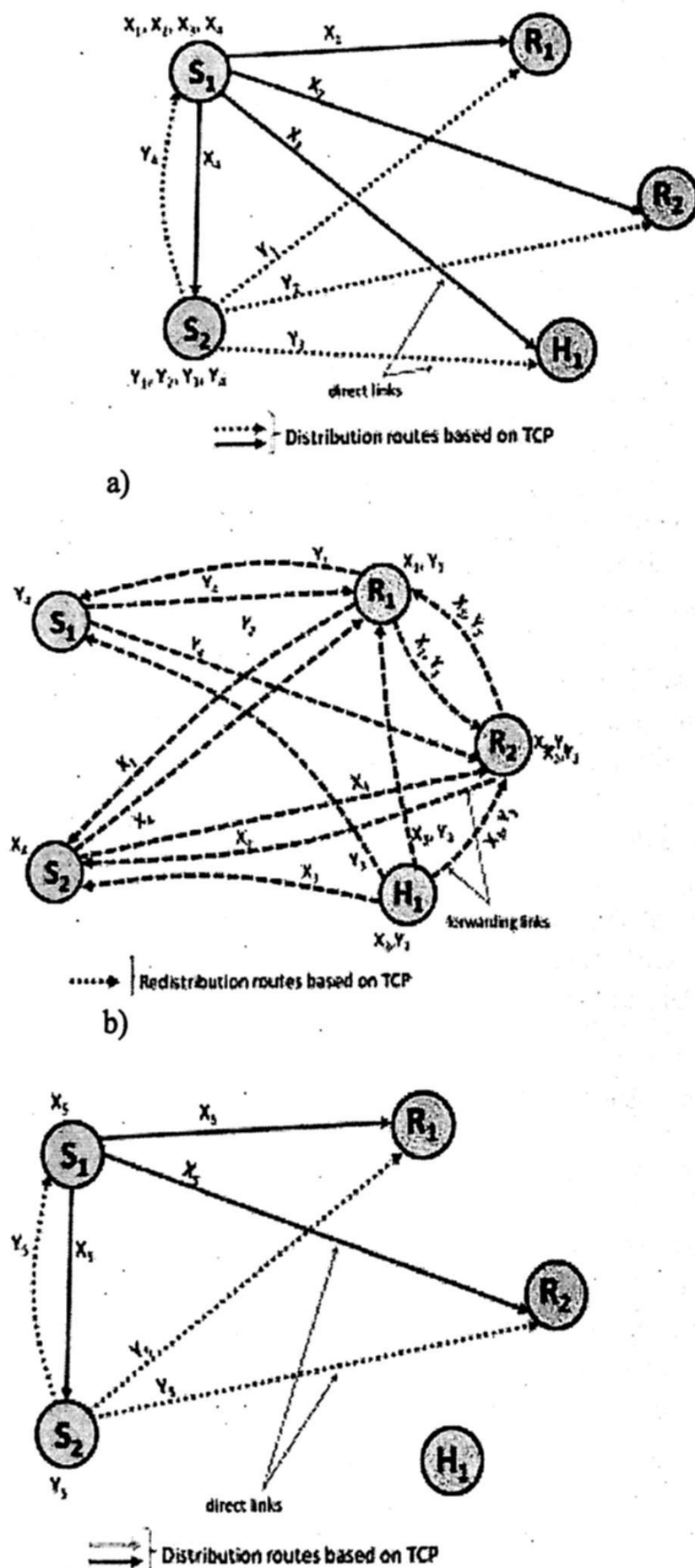
- **Combined scalability:** is a combination of all three or two modalities above. Scalable video coding (SVC) starts with the base layer, which contains the lowest level of spatial, temporal and quality perspective detail. Additional layers called enhancement layers can increase the quality of the video stream.

An enhancement layer is called a spatial enhancement layer when the spatial resolution changes with respect to its reference layer, and it is called a fidelity enhancement layer when the spatial resolution is identical to that of its reference layer [6]. SVC introduces new video coding techniques which provide the following features: reduced bit-rate, reduced spatial temporal resolution and coding efficiency comparable to non-scalable video systems. SVC extends the H.264/AVC (Advanced Video Coding) standard to enable video transmission to heterogeneous clients. To achieve it, SVC uses available system resources [6], in the case of the lack of a prior knowledge of the downstream client capabilities, resources, and variable network conditions. SVC provides a higher degree of error resiliency and video quality with no significant need for higher bandwidth. Also, scalable video coding can support a broad range of devices and networks. Thus, scalable video coding is a suitable solution for adaptive content delivery to end-users having various preferences, terminal capabilities, and network conditions [10]. Scalable video coding has been used for different scenarios and applications and several works can be found in the literature [11], [12], [13]. In this paper we propose to apply scalable video coding in P2P video streaming networks. We aim to provide homogeneous quality video to different resources of the network. The idea is to encode the videos with different rates, and source peers collaborate to ensure that the requesting peers will receive the videos with the same quality. The JSVM (Joint Scalable Video Model) software [7] is used as the codec to provide SNR (Signal-to-noise ratio) scalable bit streams. We do not compare our proposed model with other methods, because we implement our SVC strategy in a specific P2P multi-source infrastructure. Our proposed strategies also can be useful in scenario where networks with varying bandwidths and loss rates.

### 3 P2P Video Streaming Model

The proposed strategy uses scalable video coding to effectively exploit the available upload capacity from each source to deliver a homogeneous video quality for all streams in a multi-source structure [9]. We use the peak signal-to-noise ratio (PSNR) as a measure of video quality. This P2P scheme uses TCP, which has several limitations for video on wide area network (WAN), such as retransmission delay, packet loss, etc. However, for video applications in LAN (local area network), retransmission delay introduced by TCP (Transport Control Protocol) could be not significant [11]. In fact, retransmission has been used very successfully for non real-time data transmission. The framework used by this strategy is illustrated in Figure 1 for two source peers, two requesting peers and one helper peer. In this example, the peers S1 and S2 are the sources, which contain the video sequences X and Y to be distributed. Peers R1, R2 are the requesting peers, and peer H1 is a helper peer. The peer H1 does not request the videos, but contributes its upload capacity to help distributing the videos to the other peers. Here, we can see that all the peers are in fact receivers and senders

at the same time, as it is for instance the case in a multipoint video conferencing scenario.



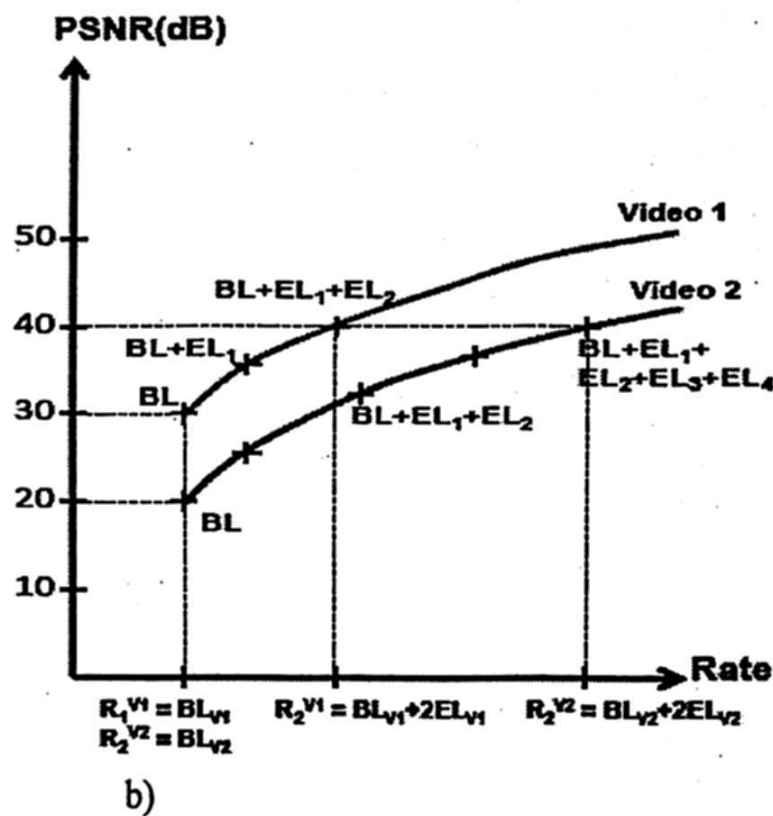
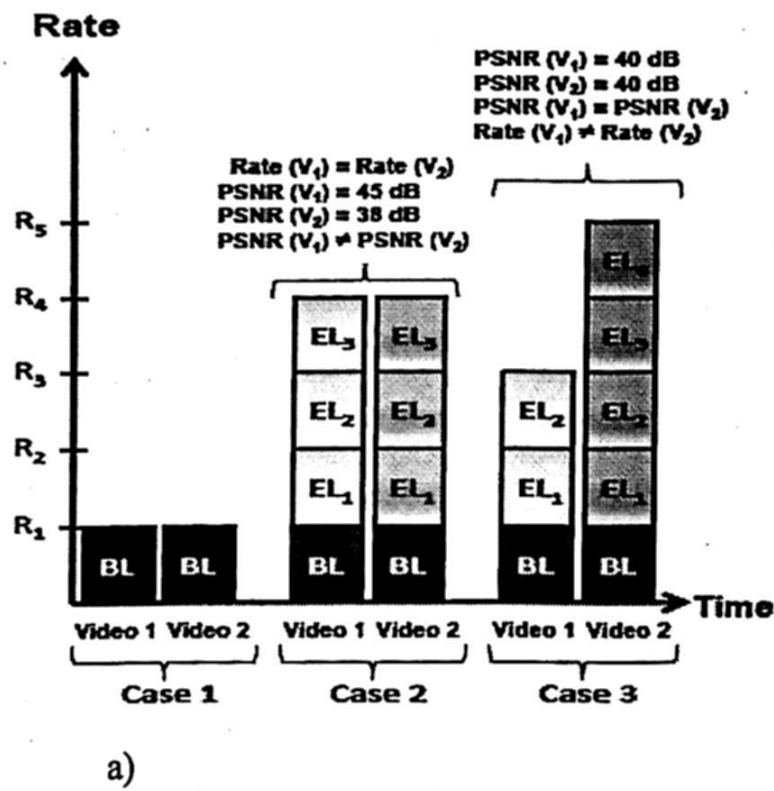
**Fig. 1.** Multi-source scheme based on a mesh topology. a) initial distribution, b) redistribution from all peers, c) one block is directly sent to each receiving peers



Each source splits the original content into small blocks and one unique peer is selected to distribute a block to the rest of the peers. The requesting peers and helper peers forward the received video block to the other requesting peers and the other source peers. For our example, the source S1 divides the video X into the blocks X1 to X4, while the source S2 divides the video Y into the blocks Y1 to Y4 (Figure 1a). Because our approach is based on collaboration among sources, each source distributes its own video while additionally forwarding the block of video received from the other source to the rest of the requesting peers. At the same time, each requesting peer forwards the blocks directly received from a source to the rest of the participating peers. Thus, the blocks (X1, Y1), and (X2, Y2) are assigned to the requesting peers R1 and R2, respectively, while the block (X3, Y3) is assigned to the helper peer H1, the block X4 is assigned to the source peer S2 and block Y4 is assigned to the source peer S1 for distribution. This scenario is shown in Figure 1a. Peers with different upload capacity distribute a different amount of content. The block size assigned to each requesting peer is proportional to its upload capacity. Thus, source S1 sends one block to each participating peer for redistribution, one block in parallel to all requesting peers, and forwards one block of the video Y received from the source S2 to each requesting peer  $R_i$ . The source S2 behaves similar as source S1, but in a complementary way. It sends the video Y and forwards the video block X4. Figure 1b shows a redistribution scenario. In this case, each requesting peer forwards the blocks received from the sources S1 and S2 to the other requesting peers and the other sources. Peer R1 receives the blocks X1 from source S1 and the block Y1 from the source S2. After this, peer R1 forwards the block X1 and Y1 to the rest of the participating peers except to the source where the block was originally generated and the helper peer H1. The blocks X3 and Y3 are sent by the sources S1 and S2, respectively to the helper peer H1, which forwards the blocks to all participating peers except to the source where the block was originally generated. When the source peers have abundant upload resources, each source additionally sends one block directly to the video receiving peers. Figure 1c shows this case. Here, source S1 directly sends block X5 to each video requesting peer and source S2 directly sends block Y5. In this strategy, the sources jointly decide the rate allocation for all participating peers, but additionally enforce the same video quality for all video streams by using scalable video coding techniques. We assume that all participating peers are fully connected and all of them need to receive all videos. Our scenario is described for two sources (S1 and S2) distributing two different video sequences with same PSNR (Peak Signal-to-Noise Ratio) to all participating peers.

The same quality for all videos is obtained if both videos have the same RD-function. However, when the same rate for all video sequences is not enough to obtain a similar video quality among them, a same PSNR need to be enforced. The PSNR enforcement is possible, when the sources have abundant upload capacity. To this end, the broadcast links in each source are manipulated, and the rate of the sequence with the largest rate is reduced. We effectively use the available upload capacity from each source to deliver a homogeneous video quality for all streams. To this end, each source schedules the distribution according to the ratio of the video bit rates based on scalable video coding techniques. We assume that the quality requirements are known. Then, a number of layers to reach this quality level are determined for each video in each

source using scalable video coding. Determining the number of layers and the coding rate for two different video sequences is illustrated in Figure 2.



**Fig. 2.** Enforcement of the same video quality for two different videos using scalable video coding. a). Redistribution of layers, b). PSNR comparison

In Figure 2a), the sources send the base layer of their videos in Case 1. In Case 2, both sources send three enhancement layers of their videos, and the rate  $R_4$  for both sequences is the same. The example assumes that video 1 and video 2 are different and they have been encoded with different bit rates. Thus, using the rate  $R_4$  for both videos,



a PSNR (Peak Signal-to-Noise Ratio) of 45 dB and 38 dB for video 1 and video 2, are obtained respectively. In order to enforce the same PSNR for both sequences scalable video coding is used in Case 3. Then, the source 1 sends two enhancement layer of video 1, while the source 2 sends four enhancement layers of video 2. Figure 2b) shows how both videos sequences can reach the same PSNR using different number of enhancement layers and different rate. In this paper, the PSNR's measurements are obtained experimentally through the JSVM (Joint Scalable Video Model) software [8].

Once the number of required layers and the coding rate are known in each source and before starting the distribution, the sources exchange the coding rate of their videos. Each source computes a local distribution ratio by using these coding rates. This ratio is used in each source to determine the number of required packets for each video. In an ideal situation it is desirable that the throughput is the same as the playback bit rate of the videos in order to obtain a short initial waiting time and a minimal size of buffers. In contrast, if the upload capacity of the sources is not enough to satisfy the requested throughput, the initial time and the buffer usage is increased. Additionally, when different videos are distributed from different sources, the sources need to synchronize the playback of the videos and adapt their upload allocation for the distribution, so that all videos can be received with adaptive throughput and have similar initial waiting time or video quality. The sources use the distribution ratio to adapt the distribution throughput of streams and each source can schedule the number of packets to distribute from itself and the number of packet to distribute from the other sources. The number of distributed packets for each video is proportional to its coding rate.

#### 4 Evaluation

In order to evaluate the performance of our proposed solution, we implemented a prototype of this in Linux (Fedora distribution). Our implementation consists of different programs written in the C/C++ language. This implementation includes a server module run by the source peers and a receiver module run by each requesting peer. Both modules have been enabled with a sender/receiver mode. All links among the participating peers are established using TCP (Transport Control Protocol) connections. Reliable data delivery, flow-control and handling of node leave events are automatically supported by the TCP protocol. Each requesting peer runs a receiver module which receives the video blocks from the sources for its playback and forwards these blocks to the rest of the requesting peers that need to receive this content. We have used a LAN (Local Area Network) infrastructure to evaluate the performance of our scheme implementation. Our proposed scheme is evaluated in terms of throughput, PSNR, and delay for all video streams. We use the peak-signal-to-noise-ratio (PSNR) as the quality metric. Our experiments are based on a joint rate allocation decision and we concurrently control the bit rate for both sources.

The videos sequences used to evaluate this strategy are: Mother and Daughter (M&D), and Foreman. Foreman sequences shows a man talking in a construction site, while M&D shows two persons in the foreground. A person is talking, while the other persona is fixed. The short Foreman sequence is concatenated to a long test sequence with 3000 frames. The same is done with the M&D sequence. Both video streams are encoded with the JSVM software [8] with the same video quality PSNR around 42 dB,

but using different encoding rate and different number of layers for each sequence. To achieve this video quality the Foreman sequence needs a bit rate of 1600 kbps, while Mother and Daughter sequence is encoded at 230 kbps. Our Foreman sequence uses one base layer and two enhancement layers, while M&D sequence uses the base layer and one enhancement layer. Both videos have the same duration (60 seconds), but different size. Foreman file is 10 MB, while M&D file is 1.5 MB.

In these experiments, we considered that all the participating nodes are fully interconnected, including the sources nodes. We compare the time required by both videos to be received in a requesting peer. To this end, two test videos with the same PSNR are allocated as video 1 and video 2 at the sources S1 and S2, respectively. The initial rates are fixed to 230 kbps and 1600 kbps for sources S1 and S2, respectively. M&D video is located at source S1, while Foreman video is located at source S2. Each source delivers its video files to all participating peers, including the sources. Figure 3 shows the distribution throughput on source peers, while Figure 4 displays the receiving throughput of M&D and Foreman videos on a requesting peer.

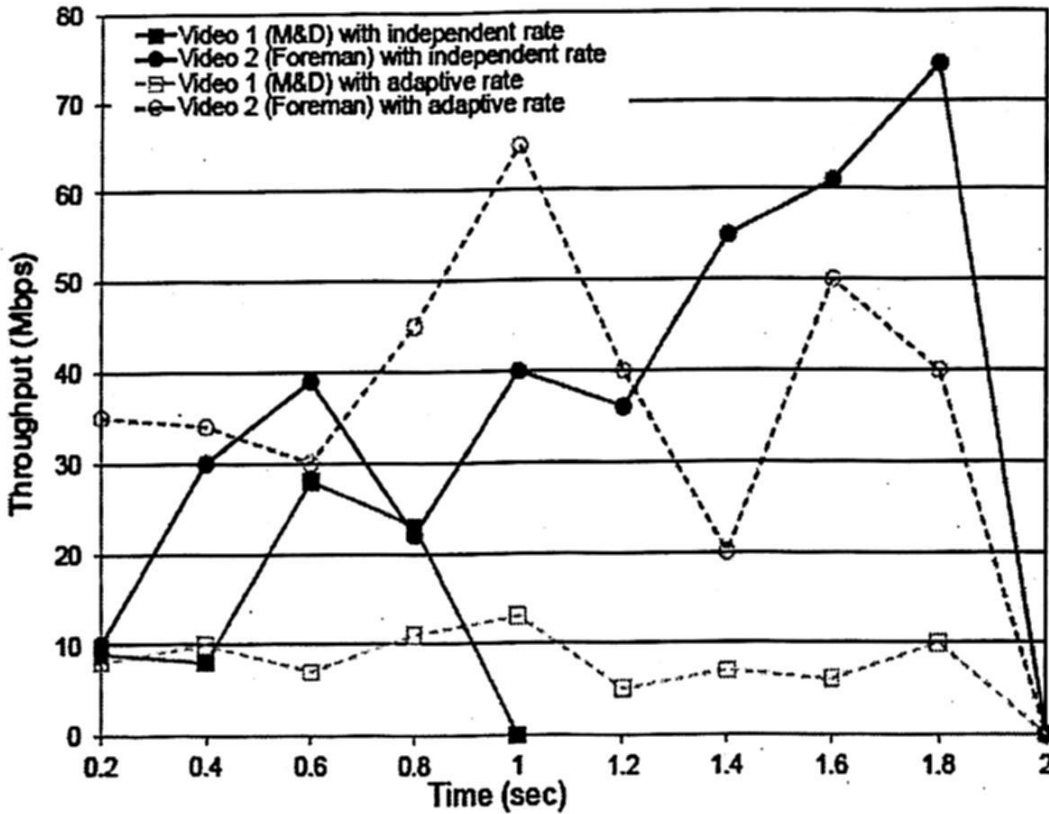


Fig.3. Distribution throughput on a source peer for two different video streams

We can see in Figure 3 that without our strategy of distribution control at the source peers, the distribution throughput of Foreman video on source peer S2 is larger than the distribution throughput of M&D video on source peer S1. The M&D video is distributed very quickly. Contrary, we can see that, with scheduling the distribution, each source peer can regulate the distribution throughput of its own video and the other video in proportion. Thus, the source ends to provide both videos at same time, but Foreman video requires greater source's throughput than M&D video. In Figure 4 we can easily see that without regulating the distribution throughput at the source peers,

M&D video is received very quickly on the requesting peer. Most probably, the initial delay of playback for this stream is shorter. In contrast, for Foreman video, because it is slowly received, its initial delay may be longer than D&M video. In order to be able to synchronously play out these two streams the buffer demands on the nodes is very high. However, if both source peers schedule distribution, the delay of receiving both video streams on a requesting peer takes almost same time. Therefore, the playback during receiving can be synchronously with low buffering demands.

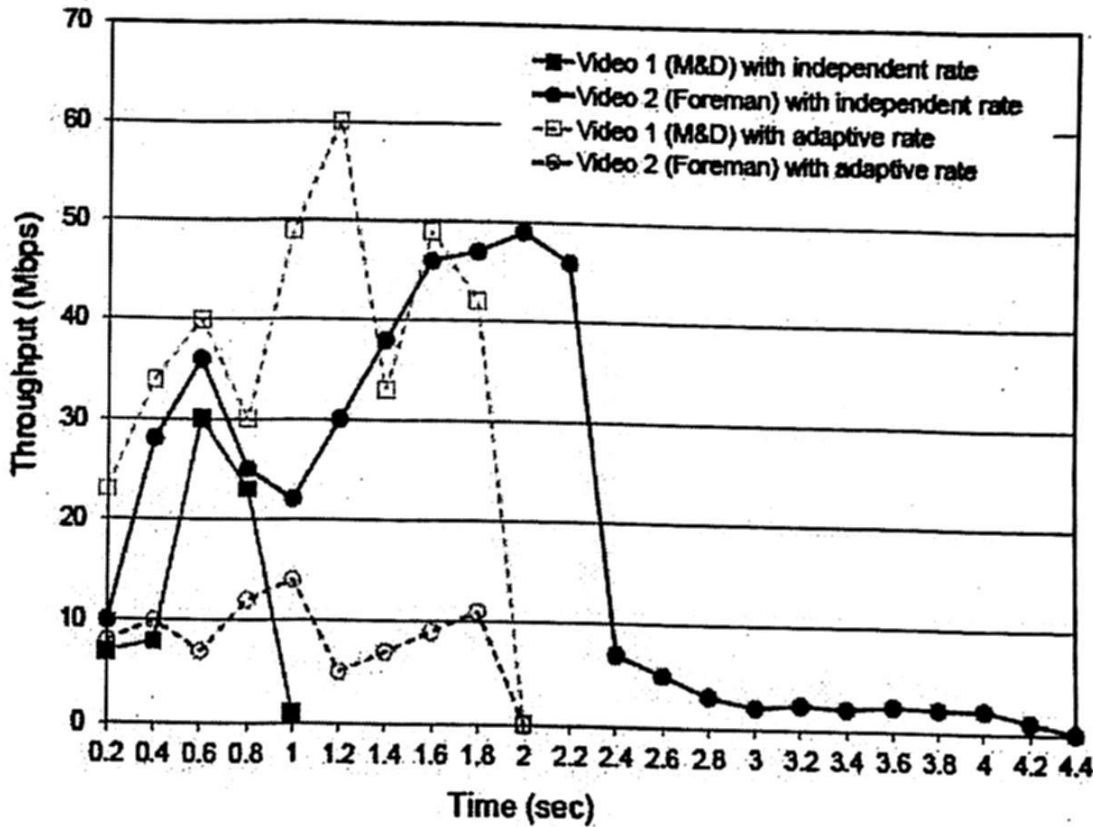


Fig.4. Receiving throughput on a requesting peer for two different video streams

Based on these results, we can see that an adaptive rate control and more collaboration between the sources allow that all video sequences achieve a similar PSNR (Peak Signal-to-Noise Ratio) quality. Thus, the average throughput can be used to recalculate the PSNR using JSVM (Joint Scalable Video Model) software, and we can obtain PSNR values of 43.4 dB and 43.9 dB for Foreman and Mother and Daughter, respectively.

## 5 Conclusions

Video applications are characterized by different resolutions designed for devices with different computational capabilities. In this paper, we have presented and evaluated a multi-source scheme to video streaming based on scalable video coding (SVC). Our proposed solution helps to reach similar video quality for all streams from multiple sources. Our proposed strategy was evaluated in a LAN infrastructure using simple experiments with four nodes. Our experiments show that we can reach a similar

video quality from multiple peers by using a strategy based on scalable video coding. Therefore, different videos can be received in similar times by a requesting peer independently of their size. Thus, our proposed scheme with SVC shows more effectively our scheme where SVC is not used. We believe that our proposed solution with SVC could be ideally used for peer-to-peer (P2P) video streaming scenarios with few participants such as video-conference or surveillance systems. As future work, some important properties of P2P networks, such as scalability and churn can be incorporated in our proposed model.

## References

1. Tran, D. A., Hua K. A Do, T. T.: A Peer-to-Peer Architecture for Media Streaming. In: IEEE Journal on Selected Areas in Communication, Special Issue on Advances in Overlay Networks, (2003).
2. Zhang, X., Liu, J., Li, B., Yum, P.T-S.: CoolStreaming/DONet: A Data-driven Overlay Network for Efficient Live Media Streaming. In: 24<sup>th</sup> IEEE INFOCOM, Miami, FL, USA, pp.2102--2111, (2005).
3. Li, J., Chou, P.A., Zhang, C.: Mutualcast: An Efficient Mechanism for One-To-Many Content Distribution. In: ACM SIGCOMM ASIA Workshop, (2005).
4. Schwarz, H., Marpe, D., Wiegand, T.: Overview of the Scalable Video Coding Extension of the H.264/AVC Standard. In: IEEE Transactions on Circuits and Systems for Video Technology, 17(9), pp.1103--1120, (2007).
5. Wien, M., Schwarz, H., Oelbaum, T.: Performance Analysis of SVC. In: IEEE Transactions on Circuits and Systems for Video Technology, 17(9), pp. 1194--1203, (2007).
6. Schwarz, H., Wien, M.: The Scalable Video Coding Extension of the H.264/AVC Standard. In: IEEE Signal Processing Magazine, 25(2), pp. 135--141, (2008).
7. Reichel, J. Schwarz, H., Wien, M.: Joint Scalable Video Model (JSVM) 11, Doc. JVT-X202, Joint Video Team, Video Coding Experts Group, (2007).
8. Joint Video Team (JVT); JSVM software manual, version 9.12.2, Heinrich-Hertz-Institute, <http://ip.hhi.de/imagecomG1/savce/downloads/SVCReferenceSoftware.htm>.
9. López-Fuentes, F. A., Steinbach, E.: Multi-source video multicast in peer-to-peer networks. In: 22<sup>nd</sup> IEEE International Symposium on Parallel and Distributed Processing, Miami, Florida USA, (2008).
10. Lee, J. S., De Simone, F., Ebrahimi, T.: Subjective Quality Evaluation via Paired Comparison: Application to Scalable Video Coding. In: IEEE Transactions on Multimedia, Vol. 13, No. 5, pp. 882--893, (2011).
11. Wang, Y., Ostermann, J., Zhang, Y.- Q. In: Video Processing and Communications, Prentice-Hall, (2002).
12. Mirshokraie, S., Hefeeda, M.: Live Peer-to-Peer Streaming with Scalable Video Coding and Networking Coding. In: ACM SIGMM conference on Multimedia systems, MMSys '10, ACM, New York, NY, USA, 2010, pp. 123--132 (2010).
13. Abboud O., Zinner, T., Pussep, K., Oechsner, S., Steinmetz, R. and Tran-Gia, P. In: A QoE-Aware P2P Streaming System Using Scalable Video Coding. In: IEEE International Conference on Peer-to-Peer Computing (P2P), Delft, Netherlands. (2010).



# Circular Monopole Antenna with defected ground plane for UWB applications

Cruz Ángel Figueroa Torres<sup>1</sup>, José Luis Medina Monroy<sup>1</sup>, Ricardo Arturo Chávez Pérez<sup>1</sup>, Andrés Calvillo Tellez<sup>2</sup>

<sup>1</sup>Centro de Investigación y Educación Superior de Ensenada, Ensenada, México  
{cfiguero, jmedina, chavez} @cicese.mx

<sup>2</sup>Centro de Investigación y Desarrollo de Tecnología Digital del IPN, Tijuana, México  
{calvillo@citedi.mx}

*Paper received on 11/20/13, Accepted on 01/19/14.*

**Abstract.** In this paper a circular monopole antenna for Ultra Wide Band (UWB) applications is proposed. The structure is a simple circular shape monopole antenna designed on FR-4 substrate and fed by a  $50\Omega$  microstrip line. The ground plane of the antenna has been modified including several defects to improve its behavior in matching and wideband. A parametric study has been performed to the semicircular slots located in the ground plane area. The frequency range measured for  $S_{11} < -10\text{dB}$  was from 1.66 - 20 GHz. The total dimensions of the antenna are 55x79mm (W x L).

**Keywords:** circular monopole antenna, defected ground plane, UWB.

## 1 Introduction

Microstrip antennas are investigated intensively due to their properties, such as low profile, low cost, conformability and ease of integration with active devices [1]. Nowadays, printed ultra wideband (UWB) antennas have been attractive for researchers due to their small size, low cost and high data rate features. The Federal Communications Commission has allocated the 3.1 GHz to 10.6 GHz frequency band for unlicensed ultra wideband applications. Most of the ultra wideband antennas are either microstrip fed or coplanar waveguide fed monopoles or slots [2]. By using different shapes for the patch and the slot, several ultra wideband antennas have been proposed [3-12]. In [4], a comparison of the different shapes such as rectangular, circular, square, elliptical and triangular shape for the patch as well as the slot was made. In [2] is shown that adding a slot to an antenna can improve the impedance bandwidth, compared to a simple patch antenna, due to the coupling between the slot and the feed line.



UWB systems have many advantages such as their wide bandwidth and low cost, being suitable to telecommunications, medical imaging and biomedical systems [13-15]. Printed monopole antennas fabricated on a substrate, offer wide impedance bandwidth that can cover the complete UWB range [16]. In [16] a circular monopole antenna is presented, which have an L-shape bended ground plane and was fabricated in FR-4 with a size of 56 x 60 mm for operation from 1.3-12 GHz having a bandwidth of 10.7 GHz. Furthermore, a circular monopole with a ground plane only below the feed line has been designed in [17] and built in a substrate with approximate dimensions of 40 x 30 mm, to work in the 3.1-11 GHz frequency range with a bandwidth of 7.9 GHz. In this paper, we propose a monopole circular antenna with a defected ground plane having a semicircular shape below the  $50\Omega$  feed line, and two quarters of a circle on the edges of the ground plane. A parametric study performed in this work, also confirm that the use of slots in the ground plane of the antenna, improve its impedance bandwidth characteristics. The proposed structure compared with the designs [16-17] shown an improvement in bandwidth, from 1.66 GHz up to 20 GHz. The advantage of the proposed antenna is the extremely wide bandwidth obtained due to the defected ground plane with semicircular slots.

## 2 Antenna structure

The geometry of the proposed antenna is shown in Figure 1. It is fed by a  $50\Omega$  microstrip line and fabricated on a FR4 substrate of size  $W=55$  mm by  $L=79$  mm and a thickness of 1.5778 mm. The relative dielectric constant and dissipation factor of the substrate are 4.08 and 0.019 respectively.

On the top side of the substrate, there is a circular patch with radius  $R_p=27.5$  mm and a microstrip  $50\Omega$  feed line with dimensions  $W_f=3.1$  mm by  $L_f=24$  mm as it is shown in Figure 1a.

On the bottom side of the substrate, there is a defected ground plane with sizes  $W_g=55$  mm and  $L_g=23.5$  mm. A semicircular shape with a radius  $y_c$  has been removed from the ground plane below the  $50\Omega$  feed line, and two quarters of a circle, with a radius  $x_t$  from the edges of the ground plane, as it is shown in Figure 1b.

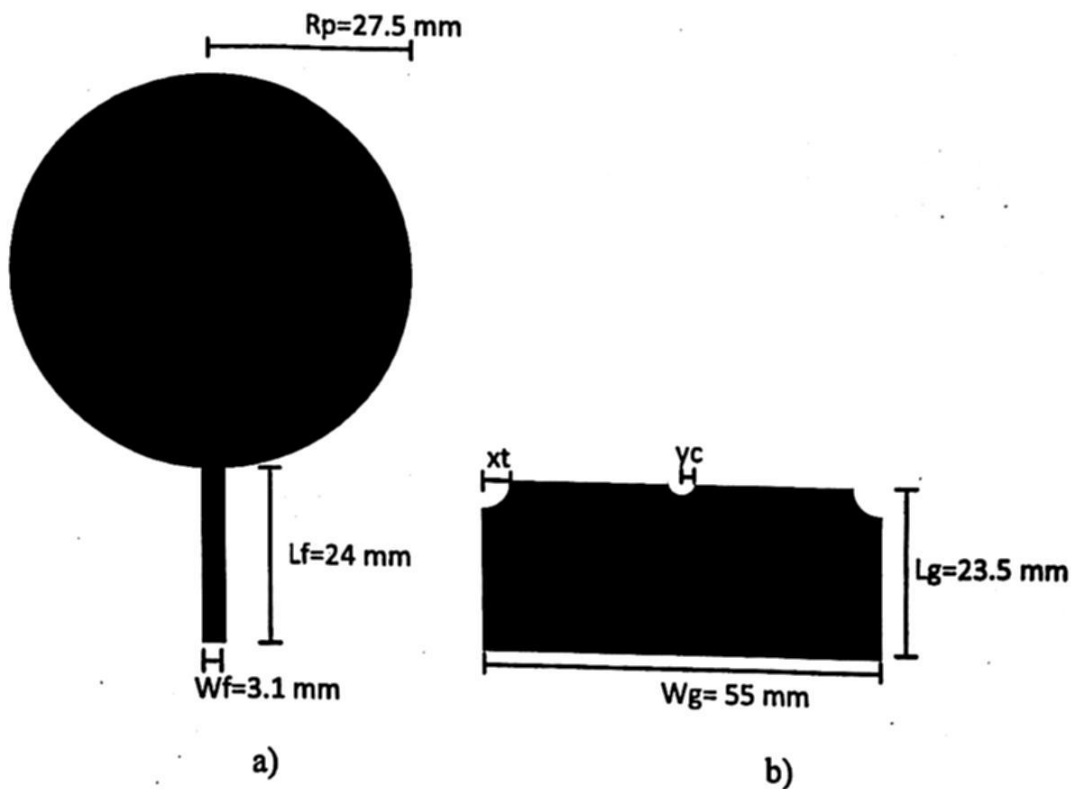


Fig. 1. Circular Monopole Antenna with defected ground plane: a) Top side, b) Bottom side.

The design expression of a simple circular microstrip antenna [13-14] to calculate the resonant frequency is

$$F_r = \frac{1.841c}{2\pi r_{eff} \sqrt{\epsilon_{eff}}} \quad (1)$$

Where  $c$  is the velocity of light,  $r_{eff}$  is the effective radius of the patch and  $\epsilon_{eff}$  is the effective dielectric constant. Using the equation (1), the circular patch has been designed for operation at 2 GHz, giving a radius  $r_{eff}$  of 27.5 mm.

### 3 Parametric study of the antenna

Before initiating the parametric study in the ground plane, the parameters  $W_f$ ,  $L_f$  and  $L_g$  were optimized for a reduction of the return losses ( $S_{11} < -10$  dB) and to increase the frequency bandwidth. The electromagnetic analysis and optimization of the antenna was performed using the *CST Microwave Studio* software from 1 to 20 GHz. The variation effect in the slots of the ground plane is given in Figures 2 and 3, which clearly show the effect of the parameters ( $y_c$ ) and ( $x_t$ ) on the impedance bandwidth of the antenna. Figure 2 presents the return losses of the antenna in function of the variation of the radius ( $y_c$ ) of the semicircular slot, located in the center of the ground plane below the feed line. The final value of the radius ( $y_c$ ) was 2 mm, which reduces the return losses  $S_{11}$  to approximately -12dB in the frequency range. On the other hand, the  $S_{11}$  results in function of the variation of the radius ( $x_t$ ) of the two slots in form of a quarter of a circle, removed from the edges of the ground plane is shown in Figure 3. It can be noted clearly the improvement of  $S_{11}$  in the whole frequency range when the radius is increased to 4 mm, showing that the return losses are lower than -12 dB at 2 GHz and -15dB from 2.5 GHz to 20 GHz.

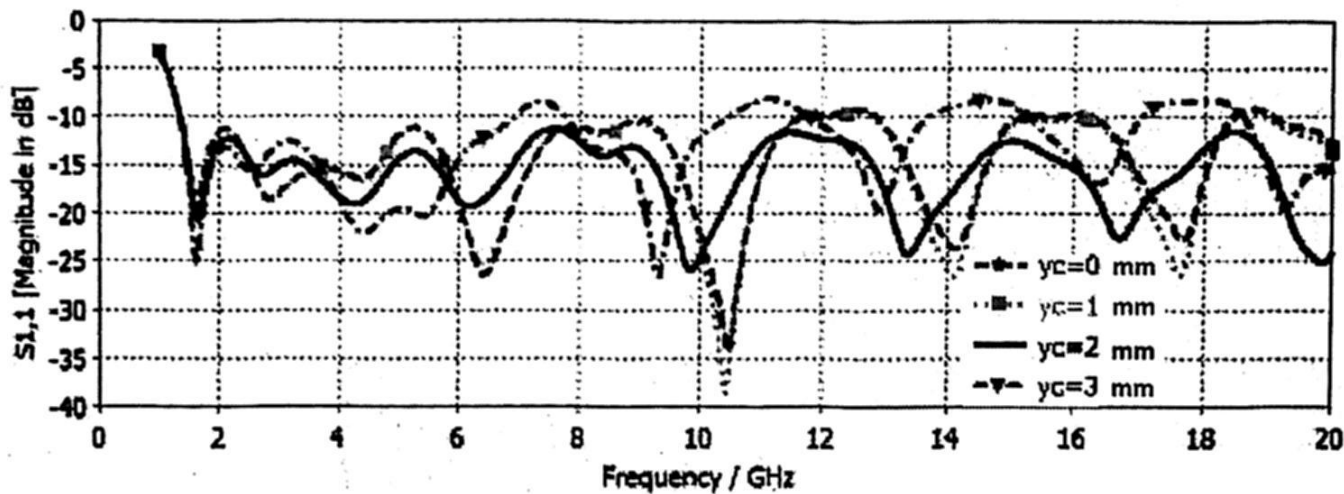


Fig. 2. Effect on the S11 due to the semicircular slot in the ground plane, below the feed line.

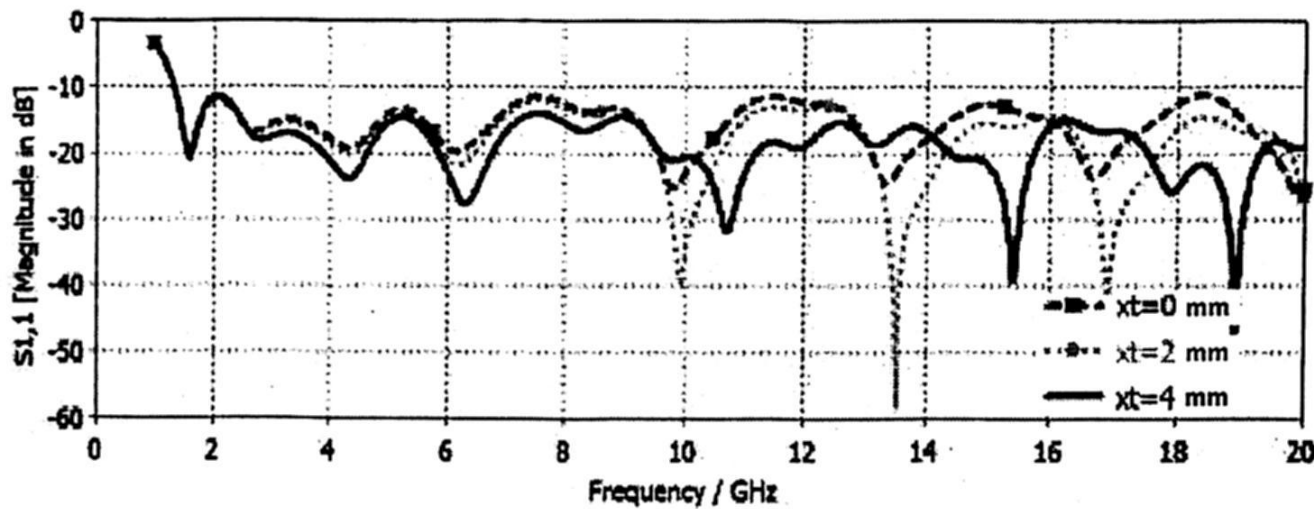
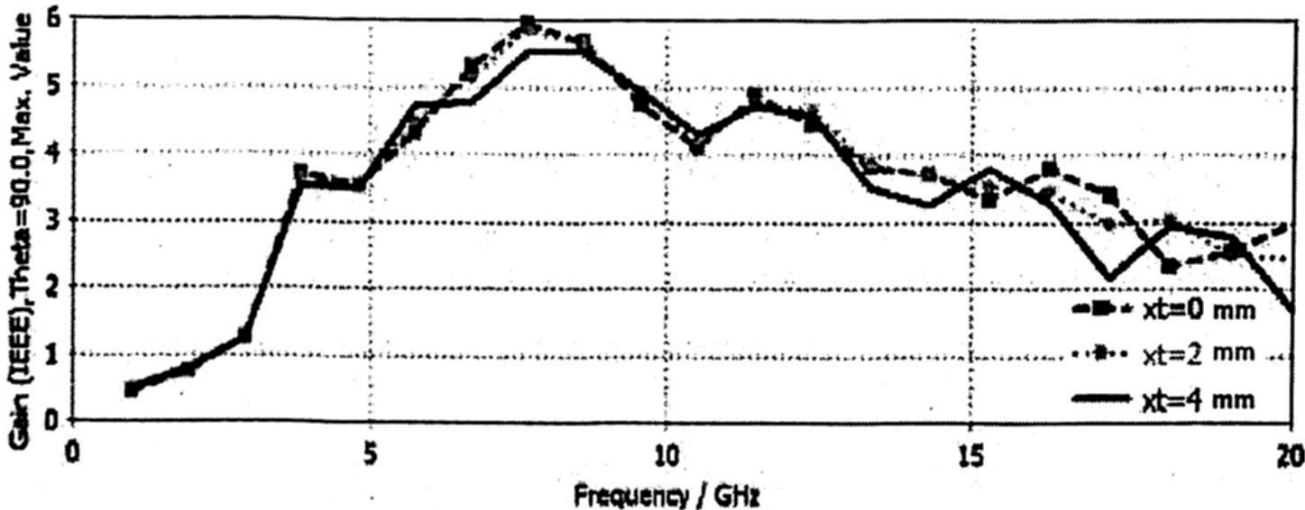


Fig. 3. Effect on the S11 caused by the two quarters of a circle removed from the edges of the ground plane.

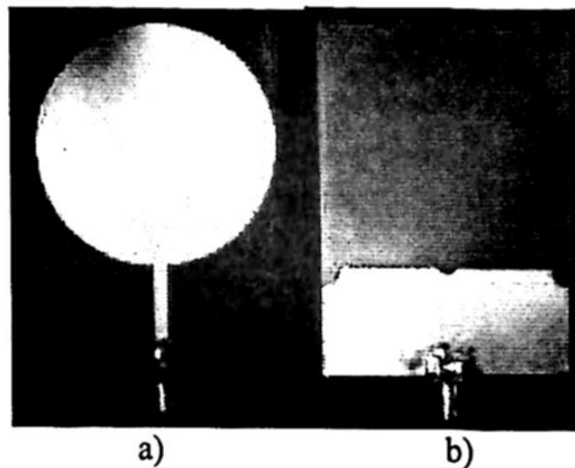
The effect of the variation of the slots 'xt' in the gain of the antenna is given in Figure 4. It can be noted that the gain is slightly reduced when xt is increased to 4 mm. The gain varies between 0.8 to 5.6 dB from 1.66 to 20 GHz, giving the maximum from 7.5 GHz to 8 GHz.



**Fig. 4.** Effect on the gain due to the two quarters of a circle removed in the edges of the ground plane.

## Results

The circular monopole antenna was analyzed using the electromagnetic software *CST Microwave Studio*, giving the simulation results already shown in Figures 2 to 4 with dimensions shown in Figure 1 and with  $y_c = 2$  mm and  $x_t = 4$  mm. The antenna was fabricated on the two sides of a substrate FR-4 with 55 x 79 mm. In the microstrip feed line was soldered a SMA connector, as it is shown in the Figure 5. Figure 5a, present the front view showing the circular monopole, and Figure 5b the back view, where are shown the semicircular slots in the ground plane.



**Fig. 5.** Fabricated antenna: a) Front, b) Back

The fabricated prototype was measured on a HP Vector Network Analyzer (8510A) calibrated from 1 to 20 GHz. The measured results are presented in Figure 6, compared to the theoretical results (Simulated) obtained with the EM software CST. In this Figure, it can be noted that the simulated parameter  $S_{11}$ , is lower than the measured, particularly at high frequencies. Theoretical results show very good behavior in the 1.35 GHz to 20 GHz, while the measured ones provide a good performance from 1.66 GHz to 20 GHz.

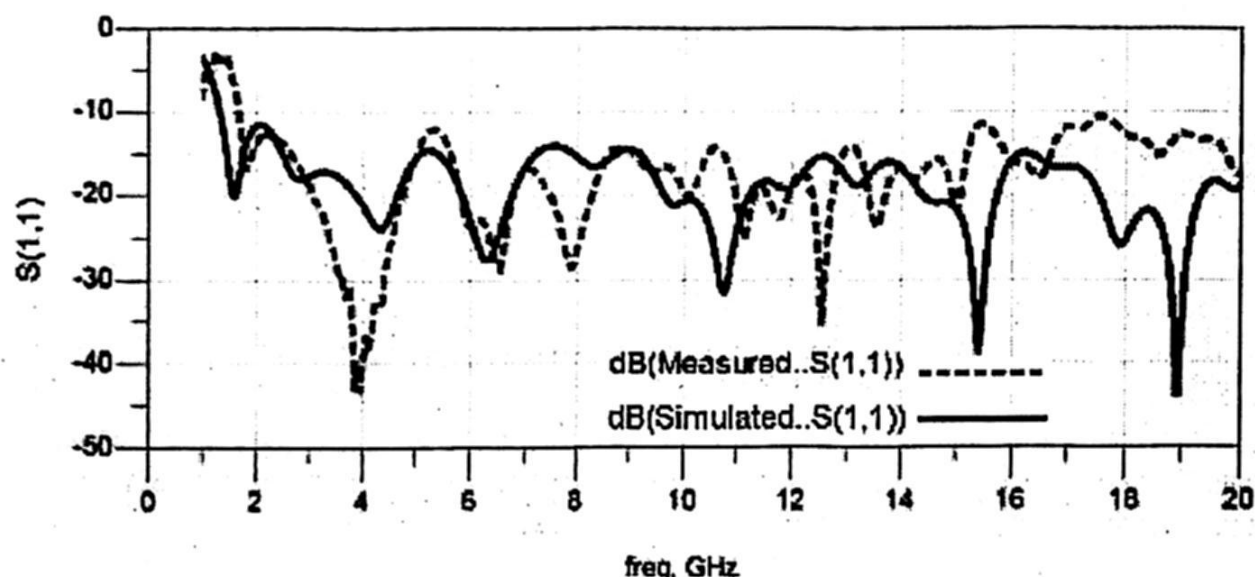


Fig. 6. Simulated and Measured Results.

As can be noted in Figure 6, the impedance bandwidth of this antenna is obtained when the  $S_{11}$  parameter  $< -10$  dB.

Difference between theoretical and experimental results is attributed to fabrication errors in the construction process, where the dimensions of the antenna elements have changed slightly. The antenna dimensions were measured with a microscope, showing that two key parameters have changed, and caused a variation in the measured results respect to theoretical ones. These are  $y_c$  (from 2mm to 1.6mm) and  $W_f$  (from 3.1 to 3.32mm).

It is worthy to mention that theoretical and measured results never will be exactly the same, because the electromagnetic analysis methods are only an approximation that can be more accurate when the number of cells used tends to infinity. This also can be seen in the comparison made in [2].

Figure 7 shows the theoretical and measured behavior, including a new EM analysis response using the physical dimensions measured, showing that the new results are closer to the measured ones.

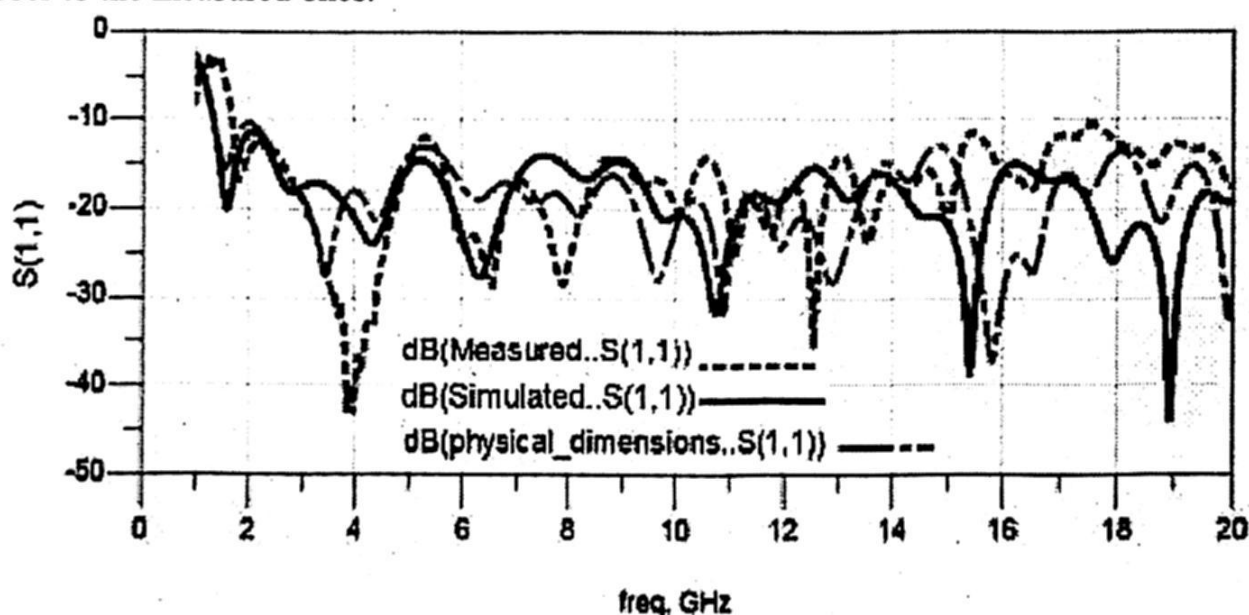


Fig. 7. Analysis of the structure with physical dimensions



Compared with previous work, in [2] an antenna design using a defected ground plane was presented, which has an impedance bandwidth of 8.2GHz. The proposed structure has the clearly advantage of a very large bandwidth behavior of 18.34 GHz compared with references [2], [16] and [17].

## 5 Conclusions

A circular monopole antenna was proposed and optimized for ultra wide band applications, based on a defected ground plane. To improve the impedance bandwidth and reduce the return losses, three semicircular slots in the ground plane were proposed. This antenna operates with a very good behavior from 1.66 GHz to 20 GHz, giving an extremely wide band allowing to be used in many applications, making possible its integration into mobile phones and other portable devices, due to the very small dimensions of the antenna. In addition to the presented results, the proposed antenna was analyzed to obtain other characteristics like the radiation pattern, beamwidth and efficiency at some frequencies. At 10.5 GHz the antenna shows an omnidirectional radiation pattern with a gain of 4.5 dB and 73% efficiency. The maximum efficiency was 88% at 3.1 GHz, and the minimum efficiency was 52% from 18 to 20 GHz where gain decreases to 2dBi.

## 6 Acknowledgements

The authors want to acknowledge to Ing. René Torres Lira for his technical support in the fabrication of the antenna.

## References

1. Kordzadeh, A., Hojat, F.: A new reduced size microstrip patch antenna with fractal shaped defects. In: Progress in Electromagnetics Research B, Vol. 11, 29-37 (2009)
2. Kushwaha, N., Kumar, R.: An UWB fractal antenna with defected ground structure and swastika shape electromagnetic band gap. In: Progress in Electromagnetics Research B, Vol. 52, 383-403 (2013)
3. Sadat, S., Fardis, M., Geran, F., G., Dadashzadeh, G., R.: A compact microstrip square-ring slot antenna for UWB applications. In: Progress in Electromagnetics Research, Vol. 67, 173-179 (2007)
4. Azim, R., Islam, M., T., Misran, N.: Compact tapered-shape slot antenna for UWB applications. In: IEEE Antennas and Wireless Propagation Letters, Vol. 10, 1190-1193 (2011).
5. Krishna, R., V., S., Raj, K.: Design of temple shape slot antenna for ultra wideband applications. In: Progress in Electromagnetics Research B, Vol. 47, 405-421 (2013)
6. Chen, H., D.: Broadband CPW-fed square slot antennas with a widened tuning stub. In: IEEE Transactions on Antennas and Propagation, Vol. 51, No. 8, 1982-1986 (2003)
7. Dastranj, A., Imani, A., Naser-Moghaddasi, A.: Printed wide-slot antenna for wideband applications. In: IEEE Transactions on Antennas and Propagation, Vol. 56, No. 10, 3097-3102 (2008)

8. Fallahi, R., Kalteh, A., A., Roozbahani, M., G.: A novel UWB elliptical slot antenna with band-notched characteristics. In: *Progress in Electromagnetics Research*, Vol. 82, 127-136 (2008)
9. Bahl, I., J., Bhartia, P.: *Microstrip Antennas*. Artech House, Dedham, MA (1980)
10. Kumar, G., Ray, K., P.: *Broadband Microstrip Antennas*. Artech House, Norwood, MA (2003)
11. Balanis, C., A.: *Antenna Theory Analysis and Design*. Wiley Publication, New Jersey (2005)
12. Garg, R., Bhartia, P., Bahl, I., J., Ittipiboon, A.: *Microstrip Antenna Design Handbook*. Artech House, Norwood, MA (2001)
13. Quintero, G., Skrivervik, A., K.: Analysis of Planar UWB Elliptical Dipoles fed by a Coplanar Stripline. In: *IEEE International Conference on Ultra Wideband*, Vol. 1, 113-116 (2008)
14. Powell, J., Chandrakasan, A.: Differential and Single Ended Elliptical Antennas for 3.1-10.6 GHz Ultra Wideband Communication. In: *IEEE Antennas Propagation Society Int. Symp.*, Vol. 3, 2935-2938 (2004)
15. Liang, J., Guo, L., Chiau, C., C., Chen, X.: CPW-Fed Circular Disc monopole Antenna for UWB Applications. In: *IEEE International Workshop on Antenna Technology*, 505-508 (2005)
16. Shashwat, C., Sanjay, S., Shashank, C., Sachin, S., Sachin, S.: Ultra-Wide bandwidth Circular Monopole Antenna. In: *International Journal of Scientific Research Engineering & Technology (IJSRET)*. Vol. 1, 279-282 (2012)
17. Sonar, A., F., Mishra, R., P., Mhaskar, J., Kharche, S.: UWB Circular Monopole Antenna. In: *ITSI Transactions on Electrical and Electronics Engineering*. Vol. 1, 2320-8945 (2013)

# Face detection method based on nonlinear composite correlation filters

Everardo Santiago-Ramírez<sup>1</sup>, José-Ángel González-Fraga<sup>1</sup>, Everardo Gutiérrez-López<sup>1</sup>, Omar Álvarez-Xochihua<sup>1</sup>, and Sergio-Omar Infante-Prieto<sup>2</sup>

<sup>1</sup> Facultad de Ciencias, Universidad Autónoma de Baja California, Carretera Transpeninsular Ensenada-Tijuana Núm. 3917, Colonia Playitas, Ensenada, Baja California, C.P. 22860  
{everardo.santiagoramirez, angel\_fraga, everardo.gutierrez, aomar}@uabc.edu.mx

<sup>2</sup> Facultad de Ingeniería, Arquitectura y Diseño, Universidad Autónoma de Baja California, Carretera Transpeninsular Ensenada-Tijuana Núm. 3917, Colonia Playitas, Ensenada, Baja California, C.P. 22860  
sinfante@uabc.edu.mx

*Paper received on 11/30/13, Accepted on 01/19/14.*

**Abstract.** Face detection is an important first step in a fully automatic face processing system. Current algorithms are able to detect faces that are easily distinguishable. However, most of these algorithms perform poorly when used to process images taken under conditions with non-uniform illumination and the face present variations in pose. In this paper, we present a face detection algorithm that uses nonlinear composite correlation filters, designed with strong classifiers. For the design of strong classifiers, a set of transformations were applied to original training images. In order to improve the discrimination capacity and robustness in conditions with homogeneous and structured backgrounds, the training images for the filters were selected by an algorithm from a face database. The performance of the proposed algorithm was evaluated in terms of its ability to determine the location of a single face under conditions with non-uniform illumination and slight variations in pose.

## **Keywords:**

Face detection, nonlinear composite correlation filters, correlation pattern recognition

## 1 Introduction

The need for reliable face detection systems, that function in both indoor and outdoor environments has caught researchers and technologists' interest in developing facial-distortion invariant algorithms. While face detection has a wide range of applications, its principal application has been in automatic face recognition systems. The accuracy of a face detection algorithm is important for the performance of subsequent face processing tasks in a system. Given an arbitrary

image, the objective of face detection is to determine whether or not any faces are found in the image, and, if detected, to return the location and dimension of the face [1]. This can be achieved using the information provided by several cues, such as skin color (for color images), movement (for faces appearing in a video), face or head shape, facial appearance, or a combination of these parameters [2].

Face detection methods can be classified into four main categories [1]: 1) Knowledge-based methods, which encode human knowledge about face components; 2) feature invariant approaches, which are mainly based on the face's structural features; 3) template matching methods, where several previously stored standard patterns are correlated with an input image in order to detect a face and; 4) appearance-based methods, where facial models are learned from a set of training images which should capture the representative variability of facial appearance. The most successful face detection methods are based on appearance and are able to detect all faces in an image with great accuracy, independently of their position, dimension, orientation, age and expression [2].

However, these face detection algorithms only perform well under conditions where the facial regions are easily distinguishable. Composite correlation filters are able to combine the characteristic of both template matching and appearance based methods as they use both face shape (structure) and face content (appearance). A composite correlation filter is designed by combining training images that are representative of the expected distortions for the reference object. For this reason, the performance of the composite filters depends largely on an appropriate selection of training images. A face detection method that employs correlation filters does appear in the literature [3]. An important aspect to consider in the use of this method is that it requires a large amount of training images, however, it not consider a method for selecting solely the most suitable face images.

This paper presents a face detection algorithm based on nonlinear composite correlation filters which was designed using strong classifiers that emphasize facial features. Given a face database, a simple algorithm selects only those face images that produce a sharp and high peak for the training set. In order to increase the amount of data and model highly representative distortions, different versions of the training set were obtained by applying image transformations. Each version of the training set was used to design strong classifiers, which were then used to design a robust nonlinear composite correlation filter for detecting faces in scene images with homogeneous and structured backgrounds.

The rest of the paper is organized as follows. Section 2 presents the theoretical foundation for composite correlation filters. Section 3 then describes the proposed face detection method. The results of the experiment and a discussion of them are presented in Section 4. Finally, Section 5 presents the main conclusions of this work.



## 2 Composite correlation filters

Correlation pattern recognition (CPR) is based on selecting or creating a reference signal  $h(x, y)$ , called a correlation filter, and then determining the degree of similarity between the reference and test signals [4]. Correlation filters can be designed in either the spatial or frequency domain. The correlation process using the Fourier Transform (FT) is given by:

$$g(x, y) = \mathcal{F}^{-1}\{S(u, v) \cdot H^*(u, v)\}, \quad (1)$$

where  $g(x, y)$  is the correlation output,  $\mathcal{F}^{-1}$  is the inverse Fourier transform, and  $S(u, v)$  and  $H(u, v)$  are the Fourier transforms of the test signal  $s(x, y)$  and reference signal  $h(x, y)$ , respectively. The symbol  $\cdot$  indicates that  $S$  and  $H$  are multiplied element by element, and  $*$  represents the complex conjugate of  $H$ . In ideal circumstance, when  $s(x, y)$  contains multiple objects that are similar to  $h(x, y)$ ,  $g(x, y)$  should exhibit large correlation peaks for each object in the scene that matches with  $h(x, y)$ .

The most basic correlation filter is the Matched Filter (MF), which is robust in recognizing reference images affected by additive white noise [5]. However, it is very sensitive to distortions such as rotation and scale. MF is given by:

$$H(u, v) = \alpha T^*(u, v), \quad (2)$$

where  $T^*(u, v)$  is the is the FT of the reference image  $t(x, y)$ .

### 2.1 Nonlinear composite filter

A Synthetic Discriminant Function (SDF) filter is a linear combination of Matched Filters [6]. This filter is designed using a training set  $T$  composed images containing the principal distortions expected for the reference image, and is, therefore, robust in recognizing an object that presents distortions similar to those found in  $T$ .

Let  $T = \{t_1(x, y), t_2(x, y), \dots, t_N(x, y)\}$  be the training image set and  $\mathbf{x}_j$  the column-vector form of  $t_j(x, y)$ .  $\mathbf{x}_j$  is created by lexicographic scanning, in which each image is scanned from left to right and from top to bottom. Each vector is a column of the training data matrix  $X = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N]$ . The SDF correlation filter is given by:

$$\mathbf{h} = X(X^+X)^{-1}\mathbf{u}, \quad (3)$$

where  $+$  is the complex conjugate transpose and  $\mathbf{u} = [u_1, u_2, \dots, u_N]^+$  is a vector of size  $N \times 1$  that contains the expected values at the origin of the correlation output for each training image. Typical values for  $u$  are 1 for images belonging to the true class and 0 for those in the false class. The  $-1$  indicates the inverse of the matrix. Although this filter is tolerant to distortions, the correct location of the object is complicated by the wide peak produced in the output correlation. Therefore, the FT of each training image can be filtered with the *kth - Law*



nonlinear factor, as shown in expression 4, thus improving the sharpness of the peak.

$$\hat{T}_i^k(u, v) = |\hat{T}_i(u, v)|^k \exp(i\varphi(u, v)), \quad (4)$$

where  $0 < k < 1$  is the nonlinear factor.  $\hat{T}_i(u, v)$  is the FT of  $t_i(x, y)$ ,  $|\hat{T}_i(u, v)|$  is the module of  $\hat{T}_i(u, v)$ , while  $\varphi(u, v)$  is its phase. As can be observed, the nonlinearity factor raises the magnitude of the Fourier spectrum to the  $k$ th power, while the information of the phase remains intact. In the correlation process, the same nonlinear factor  $k$  used in the filter training must be applied to the test image. Applying the nonlinear factor to training images for the SDF filter in expression 3, a nonlinear composite correlation filter is given by [7]:

$$\mathbf{H}^k = \mathbf{X}^k ((\mathbf{X}^k)^+ \mathbf{X}^k)^{-1} \mathbf{u}. \quad (5)$$

## 2.2 Average of synthetic exact filters

Let  $t_i(x, y) \in T$  be the training image and  $g_i(x, y)$  its desired correlation output, which can be synthetically generated by a Gaussian functions as follow:

$$g_i(x, y) = \exp \left( -\frac{(x - x_i)^2 + (y - y_i)^2}{\sigma^2} \right), \quad (6)$$

where  $\sigma^2$  is the radius of the Gaussian at the center of the object. Each pair  $t_i(x, y)$ ,  $g_i(x, y)$  is used to construct an exact filter with the following expression:

$$H_i^*(u, v) = \frac{G_i(u, v)}{\hat{T}_i(u, v)}, \quad (7)$$

where the division is element by element.  $G_i(u, v)$  and  $\hat{T}_i(u, v)$  are the FT's of  $g_i(x, y)$  and  $t_i(x, y)$ , respectively. The ASEF filter is obtained by averaging  $N = |T|$  exact filters [3], such as is shown in the following expression:

$$H(u, v) = \frac{1}{N} \sum_{i=1}^N H_i^*(u, v). \quad (8)$$

## 3 Face detection method using nonlinear composite filters

The algorithm proposed in this work exploits both face shape and face content in a nonlinear composite correlation filter that contains enough information to detect faces in a scene. This algorithm comprises of the following steps: *Training set selection*, *nonlinear composite correlation filter design* and *face detection*.

### 3.1 Training set selection

The success of face detection by correlation filters depends largely on the training set  $T$ , which has to describe the expected distortions of a human face. This set  $T$  must be small enough for computational convenience and contain only those images suitable for the design of a filter for reliable face detection. A simple correlation-based strategy, described by Algorithm 1, was designed for selecting the suitable images for filter design.

---

#### Algorithm 1: Training set selection.

---

**Data:** Whole face images set  $F_{DB}$ , initial images training  $S_f$ , threshold value  $\tau$

**Result:** Training set  $T$

```

1  $H(u, v) \leftarrow$  Design initial filter with  $S_f$ 
2  $N \leftarrow 1$ 
3  $T \leftarrow \{\}$ 
4 while  $N \leq |F_{DB}|$  do
5    $t_N(x, y) \leftarrow$  Read  $t_N(x, y)$ , such that  $t_N(x, y) \in F_{DB}$ 
6    $\hat{T}(u, v) \leftarrow \mathcal{F}\{t_N(x, y)\}$  //  $\mathcal{F}$  is the Fourier transform
    $\hat{T}^k(u, v) \leftarrow |\hat{T}(u, v)|^k \exp(j * \varphi(u, v))$ 
7    $g(x, y) \leftarrow \mathcal{F}^{-1}\{\hat{T}^k(u, v) \cdot H^*(u, v)\}$ 
8    $psr \leftarrow PSR(g(x, y))$ 
9   if  $psr \geq \tau$  then
10     $T \cup t_N(x, y)$ 
11     $H(u, v) \leftarrow$  Update  $H$  using  $t_N(x, y)$ 
12   $N \leftarrow N + 1$ 

```

---

Algorithm 1 receives a face images database as first parameter. As the goal is to maximize the performance of a filter which averages the training images (see Subsection 3.2), the database must contain as many images as possible. The second argument  $S_f$  contains some ideal images for building an initial filter. So, a first arbitrary face image can be included in the training set only if it is similar to the initial filter. The arbitrary face image is correlated with the filter  $H(u, v)$  and, if its sharpness is equal to or greater than the third argument  $\tau$ , then it is added to the training set  $T$  and used to update  $H(u, v)$ . Both the initial and updated filter  $H(u, v)$  in Algorithm 1 are designed by an average accumulator function  $H_i(u, v) = \frac{N-1}{N} H_{i-1}(u, v) + \frac{1}{N} \hat{T}_{current}^k(u, v)$ . Where  $\hat{T}_{current}^k(u, v)$  is the FT of a face image with  $k$ th-Law nonlinear filtering. The Peak-to-Sidelobe Ratio ( $PSR$ ) measures the peak sharpness in the correlation output; therefore, the larger the  $PSR$  the more likely the test image belongs to the true class [4]. The threshold value  $\tau$  is the minimal  $PSR$  that assures that an image region corresponds to a face. The threshold was experimentally determined and fixed at  $\tau = 10$ . The peak-to-sidelobe ratio ( $psr$ ) measures the number of standard deviations at which the peak is found to be above the mean value in the correlation output. The  $PSR$  metric is given by the expression 9, where  $\mu_{area}$  and  $\sigma_{area}$  respectively

are the mean and standard deviation of some area or neighborhood around, but not including, the peak.

$$psr = \frac{(peak\ value - \mu_{area})}{\sigma_{area}}. \quad (9)$$

### 3.2 Nonlinear composite filter design

In order to produce a sharp and high peak on a less noisy correlation plane, our proposed algorithm is based on the Nonlinear Composite Correlation Filter and ASEF filters. An exact filter is considered as a weak classifier because it only matches the training image. However, averaging many weak classifiers, as ASEF filters do, yields a robust classifier that will match with many objects of the same class, even if they do not belong to the training set. For the modeling of some principal distortions not contained in the training set, each training image for Algorithm 2 is processed by the following operations. For in-plane rotation, each image is rotated  $+15$  and  $-15$  degrees. Zero mean Additive White Gaussian Noise (AWGN), with variances of 0.1 and 0.2, was added to each face image for noise modeling. There are two main face shapes, rounded or elongated, which is an important issue that must be taken into account by approaches based on template such as that which is presented in this paper. For this reason, the images in  $T$  were scaled in width to  $\frac{2}{3}$  and  $\frac{3}{4}$ . Finally, each training image was flipped from left to right. These image transformations are summarized in Table 1.

**Table 1.** Image transformations to model distortions not contained in  $T$ .

Number	Image transformation
0	Original images
1	In-plane rotation of 15 degrees
2	In-plane rotation of $-15$ degrees
3	AWGN with mean 0 and variance 0.1
4	AWGN with mean 0 and variance 0.2
5	Scale in width to $\frac{2}{3}$
6	Scale in width to $\frac{3}{4}$
7	Flipping left to right

The input argument for Algorithm 2 is the training set  $T$  generated by Algorithm 1, while the output is a simple Nonlinear Composite Correlation Filter. First, eight training sets are derived from the application to the input training set of the transformations discussed in Table 1. Second, each training set generated is used to build a strong classifier  $H_{ASEF}(u, v)$ , based on the design of an ASEF filter. For computational convenience, ASEF filters take on the original size of the images and are then synthesized in the spatial domain to enable a

**Algorithm 2:** Nonlinear composite correlation filter design.**Data:** training set  $T$ **Result:** Correlation filter  $H(u, v)$  for face detection

---

```

1  $T_{trainset} \leftarrow \{\}$ 
2 for  $i \leftarrow 0$  to 7 do
3    $T_i \leftarrow$  Apply the  $i$ th image transformation of Table 1 to images in  $T$ 
4    $H_{ASEF}(u, v) \leftarrow$  Design an ASEF filter (strong classifier) with equation 8
      using  $T_i$  as training set
5    $h_i(x, y) \leftarrow \mathcal{F}^{-1}\{H_{ASEF}(u, v)\}$ 
6    $h'_i(x, y) \leftarrow$  Padding  $h_i(x, y)$  with  $mean\{h_i(x, y)\}$ 
7    $T_{trainset} \cup h'_i(x, y)$ 
8  $H(u, v) \leftarrow$  Design a nonlinear composite correlation filter with equation 5 using
   the training set  $T_{trainset}$ .
```

---

padding operation with their mean values. Finally, each ASEF filter in the spatial domain is taken as training datum in the design of a nonlinear composite filter  $H(u, v)$ , as described in section 2.1.

### 3.3 Face detection algorithm

This work used a bank of filters of different dimensions, which is correlated with an input scene image  $s(x, y)$  for detecting human faces. This proposal is composed of two main steps: 1) Construction of the bank of nonlinear composite correlation filters and 2) the face detection algorithm. Correlation filters for face detection are designed in the step 1, and must contain enough facial information to produce a high, sharp peak centered at the face location. These filters are stored in order to use them each time that face detection is performed on a test image. Step 2 corresponds to the face detection algorithm, which is given by Algorithm 3. A given test image  $s(x, y)$  is improved by function  $s'(x, y) \leftarrow preprocessing(s(x, y))$ , which performs the following operations. First, retina filtering is applied to images for improving the quality of images with non-uniform illumination [8]. Although we could use logarithmic transformation for this first operation, experimentally it was observed that retina filtering is best for retrieving the edge information in a scene image. An energy normalization operation is then applied to the images, after which, a Cosine window is applied to the image to reduce the frequency effects of the edges of the image when transformed by FT. Lastly, the processed input image is padded using its mean value to the same dimension as the correlation filter. The improved image  $s'(x, y)$  is transformed by FT and  $k$ th - Law nonlinear filtering to obtain  $S^k(u, v)$ . The detection process iteratively correlates  $S^k(u, v)$  with the stored filters  $H(u, v)$  whose dimension is less or equal to the dimension of the input test image. Each correlation output  $g(x, y)$  is examined for search peaks with  $psr$  values greater than or equal to the threshold value  $\tau$ . The threshold value  $\tau$  indicates that the object has been located and recognized as an authentic face. If the algorithm



proposed in this work detects a face in an iteration, then the coordinates  $(x, y)$  and dimension *height, width* of each detected face is added to a subdetection vector  $D_{det}$ . Finally,  $D_{det}$  is filtered by the  $filterdetections(D_{det})$  function in order to merge detections over a same face region. This is because two filters with nearby dimensions can produce the peak in the same location or nearby locations that include the same portion of the face.

---

**Algorithm 3:** Face detection by correlation filters.

---

**Data:** Test image  $s(x, y)$ , detection threshold  $\tau$   
**Result:** Vector  $D_{det}$  with location and extent of each face detected in  $s(x, y)$ .

- 1  $s'(x, y) \leftarrow preprocessing(s(x, y))$
- 2  $S(u, v) \leftarrow \mathcal{F}\{s'(x, y)\}$
- 3  $S^k(u, v) \leftarrow |S(u, v)|^k \exp(j\varphi(u, v))$
- 4  $D_{det} \leftarrow []$
- 5  $count \leftarrow 1$
- 6 **while**  $dim\{s(x, y)\} \leq dim\{H_{count}(u, v)\}$  **do**
- 7      $H_{count}(u, v) \leftarrow$  Read the stored filter number  $count$
- 8      $g(x, y) \leftarrow \mathcal{F}^{-1}\{S^k(u, v) \cdot H_{count}^*(u, v)\}$
- 9      $subdetections \leftarrow searchdetections(g(x, y), \tau)$
- 10     $D_{det} \leftarrow [D_{det}; subdetecciones]$
- 11     $count \leftarrow count + 1$
- 12  $D_{det} \leftarrow filterdetections(D_{det})$

---

In this section, an algorithm that combines the principles of nonlinear SDF and ASEF correlation filters is proposed. An important feature of this algorithm is that it takes advantage of the facts that the correlation filters are shift-invariant and they allows multiple-object detection with only one operation.

## 4 Experimental results

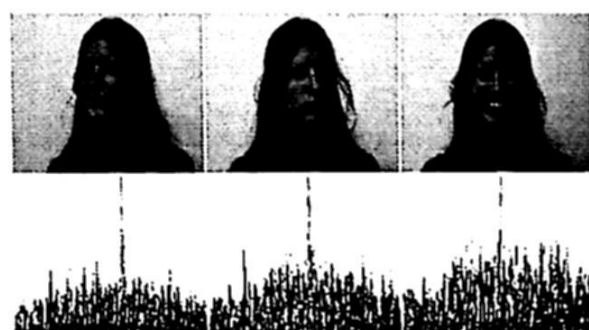
In order to design the nonlinear composite correlation filters, the training set  $T$  was generated by Algorithm 1 using the Yale B Face database [9]. This database contains 2414 face images of 38 people, each of whom has 60, 63 or 64 gray-scale face images with a resolution of  $168 \times 192$  pixels and taken under non-uniform illumination conditions. From this set only two face images of person 1 were selected for the initial filter. During the execution of the algorithm, it was observed that face images that are similar to the initial filter produce high  $psr$  values, while  $psr$  values were low for other facial classes. Whenever  $psr$  values are equal to or greater than  $\tau$ , the test images are included in  $T$ . At the output of the Algorithm 1,  $T$  contains 1795 facial images.

Two experiments were conducted. In the first experiment, two composite correlation filters were designed in order to locate human faces in two test images. The first filter was designed using the Algorithm 2 with input training set  $T =$



**Fig. 1.** Correlation output for images with homogeneous and structured backgrounds.

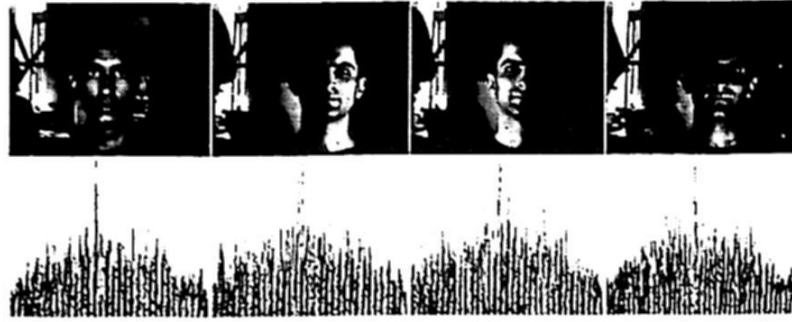
[1795], so the nonlinear composite correlation filter was designed with 14360 face images. The second filter was designed by the ASEF algorithm using the same input training set  $T = |1795|$  as the first filter and applying the image transformations from Table 1. Figure 1 shows the results of this experiment. In the first image, the person appears over a homogeneous background, while in the second image the person appears in a structured background. The second and third rows depict the improvements in illumination while retaining the edge information. The correlation outputs of the filters for each image are shown in the fourth row. As can be seen, the height of the peaks for images with homogeneous background are similar. However, the peak yield with the proposed algorithm is sharper than the peak produced by ASEF in both the homogeneous and structured background.



**Fig. 2.** Correlation output of the proposed strategy when test images contain faces with slight variations in pose and expressions.

The proposed method is able to detect faces that present slight variations in pose as shown in Figure 2. As facial expressions modify the facial structure, a

noisy correlation output with a sharp and low peak is produced. In Figure 3, the face in the scene presents non-uniform illumination. In this case, the proposed algorithm responds suitably with a peak in the center of the face region.



**Fig. 3.** Correlation output of the proposed strategy when the face contains different conditions of illumination.

Eleven test sets were created in order to perform a more intensive evaluation, and were, in turn, classified into two test sets according to their background. The first set contains images with homogeneous backgrounds from the FEI database [10], and was partitioned into two subsets: 1) Frontal, which consists of 800 scene images with frontal faces, and 2) Pose, which consists of 1000 images where the face is presented in different poses. Figure 4 shows a sample of these sets. The other nine sets contain a total of 576 scene images with a structured background, which correspond to 9 people from the Yale B database [9]. The image set for each person is identified as YaleB11, YaleB22, YaleB27, YaleB28, YaleB29, YaleB30, YaleB32, YaleB33 and YaleB34. Each set contains 64 images taken under non-uniform illumination and presents a pose different from the rest. Figure 5 shows the pose selected for each person. Using the training set  $T$ , two banks of filters were created for this evaluation. The first bank contains filters designed by the proposed algorithm, while the second bank contains ASEF filters.



**Fig. 4.** Sample of Frontal and Pose test sets with homogeneous background.

The performance evaluation was conducted in terms of the following metrics [11]: a) Localization and Recognition Rate  $LRR$ , and b) Recognition and Deviated Localization Rate  $RDLR$ . The  $LRR$  metric searches for a value close to



Fig. 5. Sample of pose selected for each person with structured background.

100, which indicate good performance, while in *RDLR* an optimum performance is given by a value close to 0. The use of these metrics allow the analysis of any correlation filter's capacity to locate and recognize the target object in scene images. The results of the evaluation are shown in Table 2. As can be observed, the proposed algorithm obtains the best performance, in terms of *LRR* metric, with test sets Frontal, YaleB11, YaleB22 and YaleB32, in which the facial images are frontal or with slight variations in pose. A greater variation in pose causes the algorithm to perform poorly, as shown in the performance achieved with YaleB30 and YaleB333. In terms of this metric, the proposed algorithm outperformed the ASEF algorithm in all test sets. The metric *RDLR* denotes the percentages of images where the filter produced the peak in a location different from the center of the face. In some cases, the detection window captures a face region that could be processed as a partial face. It was noted in the experiments that this is mainly due to non-uniform illumination. An important issue to note in the results obtained with this metric is that the proposed algorithm performs better than the ASEF algorithm in scene images with a structured background, while the ASEF algorithm performed best in those images with homogeneous backgrounds from Frontal and Pose subsets.

Table 2. Performance of the proposed algorithm and ASEF filter in the face detection task.

Background	Training set	Proposed algorithm		ASEF Algorithm	
		LRR	RDLR	LRR	RDLR
Homogeneous	Frontal	65.5	8.75	62.87	6
	Pose	51.50	20.20	42.80	18.40
Structured	YaleB11	53.12	7.80	39.06	4.68
	YaleB22	73.84	7.69	44.61	13.84
	YaleB27	41.53	1.53	9.23	9.23
	YaleB28	46.15	7.69	12.30	33.81
	YaleB29	52.30	15.38	12.30	9.23
	YaleB30	26.15	0.00	12.30	12.30
	YaleB32	92.30	3.07	32.30	36.92
	YaleB33	29.23	3.07	7.69	3.07
	YaleB34	41.53	10.76	9.23	7.69



## 5 Conclusions

A face detection algorithm based on nonlinear composite correlation filters is presented in this paper. Averaging of training images emphasizes common facial features, which gives a greater robustness to the nonlinear composite correlation filter for detecting face regions in images of real-world scene. The proposed algorithm uses strong classifiers designed with distorted versions of a training set for obtaining tolerance to scale, small variations in rotation and pose, and non-uniform illumination. Topics for future research include the application of optimization techniques for selecting the training images from a face database.

## 6 Acknowledgments

This work was partially supported by Consejo Nacional de Ciencia y Tecnología (CONACYT), with the scholarship number 344833/239152 for author whose name is given first above.

## References

1. Yang, M.H., Kriegman, D.J., Ahuja, N.: Detecting faces in images: a survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **24** (2002) 34–58
2. Z.-Li, S.: 2. In: Face detection. Springer (2005) 13–37
3. Bolme, D.S., Draper, B.A., Beveridge, J.R.: Average of synthetic exact filters. *IEEE Conference on Computer Vision and Pattern Recognition* (2009) 2105 – 2112
4. Vijaya-Kumar, B.V.K., Mahalanobis, A., Juday, R.: Correlation pattern recognition. Cambridge University Press (2005)
5. VanderLugt, A.: Signal detection by complex spatial filtering. *IEEE Transactions on Information Theory* **10** (1964) 139–145
6. Casasent, D., Chang, W.T.: Correlation synthetic discriminant functions. *Appl. Opt.* **25** (1986) 2343–2350
7. Javidi, B., Wang, W., Zhang, G.: Composite fourier-plane nonlinear filter for distortion-invariant pattern recognition. *Society of Photo-Optical Instrumentation Engineers* **36** (1997) 2690–2696
8. Vu, N.S., Caplier, A.: Illumination-robust face recognition using retina modeling. In: *Proceedings of the 16th IEEE international conference on Image processing. ICIP'09*, Piscataway, NJ, USA, IEEE Press (2009) 3253–3256
9. Georgiades, A., Belhumeur, P., Kriegman, D.: From few to many: Illumination cone models for face recognition under variable lighting and pose. *IEEE Trans. Pattern Anal. Mach. Intelligence* **23**(6) (2001) 643–660
10. de Oliveira Junior, L.L., Thomaz, C.E.: Captura e alinhamento de imagens: Um banco de faces brasileiro. Technical report, Department of Electrical Engineering, FEI, Sao Bernardo do Campo, So Paulo, Brazil, (june 2006)
11. Santiago-Ramírez, E., González-Fraga, J.A., Lázaro-Martínez, S.: Face recognition and tracking using unconstrained non-linear correlation filters. *Procedia Engineering* **35**(0) (2012) 192–201 International Meeting of Electrical Engineering Research 2012.

# Design and Hardware Implementation of Digital Amplitude Modulation on FPGA

J.A. Galaviz-Aguilar<sup>1</sup>, J.R. Cárdenas-Váldez<sup>1</sup>, J.A. Reynoso-Hernández<sup>2</sup> and J.C. Núñez-Pérez<sup>1</sup>

<sup>1</sup> National Polytechnic Institute, Research and Development Center of Digital Technology (CITEDI-IPN).

Av. Instituto Politécnico Nacional 1310, Mesa de Otay, C.P. 22150, Tijuana, B. C., México.

<sup>2</sup>Centro de Investigación Científica y de Educación Superior de Ensenada (CICESE), Carretera Ensenada-Tijuana No. 3918, Zona Playitas, C.P. 22860, Ensenada, B. C., México.  
{jgalaviz,jcardenas,nunez}@citedi.mx, apolinar@cicese.mx

*Paper received on 12/11/13, Accepted on 01/19/14.*

**Abstract.** This paper presents a novel design model of the basic principle of digital amplitude modulation implemented over a DSP-FPGA board. The design requirements are based in a sequential Top-Level methodology using VHDL. In the signal generation is used a Direct Digital Synthesis approach to control the accuracy of the carrier and modulated signal frequencies. The results are presented with simulations in Matlab and using a testbench in Modelsim to functional design verification. The experimental tests show the output modulated waveforms in order to evidence the correct implementation of the design.

**Keywords:** Digital Amplitude Modulation, Direct Digital Synthesis, FPGA, Look up table, Matlab/Simulink, VHDL.

## 1 Introduction

The modern mobile communication systems are based in digital schemes of modulation. Amplitude Modulation (AM) principle is the process where the information is carried via a fast frequency signal in the high frequency (HF) band. In AM the modulation signal controls the carrier amplitude, causing a linear change, but maintains the carrier frequency. Digital Amplitude Modulation (DAM) is a method commonly used in radio communication and presents advantages of accuracy and control of the signal compared with analog AM [1-2]. In this work the advantages of Field Programmable Gate Array (FPGA) based on flexibility for developing of digital hardware implementation is exploited, allowing a rapidly prototyping of the overall circuit.

Accordingly, in this work the DAM design is based in a direct digital synthesis (DDS) technique; which permits a digital controllability in the carrier frequency and modulating waves. This paper is organized as follows: Section 2 describes the whole

model divided into designs blocks explaining the design and operation of the signal generator and digital amplitude theory. In Section 3 hardware simulation and verification of design are presented. In the Section 4 the testing results are depicted. Finally in Section 5 the conclusions are presented.

## 2 Model design and simulation

The implementation design is described in a high level model in the Matlab-Simulink environment. Furthermore, it is described with blocks that all belong to the DSP Builder Blockset Library. The particularity of this library is that Hardware Description Language (HDL) can be directly generated from the model description block and included in a Quartus II project. The project contains all the source code using VHDL Very High Speed Integrate Circuit Hardware Description Language (VHDL) and generated by the Signal Compiler tool. For this purpose is necessary to make some tasks in Quartus II software in order to fit the design project, pins assignments and timing constraints. Fig. 1 shows the overview model of the entire implementation using DSP Builder blocks.

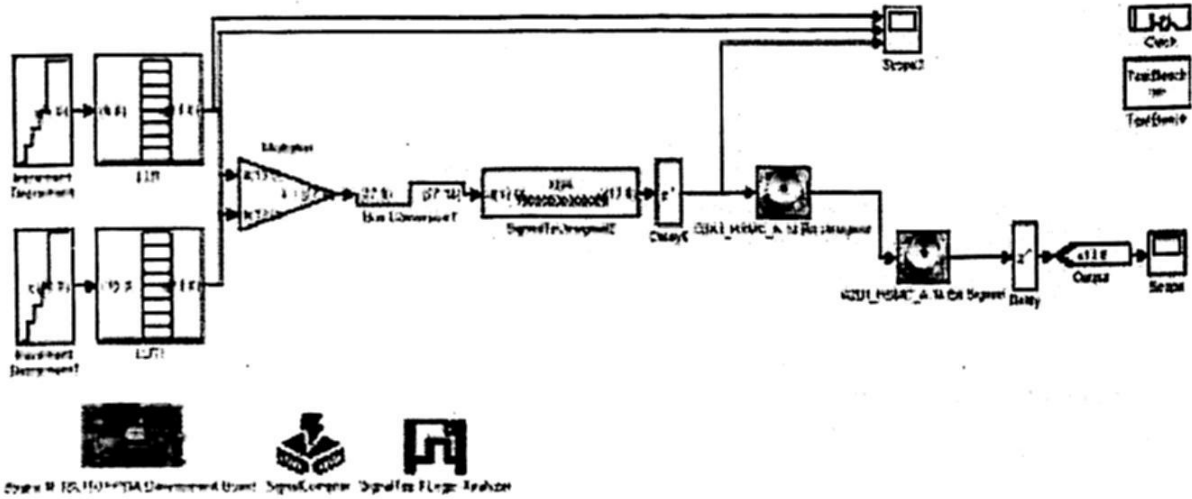


Fig. 1. Overall Design Model in Simulink.

### 2.1 Signal Generator using DDS Approach

The signal generator block is based on DDS methodology, which consists in generating a periodic and discrete-time waveform of known frequency  $F_0$ [3]. DDS is a total digital frequency synthesizer, this includes an increment accumulator, wave store using a look-up table (LUT) and digital to analog conversion (D/A). This principle is shown in Fig. 2, where the phase locked-loop (PLL) component represents the sampling clock time in the signal generator and the overall design logic into FPGA.

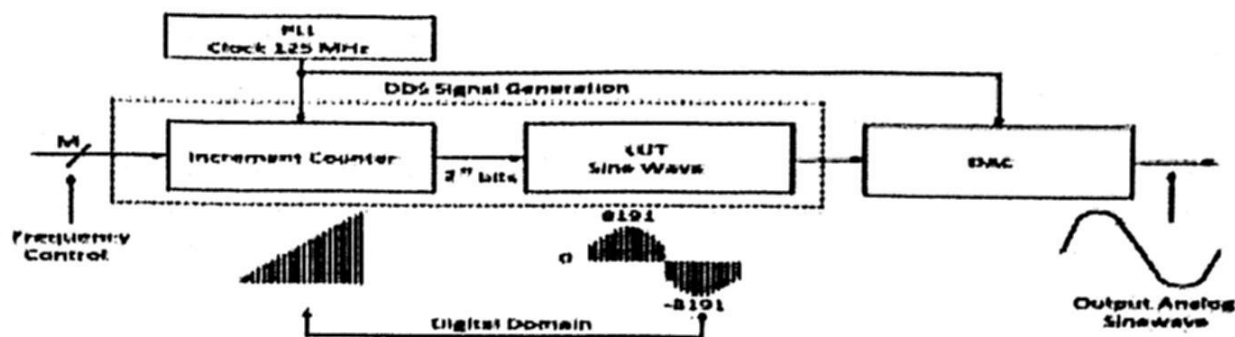


Fig. 2. DDS principle for digital signal generator.

The DDS signal generation shown in Fig. 2 is based on an increment block of  $2^n$ -bits, which are addressed with a sine LUT. The LUT size has to correspond to the counter resolution. The output steps counter are addressed to the LUT, where is stored the sample data computed of sine function. The counter output value augments on each rising edge of the clock; the DDS block model is shown in Fig. 3.

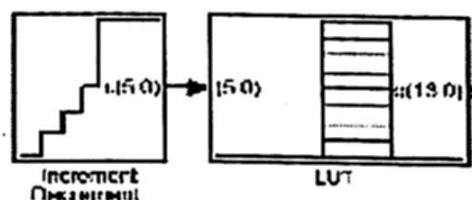


Fig. 3. DDS component blocks.

There are three steps to obtain a proper signal generation output. First, the number of samples in a signal period has to fit the size of the LUT [4]. Secondly, the LUT resolution is defined by the number of bits related to the counter depth. Finally, the maximum frequency of the signal has to be theoretically 2 times considering sampling Nyquist's criterion, but practically 10 times lower than the frequency of the internal clock established in 125 MHz. The following expression determinates the output sinewave frequency in the DDS component.

$$F_{out} = (M/2^n)f_{clock} \quad (1)$$

Where  $M$  represents a digital control to determinate the output frequency [5-6]. The values of the LUT are coded with a word of  $n$  bits signed to compromise between precision and the bus size conversion at the input of a multiplier. It is necessary a Matlab array with a smaller length than  $2^{(address\ width)}$ , which represents one cycle of a calculated sinewave data stored in the LUT, consider the

$$8191 * \sin([0:2 * \pi/(2^5):(2 * \pi)]) \quad (2)$$

## 2.2 Digital Amplitude Modulation Principle

The aim of this work is a fully DAM implementation method based in a digital control amplitude values of two LUT sinewaves. For this purpose a carrier discrete signal and load the signal modulation are considered [7], this process is realized using



a multiplier; which the product represents the amplitude modulate signal, in Fig. 4 are shown the input data of 14-bit that represents the frequencies to multiply. In the output is necessary a truncation of bits in order to send only the 14-bit most significant bit according with the DAC resolution.

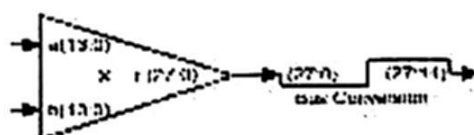


Fig. 4. Product multiplication block with 14-bit truncation.

### 2.3 Simulation of the Model in Simulink

The simulation results in Simulink are able to verifications in generated signals into the DDS blocks; which permits improving the functionality before synthesis stage of the model into a VHDL. The simulation shown in Fig. 5 system verification to checks the carrier frequency and the modulated signals generated by the DDS component and the output modulated wave respectively.

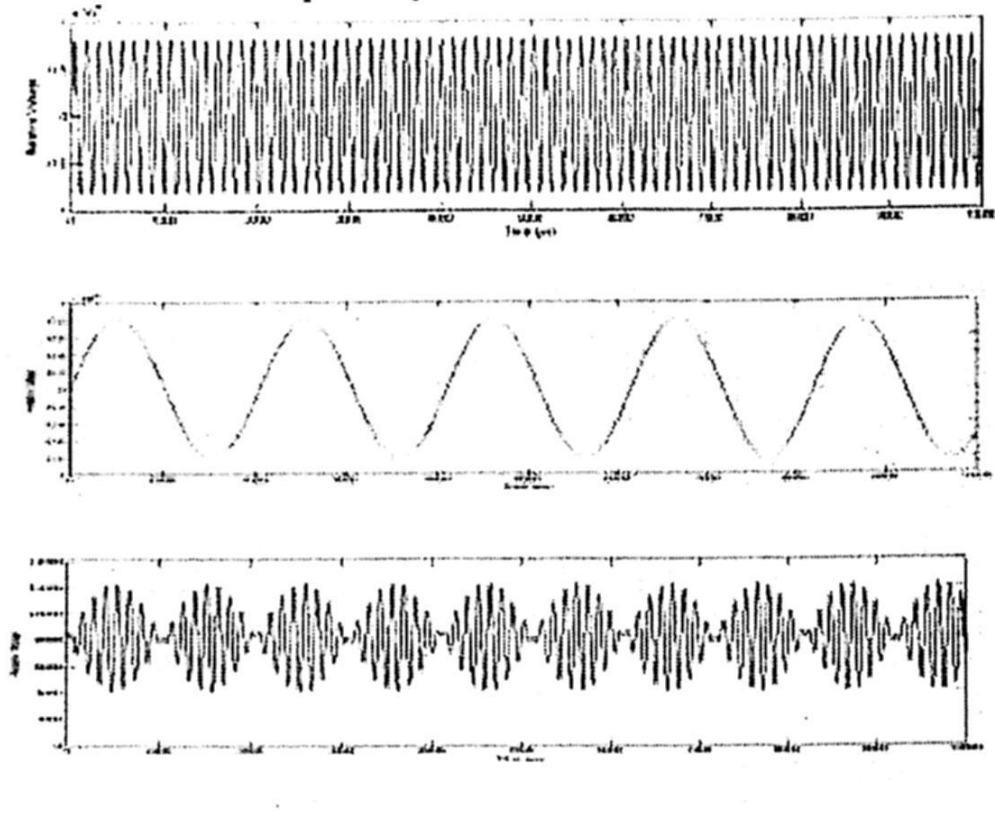


Fig. 5. Simulation results into Matlab (a) RF carrier (b) Modulating wave (c) Modulated result.

### 2.4 Hardware Simulation and Verification of Design

After simulation, the model is verified using a test bench on Modelsim which allows debugging tasks of the design and checking the performance in FPGA device previous the synthesis stage and program. Fig. 6 represents the output analog modulated waveform in time using the sampling clock of 125 MHz, also can check the reset signal into the system.

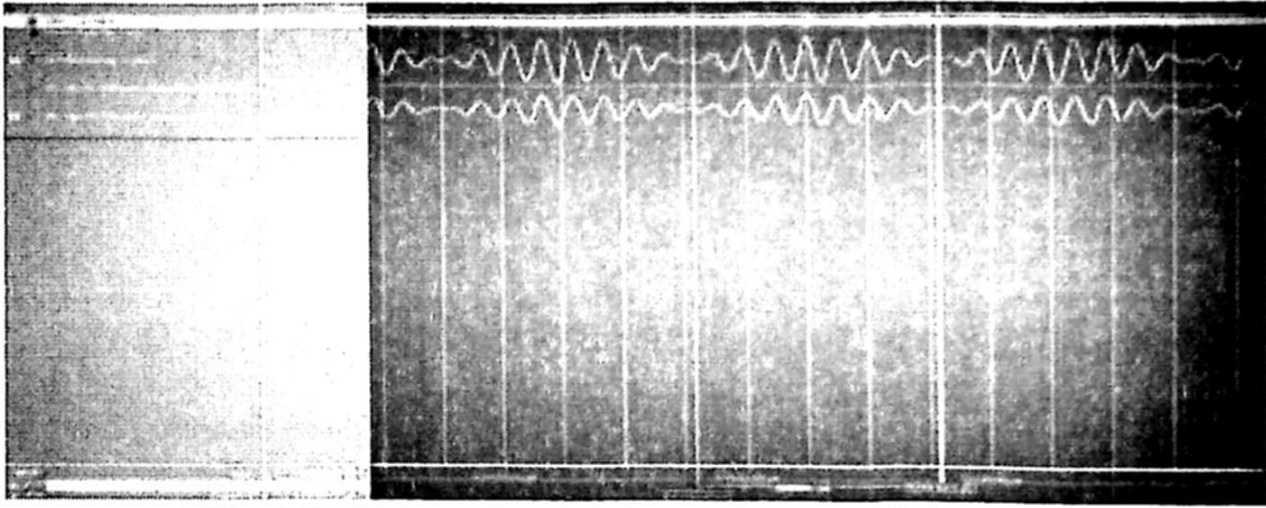


Fig. 6. Modelsim test bench design amplitude modulation output.

The design synchronization is based in a unique clock of 125 MHz which is shown in Fig. 7. According with this clock the internal data sampling is accomplished each cycle clock (rising edge) generating the sample data, the reference clock cycle has a period of 8 ns. The signal of the reset internal is running in synchronization with the clock.

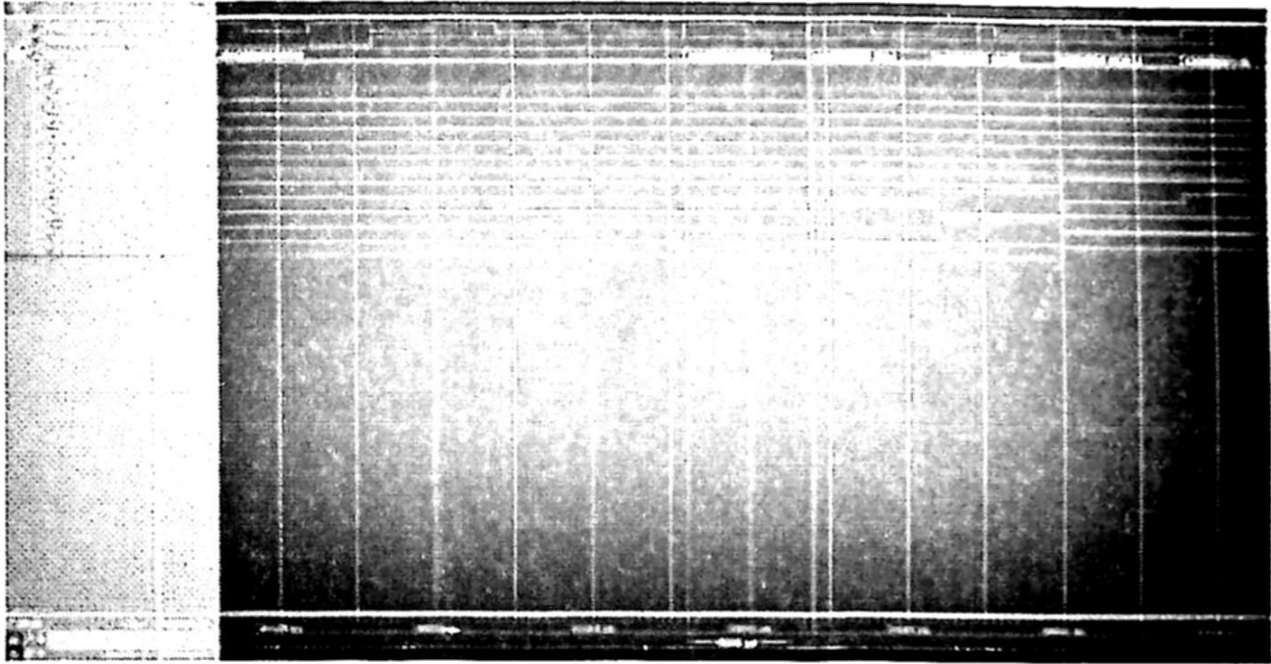


Fig. 7. Modelsim test bench design simulation.

### 3 Implementation design

For this design all the internal circuitry is controlled using a synchronous clock of 125 MHz, with an active low reset implemented through a PLL circuit; this clock also acts as the reference DAC sampling clock DAC. The design source code generated by the signal compiler tool which is processed, compiled and analyzed using the Quartus II tools several times to debugging and synthesize the overall circuit, before the implementation of the design in the FPGA device.

To check the VIIDL of the design is necessary to open in a new project in order to complement and debug the code. Indeed, the DSP Builder does not allow the I/O pin

assignments, it has to be made manually using the pin planner tool of Quartus II Software. In addition, the interface between the FPGA and the DAC has to be created. The implementation structure of amplitude modulating design is integrated into the architecture of DSP Stratix III 3SL150 Development Board of Altera and a data acquisition card.

3.1 Conversion to Data Acquisition Card

The last step of the implementation consists in a conversion process that aims acquiring the data with the DAC. An overview of the conversion process can be seen in the Fig 8. For the design are used the channel A of the DAC where is necessary converts the representation signed to an unsigned type of 14-bit to send an analogic signal observable with an oscilloscope. The signal directly obtained at the output of bus type with the conversion XOR is used for the interpretation.

However, the real signal that would be used in case of including the implementation in a test bed is the one provided by the DAC.

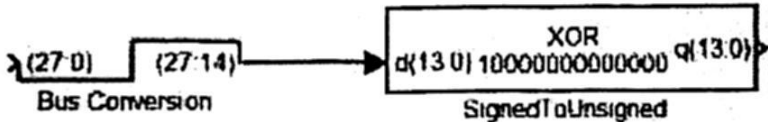


Fig. 8. Data type conversion chain.

3.2 Overall FPGA Resources Utilization

Total resources consumption by the implementation of model is shown in chip floor pan in the Fig. 9 where the significant usage is described by the utilization of embedded adaptive logic modules. In the design, implementing ROM resources of the FPGA for store the LUT sinewave sample data is needed, where is implemented over embedded memory block resources of 144 Kbits on chip.

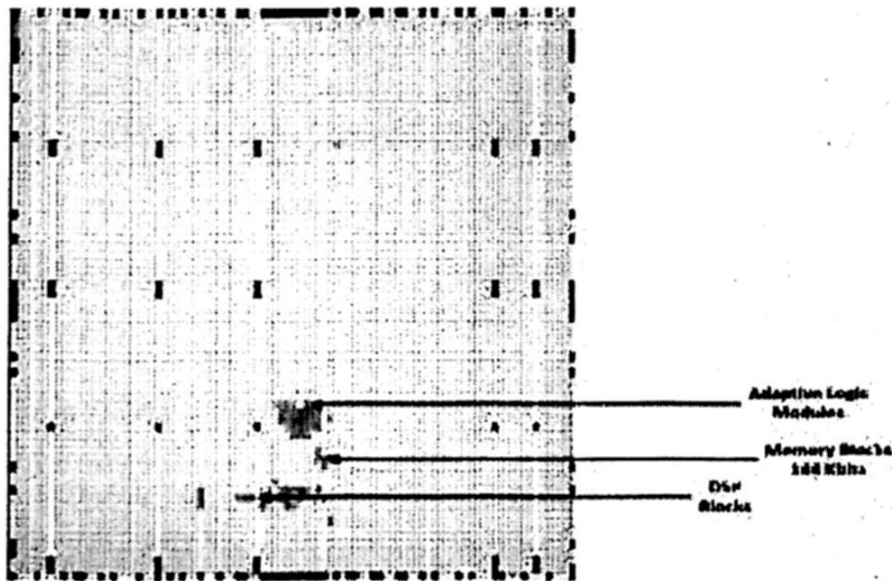


Fig. 9. Chip planner with overall resource usage.

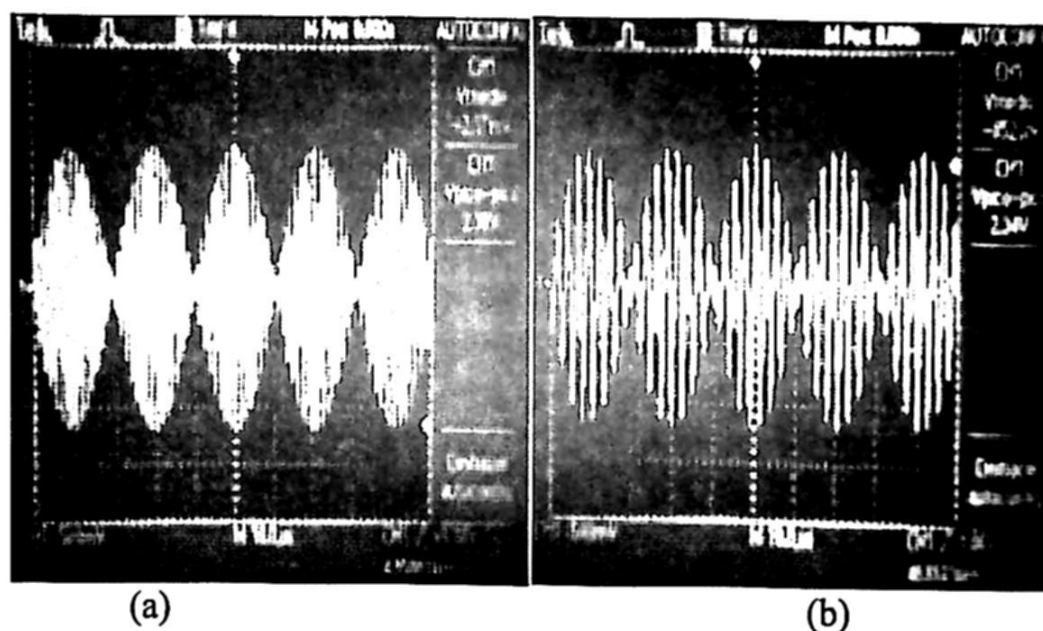
The overall resources used in the FPGA for the digital amplitude modulation design are summarized in Table 1; the consumed resources necessary by the whole design implemented are denoted in detail.

**Table 1.** Overall resources consumed in FPGA.

FPGA Resources	DSP Blocks (Multipliers)	Dedicated Logic Registers	Memory bits	Adaptive Logic Modules
Total used	2	738	43,008	330
FPGA Available	384	113,600	5,630,976	113,600
Percent used	<1 %	<1 %	<1 %	<1 %

## Testing Results

The testing results are depicted in Fig. 10 generating two different modulated waveforms. The frequency of carrier shown in Fig. 10 (a) is established in 4 MHz and the modulated waveform in 125 KHz and (b) carrier is 2 MHz and 125 KHz modulated waveform.



**Fig. 10.** Modulated transient waves in oscilloscope.

## 4 Conclusions

The design model presented in this paper basically demonstrates a short time frame implementation of a digital Amplitude Modulation system into the FPGA. Through the experiment of an implementation into a Stratix III FPGA device was presented hardware architecture to a digital amplitude modulation model. Given the complexity and time invest to design in VHDL; DSP Builder provides an efficient tool to design practical circuits applications based in digital amplitude modulation.



A general design method for an amplitude modulation model applicable in any other work has been described. Finally a precise description and justification of the components used in the hardware implementation had been given. Also in order to validate the improved performance of the design hardware simulations using a test bench in Modelsim were presented. In addition to that, the feasibility of the design theory was also compared with the design proposed method, based on DDS to provide accuracy in the output waveforms.

## References

1. D. Xie, S. Tian and K. Liu "Design and Implementation of DDS based Digital Amplitude Modulation," in IEEE Circuits and Systems International Conference on Testing and Diagnosis, ICTD 2009, pp. 1-4, July 2009.
2. J. Gao, N. Liu, and X. Xu "Design of AM modulation signal generator based on Matlab/DSP Builder," in 2nd International Conference on Industrial and Information Systems (IIS). Vol. 1, pp. 527 - 530, 2010.
3. L. Cordesses, "Direct digital synthesis: a tool for periodic wave generation (part 1)," in IEEE Signal Processing Magazine, Vol. 21, No. 4, pp. 50 - 54, July 2004.
4. M. Butt and S. Masud. "FPGA based bandwidth adjustable all digital direct frequency synthesizer," in 9th International Symposium on Communications and Information Technology, ISCIT 2009. pp. 1399-1404, 2009.
5. G. Xiong, X. Zhou, and P. Ji "Implementation of the Quadrature Waveform Generator Based on DSP Builder," in International Symposium on Intelligent Information Technology Application Workshops, 2008. IITAW '08. pp. 773-776, 2008.
6. Z. Chail, and J. Shen "The Application of System Generator in Digital Quadrature Direct Up-Conversion," in Communications in Information Science and Management Engineering (CISME). Vol. 3, No. 4, pp. 192-197, ISSN: 2222-1859, 2013.
7. C. Erdoğan, I. Myderrizi, and S. Minaei "FPGA Implementation of BASK-BFSK-BPSK Digital Modulators," in IEEE Antennas and Propagation Magazine, Vol. 54, No. 2, pp. 262 -269, April 2012.

# FPGA Implementation of behavioral Models for RF Power Amplifiers

J. R. Cárdenas-Valdez<sup>1</sup>, J. A. Galaviz-Aguilar<sup>1</sup>, J. C. Núñez-Pérez<sup>1</sup>, C. Gontrand<sup>2</sup> and A. Calvillo<sup>1</sup>

<sup>1</sup>National Polytechnic Institute, Research and Development Center of Digital Technology (CITEDI-IPN). Tijuana, México.

<sup>2</sup>Institut des Nanotechnologies de Lyon (INL), Institut National des Sciences Appliquées de Lyon (INSA-Lyon). Villeurbanne, France.

{jcardenas, jgalaviz, nunez, calvillo}@citedi.mx;  
christian.gontrand@insa-lyon.fr

*Paper received on 12/01/13, Accepted on 02/19/14.*

**Abstract.** This paper presents experimental results of Power Amplifiers modeling controlled in Matlab environment. The model is implemented firstly through VHDL in Hardware and improved then by DSP Builder, this emulation is made in FPGA DSP Development Kit, Cyclone® III Edition-ALTERA proving a proper behavioral modeling for RF or high frequency applications. The whole system is able to consider the memory depth and nonlinearity order based on real PA measurements; this work leaves the option for future implementations as Digital Predistortion as linearization technique of PA behavior. The obtained VHDL results are based on a MPM, compared with DSP Builder and an ANN is trained showing a proper behavior.

**Keywords:** ANN, FPGA, GUI, Memory, Power Amplifier, VHDL.

## 1 Introduction

Nowadays, wireless communications scenario demands higher data-rate transmissions, complex envelop techniques like wide code division multiple access (WCDMA) and orthogonal frequency division multiplexing (OFDM) techniques are employed because of their high spectral efficiency, should be noted that to achieve this high data rata transmissions the schemes impose strict linearity requirements.

The Power Amplifier (PA) as the main device in the transmission chain involves undesirable effects as memory and plays a role in the transmitter nonlinearities creation, due that behavioral models have been proposed to represent PAs [1-2]. The PA modeling used for very high frequency (VHF) became an important issue before the fabrication, especially if the undesirable effects are considered for wideband applications in order to predict unexpected results, examples include techniques based on polynomial models trying to reproduce real PA measurements, others taking into account memory depth and nonlinearity order [3-4].

Volterra Series formalism is a proper and well known technique where the memory depth and nonlinearity order can be considered during the amplification process [5]. However, the nonlinearity is an inherent property of PAs, leading not only to inband signal distortion but also to outband spectral regrowth, which are strictly regulated especially with the current wireless communication systems which are continuously growing.

PAs exhibit memory effects as well, which mean the current output is stimulated by not only the input current but also by the previous input states, based on this affirmation a special truncation as Memory Polynomial Model (MPM) is used in this paper and compared with a model based on Artificial Neural Network (ANN)

The aim of this work is addressed to model and create a link between a software system as Matlab and hardware implementation through Very High Speed Integrated Circuit Hardware Description Language (VHDL) demonstrating a fully PA digital behavior leaving the alternative for a further linearization stage or specific application into Very High Frequency (VHF) band.

Modern Field Programmable Gate Array (FPGA) devices offer a multitude of resources. The FPGA characteristic is the possibility of implementing algorithms directly into hardware, maintaining the parallelism of the functioning in the implementation and thus minimizing the execution time, several implementations in the field are related to FPGA devices offering advantages like reprogramming, tolerance and debugging errors, reducing nonrecurring engineering cost and shorter processing time [6-8]. In this context, behavioral models implemented in hardware have been explored even for dual band and modern transmission standards as LTE [9-11]. This work shows flexibility between the modeling based on VHDL and DSP Builder through MPM and ANN as modeling technique.

This paper is organized as follows. First in Section 2 the description of the Volterra Series, MPM and ANN are presented. The Section 3 shows the results for the implementation made for MPM and ANN. Finally the Section 4 the conclusions are summarized.

## 2 Description

### 2.1 Memory Polynomial Model

The MPM is a subset of the Volterra series, it consists of several delay taps and non-linear static functions; and it represents a truncation of the general Volterra series where only the diagonal terms in the Volterra kernels are considered [9]. Thus, the number of parameters is significantly reduced compared to the original series.

The Volterra Series can be defined in discrete domain as equation (1), they can be used to describe the input-output relation considering the undesirable effects as memory:

$$y(n) = \sum_{m_1=0}^M h_1(m_1)x(n-m_1) + \dots + \sum_{m_1=0}^M \sum_{m_2=0}^M h_2(m_1, m_2)x(n-m_1)x(n-m_2) + \dots + \sum_{m_1=0}^M \sum_{m_2=0}^M \sum_{m_3=0}^M h_3(m_1, m_2, m_3)x(n-m_1)x(n-m_2)x(n-m_3) \quad (1)$$

where  $x(n)$  is the complex input base-band signal,  $y(n)$  is the complex output base-band signal,  $h_k$  are complex valued parameters and  $M$  is the memory depth. The MPM can be represented in equation (2):

$$y(n) = \sum_{q=0}^Q \sum_{k=1}^K a_{2k-1} |x(n-q)|^{2(k-1)} x(n-q) \quad (2)$$

where  $a_{k,q}$  are complex valued parameters,  $Q$  is the memory depth, and  $K$  is the polynomial order. Equation (2) can be rewritten just in terms of the memory depth, by the following equation (3):

$$y(n) = \sum_{q=0}^Q F_q(n-q) = F_0(n) + F_1(n-1) + F_2(n-2) + \dots + F_q(n-q) + \dots + F_Q(n-Q) \quad (3)$$

where  $F_q(n)$  can be expressed as equation (4) :

$$F_q(n) = \sum_{k=1}^K a_{2k-1} |x(n-q)|^{2(k-1)} x(n-q) \quad (4)$$

Each MPM stage can be subdivided in terms of the desired memory depth during the modeling; this can be represented in Fig. 1 showing the internal structure and the delays made for each step of the input signal. Each stage can be subdivided by internal processes related to the nonlinearity order where the calculated constants must be inserted. In Fig. 2 the internal stage for each delay is represented.

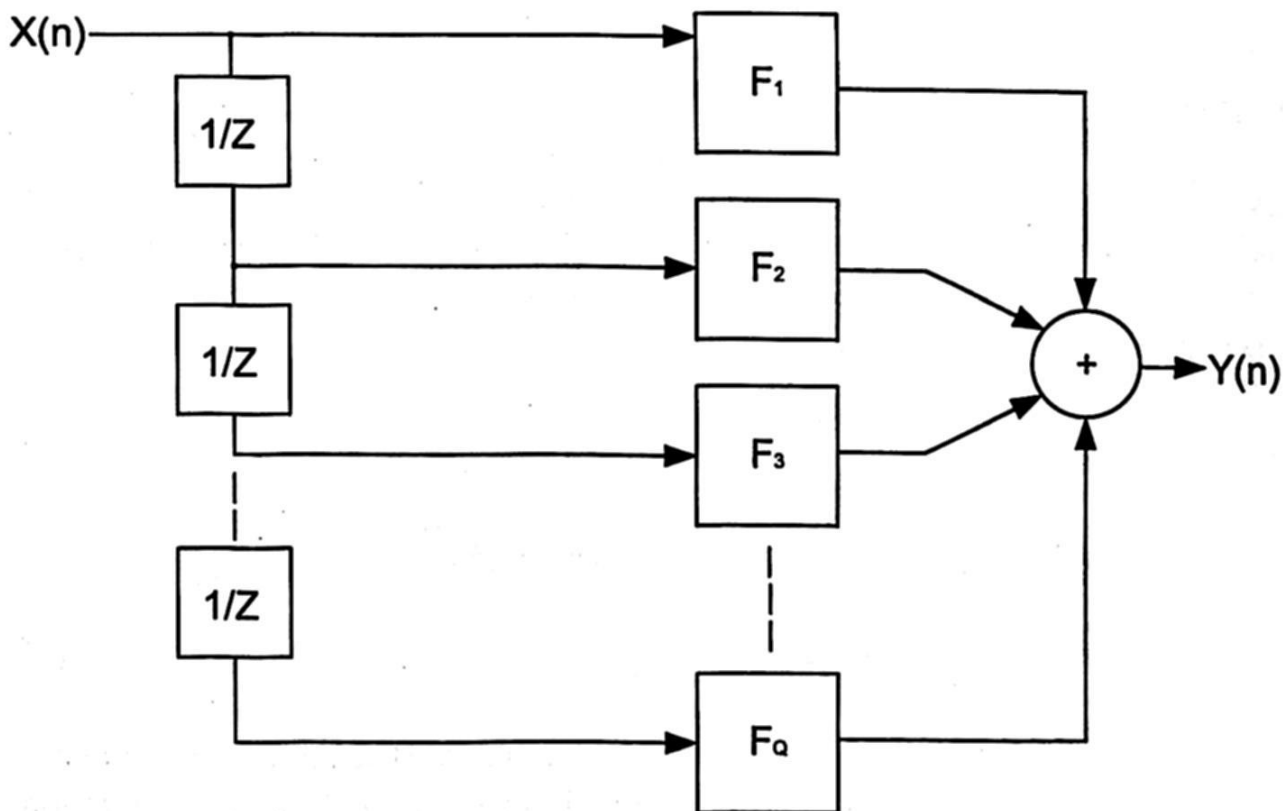


Fig. 1. General structure of the MPM



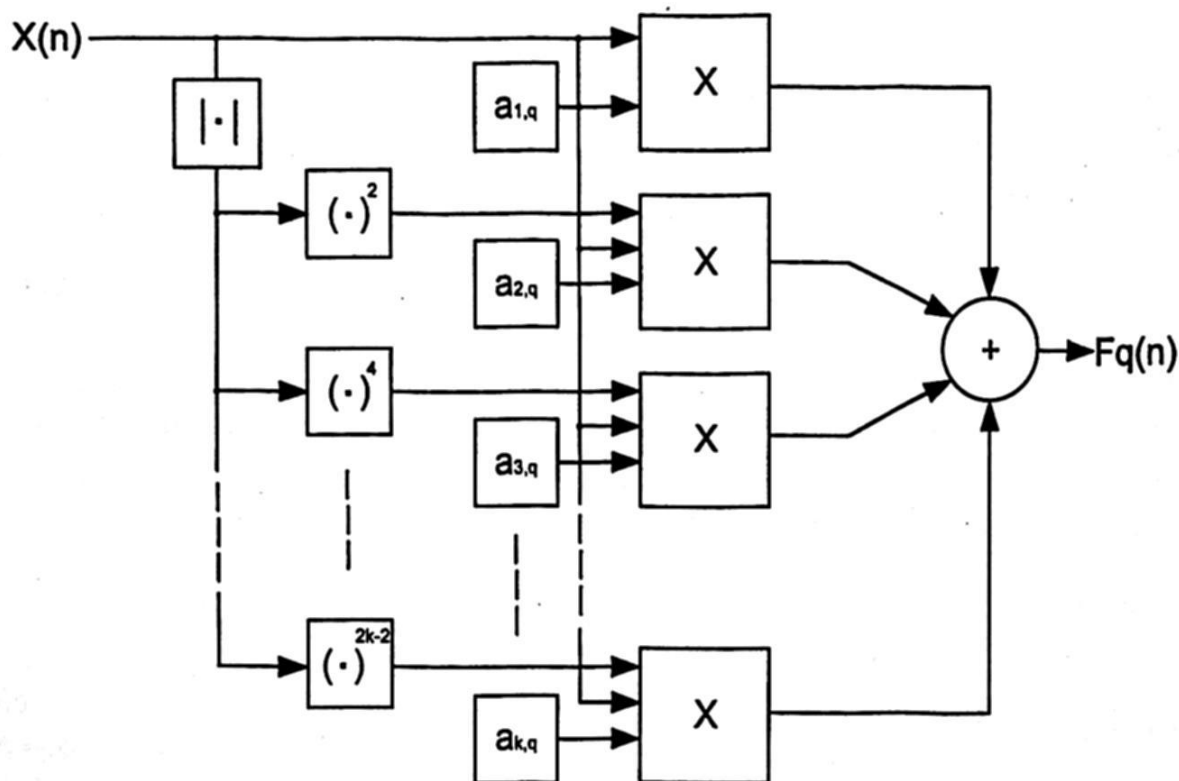


Fig. 2. Internal structure of each stage of the MPM.

The modeling based on MPM can be reproduced through ANN, this work makes use of its advantages containing five neurons, should be noted that the accuracy can be improved if more neuron are added.

## 2.2 Artificial Neural Network

A neuron is an element based on a biological model. It consists in a system with input signals with a unique output. The artificial neuron structure can be observed in Fig. 3. Fig. 4 represents an ANN where the weights can be denoted by  $w(i,j)$  and a propagation law is expressed by the equation (5):

$$h_i(t) = \sigma(w_{ij}, x_j(t)); h_i(t) = \sum w_{ij} x_j \quad (5)$$

The activation function can be denoted in the equation (6):

$$y_i(t) = f_i(h_i(t)) \quad (6)$$

Based on the biological model, the neuron has a threshold  $\theta$ . The inputs summatory is multiplied by its weight and the activated neuron has a response denoted by equation (7).

$$y(t) = f(\sum_j^n w_{ij} x_j - \theta_i) \quad (7)$$

$$y_i = 1, \text{ si } \sum_j^n (w_{ij}x_j) > \theta_i \text{ o } 0, \text{ si } \sum_j^n (w_{ij}x_j) < \theta_i \quad (8)$$

A neuron is an information-processing unit that is fundamental to the operation of a neural network. The Fig. 3 shows the model of a neuron, which forms the basis for designing neural networks.

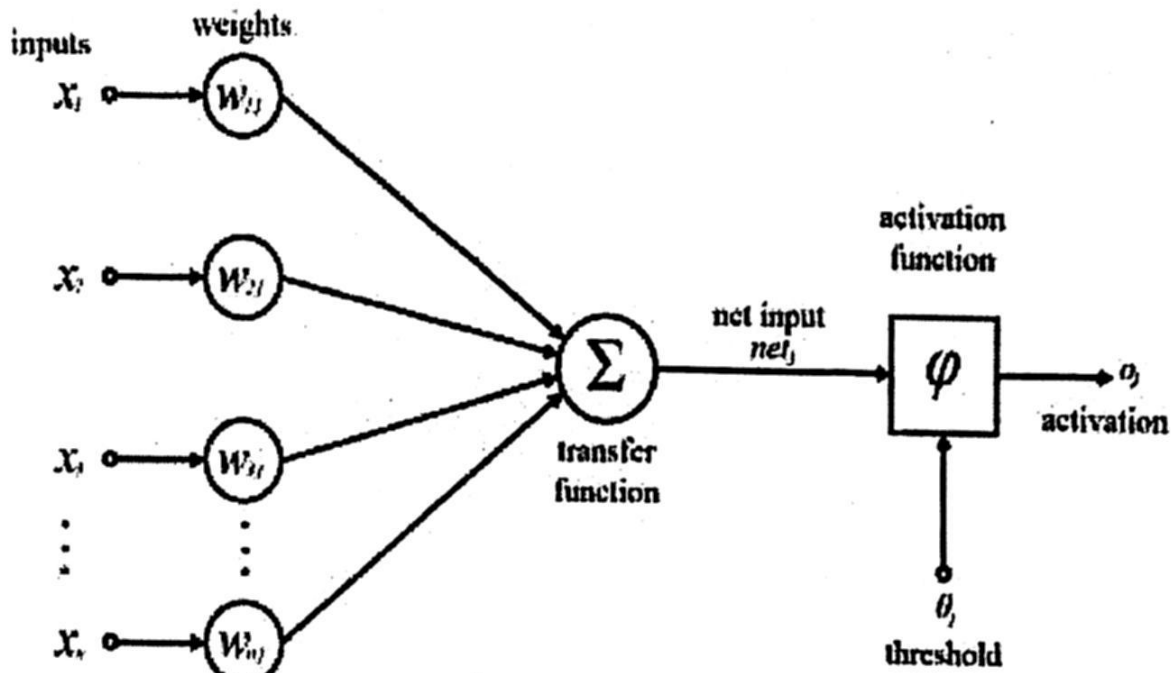


Fig. 3. Overview of a ANN.

Using just a neuron cannot be enough accurate, even more if the system is involving multiple inputs. Associating some neurons as a network, allows an easier emulation of complex functions. Fig. 4 is a representation of an ANN, is said to be fully connected in the sense that every node in each layer of the network is connected to every other node in the adjacent forward layer.

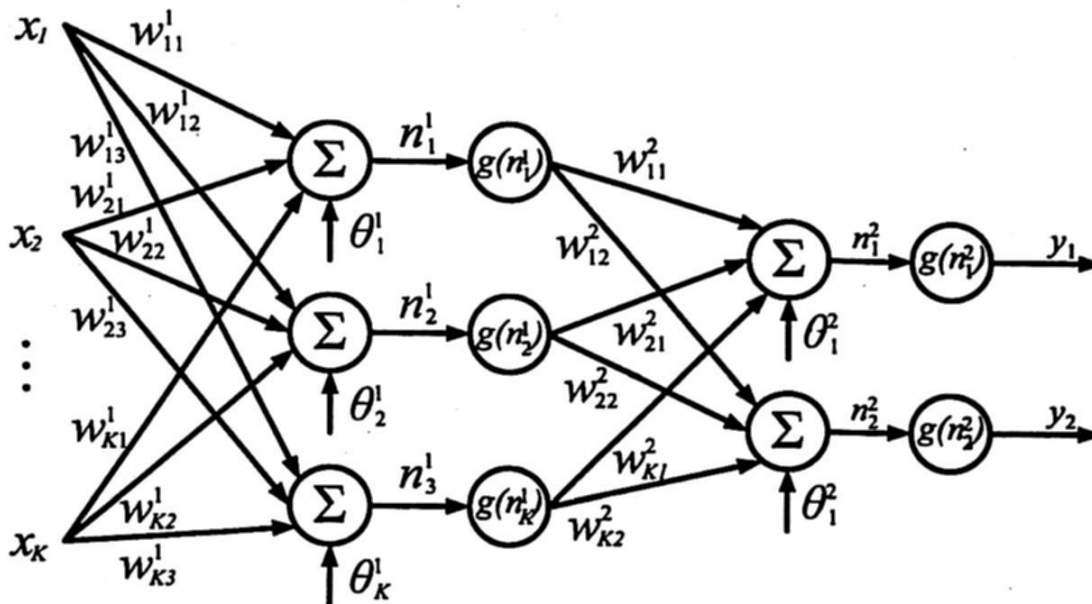


Fig. 4. Representation of an Artificial Neural Network.

ANN has the ability to learn based on an example, making themselves very flexible and powerful for modeling nonlinearity order and memory effect. Furthermore, it has

the capability to learn a complex nonlinear behavior of the dynamic system without the need to understand the internal mechanisms of the system and can be an alternative and attractive approach for PA modeling with memory effects.

### 2.2.1 The Multilayer Perceptron (MLP)

The simpler architecture is constituted by one input layer, one hidden layer containing a given number of neurons and an output layer. The transfer function of all the neuron is tangent sigmoid type given by the following equation (9):

$$f(x) = \frac{1 - e^{-2x}}{1 + e^{-2x}} \quad (9)$$

The transfer equation of the MLP with one single input layer is quite easy to formulize in equation (10):

$$y = \sum_{k=1}^K w_{kj} y_k = \sum_{k=1}^K w_{kj} f \left( \sum_{j=1}^J w_{ji} x_i \right) \quad (10)$$

Where  $J$  is the number of neurons in the hidden layer and  $K$  the number of output neurons. The MLP generally uses the back propagation algorithm that consists in actualizing the parameters of the output layer first, then the parameters of the last hidden layer and so. The back propagation algorithm is generally coupled with the Levenberg-Maquart algorithm. This algorithm consists in applying a second derivative to the cost function  $E$  to find the maximum gradient.

## 3 FPGA Implementation

VHDL code is used in electronic design automation to describe digital and mixed-signal systems such as DSP and integrated circuits. The DSP Development Kit Cyclone III Edition-Altera delivers a complete digital signal processing development environment; it includes the Cyclone III development board and Quartus II development software. The system operates with an internal clock of 125 MHz provided by a Phase-Locked Loop (PLL) circuitry. The overall view through blocks of the whole system is showed in Fig. 5.

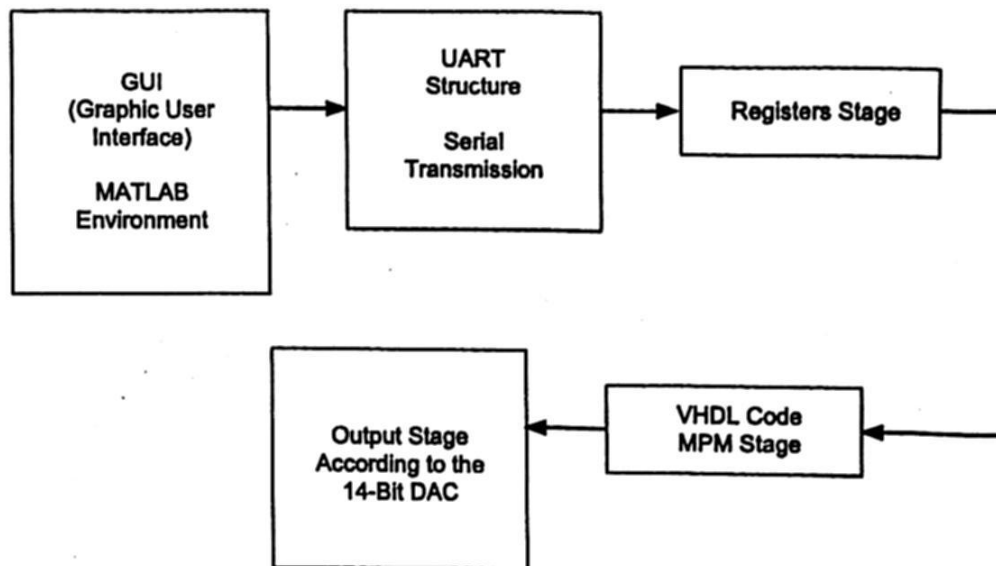


Fig. 5. Block view of the developed VHDL system.

The modeling based on VHDL can be accurate, unfortunately the required code and translation of the MPM coefficients to VHDL environment requires design time.

This Universal Asynchronous Receiver-Transmitter (UART) showed by the VHDL Register-Transfer Level (RTL) is exhibited. In Fig. 6 the UART structure is able to manipulate until 128 calculated records obtained from the workspace in Matlab. There is an internal function in Matlab able to calculate the MPM coefficients based on the LSE technique.

Fig. 7 shows synchronization based on D-Flip Flops during the modeling stage of the send data from Matlab. All these registers are controlled by the same clock provided by the internal clock with frequency 125 MHz.

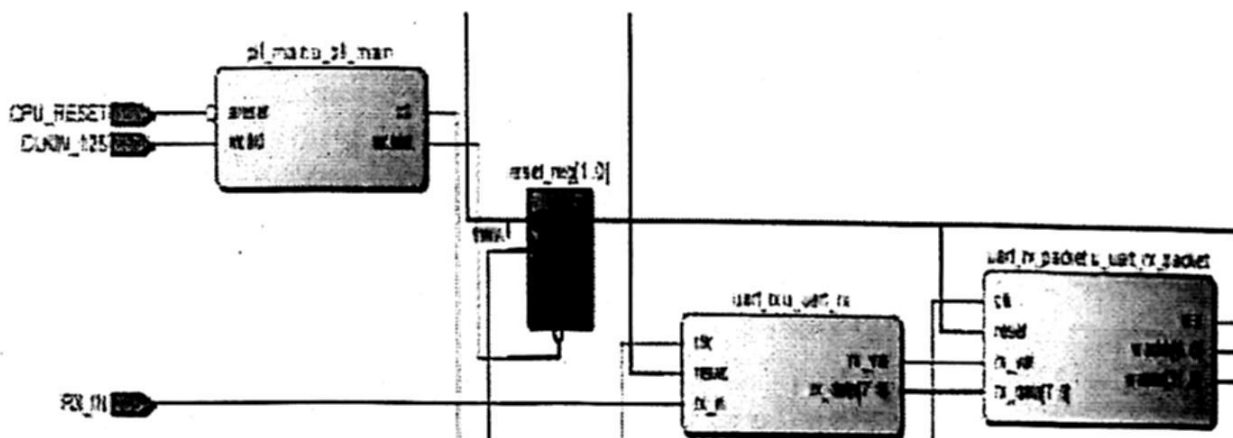


Fig. 6. UART-RTL provided by the Quartus II- ALTERA Software.



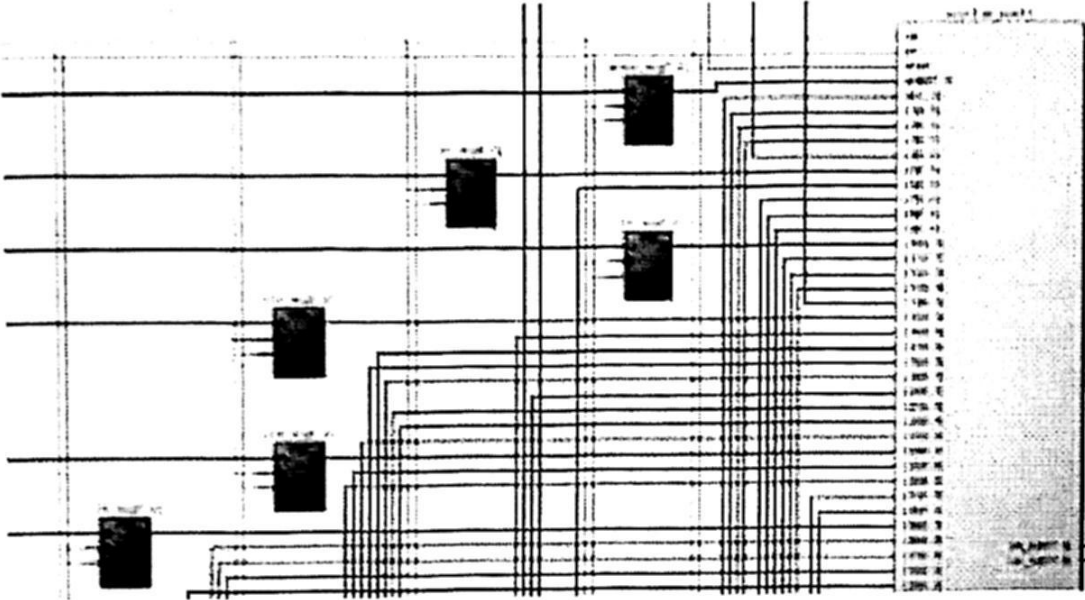


Fig. 7. Coefficients stored as registers and inserted into the Amplification Stage.

In this paper was used the DSP Builder Tool as a faster alternative to run the behavioral modeling. DSP Builder includes basic FPGA block models and enables to build and verify the user’s digital system with real hardware parameters without requiring FPGA implementation in the system. Once the system was built, the generated outputs are tested by transferring them to the FPGA and emulated them to ensure that they provide similar performance to the MATLAB implementation. Moreover, these high level synthesis tools are able to translate to RTL code directly via an automated process. The developed system based on DSP Builder is shown in Fig. 8.

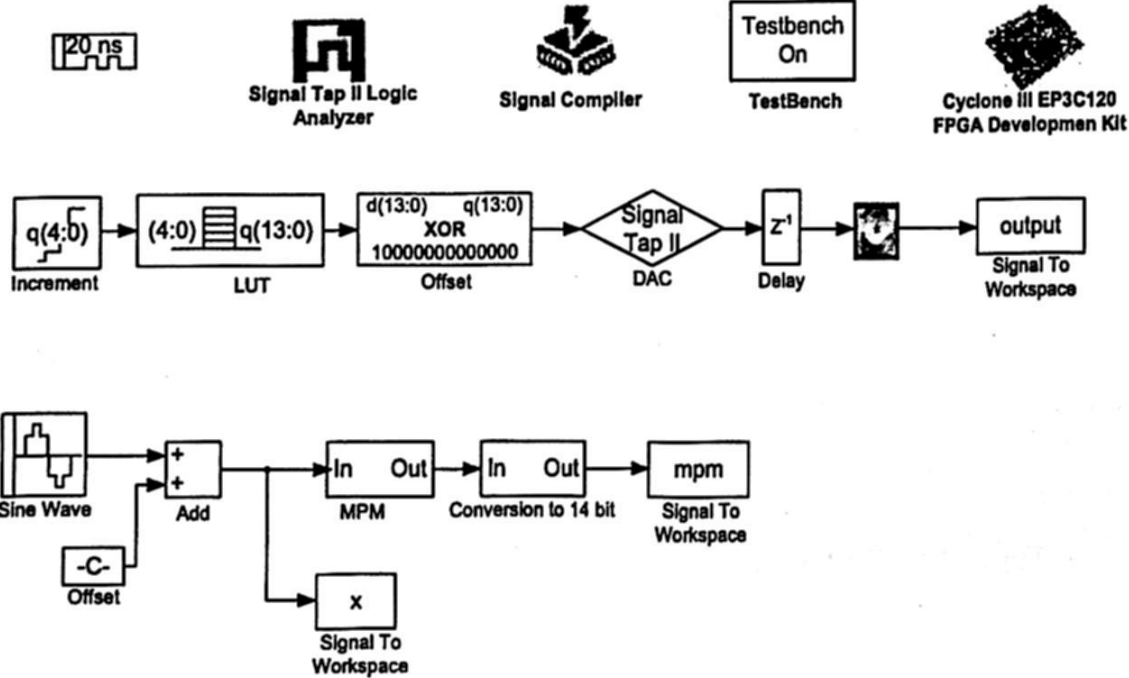


Fig. 8. Overview of MPM implemented using DSP Builder.

The experimental verification of the MPM allows representing any real PAs measurements which closely resembles a real PA behavior. The MPM is exploited during the ANN training including 5 neurons is depicted in Fig. 9.

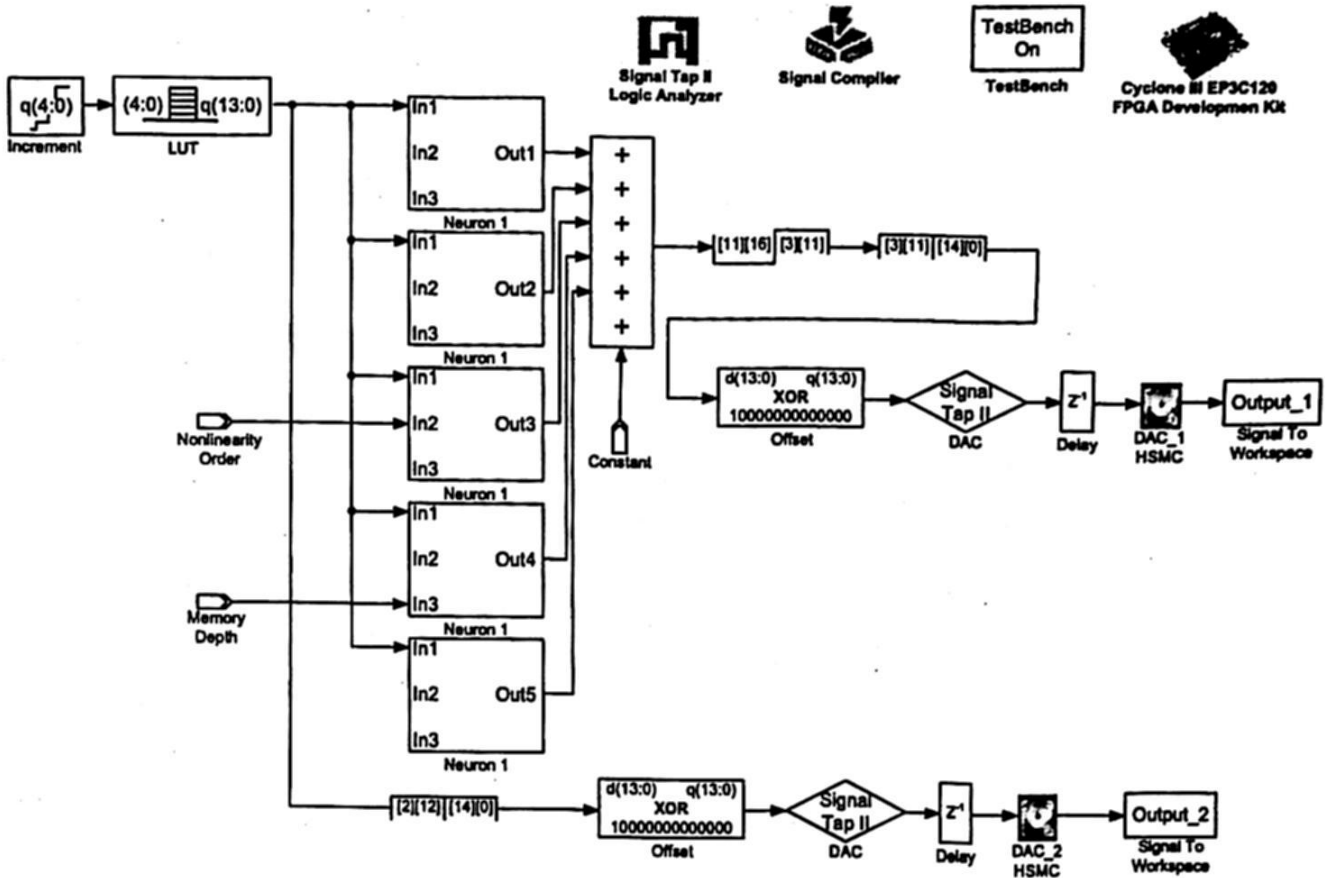


Fig. 9. Overview of the ANN model based on the MPM performance.

## 4 Results

Fig. 10a shows amplification without memory effects and considering a basic nonlinearity order. In Fig. 10b the memory effects that can be considered for linearization stage are observed, these both results can be achieved through VHDL or DSP Builder. The first part is representing a system with nonlinearity order  $n=1$  and memory depth  $m=0$ , should be noted that the DAC is bounded due to the 14-Bit resolution, for a practical application a power stage must be added in order to increase the signal level and protect the High Speed Mezzanine Card HSMC. The second part shows the modeled output for an input signal with memory depth  $m=1$ , it can be noted the undesirable memory effect attenuates the output signal. The last emulation depicted in Fig. 11 in hardware was created for a signal with nonlinearity order  $n=4$  reaching an output frequency of 1.53 MHz represented in Fig. 11: the sampling frequency was set to 125 MHz.

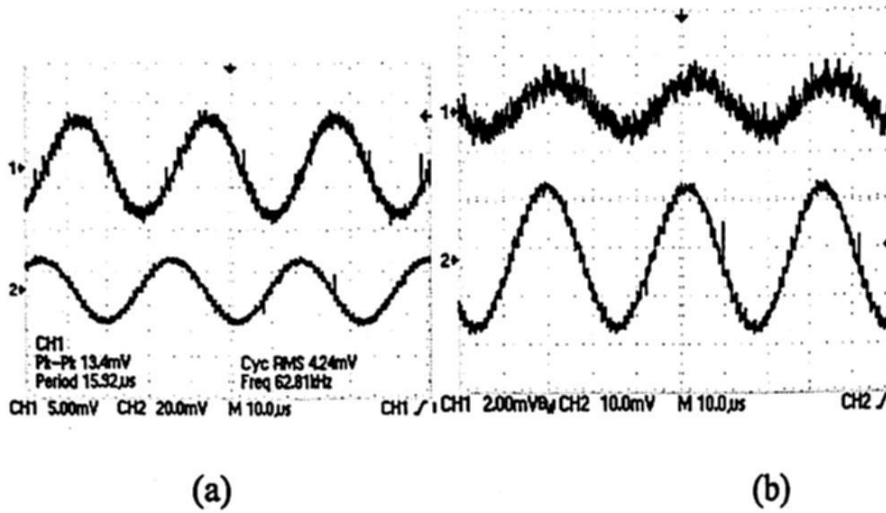


Fig. 10. MPM emulated in hardware for a system with (a) nonlinearity order  $n=1$  and memory depth  $m=0$  and (b) nonlinearity order  $n=1$  and memory depth  $m=1$ .

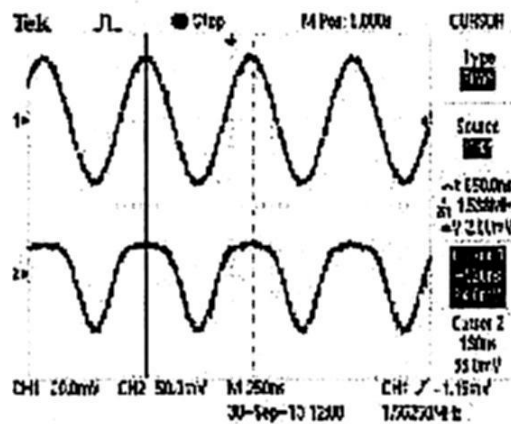


Fig. 11. MPM emulated in hardware for a system with nonlinearity order  $n=4$  and memory depth  $m=0$ .

The modeling for the signal without memory and nonlinearity order  $n=1$  has a frequency of 33 KHz, the output frequency can be defined through the sequence and processing in the internal iterations and the desired resolution for each case, this affirmation was done for the case with memory depth  $m=1$  and nonlinearity order  $n=1$ .

## 5 Conclusion

In this paper, a behavioral modeling based on a special case of Volterra Series modeled with Matlab environment and emulated in Hardware through VHDL using the ALTERA Cyclone III FPGA was presented. The goals of this work have been successfully achieved and show clearly the undesirable effects as memory and nonlinearity order representing closely a real PA behavior. Throughout the project, lots of research regarding modeling PA devices has been gained, all this knowledge can be applied when dealing with nano-devices fabrication when the intermodulation products and memory effects must be considered or they are monitored for a linearization stage. The sampling rate must be considered based on the development board; in this case

the reached frequency of 1.53 MHz was established based on the available development board. The maximum frequency of the signal has to be theoretically 2 times considering the Nyquist sampling criterion, but practically 10 times lower than the frequency of the internal clock established in 125 MHz. This developed work based on VHDL and improved by the DSP Builder tool provides significant advantage related to flexibility and integration about modeling with memory effects, even additional kind of modeling can be added to the system.

## References

1. H. Ku and J.S. Kenney, "Behavioral Modeling of Nonlinear RF Power Amplifier Considering Memory Effects," in *IEEE Transactions on Microwave Theory and Techniques*, Vol. 51, Issue: 12, pp. 2495-2504, 2001.
2. A. Zhu, "Simplified Volterra Series Based Behavioral Modeling of RF Power Amplifiers Using Deviation-Reduction," in *IEEE 6th International Conference on Wireless Communications Networking and Mobile Computing (WiCOM 2010)*, Chengdu, China, pp. 1-4, 2010.
3. A. Zhu, et al., "Dynamic deviation reduction-based Volterra behavioral modeling of RF Power Amplifiers," in *IEEE Transactions on Microwave Theory and Techniques*, Vol. 54, Issue: 12, pp. 4323-4332, 2006.
4. H. Zhou, et al., "A nonlinear memory power amplifier behavior modeling and identification based on memory polynomial model in soft-defined shortwave transmitter," in *IEEE 6th International Conference on Wireless Communications Networking and Mobile Computing (WiCOM 2010)*, Chengdu, China, pp. 1-4, 2010.
5. J. Núñez-Perez et al., "Flexible Test Bed for The Behavioral Modeling of Power Amplifiers," in *International Journal for Computation & Mathematics in Electrical & Electronic Engineering (COMPEL)*, Emerald, 2013. In Press.
6. H. M. Muhammad, N. Shaikh and S. Kamilah, "Implementation of an Inter-carrier Interference Self-Cancellation Technique for OFDM System in Altera CYCLONE II FPGA," in *International Conference on Electronic Design*, Penang, Malaysia, pp. 1-6, December 2008.
7. H. Alasady and M. Innkahla, "Design and Hardware Implementation of Look-Up Table Predistortion on ALTERA Stratix DSP Board," in *Canadian Conference on Electrical and Computer Engineering*, Canada, pp. 1535-1538, May 2008.
8. J.P. Oliver and E. Boemo, "Power estimations vs. power measurements in Cyclone III devices," in *VII Southern Conference on Programmable Logic (SPL)*, pp. 87-90, 2011.
9. C. Quindroit, et. al, "Concurrent Linearization: The State of the Art for Modeling and Linearization of Multiband Power Amplifiers," in *IEEE Microwave Magazine*, pp. 75-91, Vol. 14, Issue: 7, 2013.
10. C. Quindroit, et. al, "Concurrent Dual-Band Digital Predistortion for Power amplifier based on Orthogonal Polynomials," in *IEEE International Microwave Symposium Digest (IMS)*, pp. 1-4, Seattle, USA, 2013.
11. N. Narahariseti, et. al, "2D cubic spline implementation for concurrent dual-band system," in *IEEE MTT-S International Microwave Symposium Digest (IMS)*, pp. 1-4, Seattle, USA, 2013.
12. H. Ku and J. S. Kenney, "Behavioral modeling of nonlinear RF power amplifiers considering memory effects," *IEEE Transactions on Microwave Theory and Techniques*, Vol. 51, No. 12, 2003, pp. 2495-2504.

# Adaptation of a PCA fusion stage to improve accuracy in a biometric iris recognition system for unconstrained environments

Juan M. Colores-Vargas<sup>1</sup>, Mireya García-Vázquez<sup>1</sup>, Alejandro Ramírez-Acosta<sup>2</sup>

<sup>1</sup>Instituto Politécnico Nacional CITEDI

Avenida del Parque 1310, Tijuana, B.C. México 22510

<sup>2</sup>MIRAL R&D&I, Palm Garden, Imperial Beach, USA 91932

<sup>1</sup>{colores, msarai}@ipn.mx, <sup>2</sup>ramacos10@hotmail.com

*Paper received on 12/01/13, Accepted on 01/19/14.*

**Abstract.** Traditional iris recognition system uses a fixed image acquisition stage which requires a high collaboration of the user to be recognized, a short distance to acquire the fixed image and the correct illumination. All this characteristics are the structure of a controlled environment system. Nowadays, researchers are focused on the development of biometric recognition system which can operate under non-controlled environment, emphasizing the video management within the architectures, optimizing and proposing new stages to improve the recognition rate levels. In this paper we propose a new stage to exploit the biometric information from video-iris, the new stage generate a representative iris biometric template based on the principal component analysis (PCA) which integrates the biometric information of group of frames from a set of templates from a video-iris with optimal coefficients. The resulting representative biometric template should contain more biometric texture information as compared to individual templates that resulting in better recognition performance. The implementation of the new stage generates a new architecture based on video iris biometric recognition for unconstrained environments which achieved a recognition rate of 99.236% with respect to 84.42% of a conventional system. The new system deals with a false acceptance rate FAR=0.3288% and a false rejection rate FRR=0.7634%. The tests were performed with the information of 160 users from the MBGC.v2 database.

**Keywords:** Iris recognition, Video, MBGC, Fusion, Biometric systems, unconstrained environments.

## 1 Introduction

The biometric is the study of physical, biological or behavioral traits used to identify and verify a person [1]. Some commonly used biometric features include fingerprints, face, hand geometry, voice, iris, signature, DNA, Palm, Iris, body odor. The biometric recognition systems offer the advantage over information security problem ensuring that only authorized users are able to access the information. Each biometric feature



has its strengths and weaknesses. Hugo Proenca [2] conducted a research where analyzed the different biometric identifiers, evaluating them according with 7 properties including: universality, uniqueness, permanence, performance and Measurability. The obtained values were performed by averaging and subjectively weight the classification exposed in at least 10 scientific papers of other authors. According with results obtained by Hugo Proenca, signature and voice have the lower values for the uniqueness property, these biometric identifiers are easy to manage and suitable for low security applications. On the other hand, the more distinctive biometric features such as the retina and iris involve a large number of processes. In his analysis, the iris had the highest average value for the seven properties (84.43 %).

The iris has a diameter of 11 mm approximately, on the outside is surrounded by a white area called sclera. The iris is the pigmented area surrounding the pupil and has a fibrous tissue called stroma, iris connects muscles to contract and dilate the pupil in response at changing light, the fibrous constitution has over 400 features used to recognize a person. The advantages of the iris over others biometric features: It is highly protected, the iris patterns can be captured from distance, encoding and decision analysis not require a computationally complex algorithm and can be made in less of one second, it possesses high degree of randomness (John Daugman [3] probe the uniqueness and randomness of Iris over 200 billion cross comparisons of irises, he reports a theoretical false match rate of 1 in  $5 \times 10^{15}$ ). However, the iris is small (approximately 11 mm diameter) which makes it difficult to acquire the biometric information at long distances. Moreover, the iris is a moving object located on a curved surface, wet and reflective, making it difficult to acquire the biometric information.

The traditional iris recognition systems based on still images [4] are designed to work with special or restricted conditions; this means that they require an ideal environment and cooperative user's behavior during the eye image acquisition stage to obtain high quality images. Therefore, if any of these requirements are not met, it can cause a substantially increase of error rates, specially the false rejections. Many factors can affect the quality of an eye image, including defocus, motion blur, dilation and heavy occlusion. Naturally, poor image's quality cannot generate satisfactory recognition because they do not have enough feature information, in this regard; iris recognition is dependent on the amount of information available in two iris images being compared. A typical iris recognition system commonly consists of four main stages [5, 6]; *Acquisition* the aim is to acquire a high quality image. *Preprocessing*, involves the segmentation and normalization processes. The segmentation consists in isolating the iris region from the eye image. The normalization is used to compensate the varying size of the pupil. *Feature encoding*, uses texture analysis method to extract features from the normalized iris image. The significant features of the iris are extracted as a series of binary codes known as digital biometric template. *Matching* compares the user digital biometric template with all the stored templates in the database. The matching metric will give a range of values of the compared templates from the same iris. In recent years, with an increasing of new massive biometric security demands around the world, it seems difficult to fulfill the conditions mentioned above in order to have a reliable iris recognition system [7]. Thus, with the aim of overcoming these drawbacks, news approaches are being developed in an attempt to improve iris recognition performance under non ideal situations i.e. unconstrained environments. These biometric recognition systems are more flexible, the aim has been to achieve automatic acquisition system, where the image acquisition process is transparent to the user, and this has been achieved using the video acquisition system. The development of these

systems has involved the redesign of the traditional architecture of a biometric system based on still images. These new architectures are part of the multi-biometric systems which are the current trend in biometric systems. The term multi-biometric denotes the fusion of different types of information (e.g., fingerprint and face of the same person, or fingerprints from two different fingers of a person).

Among these approaches, the video-based eye image acquisition for iris recognition seems to be an interesting alternative [8, 9, 21] because it can provide more information through the capture of a video iris sequences. Besides that, it is a friendly system because it is not intrusive and requires few users' cooperation.

In this paper, we propose to exploit the video-iris; it contains information related to the spatio-temporal activity of the iris and its neighbor region over a short period of time. Therefore, the information from individual iris images can be fused into a single composite iris image with higher biometric texture information, resulting in better recognition performance and reducing the error rates. The idea of fusing iris biometric templates to perform biometric recognition has been recently described in the literature. Zhao and Chellappa [10] suggested averaging to integrate texture information across multiple video frames to improve face recognition performance. Hollingsworth et al. [11] improve the matching performance using signal-level fusion, taking advantage of the temporal continuity in an iris video sequence to create a single average image from multiple frames, but they suppose an ideal situation, in which bad segmented iris frames are manually discarded, that may limit its application. There are several methods of fusion, Colores et al. [12] analyzed some fusion methods to determine the most suitable to use in biometric applications. Indeed, the main objective in this work is to determine the effectiveness of including a new stage that implements a fusion method within architecture for a biometric recognition system based on iris. This paper is organized as follows: section 2 explains the basics of the Iris recognition system based on video. Section 3 describes the new stage. In section 4 presents the evaluation of the new architecture, and finally in section 5 presents the conclusions.

## 2 Iris recognition system based on video

The scheme, shown in figure 1, is based on the conventional iris recognition system [5, 6] modified to operate with video captured on unconstrained environments. Thus, in the modified system firstly the eye frames are captured by a proper video camera in the video acquisition stage.

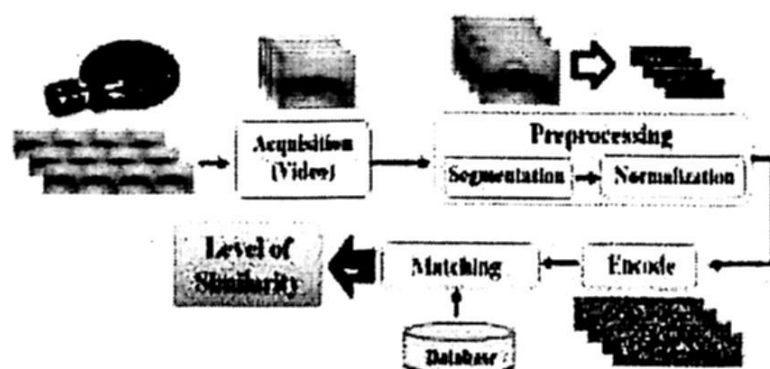


Fig. 1. System 1: Modification of Iris recognition system to use video.

Ideally, the captured eye image should be centered in the frame, free of defocus and aberration errors. It is possible to achieve it by forcing to the user to remain perfectly static and looking to the camera while the video is taken. However, the main purpose of any unconstrained scheme is to be minimally invasive or restrictive with the users. In the case of moving subjects, the images may be of poor quality due to improper illumination, motion blur, occlusions, etc. Moreover, increasing the distance of the subject from the camera causes a decrease in the resolution of the eye recorded in the image; the performance of an iris recognition system is negatively impacted when the spatial resolution of the acquired iris images is low. This fact may result, sometimes, in the introduction of frames with few, or even without any iris's texture information, which can in turn affect the performance of the recognition system. In other cases, even though the image had high quality, may not enough contain information needed to achieve recognition. Therefore, the processing of each of the images will result in a higher processing time. A sample set of all these problems in capturing iris images are shown in figure 2.



**Fig.2.** Video sequences depicting various problems during capturing iris images.

Some studies report that using high quality image affects recognition accuracy and can improve system performance [7]. Then, it is necessary to select suitable images with high quality from an input video sequence before the next recognition processing. Choosing an eye frame with an appropriate image quality seems to be a challenge. Defocus-blur and motion-blur are the major source of eye frame quality degradation. Indeed, in a less restrictive environment, the users are free to move outside the optimal distance from the camera during video capture process, which means that they may move outside the optimal "depth of field" of the system causing the blurring effects in the captured frames. In general, a focused eye image has a relatively uniform frequency distribution in the 2D Fourier spectrum, while the power spectrum of a defocused or blurred image is concentrated on the lower frequencies. This fact suggests that a spectral analysis of the frequency distribution may be an effective way to estimate the image quality of eye frames for discriminating the distorted frames from the clear ones. Consequently, several discrete formulations have been proposed and that allow obtaining only the power of high frequencies by attenuating the low frequency components of the eye frame and calculating the power spectrum of the high-pass filtered eye image. Colores et al. [13] analyzed four methods used to obtain the high frequency power spectrum of the eye frames for image quality assessment. Evaluation results show that the Kang and Park convolution kernel provides better performance than the other kernels in terms of speed and accuracy [14]. Therefore, in the modified scheme (System 1), the acquisition stage employs the Kang and Park convolution kernel.



## 2.1 Preprocessing video stage

The preprocessing stage of the modified scheme performs the iris segmentation and normalization tasks whose main purpose is to provide a good enough segmentation of the iris region for each frame obtained from video-iris, to enable the encoding and matching stage to perform accurate iris recognition. The *segmentation module* isolates the iris region from the eye frames using the segmentation algorithm proposed by Wildes [6], which is based on the circular Hough transform combined with a Canny edge detector to obtain the iris region. The goal of edge detection algorithms is to produce an image containing only edges of the original image. However, most edge detection algorithms produce an image containing fragmented edges; then in order to turn these fragmented edge segments into useful lines, circles and object boundaries, an additional processing is needed. To this end, the circular Hough transform is used to find circles in eye frame and deduce the radius  $[r_p, r_i]$  and centres  $[(x_{cp}, x_{cp}), (x_{ci}, x_{ci})]$  corresponding to the pupil and iris regions. This stage plays a very important role because if the segmentation process is not performed with enough precision, the segmentation error will further propagate to the encoding and matching steps. The *normalization module* is used to compensate the size variation of the iris region, in the eye frames, mainly because the stretching of the iris caused by pupil dilatation due to varying illumination levels. This process is done using the linear rubber sheet model proposed by Daugman[5]. This transformation maps each point within the iris region to polar coordinates  $(r, \theta)$  where  $r$  and  $\theta$  are in the intervals  $[0, 1]$  and  $[0, 2\pi]$ , respectively. The mapping of the iris region from Cartesian representation  $I(x, y)$ , to the normalized non-concentric polar representation,  $I(r, \theta)$  is given by equation 1,  $I(x(r, \theta), y(r, \theta))$  is the segmented eye image,  $(x, y)$  are the original Cartesian coordinates,  $(r, \theta)$  are the corresponding normalized polar coordinates.

$$I(x(r, \theta), y(r, \theta)) \rightarrow I(r, \theta) \quad (1)$$

where

$$x(r, \theta) = (1 - r)(x_{cp}(\theta) + r_p \cos \theta) + r(x_{ci}(\theta) + r_i \cos \theta) \quad (2)$$

$$y(r, \theta) = (1 - r)(y_{cp}(\theta) + r_p \sin \theta) + r(y_{ci}(\theta) + r_i \sin \theta) \quad (3)$$

## 2.2 Feature encoding video stage

The extracted features are fed into the encoding stage which is used to obtain the digital biometric template [15]. This process has two components: firstly, the filter component is applied in each normalized iris region from video-iris frames using a predefined complex filter or operator to extract the most discriminating information present in the iris region. Secondly, the phase quantization where the resulting complex array is translated into a binary code that constitutes the digital biometric template. The feature encoding stage, then, was implemented by convolving the normalized iris region with a 1D Log-Gabor wavelets, where each row of  $I(r, \theta)$  corresponds to a particular circle extracted from the iris rim. Some enhancements are then performed on the extracted signals, such that the intensity values at distorted areas in the normalized iris region are filled with the average intensity of surrounding pixels. Finally, the filter

output is transformed into a binary code using the four quadrant phase encoder, with each filter producing two bits of data for each phasor [16].

### **2.3 Matching video stage**

The operation of this stage consists of the comparison of digital biometric templates, producing each a numeric dissimilarity value. In this scheme, the Hamming distance (HD) was employed for Daugman [5]. The HD measure can be used to make a decision whether the digital biometric template is produced by the same or different users.

## **3 New fusion video stage**

The image fusion tries to solve the problem of combining information from several images taken the same object to get a new fused [17]. In this paper, each frame of video-iris is first pre-processed in order to obtain the normalized iris region templates. Then, a fusion method is applied to provide a representative fused normalized iris region template from these individual templates. The resulting template should be contains more iris biometric texture information as compared to individual templates. We analyzed the image fusion methods to determine the most suitable to achieve greater extraction of iris biometric information [12], the result show that the principal component analysis (PCA) method presents the best performance to improve recognition values according to the Hamming distances. The PCA fusion method transforms the features from the original domain to the new domain (known as PCA domain). Here the features are arranged in order of their variance. Fusion process is achieved in the PCA domain by retaining only those features that contain a significant amount of information. The main idea behind PCA is to determine the features that explain as much of the total variation in the data as possible with as few of these features as possible. Image fusion based on PCA has advantages in maintaining image information, reduce redundant information and highlight the components with biggest influence, can be performed by parallel computing, the spectral information loss is slightly better than others methods of fusion.

Thus, the results suggest that adding a fusion video stage to the architecture of the unconstrained environment iris recognition, it could increase the system performance. Thus, we have a new architecture for a system based on video iris biometric recognition for unconstrained environments. Figure 3 shows the new architecture, the added stage is based on the PCA fusion, operates fusing the normalized templates to generate a single digital normalized template. This new stage will provide to the matching stage a digital template that contains more biometric texture information of the iris region.



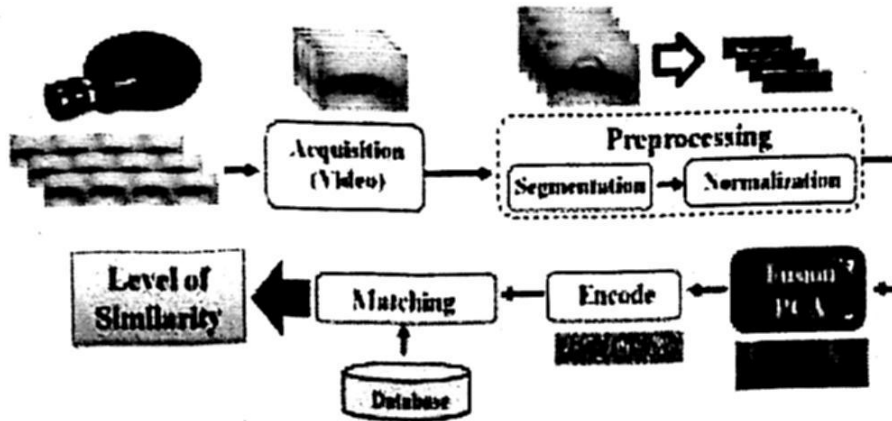


Fig. 3. System 2: Iris recognition system with a fusion video stage.

### 3.1 Image fusion using principal component analysis

The fusion method based on the principal component analysis [17] is a straightforward way to build a fused image as a weighted superposition of several input images. The optimal weighting coefficients, with respect to information content, can be determined by a principal component analysis of all input intensities. By performing a PCA of the covariance matrix of input intensities, the weightings for each input image are obtained from the eigenvector corresponding to the largest eigen-value.

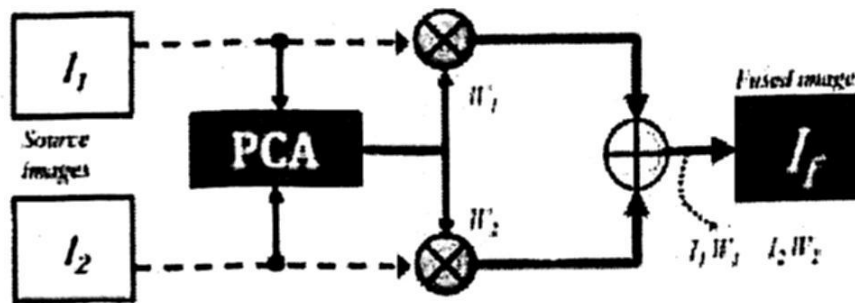


Fig. 4. PCA operation to fuse two images.

Figure 4 shows the basic fusion scheme, where two images  $I_1$  and  $I_2$  are fused to obtain a resultant image  $I_f$  given by equation 4,  $W_1, W_2$  are the weights coefficients.

$$I_f(x, y) = W_1(x, y)I_1(x, y) + W_2(x, y)I_2(x, y) \quad (4)$$

$$X_k = I_k - \Psi \quad (5)$$

$$C = \frac{1}{2} (X_1^T X_1 + X_2^T X_2) \quad (6)$$

$$W_1 = U^T X_1 \quad (7)$$

$$W_2 = U^T X_2 \quad (8)$$

The weights for each source image are obtained from the eigenvector corresponding to the largest eigen-value of the covariance matrix of each source. Arrange source images in two-column vector.

- Organize the data, let  $S$  be the resulting column vector.
- Compute empirical mean ( $\Psi$ ) along each column.

- Subtract  $\Psi$  from each column of  $S$ , the resulting is a matrix  $X_k$ . (eq. 5)
- Find the covariance matrix  $C$  of matrix  $X_k$ . (eq. 6)
- Compute the eigenvectors and eigen-value and sort them by decreasing eigen-value.
- Consider first column  $U$  which correspond to larger eigen-value to compute normalized component  $W_1$  and  $W_2$ . (eq.7 and eq.8 respectively)

## 4 Experimental results

To evaluate the performance of the proposed scheme shown in Figure 5, we selected the "MBGC.v2" dataset [4, 18], which presents several noise factors, especially those related to reflections, contrast, luminosity, eyelid and eyelash iris obstruction and focus characteristics. These facts make it the most appropriate to study the iris recognition system for uncontrolled environments. Regarding to the images size, each eye frame is 480 by 640 pixels in 8 bits-gray scale at 30frames per second (fps). This database has been distributed in MPEG-4 format to over 100 research groups around the world. For experiments purposes, iris-videos from the MBGC.v2 database were selected from 131 users to generate the testing dataset. For each user, we select a reference eye frame. The testing eye frames were selected sub-sampling the video-iris at 1/10 frames, although the reference eye frame was chosen according to the characteristics of high-frequency concentration previously described. As shown in Figure 5, the recognition tests were conducted on 3000 eye frames with 131 reference eye frames, allowing the generation of the distributions inter-class and intra-class to compare the performance of proposed and conventional systems. Each frame in the set of test was segmented and normalized using a modified version of the Libor Masek [19] algorithms for iris recognition (based system 1), improvements in the algorithms allow operating with video-iris, obtaining for each frame a normalized template; this template contains only the texture information of the iris region.

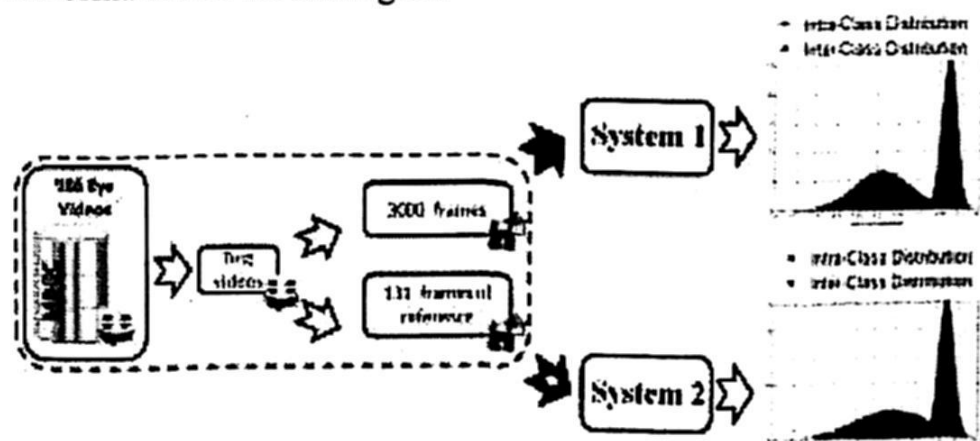


Fig. 5. Scenario of the testing systems.

To evaluate the performance of proposed scheme in verification mode, the equal error rate (EER) and the receiver operating characteristics (ROC) curve [20] were used. Figure 1, shows the iris recognition scheme used in unconstrained environments modified to process video (System1) and Figure 3 illustrate the new scheme called System 2 which integrates the fusion stage. Figure 6.a shows the false acceptance rate (FAR) and false rejection rate (FRR) achieved by System 1 which provides an EER equal to 13.1771%, with a threshold (Th) of about 0.4586. Figure 6.b shows the FAR and FRR

achieved by the new scheme (System 2) which achieves an EER equal to 0.7751% with a  $Th$  equal to about 0.44074.

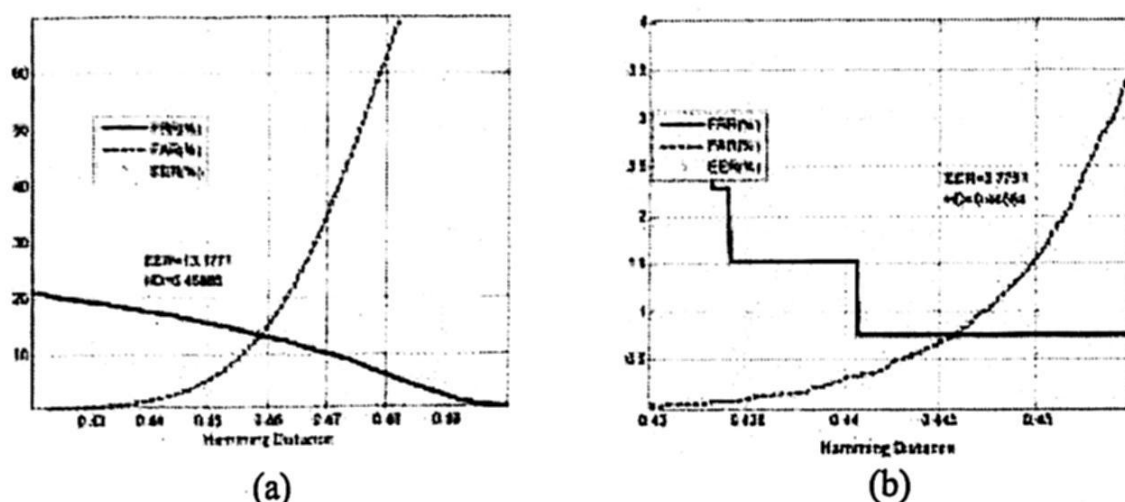


Fig. 6. The crossover point between the curves FRR and FAR, EER for systems. a) System 1, b) System 2.

The ROC curves, shown in Figure 7, plot the FAR as a function of the FRR, are useful to compare the performance of proposed and based systems. To confirm the accuracy of iris matching process and to show the overall performance of proposed new scheme, independently of the threshold value, the ROC curves were used. From experimental curves, it follows that the proposed new scheme (system 2) provides a better performance since the ROC curve is much closer to the origin than the proposed based system (system 1). Finally, using properly selected  $Th$ , the System 1 may achieve a FAR equal to 4.86%, which is significantly slower than the EER, although in this situation the FRR increases to 15.57% which is much higher than the EER. On the other hand, using the same threshold, the proposed new scheme (system 2) achieves a FAR equal to 0.3288% and a FRR equal to 0.7634%. In addition, in this situation, the genuine acceptance rate (GAR) for System 1 is 84.42% while for the proposed system (system 2) is about to 99.236%.

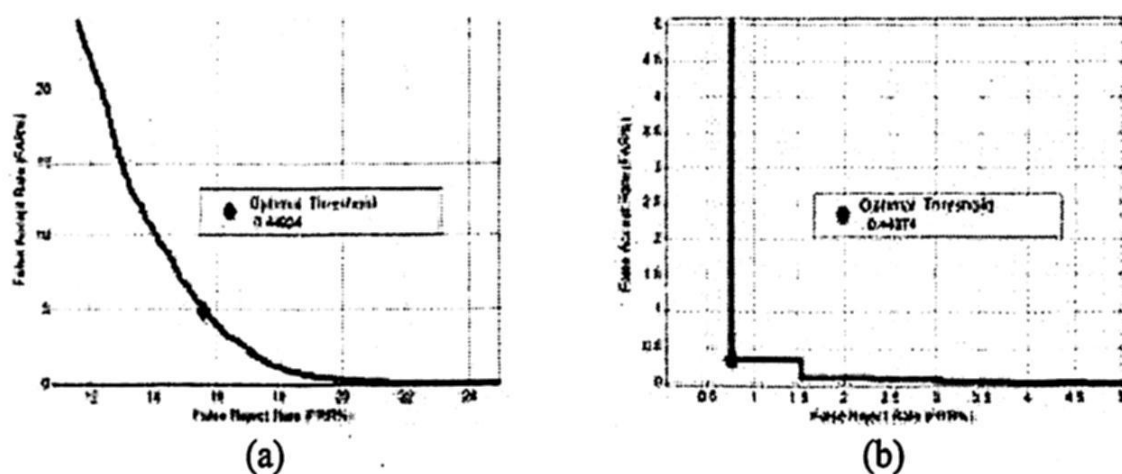


Fig. 7. ROC curves (a) Cut-off point optimal, system 1, (b) Cut-off point optimal, system 2.

## 5 Conclusions

In this work, we propose an improved iris recognition system, integrating a new stage into the iris recognition system modified to use video, in order to increase its adaptability toward less constrained environments. Indeed, under less constrained environments, it is expected that the captured eye frames contain several types of noise and distortion, which affect the segmentation process and consequently impacts the recognition rate. The new stage exploits the biometric texture information from video-iris acquired under non-cooperative scheme, creating a fused normalized template through an image fusion technique based on PCA. We used the ROC curves to obtain the optimal decision threshold. The experimental results show that the proposed stage help to reduce the recognition error rates on the new proposed scheme (system 2), contributing to improve the recognition performance. It also decreases the EER by 12.4%; and for a given Th, FAR is reduced by 4.53%, while the FRR is reduced by 14.8% comparing with based iris recognition system (system 1). In addition, the GAR achieved by the proposed scheme (system 2) is 99.236%; while for the based system (system 1) is 84.42%. Thus, the results suggest that adding a fusion video stage to the architecture of the non-cooperative iris recognition, it could increase the system performance. Therefore, we can conclude that our proposal can be integrated as an optimization to the biometric recognition system based on video iris, for an application of iris recognition in uncontrolled environments.

## References

1. Anil K. Jain, Arun Ross, and Salil Prabhakar, "An introduction to biometric recognition," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 14, no. 1, pp. 4-20, 2004.
2. Hugo Pedro Proenca, "Towards Non-Cooperative Biometric Iris Recognition," Department of Computer Science - University of Beira Interior, Portugal, 2006.
3. John Daugman, "Probing the Uniqueness and Randomness of Iris Codes: Results From 200 Billion Iris Pair Comparisons," *Proceedings of the IEEE*, vol. 94, no. 11, pp. 1927-1935, 2006.
4. Phillips P., Scruggs W., Toole A., Flynn P., Bowyer K., Schott C., Sharpe M. : FRVT 2006 and ICE 2006 large-scale results, Technical Report, National Institute of Standards and Technology, NISTIR 7408, 2007.
5. Daugman J.: High confidence visual recognition of persons by a test of statistical independence. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 15: 1148-1161, 1993.
6. Wildes, R.: Iris Recognition: An Emerging Biometric Technology. *Proceedings of the IEEE*, 85-9, pp.1348-1363, 1997.
7. Gamassi, M., Lazzaroni, M., Misino, M., Piuri, V.: Quality assessment of biometric systems: a comprehensive perspective based on accuracy and performance measurement. *IEEE Transactions on Instrumentation and Measurement*, 54, pp.1489-1496, 2005.
8. Lee Y., Phillips P., Michaels R.: An automated video-based system for iris recognition. *Proc. of Int. Conf. on Biometrics*. 1-8, 2009.



9. Wheeler F., Perera A., Abramovich G., Bing Y., Tu P.: Stand-off iris recognition system. Proc. IEEE International Conference on Biometrics: Theory, Applications and Systems. 1:1-4, 2008.
10. Wenyi, Z., Rama, C.: Face Processing: Advanced Modeling and Methods, chapter 17: Beyond one still image: Face recognition from multiple still images or a video sequence by S.K. Zhou and R. Chellappa, pages 547-567. Elsevier, 2006.
11. Hollingsworth K., Peters T., Bowyer K.: Iris recognition using signal-level fusion of frames from video. IEEE Trans. Inform. Forensics Secur. 4(4):837-848, 2009.
12. Colores-Vargas, J., García-Vázquez, M., Ramírez-Acosta, A.: Evidencia de mejora en los sistemas de reconocimiento basados en iris, utilizando esquemas adaptados de fusión de imágenes, Research in computer science; advances in computing science and control, ISSN:1870-4069, 2012.
13. Colores-Vargas, J., García-Vázquez, M., Ramírez-Acosta.: Measurement of defocus level in iris images using convolution kernel method. *Lect. Notes Comput.* 6256:164-170, 2010.
14. Kang, B., Park, K.: A study on iris image restoration. *Lecture Notes in Computer Sciences*, 3546: 31-40, 2005.
15. Daugman, J.: The importance of being random: statistical principles of iris recognition. *Pattern Recognition*. 36: 279-291, 2003.
16. Daugman, J.: How Iris Recognition Works. *IEEE Transactions on Circuits and Systems for Video Technology*, 14, pp.21-30, 2004.
17. Haeberli, P., Singh, R., Gupta, P.: Image Fusion: Theories, Techniques and Applications. *Springer-Verlag Berlin Heidelberg* 2010.
18. Multiple Biometric Grand Challenge. [face.nist.gov/mbgc/](http://face.nist.gov/mbgc/).
19. Masek, L.: Recognition of human iris patterns for biometric identification. Master's thesis, University of Western Australia, 2003.
20. Zweig, M., Campbell, G.: Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. *Clinical Chemistry* 39: 561-577, 1993.
21. Mireya Sarai García-Vázquez, Alejandro Álvaro Ramírez-Acosta: Avances en el reconocimiento del iris: perspectivas y oportunidades en la investigación de algoritmos biométricos. *Computación y Sistemas* Vol 16, No 3, pp. 267-276. July-September 2012.

## Reviewing Committee

Ana Aguilar  
Luis Aguilar  
Alma Y. Alanis  
Joaquin Alvarez  
Calvillo Andres  
Eusebio Bugarín  
Hiram Calvo  
Ricardo Campa  
Nohe Ramon Cazarez-Castro  
Luis N. Coria  
Victor Diaz  
Jorge Dávila  
Adolfo Esquivel  
Diana Gamboa  
Mireya García  
Victor Manuel Hernandez Guzman  
Vitaly Kober  
Luis Alejandro Marquez-Martinez  
Fabiola Martínez  
Nataly Medina  
Javier Moreno-Valenzuela  
Juan Ivan Nieto Hipólito  
Jose Cruz Nunez  
Alejandro Rodriguez Angeles  
Manuel Rodríguez  
Julio Rolon  
Rodolfo Romero  
Javier Rubio Loyola  
Grigori Sidorov  
Juan Humberto Sossa Azuela  
Moisés Sánchez-Adame  
Juan Tapia  
Leonardo Trujillo  
Antonio Villegas  
Ricardo Campa  
Selene L. Cardenas Maciel  
Francisco Jurado  
Elvia Palacios  
Eduardo Rodriguez Angeles  
Hugo Rodriguez Cortes  
Jesús Sandoval Galarza  
Marcos Ángel González Olvera  
Eusebio Eduardo Hernández  
Alberto Luviano Juárez

**Impreso en los Talleres Gráficos  
de la Dirección de Publicaciones  
del Instituto Politécnico Nacional  
Tres guerras 27, Centro Histórico, México, D.F.  
abril de 2014  
Printing 500 / Edición 500 ejemplares**





[www.ipn.mx](http://www.ipn.mx)  
[www.cic.ipn.mx](http://www.cic.ipn.mx)



ISSN: 1870 4069

**RCS**  
Research in Computing Science