

Ciudad Politécnica de la Innovación

WIICT 2014

Proceedings of the
Workshop on Innovation on Information
and Communication Technologies 2014

 Institute
ITACA
Information and Communication Technologies

Editors:

Dr. Carlos Fernandez-Llatas
Dr. Alberto Bonastre

Committees

Organizing Committee

- **Chair: D. Alberto Bonastre, Universitat Politècnica de València, Spain**
- Dña Maria Guillem, Universitat Politècnica de València, Spain
- D. Jose Manuel Catala, Universitat Politècnica de València, Spain
- Dña Montserrat Robles, Universitat Politècnica de València, Spain
- D. Carlos Fernández-Llatas, Universitat Politècnica de València, Spain
- D. Jose Mariano Dahoui, Universitat Politècnica de València, Spain
- Dña. Isidora Navarro, Universitat Politècnica de València, Spain
- D. Manuel Traver Salcedo, Universitat Politècnica de València, Spain

Scientific Committee

- **Chair: D. Carlos Fernandez-Llatas, Universitat Politècnica de València, Spain**
- D. Cenk Demiroglu, Ozyegin University, Turkey
- D. Johan Gustav Bellika, National Center of Telemedicine, Norway
- D. Yigzaw Kassaye Yitbarek, University of Tromso, Norway
- D. Raymundo Barrales, Universidad Autónoma Metropolitana de México
- D. Jose Carlos Campelo, Universitat Politècnica de València, Spain
- D. Juan Vicente Capella, Universitat Politècnica de València, Spain
- Dña. Sara Blanch, Universitat Politècnica de València, Spain
- D. Carlos Baraza, Universitat Politècnica de València, Spain
- D. Joaquin Gracia, Universitat Politècnica de València, Spain
- D. David De Andres, Universitat Politècnica de València, Spain
- D. Ricardo Mercado, Universitat Politècnica de València, Spain
- D. Miguel Rodrigo Bort, Universitat Politècnica de València, Spain
- D. Alejandro Liberos, Universitat Politècnica de València, Spain
- D. Jorge Pedrón, Universitat Politècnica de València, Spain
- D. Vicente Traver, Universitat Politècnica de València, Spain
- D. Antonio Martinez, Universitat Politècnica de València, Spain
- Dña. Pilar Sala, Universitat Politècnica de València, Spain
- D. Alvaro Martinez, Universitat Politècnica de València, Spain
- D. Jose Luis Bayo, Universitat Politècnica de València, Spain
- D. Juan Miguel García-Gomez, Universitat Politècnica de València, Spain
- D. Miguel Angel Rodriguez, Universitat Politècnica de València, Spain
- Dña. Elena Bernal-Mor, Universitat Politècnica de València, Spain
- D. Diego Felipe Pacheco-Páramo, Universitat Politècnica de València, Spain
- D. Julian Camilo Romero-Chavarro, Universitat Politècnica de València, Spain
- D. Diego Boscá Tomás, Universitat Politècnica de València, Spain

Table of Contents

<p>NON-INVASIVE LOCATION OF RE-ENTRANT PROPAGATION PATTERNS DURING ATRIAL FIBRILLATION</p> <p><i>Miguel Rodrigo, Andreu M Climent, Alejandro Liberos, Jorge Pedrón-Torrecilla, José Millet, Francisco Fernández-Avilés, Felipe Atienza, OmerBeren-feld and Maria S Guillem</i></p>	1 – 9
<p>NON-INVASIVE IMAGING OF THE ATRIAL ARRHYTHMIA SOURCES: PATIENT STUDY DURING ATRIAL TACHYARRHYTHMIA AND ATRIAL FIBRILLATION</p> <p><i>Jorge Pedrón-Torrecilla, Alejandro Liberos, Miguel Rodrigo, José Millet, Felipe Atienza, Omer Berenfled, Andreu M. Climent and Maria S Guillem</i></p>	10 – 17
<p>METHODOLOGY FOR DEFINING A COMMON CLINICAL INFORMATION MODEL BASED ON THE CONCURRENT USE OF ADVANCED INTEROPERABILITY STANDARDS ISO</p> <p><i>Pablo Serrano, Serapio Severiano, Jesús Castellano, Carmen Fuentes, Alfredo Payá, Ernesto Barrera, and Mariano Casado</i></p>	18 – 24
<p>IDEAS TOWARDS EARLY DETECTION OF FUGACIOUS FAULTS FOR INCREASED SAFETY OF VLSI SYSTEMS</p> <p><i>Jaime Espinosa, David de Andrés, Juan-Carlos Ruiz, and Pedro Gil</i></p>	25 – 34
<p>MICROWAVE SINTERING: RECTANGULAR AND CIRCULAR CAVITIES APPROACHES</p> <p><i>Felipe L. Penaranda-Foix, Jose M. Catala-Civera, Rut Benavente, Amparo Borrell, Maria D. Salvador, Olga Garcia-Moreno, and Pedro J. Plaza-Gonzalez</i></p>	35 – 44
<p>WOOD MOISTURE MONITORING AND DETECTION OF PESTS</p> <p><i>A. Perles, R. Mercado, J.V. Capella, and J.J. Serrano</i></p>	45 – 50
<p>LONG DISTANCE WIRELESS SYSTEM FOR RURAL ENVIRONMENTS USING WIFI TECHNOLOGY AND RADIO MOBILE</p> <p><i>Carlos A. Pérez-Cerna, and Miguel A. Rodriguez-Hernandez</i></p>	51 – 59
<p>A VISIBILIMETER BASED ON A UNIVERSAL FREQUENCY-TO-DIGITAL CONVERTER</p> <p><i>R. Barrales-Guadarrama, A. Mocholí-Salcedo, M.E. Rodríguez-Rodríguez, V.R. Barrales-Guadarrama, E.R. Vázquez-Cerón</i></p>	60 – 74

<p>LEVELS OF INDUCED VOLTAGE BETWEEN RECTANGULAR MAGNETIC LOOPS</p> <p><i>J. H. Arroyo-Núñez, A. Mocholí-Salcedo, I. Rivas-Camero, A. Arroyo-Núñez, H. Jiménez Mimila</i></p>	75 – 81
<p>ACOUSTIC COMMUNICATION SYSTEM FOR UNDERWATER WIRELESS SENSOR NETWORKS</p> <p><i>Raúl Sáez, Antonio Sánchez, Ricardo Mercado, Ángel Perles, Pedro Yuste, Juan Vicente Capella, Sara Blanc, Vicente Busquets, Juan José Serrano, and Rafa Ors</i></p>	82 – 91
<p>MODELING AND SIMULATION OF UNDERWATER WIRELESS SENSOR NETWORKS</p> <p><i>Raúl Sáez, Antonio Sánchez, Ricardo Mercado, Ángel Perles, Pedro Yuste, Juan Vicente Capella, Sara Blanc, Vicente Busquets, Juan José Serrano, and Rafa Ors</i></p>	92 – 100
<p>SIMULATION OF ALL-IP ACCESS NETWORK WITH CONFIGURABLE TRAFFIC SOURCES</p> <p><i>David Gomez-Cuadrado, Victor Sempere-Paya, Miguel A. Rodriguez-Hernandez, Angel Gomez-Sacristan</i></p>	101 – 110
<p>LOW INTRUSION ACTIVE HYBRID MONITOR FOR NODES OF SENSOR NETWORKS</p> <p><i>Marlon Navia, José Carlos Campelo, Alberto Bonastre, Juan José Serrano</i></p>	111 – 120
<p>SPRING FORCES BASED INTERACTIVE AUTOMATIC TIMED PARALLEL AUTOMATON LAYOUT ALGORITHM FOR PROCESS MINING OUTPUT REPRESENTATION</p> <p><i>Carlos Fernández-Llatas, Jose Luis Bayo Montón, Jose Miguel Benedí, and Vicente Traver</i></p>	121 – 128
<p>SELECTABLE ERROR DETECTION AND CORRECTION LEVELS ERROR CONTROL CODES</p> <p><i>Joaquín Gracia-Morán, Luis-J. Saiz-Adalid, Pedro-J. Gil-Vicente, Daniel Gil-Tomás, and J. Carlos Baraza</i></p>	129 – 136
<p>TOWARDS A SERVICE-ORIENTED PLATFORM FOR THE CUSTOMIZED ANALYSIS OF BIOMEDICAL DATA QUALITY</p> <p><i>Pablo Miñarro, Carlos Sáez, Juan Miguel García-Gómez</i></p>	137 – 143
<p>SEASONAL VARIATIONS OF THE EARTH ROTATION INPRINT INTO NOISE PROCESSES</p> <p><i>V. Milián, V.N. Lesnykh, V.A. Kolombet, A. Mocholí-Salcedo, A. A. Medvedeva, G. Verdú-Martín</i></p>	144 – 153
<p>AN AUTOMATIC ALGORITHM FOR THE SEPARATION OF LEFT AND RIGHT LUNGS FROM COMPUTED TOMOGRAPHY IMAGES</p> <p><i>Noelia Ruiz García, Roberto Sanz Requena, Montserrat Robles Viejo</i></p>	154 – 163
<p>CONTENT MANAGEMENT SYSTEM: A REVIEW.</p> <p><i>José Luis Bayo, Álvaro Martínez-Romero, Carlos Fernández-Llatas, Vicente Traver</i></p>	164 – 172

<p>IMPROVING RELIABILITY IN WIRELESS SENSOR NETWORKS: TOWARDS INTELLIGENT MONITORING PLATFORMS</p> <p>Juan Vicente Capella, José Carlos Campelo, Alberto Bonastre and Rafael Ors</p>	173 – 178
<p>LOW-POWER SENSOR CONTROL MANAGER FOR MULTIHOP UNDERWATER NETWORKS</p> <p>Paulin FAUVEAU, Radwan FANIDKI, Sara BLANC, Raul SÁEZ, Juan José SERRANO</p>	179 – 188
<p>PERFORMANCE ANALYSIS OF TELECOMMUNICATION SYSTEMS BASED ON TIME-SCALE SEPARATION</p> <p>Luis Tello-Oquendo, Vicent Pla and Jorge Martinez-Bauset</p>	189 – 198
<p>MULTI CRITERIA DECISION MAKING TECHNIQUES IN DEPENDABILITY BENCHMARKING: HOW TO PROCEED?</p> <p>Miquel Martinez, David de Andrés, and Juan-Carlos Ruiz</p>	199 – 207
<p>BEST PRACTICE IN HORIZON 2020: FROM DRAWING A PROPOSAL TO FINAL PROJECT CLOSURE</p> <p>Jose A. Morgado-Miguelés, M^a Jesús Arnal-Parada, M^a José Nodal-López, and José M. Dahoui-Obon</p>	208 – 215
<p>PROPOSAL OF A SEMIAUTOMATIC CONVERSION ALGORITHM FROM ARCHETYPES TO RELATIONAL DATABASES</p> <p>Antonio Félix, Monserrat Robles, and Diego Boscá</p>	216 – 222
<p>FAST PARAMETER ESTIMATION BASED ON B-SPLINES TO MODEL ENZYMIC ACTIVITIES: A PESTICIDE DEGRADATION APPLICATION</p> <p>Pedro Mayorga, Fernando Ramos, Edgar Dantán-González, Laura Ortiz-Hernández, Claudia Díaz-Camino, Alfonso Estudillo</p>	223 – 236

Non-invasive Location of Re-Entrant Propagation Patterns during Atrial Fibrillation

Miguel Rodrigo¹, Andreu M Climent², Alejandro Liberos¹, Jorge Pedrón-Torrecilla¹, José Millet¹, Francisco Fernández-Avilés², Felipe Atienza², Omer Berenfeld³ and Maria S Guillem¹

¹ Universitat Politècnica de València, València, Spain.

² Hospital General Universitario Gregorio Marañón, Madrid, Spain

³ University of Michigan, Ann Arbor, MI, USA

Area of interest: Biomedical engineering

Abstract. Reentrant drivers or mother rotors have been defined as a mechanism responsible of atrial fibrillation (AF) maintenance whose ablation can terminate the fibrillatory episodes. These re-entrant patterns can be identified in the atrial wall by detecting phase singularities (PS) in the epicardial phase maps. In this study, we evaluate the potential role of body surface phase maps to non-invasively locate atrial sites that may harbor rotors.

High-density surface potential recordings were obtained from 14 AF patients and singularity points (SP) were located in the torso phase maps after band-pass filtering at the highest dominant frequency. Stable SPs were found during 73.1±16.8% of time after band-pass filtering vs. 8.3±5.7% on raw signals. Surface SPs representing atrial rotors were detected at specific torso areas related with their atrial location.

Besides, in the present study we made use of mathematical models of atrial activity to evaluate the representation of SP on the torso and inside the passive volume between heart and torso. We simulated two propagation patterns led by mother rotors and the projection of this atrial activity on the virtual torso. Our results showed that even for small rotors in the atria, the reentrant patterns can be detected on the torso by filtering at the highest dominant frequency.

These results show that the electrical propagation pattern in the atria during AF is reflected in the electrocardiogram. Our simulations demonstrated that rotors that remained stable on the epicardium can be detected noninvasively. Phase analysis after band-pass filtering can detect atrial drivers, helping to plan the best therapy strategies.

1 Introduction

Atrial fibrillation (AF) is the most common arrhythmia in clinical practice, with a prevalence reaching 10% in the population older than 70 years. The main pitfall in the treatment of this cardiac disease is the lack of knowledge about their mechanisms of initiation and maintenance and so the pharmacological therapies do not have the expected success.

The ablation of certain sites of the atria has demonstrated to be a procedure which can terminate AF episodes. Several ablation methods have been developed to improve the success ratios and decreasing the impact on the patient [1]. The ablation of atrial sites responsible of AF onset and/or perpetuation is successful in terminating the arrhythmia and therefore its location in the atrial wall is of great interest [2]. Recently, novel recording techniques have demonstrated that functional re-entries, or mother rotors, can be a mechanism responsible of the maintenance of AF processes [3]. These functional re-entries are a self-sustained propagation pattern which can generate fibrillatory processes. Mother rotors can be identified in the atrial wall by studying the phase map of electrograms (EGM), since the location of re-entrances corresponds to phase singularities [4]. Previous studies have found a correlation between the dominant frequency (DF) distribution along the atria and the position of the atrial rotors [2, 5]. Besides, it has recently been shown that non-invasive mapping allows the identification of these high DF atrial sources during human AF by using surface recordings [6]. In this study, mathematical models of the atrial electrical activity have been used in order to study the relation between phase singularities that appear on the atrial wall due to the presence of mother rotors and phase singularities that can also appear in body surface phase maps recorded at the torso. Furthermore, we have analyzed the body surface phase maps looking for reentrant patterns. Whether there is any connection will be determined and if this relationship can be used to study and plan ablation procedures. Thus, the aim of this study is to investigate the potential use of surface mapping recordings for the detection of atrial drivers prior to the invasive interventions.

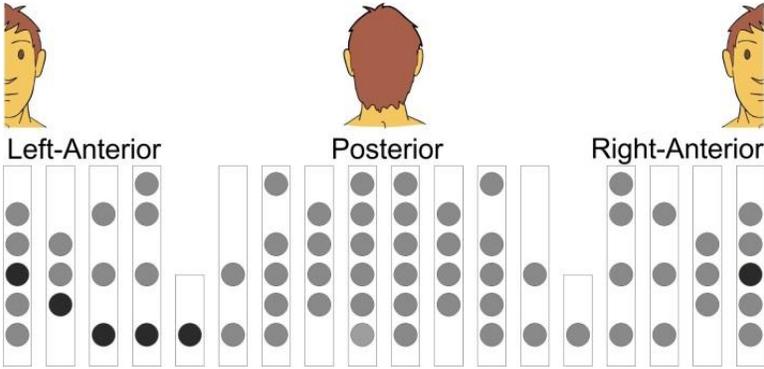


Fig. 1. Schematic geometrical configuration of the surface electrodes relative to the torso. Electrodes at locations corresponding to the standard ECG precordial leads are denoted as black circles.

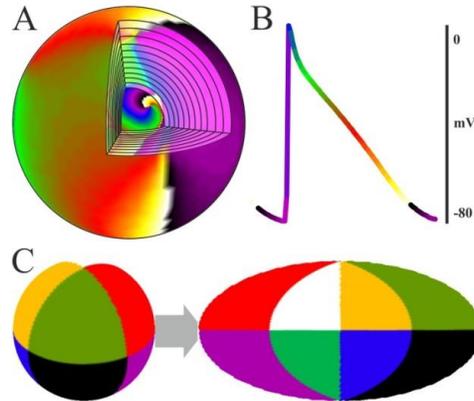


Fig. 2. (A) Scheme of the spherical atrial model (inside) and the concentric layers. (B) Color map used in phase representations. (C) Cartographic mapping of the sphere.

2 Methods

Surface electrocardiograms (ECG) from 14 paroxysmal and persistent AF patients were recorded using a grid of 67 electrodes on a vest covering the torso (see Figure 1). Ventricular activation was removed by administration of a central venous bolus of adenosine (12-18 mg) and 4-second segments of surface ECGs surrounding the longest RR interval were used for the analysis. Surface potentials were baseline-subtracted and low-pass filtered at 30 Hz [7]. Power spectral density of all signals was computed to determine the local DFs and their distribution on the body surface [6].

Intracardiac electrograms (EGMs) were simultaneously obtained from both atria during the procedure by using catheters introduced via the right femoral vein. Power spectral density of EGMs was computed to determine the DF of each atrium. Surface ECGs were then filtered at the highest DF (HDF) found on the torso surface or at the highest DF found at either left atrium EGMs (LA-HDF) or right atrium EGMs (RA-HDF) by using a 2 Hz bandwidth band-pass filter.

Due to the difficulty in obtaining simultaneous panoramic recordings from the atria and torso surface, we had to rely on mathematical models of the electrical behaviour of the atria-torso system to investigate how the atrial reentrant patterns are projected on the torso. Our atria-torso model consisted on a spherical shell of active tissue, representing the atria, within a passive torso modelled as a uniform volume conductor bounded by a spherical surface representing the torso surface [8] (see Figure 2.A).

Two atrial propagation patterns were simulated. The first simulation consisted on a uniform sphere with two stable functional rotors placed on the atrial surface. In the second simulation the atrial shell was divided into a 10% area with healthy conduction properties and a 90% area with fibrotic properties, resulting in a fibrillatory propagation pattern leaded by a stable rotor placed on the healthy area. The electric potential resulting from the atrial activity was studied everywhere on 20 concentric spheres from the atria to the torso by using the Boundary Element Method. The Aitoff carto-

graphic representation of a sphere has been used during the article to represent the complete sphere in a single map (see Figure 2.C).

Phase maps on the torso surface of patients and in each concentric layer in the computer models were obtained from the potential phase signal of each node by the Hilbert transform [9]. The phase signal ranges from 0 to 2π and represents the relative delay of each signal in one period (see Figure 2.B). A singularity point (SP) was defined as the point in a phase map which is surrounded by phases from 0 to 2π . Only those SPs that were present for the duration of at least one full rotation were considered. A filament was defined as the connection between SPs across spherical layers on the atrial-torso model at a given time.

3 Results

In Figure 3.A we can see the propagation pattern and phase maps of the EGMs from the uniform model. The action potential map (left column) exhibits the position of the two stable functional rotors and the phase map of the internal layer ($R=2.6$ cm) shows two PSs at that location. The location of both PSs is similar in intermediate and external layers ($R=5$ and $R=12.5$ cm), although the phase map does not evidence the shape of functional reentries. Fig. 3.B shows the temporal behaviour of filaments. Two filaments arise from the rotors in the atria and terminate in the torso, remaining stable in time. In Figure 3.C the temporal evolution of PSs found in the atrial wall (that defines the rotor meandering) and at the torso can be observed. In this case, the trajectory of the PSs on the torso and on the atrial wall was found to be similar.

Figure 4.A shows the phase map at increasing distances from epicardial to surface in the atrial-torso mathematical model with a reentrant driver in the 10% of healthy tissue. Note that phase maps from raw signals showed unstable patterns due to the fibrotic activity. However, after the HDF-filtering the surface phase map became stable, showing a stable SP at the nearest point to the atrial rotor. At Figure 4.B the filament distribution before and after the HDF filtering can be seen. It can be observed that HDF-filtering allows removing the activity at frequencies other than the rotor activity (fibrotic activity), projecting on the torso only the electrical activity produced by the rotor.

Measurements on the rotor occurrence before and after HDF-filtering confirmed the stabilization of the SPs (Figure 5). Stable SPs were found in unfiltered AF signals during $8.3\pm 5.7\%$ of the time vs. $73.1\pm 16.8\%$ in HDF-filtered signals ($p<0.01$) and the average SPs duration concomitantly increased following the HDF-filtering (160 ± 43 ms. vs. 342 ± 138 ms., $p<0.01$).

The band-pass filtering at the HDF found on the atrial EGM tended to concentrate the rotor apparition at certain areas of the torso. The trajectory of a surface SP that drifted during 2 seconds on the posterior torso obtained after the LA-HDF filtering in a LA-fastest patient is depicted in Figure 6.A. In Figure 6.B can be observed the trajectory of a SP that drifted during 500 ms on the right anterior torso detected after the RA-HDF filtering in a RA-fastest patient. In Figure 6.C, the 2-dimensional histogram of SP locations after LA-HDF filtering in patients with an inter-atrial DF gradient > 1 Hz

($n=10$) shows a predominant location of SPs on the posterior torso, while the histogram after RA-HDF filtering shows the predominant location on the right anterior torso. The locations of the maximal numbers of true LA or RA SPs are shown in Panels C and D to reside well within the areas demarcated by the HDFs originating either at the LA or RA (marked with dotted line), respectively, based on previous surface-atrial DF distribution correlation study [6].

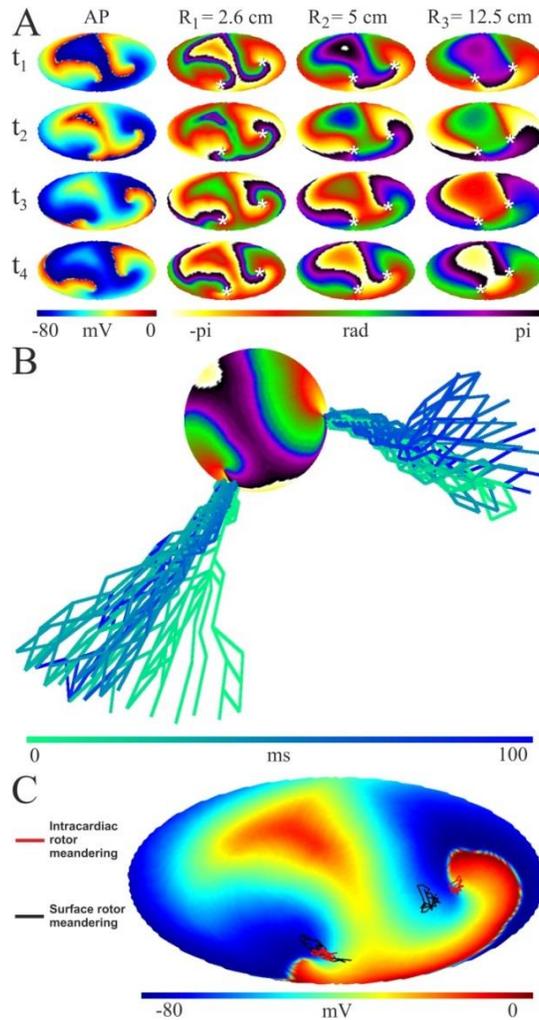


Fig. 3. Uniform model with two stable functional rotors: (A) action potential (AP) and phase maps of the electrogram at several distances from the atria and during several time instants (t_1 to t_4) (PSs are marked with white stars); (B) phase map of the inner sphere and temporal evolution of filaments; and (C) action potential map on the epicardium (red) and torso (black).

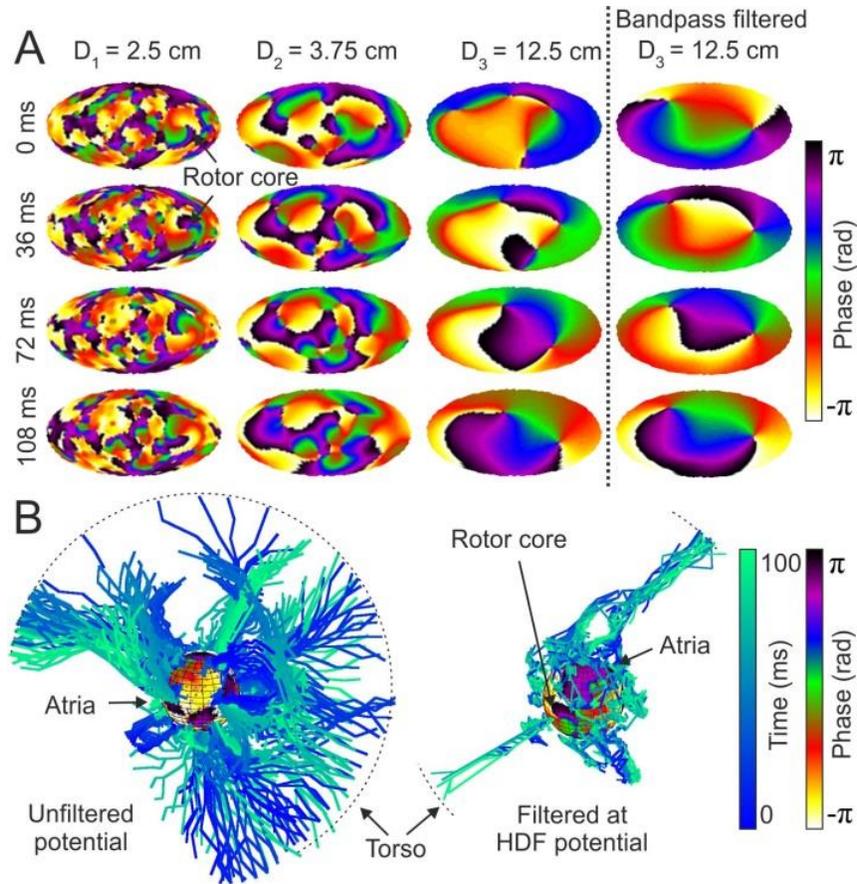


Fig. 4. (A) Phase maps at 4 time instants (top to down) in 3 concentric layers at increasing distances from the epicardium (left to right) and after HDF filtering of surface potentials. (B) Temporal evolution of filaments for unfiltered potentials and for HDF-filtered potentials.

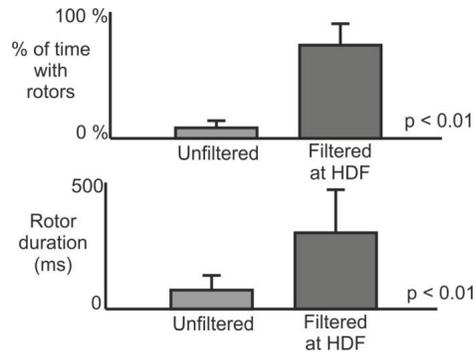


Fig. 5. Surface rotor measurements over the entire cohort.

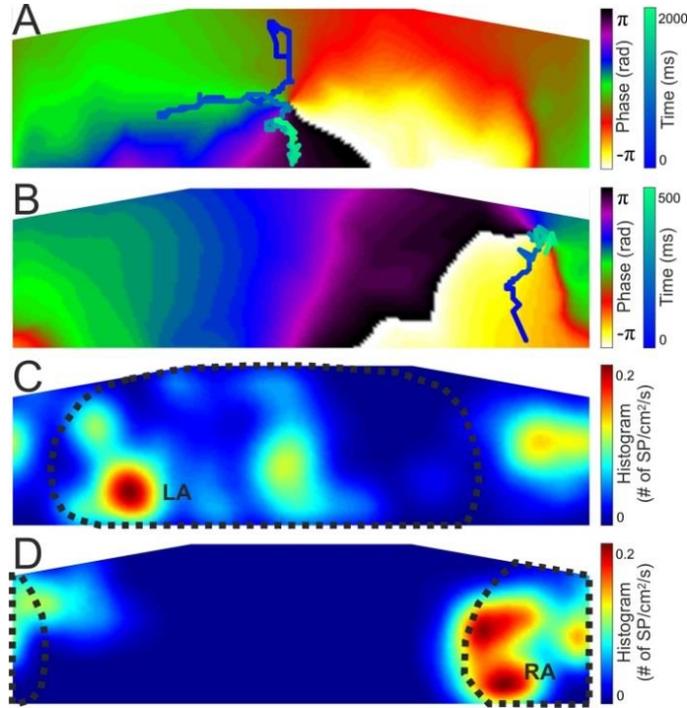


Fig. 6. (A) Phase map and rotor tracking after LA-HDF. (B) Phase map and rotor tracking after RA-HDF filtering. Histogram of the rotor position for all rotors detected in patients with an inter-atrial DF gradient after LA-HDF filtering (C) and after RA-HDF filtering (D).

4 Discussion

In this study we show that phase maps of surface potentials during AF display unstable reentrant patterns, but after HDF-filtering they allow observing reentrant patterns with spatiotemporal stability. The short-lived and unstable surface reentrant activity in the non-filtered data is suggested by computer simulations to result from superposition of irregular electrical activity at frequencies other than the HDF which may mask the presence of the more stable reentrant activation. Additionally, it has been shown as the HDF-filtering at the bands of the right and left atrium activity reflects surface rotors in different areas of the torso which coincide with the areas in which the electrical activity of the right and left atrium are projected respectively [6].

Previous studies have demonstrated that reentrant drivers can be responsible of the human AF maintenance [2, 5, 10]. Narayan et al. reported recently that rotor activity is detectable by panoramic intracardiac mapping in about 70% of 98 out of 101 AF

patients and brief ablation at the centres of those rotors was effective in terminating or slowing the arrhythmia [10].

Previous works have demonstrated that PSs in the EGM's phase signal indicates the position of functional re-entries [4]. Furthermore, several studies have shown that atrial electrical activity is reflected on the body surface potential during AF [6]. Therefore, it seems reasonable to speculate that the analysis of the surface electrical signal may contain information regarding the mechanism of maintenance of the arrhythmia and the location of the maintaining region. This work presents the possibility that these PSs provoked by stable rotors could be also observed on body surface potentials. Relation between the PSs of atrial and torso activity have been summarized into filaments, and the behaviour of filaments with different propagation patterns have been analysed. Our results showed that stable re-entrant propagation patterns exhibit filaments that remain stable on the torso.

Computer simulations have demonstrated that atrial rotors, even present in small atrial areas, can be detected on the torso surface by HDF-filtering. Activity of rotors covering a small atrial portion may be masked by the electrical activity of the rest of the atria, even if it is a highly disorganized activity, so the raw potential signal may shows unstable reentrant patterns. The HDF-filtering allows attenuating the signal component provoked by the regions working in a frequency other than the rotor, remaining after the band-pass filtering only the electrical activity provoked by the reentrant driver and thus stabilizing the surface reentrant patterns.

We cannot conclusively confirm that rotational patterns observed in patients correspond to actual atrial rotors since we do not have simultaneous epicardial and/or endocardial panoramic data. However, we have made use of mathematical models to demonstrate that in case rotors are present during AF on the atria, they can be detected on the torso surface by HDF-filtering even if the rotors are present in small areas. Although the models of the atria-torso system were simple spherical models, these simple models contain the active and passive volume conductor components needed to gain insight into the mechanisms of visual rotor stabilization by HDF-filtering.

5 Conclusion

Phase analysis of the surface potential signal after HDF-filtering and identification of singularity points can detect the presence of atrial reentrant drivers. The detection of the AF maintenance mechanism in each individual patient may help improving the diagnosis and selecting the best therapies.

References

1. Haïssaguerre, M., Jaïs, P., Shah, D.C., et al: Spontaneous initiation of atrial fibrillation by ectopic beats originating in the pulmonary veins. *N Engl J Med* (1998) 659-66
2. Aienza, M., Almendral, J., Moreno, J., et al: Activation of inward rectifier potassium channels accelerates atrial fibrillation in humans: evidence for a re-entrant mechanism. *Circulation* (2006) 2434-42
3. Jalife, J.: Déjà vu in the theories of atrial fibrillation dynamics. *Cardiovascular Res* (2011) 766-75
4. Gray, R.A., Pertsov, A.M., Jalife, J.: Spatial and temporal organization during cardiac fibrillation. *Nature* (1998) 75-8
5. Aienza, F., Almendral, J., Moreno, J., et al.: Activation of inward rectifier potassium channels accelerates atrial fibrillation in humans: evidence for a reentrant mechanism. *Circulation* (2006) 2434-42
6. Guillem, M.S., Climent, A.M., Millet, J., et al: Noninvasive localization of maximal frequency sites of atrial fibrillation by body surface potential mapping. *Circ Arrhythm Electrophysiol* (2013) 294-301
7. Guillem, M.S., Climent, A.M., Castells, F., et al: Noninvasive mapping of human atrial fibrillation. *J Cardiovasc Electrophysiol* (2009) 507-13
8. Rodrigo, M., Climent, A.M., Liberos, A., et al: Non-invasive location of re-entrant propagation patterns during atrial fibrillation. *Computing in Cardiology* (2013) 1235-8
9. Zlochiver, S., Yamazaki, M., Kalifa, J., Berenfeld, O.: Rotor meandering contributes to irregularity in electrograms during atrial fibrillation. *Heart Rhythm* (2008) 846-54
10. Narayan, S.M., Krummen, D.E., Shivkumar, K., et al: Treatment of atrial fibrillation by the ablation of localized sources: CONFIRM (conventional ablation for atrial fibrillation with or without focal impulse and rotor modulation) trial. *J Am Coll Cardiol* (2012) 628-36

Non-invasive imaging of the atrial arrhythmia sources: patient study during atrial tachyarrhythmia and atrial fibrillation

Jorge Pedrón-Torrecilla¹, Alejandro Liberos¹, Miguel Rodrigo¹, José Millet¹, Felipe Atienza², Omer Berenfeld³, Andreu M. Climent² and Maria S Guillem¹

¹Bio-ITACA, Universitat Politècnica de València, Valencia, Spain

²Hospital General Universitario Gregorio Marañón, Madrid, Spain

³Center for Arrhythmia Research, University of Michigan, Ann Arbor, U.S.A.

Area of interest: Biomedical engineering

Abstract. Ablation procedures have become one of the most efficient treatments for termination of atrial arrhythmias. The aim of the present study is the evaluation of the potential use of noninvasive imaging as a clinical tool for the identification of atrial tachycardia origin or atrial fibrillation (AF) high dominant frequency sources by solving the inverse problem prior to an ablation procedure. Simultaneous 67-lead body surface potential recordings and 15 intracardiac electrograms were obtained for two paroxysmal AF patients, who were in sinus rhythm at the beginning of the intervention. 3D atria and torso geometries were also obtained by segmenting computed axial tomography images. The first study was done during sinus rhythm and pacing the left superior pulmonary vein, reconstructing the activation sequences. For the second study, during AF, dominant frequency maps were computed. Reconstructed activation and frequency maps were consistent with the intracardiac recordings, locating accurately the source of the arrhythmia.

1 Introduction

Atrial arrhythmias are great of concern in clinical electrocardiology. Atrial fibrillation (AF) and other arrhythmias have a great clinical impact in morbidity and mortality in developed countries [1], [2]. These arrhythmias can be maintained by a small tissue area [3] and, therefore, ablation procedures have become one of the most efficient treatments since isolation of these regions usually results in termination the arrhythmia.

In the clinical settings, identification of ablation targets is achieved by electroanatomical mapping. Nowadays, electroanatomical mapping is the most common used technique in clinical practice. This technique can be only used to detect stable foci in the atria or to characterize a re-entrant pattern during atrial flutter, however, location of atrial sources during AF is more complicated since contact mapping is performed

sequentially and it is not possible to acquire a long time recording in multiple points of the atria under a non-stationary activation. A non-invasive reconstruction of the atrial activation sequence based on the recording of multiple simultaneous ECGs may overcome this limitation. Furthermore, the prior knowledge of source location may help ablation planning and reducing the procedure time.

The aim of this study is to evaluate the noninvasive electrocardiographic imaging technique as a clinical tool for (1) the identification of the origin of atrial tachycardias with stationary activation sequence by reconstructing its isochrones map and (2) locating the region with the highest dominant frequency, cause of the initiation and maintenance of the FA [4], for two patients with different frequency gradient. In this study, body surface potential mapping and computed axial tomography (TAC) images recorded in patients were used to non-invasively reconstruct the epicardial activity in the atria. Results were validated by using simultaneous intracardiac recordings.

2 Methods

Methodology section is divided as follows: (1) technical details of the inverse problem resolution are summarized, (2) clinical case studies used to illustrate the performance of the technique is described.

2.1 Inverse Problem resolution

In order to obtain the potentials on the heart surface from the potentials recorded non-invasively from the torso surface of the patient, we solved the inverse problem of the electrocardiography based on the Boundary Element Method (BEM). Briefly, according to the BEM formulation [5], [6], [7], potentials on the surface of the torso can be computed from potentials on the atrial surface by using Equations 1-3:

$$mx = b \quad (1)$$

$$m = \begin{pmatrix} D_{AA(nxn)} & G_{AA(nxn)} \\ D_{TA(mxn)} & G_{TA(mxn)} \end{pmatrix}, \quad x = \begin{pmatrix} \Phi_A \\ \Gamma_A \end{pmatrix}, \quad b = \begin{pmatrix} -D_{AT(nxm)}\Phi_T \\ -D_{AT(mxm)}\Phi_T \end{pmatrix} \quad (2)$$

$$\Phi_T = M\Phi_A = (D_{TT} - G_{TA}G_{AA}^{-1}D_{AT})^{-1} \cdot (G_{TA}G_{AA}^{-1}D_{AA} - D_{TA})\Phi_A \quad (3)$$

where Φ_A is the potential on the surface of the atria, Φ_T is the potential on the surface of the torso, Γ_A is the potential gradient in the atria, D_{TA} and D_{AT} are the potential transfer matrices from the torso to the atria and from the atria to the torso respectively, and G_{TA} and G_{AT} are the potential gradient transfer matrices from the torso to the atria and from the atria to the torso respectively.

The inverse problem could be solved by computing the inverse of matrix M (M^{-1}). However, M is ill-conditioned and, in order to overcome the ill-conditioned nature of

M , the system needs to be regularized. We accomplished this regularization by using zero-order Tikhonov's method [8], which consists of a minimization problem, according to Equation 4:

$$\min_{x \in E^n} \left\{ \|M \Phi_A - \Phi_T\|^2 + \lambda \|B \Phi_A\|^2 \right\} \quad (4)$$

where λ is a regularization parameter that we automatically obtained by the L-curve method [9] and B is a spatial regularization matrix, which is the identity matrix in zero-order formulation. Therefore, the inverse problem can be solved by using Equation 5:

$$\Phi_H(t) = (M^t M + \lambda B^t B)^{-1} M^t \Phi_B \quad (5)$$

For the study with the sinus rhythm and pacing in the left superior pulmonary vein activation pattern, in order to localize the origin of the atrial P-waves and reconstruct the activation sequence, electrograms (EGMs) were analyzed in terms of their activation times. Activation times of unipolar EGMs were computed as those instants with a maximum $-dV/dt$ and activation times of bipolar EGMs were computed as those instants with maximum amplitude.

For the AF study dominant frequency maps were non-invasively reconstructed to detect the highest frequency region AF source. Power spectral density (PSD) of all signals was computed by using Welch's periodogram with a Hamming window of 2 seconds and 50 % overlap, with 8192-point Fast Fourier Transform (FFT). The largest peak in each spectrum was defined as the dominant frequency of the node.

2.2 Clinical Cases Studies

The included cases studied were patients admitted for ablation of drug-refractory paroxysmal atrial fibrillation (males with paroxysmal AF, 49 and 50 years old). The protocol was as approved by the Institutional Ethics Committee of our institution and both patients gave informed consent. Patients were under general anesthesia and periodic heparin bolus administration. Since both patients arrived in sinus rhythm, AF was induced by electrical burst pacing.

Before AF was induced, the stationary activation pattern study was done only with the second patient during the sinus rhythm and during the left superior pulmonary vein (LSPV) pacing. Two representative P-waves were used for the noninvasive reconstruction of the isochrones maps by solving the inverse problem.

During the AF, high frequency regions were identified in two patients with different left-to-right and right-to-left frequency gradients. Once the highest dominant frequency (DF) site was identified, the navigation catheter was placed at this site. A central venous bolus of adenosine (12-18 mg) was administered to produce significant transient atrioventricular block avoiding ventricular activity for at least than 4 seconds [10].

Spectrums for all nodes were calculated, considering the largest peak in each spectrum as the dominant frequency of the node.

For the different studies, in order to reconstruct the electrical activity by solving the inverse problem of the electrocardiography, multichannel ECGs were recorded with a BSPM system. Patients wore a custom-made vest with 64 chest ECG leads (28 anterior, 34 posterior and 2 lateral leads) plus 3 electrodes used for recording the limb leads, amplified by using a commercial system (Active Two, Biosemi, TheNetherlands) [10]. Signals were acquired at a sampling rate of 2048 Hz, with a resolution of 1 μ V and a bandwidth of 500 Hz.

Intracardiac EGMs were simultaneously recorded to the surface signals by introducing the following catheters via the right femoral vein: (1) a standard tetrapolar catheter in the right atrial (RA) appendage; (2) a deflectable 4-mm mapping catheter (Marinr; Medtronic Inc., Minneapolis, MN) in the distal coronary sinus; (3) a decapolar circular mapping Lasso catheter (Biosense-Webster, Diamond Bar, CA) used to map the PV-left atrial (LA) junctions; and (4) a Navistar catheter (3.5-mm tip, 2-5-2 interelectrode distance; Thermo-Cool, Biosense-Webster, Diamond Bar, CA). Signals were acquired at a sampling rate of 977 Hz. EGMs of both atria were sequentially acquired and frequency analysis was performed in real time using a CARTO navigation system with embedded spectral capabilities (CARTO XP, version 7.7; Biosense-Webster, Diamond Bar, CA).

Prior to the intervention, computer axial tomography (CAT) images with a resolution of 0.5 mm were acquired to build the 3D atrial and torso models. Specifically, heart and torso conductive volumes were segmented from CAT images and isotropic and homogeneous conductivities of 0.6 S/m and 0.2 S/m respectively. Segmentation was accomplished by using 3D Slicer [11], then 3D surfaces were smoothed by using Matlab 2010 (The Mathworks Inc, The Netherlands).

ECG signals were processed using Matlab 2010. Baseline and high-frequency noise was reduced by applying a 0.7 to 30 Hz band-pass filter [10]. All leads were visually inspected after filtering and leads with noticeable noise or very low amplitude were excluded from further analysis.

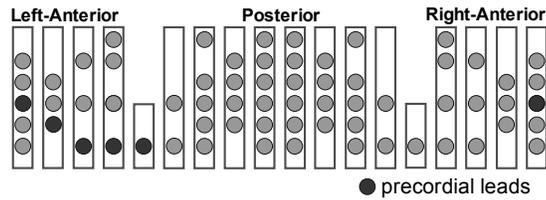


Fig 1. Electrode position in our BSPM system for atrial activity recording. Black circles correspond to the approximate location of precordial, right and left arm leads.

3 Results

3.1 Clinical case study: Atrial tachyarrhythmia

The reconstruction of the epicardial activation sequence during sinus rhythm allowed the identification of a propagation pattern from the right to the left atrium (Fig. 2A). Recorded EGMs were consistent with the computed activation sequence: the activation can be first observed in the RA catheter, followed by the CS and the LSPV (Fig. 2B). EGMs showed a difference between the activation of the RA and LSPV of 93 ms. Noninvasively computed activation times of the RA and the LA showed a maximum time difference of 102 ms.

When the LSPV was paced, EGMs showed an earlier activation in the LSPV followed by an activation in the CS and a final activation of the RA (Fig. 2D). The difference in the activation time between LSPV and RA EGMs was 49 ms. Inverse reconstruction of the epicardial propagation pattern showed one wavefront starting at the LSPV (Fig. 2C) and a depolarization of the most distal part of the RA taking place 71 ms after the stimulation.

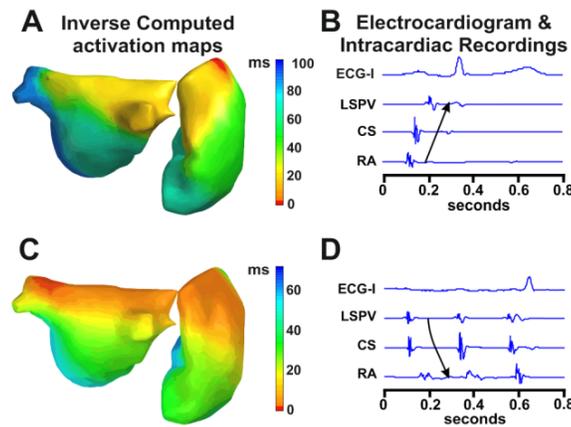


Fig 2. *Panel A:* Noninvasive reconstruction of activation sequences during sinus rhythm and *Panel C:* left atrial stimulation. *Panel B, D:* Lead I on the standard ECG and bipolar EGMs from the left superior pulmonary vein (LSPV), coronary sinus (CS) and right atrium (RA).

3.2 Clinical case study: Atrial fibrillation

Noninvasive identification and location of highest activation frequency sites was also validated by applying the inverse problem in two AF patients with different distribution patterns of activation frequency. Intracardiac EGMs evidence that patient 1 presented a left-to-right dominant frequency pattern, as depicted in Figure 3.A, with activations of the RA, LSPV and RSPV at 5.75 Hz, 7 Hz and 6.75 Hz, respectively.

Same activation frequencies were found on the RA, the LSPV and the RSPV after solving the inverse problem by using the simultaneous surface recordings, as it can be observed in Figure 3.C, which allows locating the highest dominant frequency site in the LSPV, which is consistent with the DFs measured in the intracardiac recordings.

Patient 2 presented a right-to-left dominant frequency pattern, as evidenced in the EGMs depicted in Figure 3.B, with activation frequencies of the RA, LSPV and RSPV at 8 Hz, 5.75 Hz and 6.25 Hz, respectively. Frequencies estimated from inverse-computed EGMs presented the same pattern and frequencies as those measured invasively, which allows to spatially locate the highest DF domain in the right atrium.

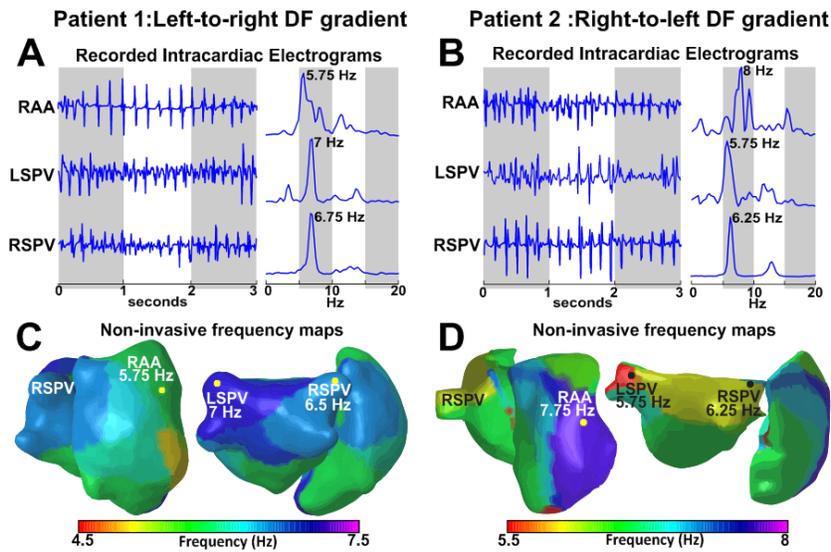


Fig 3. Human highest DF estimation during AF by solving the inverse problem. *Panel A:* Intracardiac EGMs at the right atrial appendage (RAA), Right Superior Pulmonary Vein (RSPV) and Left Superior Pulmonary Vein (LSPV) and their corresponding spectra for patient 1, with a left-to-right DF gradient. *Panel B:* Intracardiac EGMs at the RAA, RSPV and LSPV and their corresponding spectra for patient 2, with a right-to-left DF gradient. *Panel C:* Inverse-computed DF map for patient 1. *Panel D:* Inverse-computed DF map for patient 2.

4. Discussion and conclusion

In the present study, non-invasive recordings and the solution of the inverse problem of the electrocardiography were used to reconstruct (1) the activation sequence from a patient under sinus rhythm and atrial tachycardia generated by pacing the left

superior pulmonary vein and (2) the dominant frequency maps during FA. Experimental results were consistent with simultaneously recorded intracardiac EGMs.

The inverse problem of the electrocardiography was first applied for solving the electrical activity in the atria for the determination of earliest activation sites during atrial ectopies or pacing [11, 12, 13]. Solution of the inverse problem during non-fibrillating rhythms has proved to be accurate by comparing inverse-computed and intracardiac recorded EGMs. However, inverse-computed isopotential maps show paradoxically simple patterns during AF that have not been validated with simultaneous intracardiac data [11, 14].

There is experimental and clinical data supporting that in many cases AF is maintained by a region with the highest activation rate [15, 16, 17]. Ablation of the highest DF in the atria has shown to be an effective therapy for terminating with the arrhythmia [15]. In a more recent study, we have shown that an extensive recording of surface potentials by BSPM allows detecting the highest dominant frequencies and not only the global activation rate of most atrial tissue [10]. However, surface frequency maps only provide a rough estimation of the location of the highest DF site. It has already been presented how frequency maps were accurately reconstructed for different AF patterns, locating the dominant frequency regions and identifying the presence of a frequency gradient [18]. Estimation of DF maps has also shown to be less sensitive to noise than estimation of propagation patterns based on measured voltages and thus more robust for detecting electrical sources during AF [18].

Both methods demonstrated to be a promising tool for the non-invasive location of the sources of the atrial arrhythmias prior to the intervention helping in planning ablation procedures.

References

1. Fuster V, Ryden LE, Cannom DS, et al. ACC/AHA/ESC 2006 Guidelines for the Management of Patients with Atrial Fibrillation: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines and the European Society of Cardiology Committee for Practice Guidelines (Writing Committee to Revise the 2001 Guidelines for the Management of Patients With Atrial Fibrillation): developed in collaboration with the European Heart Rhythm Association and the Heart Rhythm Society. Vol. 114. *Circulation* (2006) 257-354.
2. Calkins H, Brugada J, Packer DL et al. HRS/EHRA/ECAS Expert consensus statement on catheter and surgical ablation of atrial fibrillation: Recommendations for personnel, policy, procedures and follow-up. Vol. 4 *Heart Rhythm* (2007) 816-861.
3. Haïssaguerre M, Jais P, Shah D, Takahashi A, Hocini M, Quiniou G, Garrigue S, Le Mouroux A, Le Métayer P, Clémenty J. Spontaneous initiation of atrial fibrillation by ectopic beats originating in the pulmonary veins. Vol. 339 *N Engl J Med* (1998) 659-666.
4. Atienza F, Almendral J, Moreno J, Vaidyanathan R, Talkachou A, Kalifa J, Arenal A, Vilacastin J.P, Torrecilla E.G, Sanchez A, Ploutz-Synder R, Jalife J, Berenfeld O. Activation of inward rectifier potassium channels accelerates atrial fibrillation in humans-Evidence for a Reentrant mechanism. Vol. 114. *Circulation* (2006) 2434-2442.

5. Geselowitz D.B. On Bioelectric Potentials in an Inhomogeneous Volume Conductor. Vol. 7. Biophysical Journal (1967) 1-11.
6. Horáček B.M, Clements J.C. The Inverse Problem of Electrocardiography: A Solution in Terms of Single- and Double-Layer Sources on the Epicardial Surface. Vol. 144. Mathematical Biosciences (1997) 119-154.
7. Sarvas J. Basic mathematical and electromagnetic Concepts of the biomagnetic inverse problem. Vol. 32. Physics in Medicine and Biology (1987) 11-22.
8. Tikhonov A.N. On the solution of incorrectly posed problems and the method of regularization. Vol. 4. Sov. Math. Dokl (1963) 1035-1038.
9. Hansen P.C and O'Leary D.P. The use of the L-curve in the regularization of discrete ill-posed problems. Vol. 14. SIAM J. Sci. Stat. Comput (1993) 1487-1503.
10. Guillem MS, Climent AM, Millet J, Arenal Á, Fernández-Avilés F, Jalife J, Atienza F, Berenfeld O. Noninvasive localization of maximal frequency sites of atrial fibrillation by body surface potential mapping. Vol. 6. Circ Arrhythm Electrophysiol. (2013) 294-301.
11. Cuculich P. S, Wang Y, Lindsay B.D, Faddis M.N, Schuessler R.B, Damiano Jr R.J, Li L, Rudy Y. Noninvasive Characterization of Epicardial Activation in Humans With Diverse Atrial Fibrillation Patterns. Vol. 122. Circulation (2010) 1364-1372.
12. Roten L, Pedersen M, Pascale P, Shah A, Eliautou S, Scherr D, Sacher F, Haïssaguerre M. Noninvasive Electrocardiographic Mapping for Prediction of Tachycardia Mechanism and Origin of Atrial Tachycardia Following Bilateral Pulmonary Transplantation. Vol. 23. J Cardiovasc Electrophysiol (2012) 553-555.
13. Pedrón-Torrecilla J, Climent A, Liberos A, Pérez-David E, Millet J, Atienza F, Guillem M.S. Non-Invasive Estimation of the Activation Sequence in the Atria during Sinus Rhythm and Atrial Tachyarrhythmia. Vol. 39. CinC (2012) 901-904.
14. Haïssaguerre M, Hocini M, Shah A. J, Derval N, Sacher F, Jais P, Dubois R. Noninvasive panoramic mapping of human atrial fibrillation mechanisms: A feasibility report. Vol. 24. JCE (2013) 711-717.
15. Lazar S, Dixit S, Callans DJ, Lin D, et al. Effect of pulmonary vein isolation on the left-to-right atrial dominant frequency gradient in human atrial fibrillation. Vol. 3. Heart Rhythm (2006) 889–895.
16. Atienza F, Almendral J, Jalife J, Zlochiver S, Ploutz-Snyder R, Torrecilla E, Arenal A, Kalifa J, Fernandez Aviles F, Berenfeld O. Real-time dominant frequency mapping and ablation of dominant frequency sites in atrial fibrillation with left-to-right frequency gradients predicts long-term maintenance of sinus rhythm. Vol. 6. Heart Rhythm (2009) 33-40.
17. Berenfeld O, Mandapati R, Dixit S, Skanes AC, Chen J, Mansour M, Jalife J. Spatially distributed dominant excitation frequencies reveal hidden organization in atrial fibrillation in the Langendorff-perfused sheep heart. Vol. 11. J Cardiovasc Electrophysiol (2000) 869–879.
18. Pedrón-Torrecilla J, Liberos A, Millet J, Climent A.M, Guillem M.S. Accuracy of Non-Invasive Frequency Estimation during Atrial Fibrillation. Vol. 40. CinC (2013) 1183-1186.

Methodology for defining a common clinical information model based on the concurrent use of advanced interoperability ISO standards

Pablo Serrano¹, Serapio Severiano², Jesús Castellano², Carmen Fuentes²,
Alfredo Payá², Ernesto Barrera², and Mariano Casado²

¹ Hospital Universitario de Fuenlabrada, C del Molino, nº2
28942 Fuenlabrada, Madrid, Spain
pserranob@salud.madrid.org

² Servicio Madrileño de Salud (SERMAS)
Madrid, Spain

{serapio.severiano, jesus.castellano, carmen.fuentes,
alfredo.paya, ernesto.barrera, mariano.casado}@salud.madrid.org

Abstract. Objective: For the definition of clinical information model for a shared system, from the existing EHRs at Madrid Region. Methodology: we have used concurrently the health informatics ISO advanced standards: 12967 in the service architecture, 13940 for the concepts of continuity of care and 13606 in the reference model for clinical information. It has been modeled from a use case involving the levels of primary care, hospitals and pharmacy, for chronic patients in outpatient care, applied to rheumatic disease. The use case has been transferred to workflows and clinical records and links. Results: The heterogeneous concepts of current systems have been matched through a correspondence schema with the model of 12967 and 13940. The information components have been represented with the 13606 reference model. As result we have defined the compositions, entries and links. Conclusions: the standards concepts fully cover the existing meaning for a shared repository of clinical information.

1 Introduction

The Madrid Region wants to progress in the EHR to facilitate continuity of care, based on existing EHR systems in primary care (a single system -AP Madrid-) and hospitals (more than two systems), demographic systems (Cibeles) and electronic prescription. The new shared system must evolve current regional -HORUS- and national -HCD SNS- federated viewers towards the organizational and semantic interoperability, and apply the lesson learned in the advanced research platforms (LinkERH-ISCIII).

To do this architecture is required a common information model, corporate for the regional health system, transversal to care levels and providers, and longitudinal in time, allowing continuity of care and patient-centered.

The common information model have to cover high organizational concepts, flows, reference model, clinical data structures and terminology, and it is based on the guidelines of the Madrid Region, the Ministry of Health and the EU, and is in accordance with the advanced standards ISO (Electronic Medical Record) and IHTSDO (Terminology) policy in Spain.

2 Objective

The aim was to define the model of Shared EHR for Madrid Region, according to the ISO 13606 standard: a shareable EHR with standard information model that is independent of systems EHR record and accessible by multiple authorized users.

The specification is developed from the agreement between the primary care EHR in its current model -AP Madrid-, the medication management systems with their current ambulatory and hospital prescription, and clinical history of a hospital specialty - Rheumatology - in its current departmental model of the Hospital Clinico San Carlos. The objective is the clinical modeling to design the common part of both areas, following the guideline of the HCD Summary Record of the NHS and make the detailed specification of its sections, following the methodology of clinical agreement develop in the Madrid Region for the norm "Alerts in the EHR".

3 Methods

To the design of the system we have used a methodology for the application of advanced ISO standards developed from the CEN working group named "concurrent use" and emerges as a key instrument in e-health [1]. Furthermore, experience modeling of Madrid Region points to the desirability of maintaining order in the definition from abstract to concrete concepts and from the enterprise viewpoint to information view [2]. The project has followed this sequence: first the definition of the organization information model with the intersection of the standards: organizational concepts, workflows and processes, and the reference of the clinical record context , and then in a second step the definition of clinical detailed information models with the binding of archetypes and terminology (Figure 1).

The project has a practical orientation, in order to obtain a model applicable in the short term, they fully reflect the existing meanings, that must applied to existing systems, and have to enable a services architecture accordance with standards.

The model was defined by clinical agreement, performing a first iteration between primary care, pharmacy systems and the specialty of Rheumatology, together with the existing demographic identification services. To achieve the agreement represented a great advantage both existing medical records, primary care and rheumatology are based on longitudinal models, with tracking in time, so that the high-level agreement on concepts is possible, which is necessary condition for combining information originated in primary and Rheumatology to ensure the premise of continuity of care.

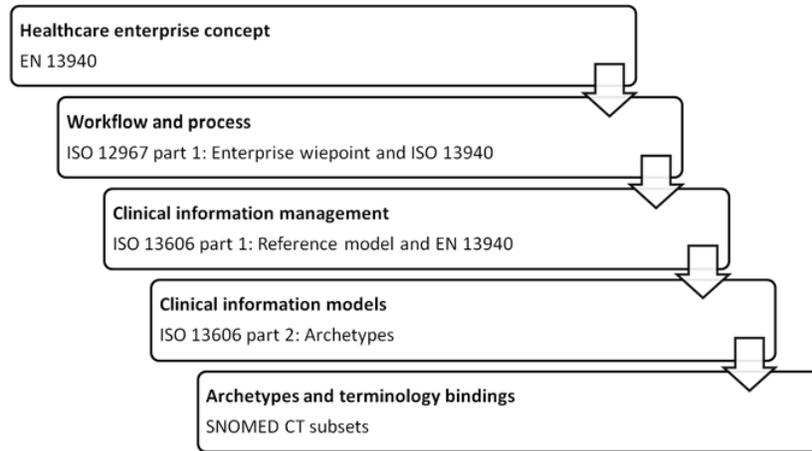


Fig. 1. Sequence for concurrent use of the standards

For this purpose we used an bottom up approach from existing use models, with the reference of complementary standards (openEHR, IHE) and more universal models of international experiences (NHS UK, NEHTA Australia) for advanced definitions and resolve discrepancies. Therefore, for the agreement were taken as a starting point the current use models of primary care, pharmacy, rheumatology, along with demographic systems, both common organizational guidelines as the systems themselves. We documented their policies and procedures, manuals, interfaces, record hierarchy, forms, data models and "local" terminology.

The clinical agreement covered the two levels: organizational (healthcare concepts, workflows and information management) and semantic (data structures and terminology) and followed three successive stages: since the enterprise policies to information models and finally terminologies.

We used concurrently the "ContSys" EN 13940 [3] and documents under development of the new standard ISO 13940 [4], "EHRCom" ISO 13606 part 1 and 3- EHR Reference Model and Term list [5, 6] and "HISA" ISO 12967 part-1-workflows [7], for agree the model of healthcare organizational concepts, the flows and the reference of clinical information (hierarchy and information context) of shared history.

The work plan followed the following tasks sequence:

1. Constitution and training the experts group (clinical and technical).
2. Document current models of primary care, pharmacy and Rheumatology.
3. Compilation of national and european guidelines and international references.
4. Agreement on model of organizational concepts for healthcare.
5. Use case description.
6. Agreement workflows and processes.
7. Agreement on management of clinical information.

We described a use case in detail, which involves all healthcare levels: primary care, specialized care and pharmacy, for patients followed over time, with orders (prescriptions, diagnostic requests, referrals) in ambulatory care, applied to the rheumatic pathology, in the current situation -as is- and in the scenario of a shared EHR -to be-. The use case has been transferred to workflows and these to clinical records components and their links, pointing the existing flows and those in development.

The concepts in the various systems have been matched through a correspondence schema with the model of 12967 and 13940. The information components have been represented with the reference model of 13606.

3 Results

The following table shows the processes and systems involved in the model.

Process	Actors	Systems
Subject of care demographic		Demographic services (Cibeles)
Identification of Actors		Demographic services (Cibeles) Hospital Information systems
Schedule	Patient relation center (CAP)	Primary care system (AP Madrid) Multi schedule system (Multicita) Hospital Information systems
Clinical observation and evaluation	Primary care Rheumatology	Primary care system (AP Madrid) Rheumatology systems
Medication prescription	Primary care Rheumatology Pharmacy office	Primary care system (AP Madrid) e-Prescription system Dispensation repository
Diagnostic tests orders	Primary care	Primary care system (AP Madrid)
Temporary occupational disability	Primary care Labor inspection	Primary care system (AP Madrid) Labor inspection system (ITwin)
Referral to specialized care	Primary care Patient relation center (CAP)	Primary care system (AP Madrid) Multi schedule system (Multicita) Hospital Information systems
Responsibility decision	Rheumatology	Rheumatology systems
Hospital medication management	Rheumatology Hospital pharmacy	Rheumatology systems Hospital pharmacy systems (Farmatools)
Care plans	Primary care	Primary care system (AP Madrid)

Were defined LINK between RECORD_COMPONENT for full semantic representation of existing meaning in current systems. Were used the coded-value attribute LINK_NATURE with the terms LINK-B0, C0 and D0.

3 Discussion

The project has demonstrated the usefulness of standardizing heterogeneous concepts of the existing systems through correspondence with the concepts of 12967 and 13940. The standards concurrent use methodology may apply from the existing systems, being useful in the short term. This approach has two consequences: that value the systems element conformed to the standards and offers a guiding the redesign of the nonconforming. The use of the three standards allows modeling at all levels of abstraction for the concepts.

We have recognized good practices in the modeling with 13606 to fully capture the existing meaning for a shared repository of clinical information. Especially relevant, the extensive use of LINK class to represent temporal sequences, flows and associations of clinical logic. The COMPOSITIONS have to reflect clinical record sessions instead of secondary composite document. In contrast to approaches based on "documents", this information model conform to the three standards, is based on the clinical information registration, according to real clinical work flow. These are "primary" compositions: those that are updated and have separate context information, while documents or extracts to communicate are secondary: views or summaries.

Additionally, we have used clinical concept for openEHR. We have differentiated persistent compositions versus event compositions and have ordered compositions sequence based on clinical reasoning: observations, assessment, instruction, and action.

Finally, have been identified the archetypes for the entries defined in the information model, to be designed in the next step.

References

1. TC251. Towards concurrent use of ContSys, 13606, and HISA. WG1 Report from 2nd Workshop [Internet]. Madrid: CEN; 2013 Mar. Available from: http://www.ehealth-interop.nen.nl/publicaties/2877&fil_Id=1260
2. Unidad de investigación en Telemedicina y eSalud, Muñoz Carrero A, Romero Gutiérrez A, Marco Cuenca G, Abad Acebedo I, Cáceres Tello J, et al. Manual práctico de interoperabilidad semántica para entornos sanitarios basada en arquetipos. Madrid: Instituto de Salud Carlos III. Ministerio de Economía y Competitividad (2013)
3. EN 13940-1:2007. Health informatics - System of concepts to support continuity of care - Part 1: Basic concepts.
4. ISO/DIS 13940: 2012. Health informatics - System of concepts to support continuity of care.
5. ISO 13606-1:2008. Health informatics - Electronic health record communication - Part 1: Reference model.

6. ISO 13606-1:2008. Health informatics - Electronic health record communication - Part 3: Reference archetypes and term lists.
7. ISO 12967-1:2009. Health informatics - Service architecture - Part 1: Enterprise viewpoint.

Ideas Towards Early Detection of Fugacious Faults for Increased Safety of VLSI Systems

Jaime Espinosa, David de Andrés, Juan-Carlos Ruiz, and Pedro Gil

Fault-Tolerant Systems (STF), Instituto ITACA
Universitat Politècnica de València (UPV), Campus de Vera s/n, 46022, Valencia,
Spain

Phone: +34 96 3877007 Ext {75774, 75752, 79707}, Fax: +34 96 3877579
{jaiesgar, ddandres, jcruizg, pgil}@disca.upv.es
Area of Fault Tolerant Systems

Abstract. As technology advances the number and profile of faults systems face has evolved quickly. Hence, traditional detection and diagnosis strategies must be revisited to guarantee such new fault profiles can be covered. In this paper we move focus from those faults manifested in system state or outputs to others happening in combinational logic, previously presented as fugacious. Though these are frequently masked either logically, electrically or temporally, their early detection could result in vital information towards failure prevention or forecasting, as they are consequence of disturbances like environmental radiation or excessive wear-out. This can be specially useful in dynamically reconfigurable FPGAs where adaptation can take place. The paper explores these ideas evaluating requirements, challenges and pitfalls of standard detection architectures, while providing insight on potential techniques to be applied towards their proper detection and diagnosis.

Keywords: Fault detection; fault diagnosis; transient faults; intermittent faults; permanent faults; fugacious faults; SET

1 Introduction

It is unquestioned that in VLSI critical systems, safety plays a key role. Likewise, commercial and high availability systems can also benefit from guaranteed service levels. The concept is based on controlled operation of both software and hardware, in order to achieve an acceptable degree of reliability. In this work we focus on the supporting hardware and the faults it suffers.

Irrespective of their originating cause, be it manufacturing defects, radiated particles, electromagnetic interference or noises, etc, faults can be classified roughly in permanent or transient categories. Research in the field has focused mainly in tackling permanent faults, disregarding transient faults whenever their effect is not visible as errors in the captured data. For instance, transient faults with short activation times (percentage of time in which it is affecting the system relative to clock period), which have been shown difficult to detect by conventional means [1], may not produce incorrect outputs at once, but are a good

indication of a problematic environment or an early warning of a future downfall. In previous work they were presented as *fugacious* [2]. When transient faults are repeated in the same area in a short period of time and due to the same cause they are classified as *intermittent* [3]. Whether the final nature of the fault is transient or intermittent will depend on several factors, namely wear-out condition, ageing, extreme temperatures etc...

Some illustrative examples on how such an early warning can be useful could include a satellite travelling along a high radiation level area, triggering a process to increase redundancy in hardware and software critical tasks and disabling non critical ones before the system is saturated and hung. Other example is early detection of intentional EMI attack to break secrecy in an encryption core, allowing to change data codification of buses, or other techniques to counter the attack. Finally, an ageing problem detected in an industry robotic arm or dangerous machine, or in a braking control system of a train could well raise service alarms before total loss of control. For that reason a proper diagnosis of faults will provide valuable information when taking decisions for the evolution of the system, thus preventing catastrophic consequences.

Previous studies devoted specifically to detection and diagnosis of fugacious faults are scarce or non-existent. For transients caused by radiation (*single events*) some works have been done, mainly focused towards characterising them along different technologies [4]. However, they are not focused towards their detection and diagnosis in working circuits. To this end, certain known detection techniques could be applied to fugacious faults with limited success [5], since only a reduced period of time is monitored.

The contribution of this paper is twofold: (i) to identify and ponder the challenges of detection and diagnosis of fugacious faults in VLSI systems and (ii) to provide insight on methods and technologies to cope with such challenges.

The rest of the paper is structured as follows. Section 2 justifies the importance of fugacious faults, and underlines the difficulties of detecting transient and intermittent faults with short activation times. Furthermore it presents the different fault models while providing an overview on diagnosis of such faults. Section 3 provides a methodology and a proposal of architecture to face the detection and diagnosis of such faults. Finally, Section 4 indicates the current implementation and following actions to be taken plus related issues.

2 Fugacious faults. Impact and challenges of detection and diagnosis

The type of fault we target in this work is the *fugacious* fault. They cannot be neglected any longer since evolution in electronics has raised them as an issue to consider. In fact, for a transient fault generated in the combinational logic to become an error, possibly causing failure and utterly a safety threat 3 conditions are to be met: generation, propagation and capture in sequential logic.

Before sub-micron technology, generation of transient upsets in combinational logic had been largely studied in literature. The best documented process

is the impact of heavy ions on silicon, where an energy transfer takes place causing a charge to be accumulated in the affected node. Because that charge had to be enough as to overtake the switching energy (critical charge Q_{crit}), this was not so frequent. Propagation of transients through logic takes 3 requirements [6]. First is to have an upset of enough energy so as to reach the voltage threshold V_{th} of the technology. Second, an open logic path needs to be ready for the upset to show in the end of the chain. Finally, the speed (slew rate) of the logic has to be fast enough to cope with the shorter pulses or they are strongly attenuated. It was then difficult to propagate quick pulses in lines mainly because the slew rates were not so sheer. As of capturing transients in the sequential logic, 2 new conditions are in the need [6], i.e. the energy of the transient is enough to switch the state of the sequential element, and more importantly the instant the capture edge of the clock arrives must be synchronised with the presence of the transient at the inputs. These conditions were rarely met.

After recent advances in technology, and with miniaturisation, the lesser amount of energy is required to accumulate in order to generate an upset thus increasing the numbers [7], and the shorter upsets it causes for the same energy [8, 9]. Also new fabrication techniques such as silicon-on-insulator (SOI) follow this trend, with values of $250ps$ pulse widths for $250nm$ feature sizes and $110ps$ for $100nm$ sized technology in available studies [8]. Along heavy ions, new materials and fabrication processes also increase the risk of generated upsets associated to miniaturisation or fabrication defects such as random telegraph noise (RTN) [10]. In the same path propagation has grown woefully because voltage thresholds have been steadily shrinking and speed has been dramatically increasing, as demonstrated in literature [8]. These effects have allowed for the propagation of shorter and shorter pulses, from a $105ps$ minimum pulse width for $350nm$ to a $40ps$ minimum width for $100nm$ sizes in SOI, and a similar trend applies for bulk silicon [8]. In the capture side, high frequency systems feature a growing weight of transient upsets linearly with the frequency [11]. However, clock frequency of latest devices has not followed a steady rise, but instead it has been kept growing much lesser than possible due to power and temperature issues.

To sum up, the situation is such that more and more ever-shorter transients reach the sequential inputs, which although replicated may suffer errors when under high transient rates [11]. For that reason detecting those transients and analysing them can bring some invaluable help in terms of (1) preventing a defeat of sequential simple replication strategies, and (2) diagnosing a premature manifestation of wear-out or poor fabrication in a certain area of the device., and (3) preventing a malicious side-channel attack by early detection.

2.1 On-line detection of faults and errors

As established by Avizienis [12] a basic criterion to catalogue faults in permanent or transient type is the persistence. This can however be an incomplete information to comprise the whole picture and thus, *activation reproducibility* is the concept introduced to better describe the observed situations. For permanent faults, different activation patterns lead to solid, hard faults when these are

systematically reproducible or to elusive, soft faults when they are not. Depending on circumstances those soft faults can be intermittent in time. For transient faults, elusive activation is the most common but certain circumstances can likewise make them manifest intermittently.

Such differentiated activation patterns require tailored techniques for detection and diagnosis for dependability threats caused by faults and errors. In several situations including high availability or high performance systems, a concurrent detection (on-line) becomes critical. So far, in order to test proper development of the systems several methods have been described. From post-manufacture checking by means of test vectors or burn-in testing used to discard flawed units, to assigning slots of regular service time for test, for instance, many off-line techniques are currently employed. But the advantage of on-line detection is clear. A loose detection or notification latency, can have disastrous consequences in certain situations [13]. Besides, the longer a fault is present in the system without detection the higher the probability of facing a multiple fault situation. Provided that the latter is a problem of increased complexity we find justified interest in early detection.

There is long tradition in the dependability community to develop on-line error detectors. In them, the relationship between a fault occurred at the processing network and an error manifested in the outputs or state variables is a limiting factor known as *observability* [14]. When the observability in an output is null for a given fault, no matter which input data combination was applied the fault will not show at the output. Therefore for detection it will be desirable to maximise the observability in order to achieve a good chance of faults coverage. However, there are more factors involved. In general, there are 3 possible causes for fault filtering: electrical, logical or temporal. Electrical causes are inherent to technology. When dealing with permanent faults, temporal causes are discarded and due to the nature of hard faults, logical filtering can only be short in time provided the observability is good enough.

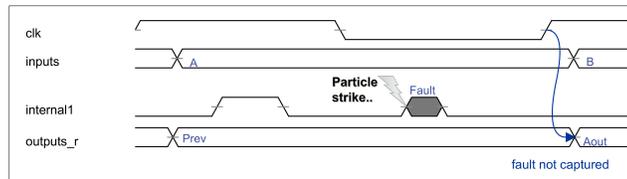


Fig. 1. Temporal filtering

Nevertheless, detection of transient/elusive and intermittent faults is not so straightforward. As explained in the beginning of this section, chances are those transient faults will not be usually captured by clock edges at the storage elements (heavy temporal filtering, see Figure 1) . The derivatives of this are that the moment a fault is captured in sequential elements and detected, many

more could be already happening and the available time for reaction could be too short. If logical filtering combines with time filtering, the detection latency is magnified. Therefore it is desirable to detect them as soon as possible, i. e. in a constant surveillance fashion, before potential mitigation strategies can be defeated. Regarding intermittent faults, tailored detection can be useful since, depending on the context of operation, random originated intermittent faults appear and disappear unpredictably, but systematic intermittent faults evolution can be numerically characterized [15]. In the latter case early detection enables a proper decision on the best moment to apply recovery actions to maximise availability, since such systematic intermittent faults start by small fluctuations which grow in time and intensity until their effect is severe.

2.2 Fault diagnosis

Multiple efforts have been conducted towards an effective diagnosis of different types of faults based on their activation reproducibility. When possible it brings important benefits for the Mission Time and Mean Time To Failure (MTTF), as demonstrated in [16]. Hence, it is clear that no equal treatment has to be given to both of them. For instance, transient faults will require no corrective action at all when hardware redundancy provides a voted fault tolerance. Disregarding the affected element for a certain period of time will negatively affect the dependability of the system. Furthermore, given the nature of intermittent faults and their proneness to become permanent, a proper distinction provides insight on the convenience to isolate or recover the functional unit. Intermittent faults diagnosis is a hot-topic in the field. An analytical model for a fault controller was presented in [17], using a thresholds-based α count methodology to discriminate transient from intermittent faults. Its Stochastic Activity Networks (SAN) analysis is specifically based on the time step, where transient faults last for less than one step and intermittent faults repeat their appearance in subsequent steps. Its drawbacks are it requires a long latency to discriminate, and infrastructure to detect and accumulate the respective faults. In the case of fugacious faults, we take into account events of a quickly 'evanescent' nature where the capture and diagnosis procedure must have intrinsically low latency. It must be able to process two or more faults per cycle in order to discriminate an intermittent activation from a transient activation, avoiding further frequency constraints.

2.3 Considered fault models

According to the presented concept of activation time, and the activation reproducibility described earlier, the framework of fugacious faults is applied to a set of 3 different types of fault models (see Figure 2). The fugacious transient faults are defined as those which remain active less than a clock cycle of the system. Likewise, fugacious intermittent faults are those intermittent faults which activate at least twice in a clock period. Finally, fugacious *permanent* faults are those faults active the whole time span of the clock period, that means for us a

fault lasting more than one clock period will be considered permanently active, or might be considered non-fugacious.

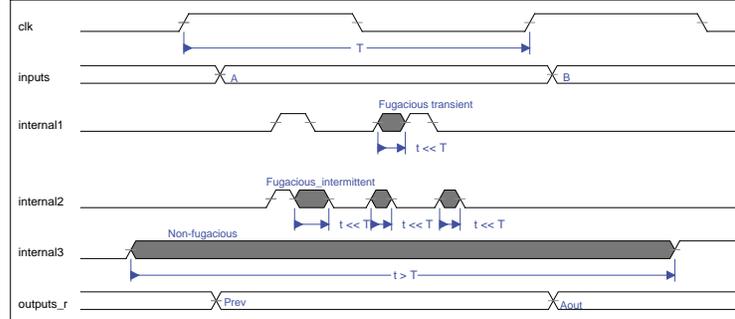


Fig. 2. Fault models and time filtering effect

3 Proposed methodology targeting fugacious faults

Our effort has been focused in two directions: determining an appropriate structure to detect and diagnose the set of faults we are targeting and defining a procedure to apply such structure to the standard design flow.

3.1 Global architecture

In every VLSI circuit we can find combinational stages separated by registers. To the primary goal of producing correct results we add another: to sense any deviations in the datapath which may be out of reach by just checking registered values.

The steps to take in order to reach these goals start by considering hardware replication and comparison. A large number of commercial systems use such effective technique at the expense of important amounts of hardware. The foremost advantage is quick on-line mitigation (when a voter is included), and usually there is no need to include voters in every stage but just in critical ones. Nevertheless, for detection and diagnosis a lighter, cheaper technique would enable the possibility to deploy detection to a larger number of partitions spread around the system. The use of codification may well fill the gap and combine with replication in a wise manner.

The use of systematic codes, which can be optimised for quick and small parity prediction functions [18] and do not alter original bits, is chosen for the envisioned topology in Figure 3. In it, a *Detection Block* would receive inputs directly from partition input registers, and also from outputs prior to registering. This block would include thus a set of Commercial Off-The-Shelf (COTS)

encoder and decoder which can be a single bit parity prediction/decoding pair in its simplest form.

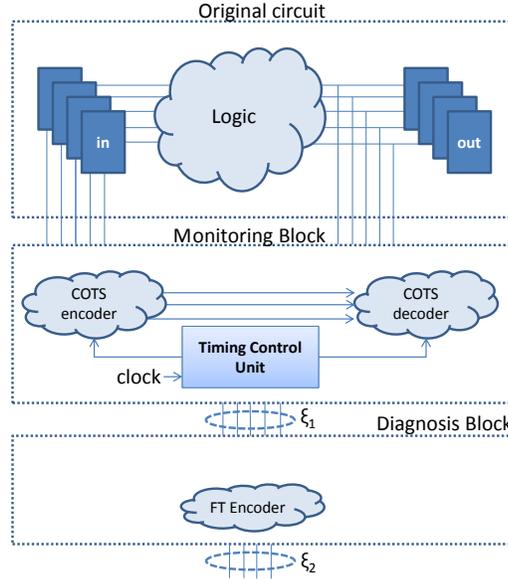


Fig. 3. Global scheme of the faults detection and diagnosis infrastructure. Timing Control Unit handles temporisation of Detection decoder

But something else is needed. Coding functions are effective against permanent errors, or transient errors which have not been time filtered. To detect transient faults of a limited activation time (smaller than a clock period), additional elements are required. An example using triplication was presented in [5], where intermittent faults were not considered at all, and the sensing time was rather reduced. An idea dealing with intermittent faults was to inject a carrier signal in the line under study and monitor the correct behaviour of it [19]. Again, the cost is rather high: an injector and receiver for the lines under analysis, plus extra wear-out due to increased switching of the primary lines. A cheaper detection can be achieved by monitoring those coded lines devoted to detection.

To avoid those shortcomings, an additional element included is the *Timing Control Unit* (TCU). Its function is to adjust the timings of detection elements with one goal in mind, i.e. to increase the *observation window*. The term refers to the percentage of the clock period when the lines under study are monitored for any potential faults. If we reduce the switching interval as opposed to the stability interval of the signals, we will have increased the available time to monitor (see Figure 4), thus gaining speed of detection.

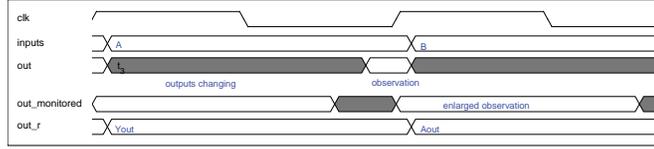


Fig. 4. Observation window enlarged by means of reducing period of signal switching

Finally, the detection information could be codified against faults using a code (ξ_1) and passed to a *Diagnosis Block*, where the same or a different code (ξ_2) can be used to notify the diagnosis output to a fault controller.

Inside the *Diagnosis Block*, inputs must be analysed and discriminated to offer 5 different output possibilities: transient fault, intermittent fault, permanent (non-fugacious) fault, no fault and error in the diagnosis. To achieve the goal, the Diagnosis Block will be built using a fault-tolerant (FT) encoder designed to minimise resources taken. By providing all these different outputs and doing so in a fault tolerant codification, the most adequate decision will be enabled to be taken at the fault controller. Hence, smart reactions can be applied well in advance to an eventual collapse of fault tolerance infrastructure.

3.2 Implementation procedure diagram and explanation of steps

In order to automate the process of deploying a detection and diagnosis infrastructure to a generic design block, a suggested procedure is shown in Figure 5. What is depicted is a typical semi-custom design flow for VLSI products, where the standard steps are on the left hand side. *Technology files* can represent a silicon foundry design kit or an FPGA manufacturer primitives library. Likewise, *Physical* element can be a layout file or a programming bitstream for an FPGA. On the right we find detail of 2 interventions in the flow.

A first intervention comes before the *Synthesis* and after *Design Entry*. This step comprises an addition of required infrastructure in the Detection Block, i.e. the COTS components and Timing Control Unit. Entry files are modified as required and new timing constraints are generated for the TCU, to drive the remainder of the design flow.

A second intervention happens in a loop between Gate-level and Physical stages of the design. The purpose is to check timings against new constraints, mainly affecting the TCU, and refine the implementation in a loop by tweaking in one of the 2 re-entry points A or B. If path B is selected, a faster process will be obtained as a result, but deep knowledge of the underlying technology will be required and we will find a side effect of loss of portability. With path A, a more general solution will be obtained at the cost of speed of implementation.

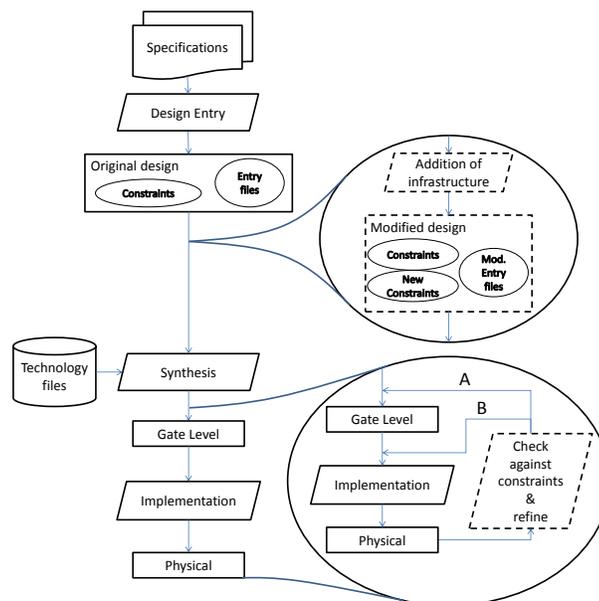


Fig. 5. Tools interaction

4 Ongoing Work

An prototype is currently under development, where an FPGA-based design flow has been chosen to support initial testing. Following the presented ideas, we have been able to develop first modification point working models. To reach optimal performance, we need first is to maximise the detection capabilities of the structure, both in area and time. This means achieving a high degree of observability at the check lines.

As second requirement is to achieve low performance penalty results and at the same time maximising the period of lines surveillance. We need the least possibly intrusive system in order not to give in too much in exchange for detection. This is vital when applied to extreme performance demanding systems.

Last but not least, keeping the additional area small can be complex in certain circuits, if a powerful logic optimisation is not wisely applied. The upper limit will be that imposed by pure replication but this should be perfectly reducible without loosing much of the observability. An associated parameter to area increase is the power drain due to new infrastructure. As usual in engineering, specifications and market constraints drive the balance between detection and diagnosis capability and power/area/performance penalty.

Acknowledgements. This work has been funded by Spanish Ministry of Economy ARENES project (TIN2012-38308-C02-01).

References

1. Gracia-Moran, J., et alt.: Experimental validation of a fault tolerant microcomputer system against intermittent faults. In: DSN. (2010) 413–418
2. Espinosa, J., Andrés, D., Ruiz, J.C., Gil, P.: The challenge of detection and diagnosis of fugacious hardware faults in vlsi designs. In Vieira, M., Cunha, J., eds.: Dependable Computing. Volume 7869 of Lecture Notes in Computer Science. Springer Berlin Heidelberg (2013) 76–87
3. Constantinescu, C.: Intermittent faults and effects on reliability of integrated circuits. In: Proceedings of the 2008 Annual Reliability and Maintainability Symposium, Washington, DC, USA, IEEE Computer Society (2008) 370–374
4. Ferlet-Cavrois, V., et alt.: Single event transients in digital cmos -a review. Nuclear Science, IEEE Transactions on **60**(3) (June 2013) 1767–1790
5. D’Angelo, S., Sechi, G.R., Metra, C.: Transient and permanent fault diagnosis for fpga-based tmr systems. In: Proceedings of the 14th International Symposium on Defect and Fault-Tolerance in VLSI Systems. DFT ’99, Washington, DC, USA, IEEE Computer Society (1999) 330–338
6. Diehl, S., et alt.: Considerations for single event immune vlsi logic. Nuclear Science, IEEE Transactions on **30**(6) (Dec 1983) 4501–4507
7. Shivakumar, et alt.: Modeling the effect of technology trends on the soft error rate of combinational logic. In: Proceedings of the 2002 International Conference on Dependable Systems and Networks. (2002) 389–398
8. Dodd, P.E., Shaneyfelt, M.R., Felix, J.A., Schwank, J.R.: Production and propagation of single-event transients in high-speed digital logic ics. IEEE Transactions on Nuclear Science **51** (dec 2004) 3278–3284
9. Ferlet-Cavrois, V., et alt.: Direct measurement of transient pulses induced by laser and heavy ion irradiation in deca-nanometer devices. Nuclear Science, IEEE Transactions on **52**(6) (Dec 2005) 2104–2113
10. ITRS: Process integration devices and structures report. Technical report, International Technology Roadmap for Semiconductors (2011)
11. Buchner, et alt.: Comparison of error rates in combinational and sequential logic. Nuclear Science, IEEE Transactions on **44**(6) (Dec 1997) 2209–2216
12. Avizienis, A., Laprie, J.C., Randell, B., Landwehr, C.: Basic concepts and taxonomy of dependable and secure computing. IEEE Trans. Dependable Secur. Comput. **1** (January 2004) 11–33
13. Johnson, C., Holloway, C.: The dangers of failure masking in fault-tolerant software: Aspects of a recent in-flight upset event. In: System Safety, 2007 2nd Inst. of Engineering and Technology International Conference on. (oct. 2007) 60–65
14. Bolchini, C., Salice, F., Sciuto, D.: Fault analysis for networks with concurrent error detection. IEEE Des. Test **15**(4) (oct 1998) 66–74
15. Correcher, A., et alt.: Intermittent failure dynamics characterization. Reliability, IEEE Transactions on **61**(3) (sept. 2012) 649–658
16. Sosnowski, J.: Transient fault tolerance in digital systems. IEEE Micro **14**(1) (feb 1994) 24–35
17. Bondavalli, A., et alt.: Threshold-based mechanisms to discriminate transient from intermittent faults. IEEE Trans. Comput. **49**(3) (mar 2000) 230–245
18. Ko, S.B., Lo, J.C.: Efficient realization of parity prediction functions in fpgas. J. Electron. Test. **20**(5) (oct 2004) 489–499
19. Kim, C.: Detection and location of intermittent faults by monitoring carrier signal channel behavior of electrical interconnection system. In: Electric Ship Technologies Symposium, 2009. ESTS 2009. IEEE. (april 2009) 449–455

Microwave sintering: rectangular and circular cavities approaches

Felipe L. Penaranda-Foix¹, Jose M. Catala-Civera¹, Rut Benavente²,

Amparo Borrell², M. D. Salvador², Olga Garcia-Moreno²

Pedro J. Plaza-Gonzalez¹

¹ITACA-UPV, Camino de Vera, s/n; 46022-Valencia (Spain)

Corresponding e-mail: fpenaran@dcom.upv.es

²ITM-UPV, Camino de Vera, s/n; 46022-Valencia (Spain)

Abstract. Microwave sintering is increasing their applications to more and more new materials or even existing ones but improving their physical and structural properties. This paper presents two different microwave cavities for sintering materials based on rectangular and cylindrical structures with two different materials sintered in each one, and some results are also presented to probe the advantages of microwave sintering.

1 Introduction

Two different materials are proposed to be sintered in two different microwave cavities based on rectangular and circular waveguides. The electromagnetic analysis of these kind of cavities is well known [1], but their use at high temperature for sintering materials in a multidisciplinary use (electrical engineering combined with material engineering) is becoming more and more popular for the excellent results of the sintered materials in terms of physical and structural results.

The two proposed materials are Alumina–zirconia nanocomposites ($\text{Al}_2\text{O}_3\text{-ZrO}_2$) and lithium alumino silicate (LAS) compositions.

Alumina–zirconia nanocomposites are considered among advanced ceramics due to their superior mechanical properties and chemical inertness, and they have been used for a wide range of applications such as biomedical implants [2 3] and structural ceramics [4,5]. One of the main reasons for improving mechanical behavior is the ability of transformation toughening which usually is induced by applied stresses to the composite structure [6,7].

Recent studies suggest that mechanical properties of $\text{Al}_2\text{O}_3\text{-ZrO}_2$ composites could be considerably increased by reducing the grain size and increasing its dispersion. The size of zirconia grain must be smaller than a critical size (0.7 μm) to ensure the stability of tetragonal phase at room temperature [8,9]. Additionally, a fine-grained Al_2O_3 matrix also results in enhanced mechanical properties. And it is also suggested

that a smaller matrix size (grain at submicron level or nanoscale) will result in a higher improvement in mechanical properties. Therefore, dense fine-grained and full-dense $\text{Al}_2\text{O}_3\text{-ZrO}_2$ composites are really desirable. Unfortunately, it is difficult to obtain $\text{Al}_2\text{O}_3\text{-ZrO}_2$ composites with high density and smaller grain size with conventional techniques. This is mainly due to the inevitable grain growth during the sintering process. Then, to minimize grain growth and to enhance mechanical properties, microwave sintering has been proposed. The significant difference between conventional sintering and microwave fast sintering is the heating rate and the heating mode. The microwave sintering method is a fastest consolidation technique with a low energetic cost [10,11].

In this case, for this material, a rectangular cavity, based on the TE_{101} mode, including a control system to follow a temperature profile to obtain $\text{Al}_2\text{O}_3\text{-ZrO}_2$ nanocomposites is proposed. Some results and comparisons between the conventional sintering and microwave sintering are then represented in the next sections.

On the other hand, the second material to be sintered is the lithium aluminosilicate (LAS) compositions, that have been extensively studied because of its very low or even negative thermal expansion that have found a wide application field including cookware, bakeware, electronic devices, telescope mirror blanks, ring-laser gyroscopes and optically stable platforms [12,13].

Sintered negative thermal expansion materials have usually low mechanical strength because the expansion anisotropy causes microcracking. This is due to different extents of thermal expansion in different crystallographic orientations, which induces internal stress with temperature change. Additionally, it has been reported [14] that the microcracking depends on the grain size. Therefore, an increasing of the β -eucryptite grain size causes a progressive microcracking and consequently a more negative bulk of thermal expansion coefficient.

In the LAS system, the high temperatures required to fully densify ceramic powders result in large grain sizes due to Ostwald ripening when traditional sintering techniques are used [15]. This makes obtaining dense materials with nanometric and submicrometric grain sizes extremely difficult and, as a consequence, the sintered materials do not achieve high mechanical properties. To overcome the problem of grain growth, non-conventional sintering methods have emerged as promising techniques [16].

Microwave heating is then proposed as the non-conventional sintering technique to solve the difficulties found with previous techniques such as SPS (spark plasma sintering). The microwave technique was specially designed to fabricate ceramic LAS bodies with high density, very low glass proportion and high mechanical properties (hardness and Young's modulus).

In this case, a cylindrical cavity, based on the TE_{111} mode is proposed. And also some results and comparisons between the conventional sintering and microwave sintering are then represented in the next sections.

2 Microwave Cavity

As mentioned above, two cavities have been designed for sintering the two described materials (Alumina–zirconia nanocomposites and LAS): rectangular and cylindrical cavities

2.1 Rectangular cavity

Fig. 1 shows the microwave system designed in rectangular waveguide. From right to left we have: the 1 kW magnetron, including the feeding system, the circulator to prevent the power source from reflected power (including a coupling system to measure the reflected power with the multimeter), the TE_{101} cavity, which will be described in detail in the following paragraphs, and a motorized short-circuit to tune the cavity and connected (like the magnetron, etc) to a laptop to allow an automatic tuning system. The rectangular cavity excites the TE_{101} mode, based on a WR340 waveguide and excited through a circular iris to maximize the H_x magnetic component of the TE_{101} mode.

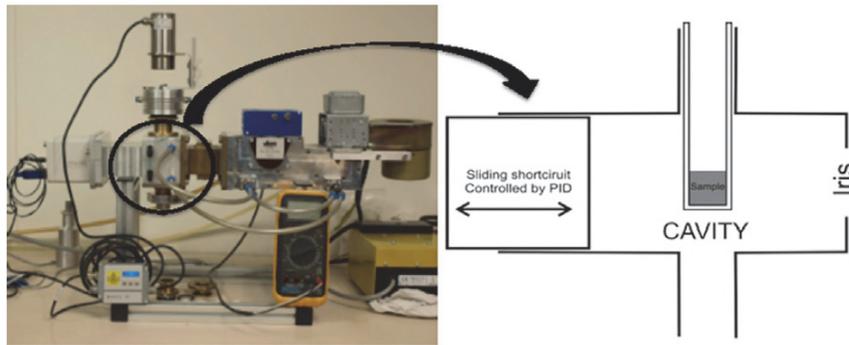


Fig. 1. The microwave system designed for sintering and scheme of the cavity

The sample is introduced in the cavity through an insertion hole located approximately in the center of the cavity (the cavity dimensions can vary with the motorized short-circuit) and on the top of it. The insertion hole has a diameter of 3 cm to guarantee that the insertion hole is under cut-off and no microwave energy leaks through, propagating the TE_{11} mode, which is the fundamental mode in the cylindrical waveguide.

In the center of the cavity the excited mode has a maximum of E_y component, and a behavior of a $\sin(\cdot)^2$ in the axial component and no variation in the vertical compo-

ment, which guarantees the homogeneity of the field in the sample, which is supposed to be about 1 cm of diameter as maximum.

To locate the sample in the center, a quartz tube is used, which is transparent to the microwave energy. This could be also used for introducing different atmosphere (or even vacuum). To monitorize the temperature, a pyrometer is located on the top, as shown in Fig. 1. This allows to include a control system to automatize the heating process, based on a PID system.

A close scheme of the cavity is shown in Fig. 1 (right), where the position of the sample is shown and also the sliding short circuit that changes the resonant frequency of the cavity and is controlled by the PID to guarantee the tuning of the cavity.

2.2 Cylindrical cavity

A single mode cylindrical cavity operating in the TE_{111} mode with a resonant frequency of 2.45 GHz was selected as the heating cell for LAS microwave sintering. The cavity has two 12 mm diameter holes in the top and lateral walls, which allows access for a quartz tube containing the specimen (radius=10 mm, height=15mm) and a temperature sensor respectively. The dimension and position of these holes were designed to ensure that there was no microwave leakage from the cavity and there was negligible perturbation of the resonant mode. The E field vectors are perpendicular to the cavity axis with the maximum electric field magnitude at the center, where the samples are located.

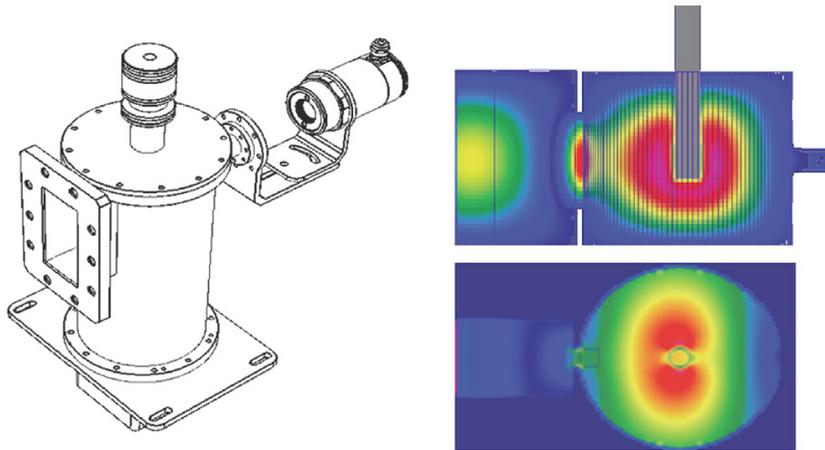


Fig. 2. Microwave E field magnitude distribution in the cavity heating cell

Preliminary dimensions of the cavity (radius=52 mm, height=85 mm) were determined analytically and afterwards optimized with the use of a commercial EM simulator QuickWave3D [17]. A movable short-circuit, as in the rectangular cavity but now at the bottom of the cavity, permits to track the cavity heating mode resonant variations caused by changes in the dielectric constant of the heated test sample during the sintering process.

Fig. 2 shows the EM fields in the cavity with holes and sample holder given by the EM simulator.

3 Results

Depending on the microwave cavity, we divided this section in two parts: rectangular and cylindrical cavities for sintering materials.

3.1 Results for rectangular cavity

Apart from other differences between the conventional sintering and the microwave sintering (like the relative the relative density or hardness and fracture toughness), we emphasize that, even knowing that the spark plasma sintering process (SPS) is shown to be a highly efficient technique for densification of fine-grained $\text{Al}_2\text{O}_3\text{-ZrO}_2$ composites at temperatures 100–200°C lower than that needed by conventional sintering (this reduction in sintering temperature is attributed to the high pressure used (100 MPa) and the efficient heat transfer from spark discharges between the particles [18]) this method has a big problem with the sintering of zirconia materials compared with microwave processing due to carbon diffusion within the zirconia sample by SPS, which is linked to the carbon rich atmosphere in which it is performed [10]. As the sintering of the compact is taking place in a graphite die, the carbon diffuses into the sample from the die and this process is affected by the applied pressure. The sample sintered by SPS shows a full black color. Eliminating this contamination is possible, but this would require high temperatures (4800°C) and a long time inside a furnace (42 h), resulting in high economic costs.

As results, the FE-SEM micrographs of the alumina–zirconia ceramics prepared by microwave and conventional sintering are shown in Figs. 3 and 4, respectively. There are obvious microstructure differences for the alumina–zirconia ceramics prepared by microwave and pressure-less or conventional sintering.

Fig. 3 shows the fully dense microstructure of microwave sintered samples, whose grain size were 200 nm and 60 to 100 nm for $\text{Al}_2\text{O}_3\text{-ZrO}_2$, respectively, and the smaller zirconia grains homogeneously dispersed in alumina matrix. During sintering, finely dispersed zirconia particles delay the motion of alumina grain boundaries, which inhibit the grain growth of alumina [19].

As shown in Fig. 3, all sintered samples have a similar grain size. Small differences in MW process samples are shown as ZrO_2 content increase. The average grain

size of Al_2O_3 and ZrO_2 in the specimen sintered at 1400°C with 5 and 10 vol% ZrO_2 reaches about 420 and 125 nm, respectively (Fig. 3a and b). While samples sintered at 1400°C with 15 vol% ZrO_2 yield a grain size of 575 and 146 nm, respectively (Fig. 3c). The final grain size of Al_2O_3 is smaller when the ZrO_2 content is lower (5 and 10 vol%). A limited amount of zirconia grains were located at the boundaries of alumina grains, and most of them are present at triple junctions, resulting in an intergranular type nanocomposite.

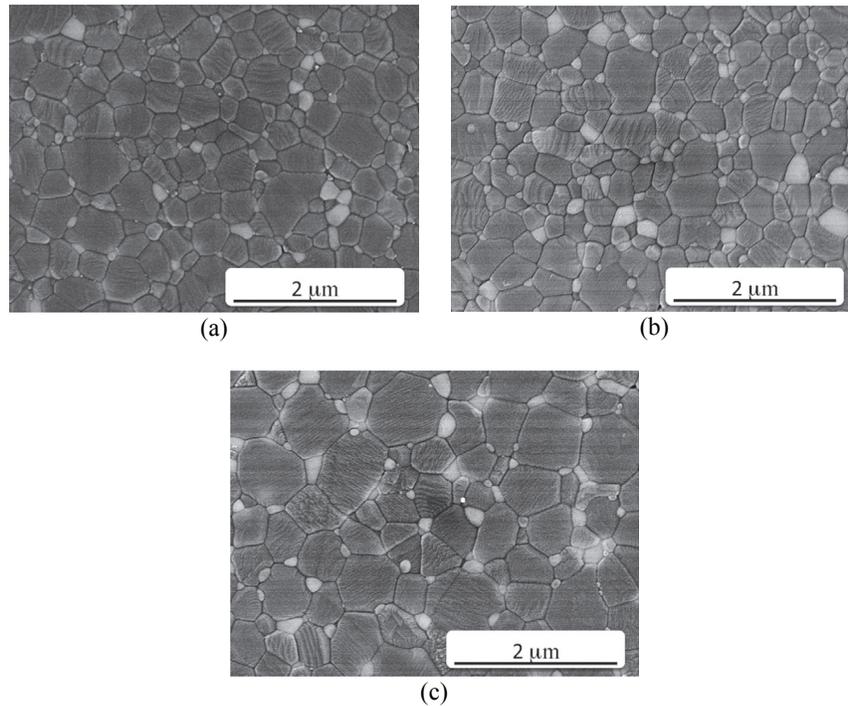


Fig. 3. The FE-SEM micrographs of nearly full dense Al_2O_3 - ZrO_2 samples sintered by MW at 1400°C : A5Z [5%] (a), A10Z [10%] (b) and A15Z [15%] (c).

The evolution of average grain size of the conventionally sintered specimens is similar to the microwave samples (Fig. 4). Therefore, the main parameter governing grain growth is the ZrO_2 content. With a lower content of ZrO_2 (A5Z and A10Z), the average grain size of Al_2O_3 and ZrO_2 reaches about 570 and 160 nm, respectively (Fig. 4a and b). But in the A15Z sample, the average grain size of Al_2O_3 and ZrO_2 reaches about 928 and 196 nm, respectively (Fig. 4c). Therefore, these data show larger grain sizes in samples sintered by the conventional process with high ZrO_2 content, this behavior suggesting that the microwaves had direct effect on cation segregation at the grain boundaries during the phase partitioning process, which is

known to be responsible for the sluggish grain growth during the final stage of sintering [20].

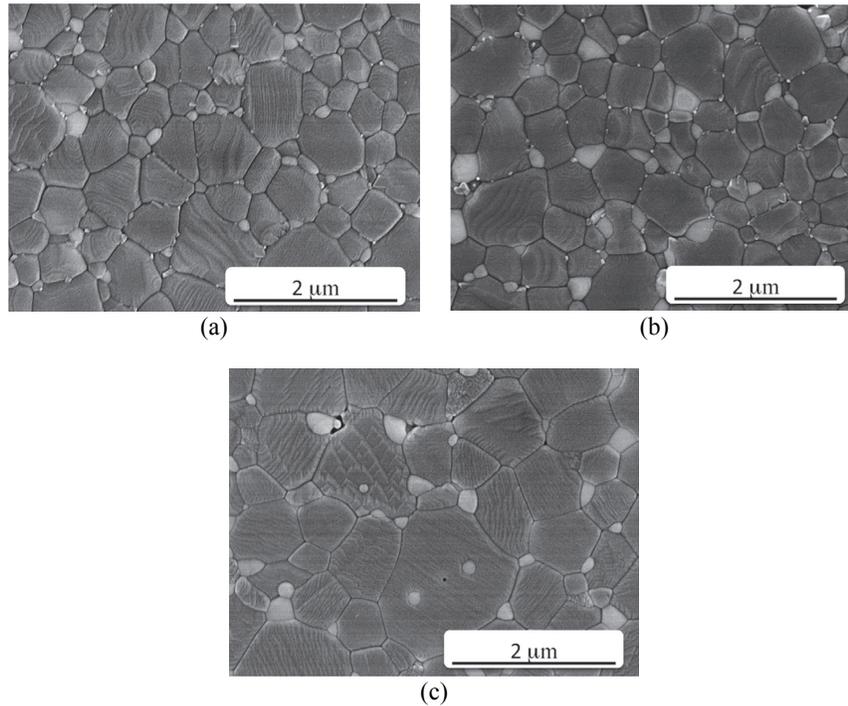


Fig. 4. The FE-SEM micrographs of nearly full dense $\text{Al}_2\text{O}_3\text{-ZrO}_2$ samples sintered by CS at 1400°C : A5Z [5%] (a), A10Z [10%] (b) and A15Z [15%] (c).

3.2 Results for cylindrical cavity

In the second case, for the cylindrical cavity, LAS ceramic powder samples (with 1 to $1.5\ \mu\text{m}$ of grain size) were placed in quartz tubes and inserted in the microwave cavity for sintering. No susceptors were required to assist the initial heating period during the trials. The samples were fired in air atmosphere at a heating rate of $100^\circ\text{C}\ \text{min}^{-1}$ to temperatures between 1200 and 1250°C with a short dwell time of 5 and 10 min. Overheating of the samples was avoided by an exhaustive control of the heating rate. The cooling rate of the microwave-heated samples was also controlled by reducing the amount of power supplied to the controller. As a result, the samples were exposed to the microwave field during the cooling process, as well as during heating. Fig. 4 shows the temperature profile and microwave absorbed power during the sintering

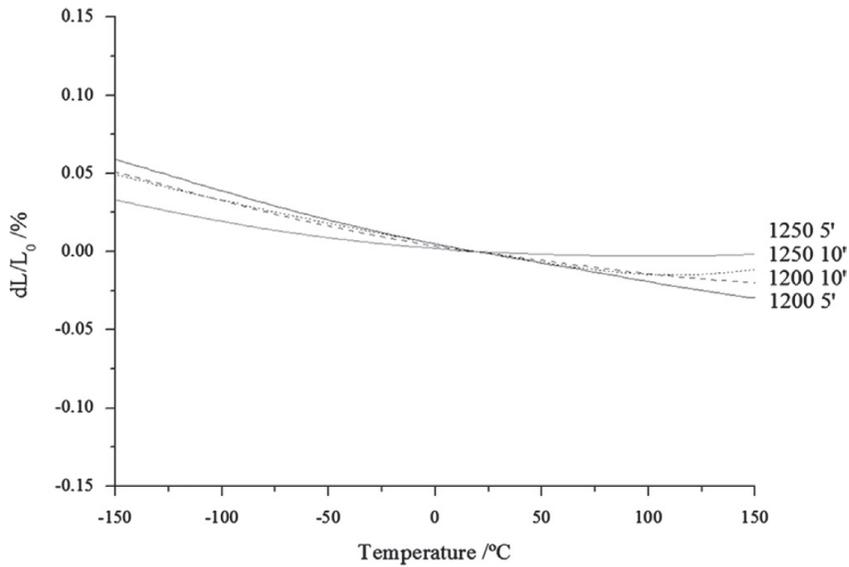


Fig. 5. Elongation vs. temperature of the sintered samples by MW at different temperatures.

process of a LAS specimen. The figure shows a microwave experiment with a resident time of the ceramic sample of 10 min around 1200°C.

Fig. 5 shows the elongation vs. temperature curves for each of the sintered samples. These data allow calculation of the thermal expansion coefficient data (or technical values).

The coefficient of thermal expansion (CTE) was very low for all the samples, with values between -3.01 and $+0.41 \cdot 10^{-6} \text{ K}^{-1}$, depending on the final sintering temperatures and the temperature interval of the measurement. In general, as can be seen in Fig. 5, the sintered LAS samples have negative expansion behavior in the low temperature range, which slightly changed to positive in the high temperature range for some samples. In sample sintered at 1250°C with 5 min, a roughly flat curve can be observed between room temperature and +150°C, with almost null elongation. In general, the CTE of all samples in range of -150 to +150°C is near-zero and controlled. This fact is necessary for a wide range of industrial applications.

4 Conclusions

This paper summarizes two papers [21,22] that show the advantages of microwave sintering, both with rectangular and cylindrical cavities. Two examples of sintering have been showed and the results probe that microwave sintering provides even better materials that those resulting from conventional sintering methods.

Acknowledgments

Authors want to thank Spanish Government (Ministerio de Economía y Competitividad (MINECO)) for financial support (project DINAWAVE, reference N° TEC2012-37532-C02-01) and co-funded by ERDF (European Regional Development Funds). And A. Borrell, acknowledges the Spanish Ministry of Science and Innovation for her Juan de la Cierva contract (JCI-2011-10498) and SCSIE of the University of Valencia.

References

- [1] Robert E. Collin, Foundations for Microwave Engineering, 2nd Edition, ISBN: 978-0-7803-6031-0, December 2000, Wiley-IEEE Press.
- [2] Z.G. Liu, J.H. Ouyang, Y. Zhou, J. Li, Effect of alumina addition on the phase evolution and thermal conductivity of ZrO₂-NdO_{1.5} ceramics, *J. Alloys Compd.* 468 (2009) 350–355.
- [3] R. Benzaid, J. Chevalier, M. Saadaoui, G. Fantozzi, M. Nawa, L.A. Diaz, R. Torrecillas, Fracture toughness, strength and slow crack growth in a ceria stabilized zirconia–alumina nanocomposite for medical applications, *Biomaterials* 29 (2008) 3636–3641.
- [4] J. Chandradass, M.H. Kim, D.S. Bae, Influence of citric acid to aluminium nitrate molar ratio on the combustion synthesis of alumina–zirconia nanopowders, *J. Alloys Compd.* 470 (2009) 9–12.
- [5] Z. Wei, H. Li, X. Zhang, S. Yan, Z. Lv, Y. Chen, M. Gong, Preparation and property investigation of CeO₂-ZrO₂-Al₂O₃ oxygen-storage compounds, *J. Alloys Compd.* 455 (2008) 322–326.
- [6] Z. Xiu, J. Laeng, X. Sun, Q. Li, S.K. Hur, Y. Liu, Phase formation of Al₂O₃/Ti(C,N)-NiTi composite, *J. Alloys Compd.* 458 (2008) 398–404.
- [7] A.Z.A. Azhar, M.M. Ratnam, Z.A. Ahmad, Effect of Al₂O₃/YSZ microstructures on wear and mechanical properties of cutting inserts, *J. Alloys Compd.* 478 (2009) 608–614.
- [8] J.R. Kelly, I. Denry, Stabilized zirconia as a structural material, *Dent. Mater.* 24 (2008) 289–298.
- [9] F.A.T. Guimaraes, K.L. Silva, V. Trombini, J.J. Pierri, J.A. Rodrigues, R. Tomasi, E.M. Pallone, Correlation between microstructure and mechanical properties of Al₂O₃/ZrO₂ nanocomposites, *Ceram. Int.* 35 (2009) 741–745.
- [10] A. Borrell, M.D. Salvador, E. Rayón, F.L. Peñaranda-Foix, Improvement of microstructural properties of 3Y-TZP materials by conventional and non-conventional sintering techniques, *Ceram. Int.* 38 (2012) 39–43.

- [11] R. Benavente, A. Borrell, M.D. Salvador, O. García-Moreno, F.L. Peñaranda-Foix, J.M. Catalá-Civera, Fabrication of near-zero thermal expansion of fully dense β -eucryptite ceramics by microwave sintering, *Ceram. Int.* 40 (2014) 935–941.
- [12] L. Xia, G. Wen, L. Song, X. Wang, The crystallization behavior and thermal expansion properties of β -eucryptite prepared by sol–gel route. *Materials Chemistry and Physics*, 119 (2010) pp. 495–498.
- [13] G.H. Beall, L. R. Pinckney, Nanophaseglass-ceramics. *Journal of the American Ceramic Society*, 82 (1999), pp. 5–16.
- [14] A. Pelletant, H. Reveron, J. Chévalier, G. Fantozzi, L. Blanchard, F. Guinot, F. Falzon, Grain size dependence of pure β -eucryptite thermal expansion coefficient. *Materials Letters*, 66 (2012), pp. 68–71.
- [15] U. Anselmi-Tamburini, J. E. Garay, Z. A. Munir, Fast low-temperature consolidation of bulk nanometric-ceramic materials. *Scripta Materialia*, 54 (2006), pp. 823–828
- [16] O. García-Moreno, A. Fernández, R. Torrecillas, Solid state sintering of very low and negative thermal expansion ceramics by spark plasma sintering. *Ceramics International*, 37 (2011), pp. 1079–1083.
- [17] P. J. Plaza-Gonzalez, A. J. Canos, J. M. Catalá-Civera, J. D. Gutierrez-Cano, Complex impedance measurement system around 2.45 GHz in a waveguide portable system. *Proceedings of the 13th International Conference on Microwave and RF Heating*, Toulouse, September 2011, pp. 447–450.
- [18] Z.A. Munir, U. Anselmi-Tamburini, M. Ohyanagi, The effect of electric field and pressure on the synthesis and consolidation of materials: a review of the spark plasma sintering method, *J. Mater. Sci.* 41 (2006) 763–777.
- [19] D. Sarkar, S. Adak, N.K. Mitra, Preparation and characterization of an $\text{Al}_2\text{O}_3\text{--ZrO}_2$ nanocomposites, Part I: powder synthesis and transformation behavior during fracture, *Compos. Part A* 38 (2007) 124–131.
- [20] K. Matsui, H. Yoshida, Y. Ikuhara, Isothermal sintering effects on phase separation and grain growth in yttria-stabilized tetragonal zirconia polycrystal, *J. Am. Cer. Soc.* 92 (2009) 467–475.
- [21] R. Benavente, et al., Fabrication of near-zero thermal expansion of fully dense β -eucryptite ceramics by microwave sintering. *Ceramics International* (2013), <http://dx.doi.org/10.1016/j.ceramint.2013.06.089>
- [22] R. Benavente et al., Mechanical properties and microstructural evolution of alumina–zirconia nanocomposites by microwave sintering. *Ceramics International*, 40 (2014), pp. 11291–11297

Wood moisture monitoring and detection of pests

A. Perles¹, R. Mercado¹, J.V. Capella¹, J.J. Serrano¹

¹ Institute ITACA, Universitat Politècnica de València,
46022 Valencia, Spain

{aperles, rmercado, jcapella,
jserrano}@itaca.upv.es
<http://www.itaca.upv.es>

Abstract. Wood is a fundamental component of a great quantity of structures with high added value e.g. historical value. However it is very prone to deteriorate in certain environmental conditions and to suffer pest attacks such as termites. This paper presents a wireless system developed in ITACA intended to continuously monitor the condition of the wood. The node implements novel analysis techniques and is capable to provide continuous operation in an autonomously manner for more than 13 years.

1 Introduction

The target problem is to monitor wood masterpieces and structures of heritage buildings. In this environment we must take into consideration that the maintenance is practically impossible, so the deployed nodes must work for years without operator's intervention [1][2].

The actual implementation of the monitoring system is based on a star configuration, where nodes send its information to a sink.

These nodes has been designed and developed with the enterprise AIDIMA (Furniture, Wood and Packaging Technology Institute) and are being used for monitoring heritage wood structures and masterpieces. As an example, the system is actually running in the cathedral of Valencia, Spain (see Figure 1).

Our wireless nodes can be utilized for the wood ambient and plagues monitoring. The ambient monitoring is done measuring relative ambient humidity and temperature to compute the equilibrium moisture content of the wood (EMC). The plagues detection is done using LEDs and light sensors to detect reflection variations when an insect such as termites, ants, cockroach, etc. crosses the detector. Most of the energy requirements of these nodes is invested in the insects detection.



Fig. 1. Altar of the Cathedral of Valencia

When it is a requirement to have a large number of nodes, the star configuration is not well suited and does not scale well. And taking into consideration the building structure of historical buildings this case is easily possible.

For example, star configuration is appropriate for an altarpiece, but not for a whole stone-made romanian church.

2 System description

As part of several joint projects with the Technological Institute of Furniture, Wood and Packaging (AIDIMA), several methods of wood monitoring have investigated in order to monitor environmental conditions and pest detection.

As a result of the project a novel technique for detecting pests and damp wood has been developed. The technique is based on the calculation of the equilibrium moisture in wood (EMC) and the optical motion detection.

As an example of the results of the second phase, Figure 2 shows a typical pattern of termite attack (in blue), where initial explorations, an onset of activity and final activity.

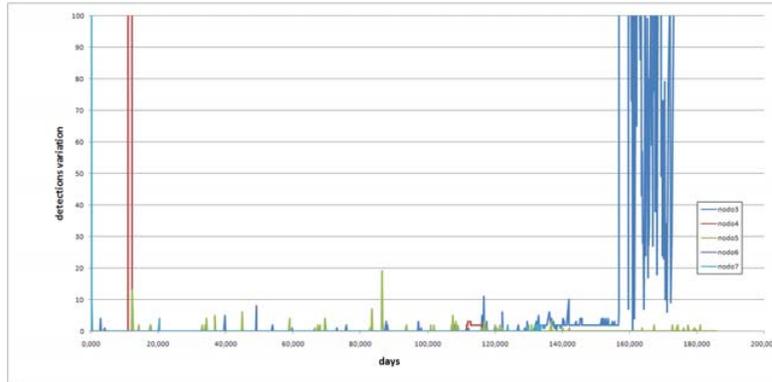


Fig. 2. Detection Pattern

This technique has been applied in nodes installed inside a cylinder of 2.5x5 cm. embedded in wood. When deployed, the node is powered and the hole sealed using a wood cap to maintain internal ambient conditions. Figure 2 shows a test network node in a block of wood attacked by termites.



Fig. 3. Node after a termites attack

This node, using a star wireless protocol, last more than 13 years[1], being the energy required for RF negligible compared with other energy requirements.

3 Implementation

The system is based on a set of small nodes inserted in wood. These nodes collect ambient information and send a digest to a base station.

A Silabs C8051F930 low power microcontroller is the heart of the node. This relatively new 8 bits 8051 derivative performs really well and the available low power modes are very flexible. Also tested that power requirements declared on the datasheet are real.

The node computes equilibrium moisture content of the wood based on the readings of ambient temperature and humidity using a Sensirion SHTx sensor.

Also, the node has an attractor for insects that are detected using light reflection variations produced with a high efficiency led from Avago and a high sensitivity sensor from Taos. This sensor is activated every 2 seconds.

The node is powered using a high energy density 1100 mAh, 3.6 V. lithium-tionyl battery.

The RF section of the node uses a Texas Instruments/Chipcon CC1101 ISM band transceiver. The chip has been configured for the 868 MHz European ISM band. This election is a trade-off between small antenna and distance reach. The 2.4 GHz band has smaller antennas but its behavior on water moisture and on crossing objects is bad. The 433 MHz European ISM band is the best election for our environment conditions, but antennas are too large.

In the star configuration, nodes only transmit data, and the receiver part is not used. Data is sent three times per day. The time instant of the transmission is calculated using a random number generator seeded with the node's id. Payload of the data packets can be reconstructed if some packet are losts, so it is not critical to lose someones.

The base station has been build assembling evaluation kits and commercial modules. In this case, the microcontroller that coordinates all the system is an evaluation kit of the Silabs C8051F120 high performance microcontroller. This microcontroller can reach a peak performance of 100 MIPS.

The RF radiomodem section is based on a TI/Chipcon CC1101EMK868 evaluation kit. To this kit we attached a $\lambda/2$ inverted dipole antenna from Antennafactor.

The collected data is aggregated and sent using a GSM/GPRS Telit GM-862 to a remote place. Is possible to apply other techniques [1][3][4] to implement a complete wireless sensor network starting from this system.

A RS-232 serial link allows the configuration of the sink using any terminal emulation software.

The collected data lets to monitor continuously the EMC of the wood and the movement of the insects. As an example, Fig. 4 shows the laboratory controlled experiment with a colony of termite (*Reticulitermes lucifugus*) and the nodes inserted in blocs of pine wood.



Fig. 4. Experimentation

As a result of this experience, we know that there is an exploratory phase and, sometimes, an abrupt increment of termites traffic until the node is damped due to termites activity. This forced us to develop an extra algorithm to detect the initialization of the insect sensing part of the node and the consequent "silence" of the detector.

6 Conclusions

This work presents the real feasibility of an ultra-low energy and long life wireless nodes applied to environments such as heritage monitoring where it is necessary to guarantee long life and no operators intervention.

A major achievement of this node is life, having achieved a good balance between analysis environment, implementation of algorithms and their lives. The most notable features are:

- Lifetime of over 13 years with lithium thinly battery.
- Optical analysis performed every second for the entire lifetime.
- Three daily transmissions of optical analysis and EMC.
- ISM Band 868 MHz balance between antenna size, scope and capacity of wave propagation in the wood.

References

1. Capella JV, Perles A, Bonastre A, Serrano JJ. Historical Building Monitoring Using an Energy-Efficient Scalable Wireless Sensor Network Architecture. *Sensors*. 2011; 11(11):10074-10093.
2. F. Wang, J. Liu, L. Sun, "Ambient Data Collection with Wireless Sensor Networks", *EURASIP Journal on Wireless Communications and Networking*, 2010.
3. P. Mil, B. Jooris, L. Tytgat, R. Catteeuw, I. Moerman, P. Demeester, A. Kamerman, "Design and Implementation of a Generic Energy-Harvesting Framework Applied to the Evaluation of a Large-Scale Electronic Shelf-Labeling Wireless Sensor Network", *EURASIP Journal on Wireless Communications and Networking*, 2010.
4. J.V. Capella, A. Bonastre, R. Ors, and S. Climent, "A New Energy Efficient, Scalable and Robust Architecture for Wireless Sensor Networks," in *Proc. of Int. Conference on New Technologies, Mobility and Security*. IEEE, Dec. 2009.

Long Distance Wireless System for Rural Environments using WIFI Technology and Radio Mobile

Carlos A. Pérez-Cerna, Miguel A. Rodríguez-Hernández

ITACA, Universitat Politècnica de València, Valencia, Spain
carpece2@teleco.upv.es, marodrig@upvnet.upv.es

Abstract. This paper presents the design of a WIFI long distance wireless system in a rural environment using Radio Mobile simulator. The communication network design is located on the eastern slope of the Andes at 3098 m.a.s.l. in the district of La Encañada, department of Cajamarca - Peru and it interconnect 14 towns in the district. The network covers a distance of approximately 63 km and will utilize the WIFI technology with 802.11n in 5 GHz ISM band, which will provide the towns of La Encañada with access to the Internet. The creation of two networks is required to accomplish this: one for transportation the other for distribution. The networks will be connected by point to point links through local and trunk repeaters. The entire system is designed using the simulation software Radio Mobile and the influence of several parameters is evaluated.

1 Introduction

More than half of the world's population in developing countries lives in isolated, rural environments characterized by a lack of access to information and communication infrastructure. The principal need in these areas, connectivity to voice and data networks, has been the foremost concern for agents of international multilateral development in the past few years, in many cases this is considered a basic service, and a critical part of human development and promotion. However, all plans to expand access to communication networks in isolated areas of developing countries usually don't get off the ground because of lack of appropriate, realistic, and sustainable technological solutions, largely due to the following reasons [1]:

- There is not only an absence of telecommunications infrastructure, but also in many cases a practically nonexistent or very low quality electrical infrastructure and lack of adequate access routes. The need to supply telecommunication systems with autonomous power supply to ensure continuous operation and durability increases the cost and difficulty of maintenance; furthermore the lack of access routes also hinder and raise the cost of both network deployment and maintenance.
- Qualified technical personnel necessary for the maintenance and operation of these technologies usually live in major cities, and it is expensive and difficult to transport them to the rural areas.

- The population of the rural areas is poor and scattered, so they cannot afford the infrastructure costs to install, maintain and operate the technologies. Nor are the governments of the developing countries in a position to subsidize the installation of rural communications networks and progress towards full coverage, due to not only their lack of resources, but also because the rural populations are not able to contribute financially.

Since 2001, one of the technologies that has been considered very seriously for long distance communications is the IEEE802.11, commonly known as WIFI; although this standard was not designed for large networks, their undoubted advantages of cost, use of license free frequencies and high bandwidth, have aroused the interest of several technological agents in developing countries. Even in urban areas of many countries there has been expansion of the distribution of WIFI Internet access to include more outdoor coverage.

Moreover, the huge success of WIFI in all areas has led to many products on the market, almost all energy efficient, low priced and great flexibility of use, especially in combination with open software developments. Regarding the use of frequencies in cases where there is a legal vacuum, most states adopt the Federal Communications Commission restrictions on the use of the ISM 2.4 GHz and 5.8 GHz bands used by this technology.

2 Technical description. Methods and Materials

The communication network design is located on the eastern slope of the Andes at 3098 m in the district of La Encañada, department of Cajamarca - Peru. The network will cover a distance of approximately 63 km and will utilize the WIFI technology with 802.11n in 5 GHz ISM band, which will provide the towns of La Encañada with access to the Internet.

The creation of two networks is required to accomplish this: one for distribution the other for transportation. The distribution network are the links for the connection between the final clients and the local repeater while the transportation network is responsible for that network reaches a certain geographic extent.

The networks will be connected by point to point links through local and trunk repeaters. The entire system is designed using the simulation software Radio Mobile. There are fourteen beneficiary towns that are part of the district that make up the telecommunications network.

2.1 Radio Mobile Simulation software

Radio Mobile is a free radio propagation simulation program developed by Roger Coudé to predict the behavior of wireless systems, simulate radio links and depict a representation of the area of coverage of a radio network, among other functions. The software works in the frequency range between 20 MHz and 40 GHz and path lengths between 1 and 2000 km., is based on the propagation model ITM (Irregular Terrain

Model) or Longley-Rice model of the Institute for Science U.S. telecommunications [2].

The model is a tropospheric prediction method for radio transmission over irregular terrain through long and medium range links, takes into account the loss of free space, which is produced by the degradation of the signal with distance, as well as diffraction losses caused by the type of obstacles that lie in the path.

Radio Mobile uses terrain elevation data that can be downloaded free from the Internet to create virtual maps of the area of interest, stereoscopic views, 3-D views and animations of flight. Elevation data can be obtained from various sources, including the NASA Shuttle Radar Terrain Mapping Mission (SRTM) project that provides elevation data with an accuracy of 3 arc sec (100 m).

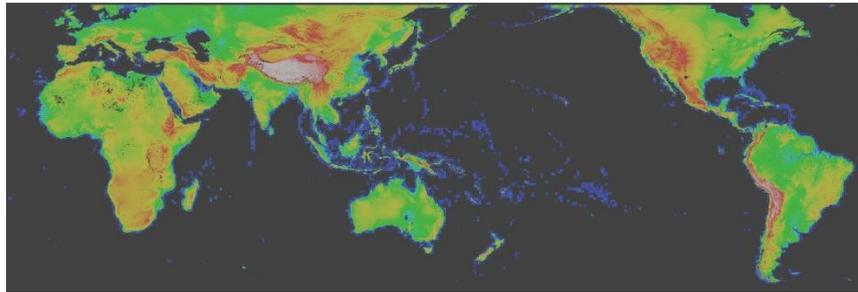


Fig.1. World Map Mobile created with Radio Mobile using SRTM elevation maps

It can create an evaluation whether a radio link between two remote stations is possible in a given frequency range, taking into consideration different geographical, climatic and statistical factors.

2.2 Equipment to be used

The simulation software Radio Mobile requires certain parameters to operate, these being the transmitter power, receiver sensitivity, antenna gain, among others that are acquired from the technical specifications of the equipment to be used to design the telecommunications network. The following equipment is considered for the design of the network:

- Wireless Routers: wireless router of the brand and model Mikrotik [3] RB433AH.
- Wireless Cards: wireless cards miniPCI mark Mikrotik [4] and model R52Hn that is compatible with the type of wireless router chosen.
- Antennas: two types of antennas of the brand Hyperlink of band 5 GHz, one of 22 dBi gain HG5822EG [5] model for the link between the stations customers and local repeaters and other antenna of 27 dBi gain HG5827EG [6] model for the link between local repeaters and trunks repeaters.

2.3 Clients stations

The clients stations are the end nodes that are located in different locations from where you will be able to access the Internet service (see Table 1).

They consist of point to point wireless links either with a local repeater station or with one other client, the links that are part of the client stations have a maximum distance of 2.8 km. in the case of the Pedregal resort and a minimum of 0.4 km for the town of Progreso La Toma, with an average distance of 1.35 km.

Clients stations	Altitude (m.a.s.l.)	Geographic coordinates	
		Latitude	Longitude
Chim Chim	3510	07°00'58.68"S	78°15'54.86"O
Chancas	3611	06°59'59.01"S	78°20'33.65"O
El Punre	3419	06°59'44.93"S	78°17'01.56"O
Michiquillay Alto	3453	07°02'31.30"S	78°19'57.65"O
Michiquillay Bajo	3437	07°02'34.68"S	78°20'01.59"O
Pampa Grande	3582	07°02'45.65"S	78°18'38.20"O
Pedregal	3508	07°00'54.02"S	78°21'05.01"O
Progreso La Toma	3582	07°03'44.22"S	78°17'10.43"O
Quinuamayo Alto	3544	07°03'46.90"S	78°17'48.91"O
Quinuamayo Bajo	3505	07°03'37.30"S	78°18'06.35"O
Quinuayoc	3594	07°01'44.29"S	78°20'15.63"O
Rodacocha	3513	07°01'40.00"S	78°20'27.57"O
Sogorón Alto	3584	07°02'32.81"S	78°21'27.86"O
Usnio	3426	07°04'15.35"S	78°19'11.60"O

Table 1: Client stations which form the telecommunications network.

2.4 Local and trunk repeaters

The local repeater stations are connected clients to access the internet service, and they connect to the trunks repeaters or locals via a wireless link point to point. All local repeaters and some trunks form the distribution network; repeater trunks are critical wireless links of the telecommunications network so that if a node breaks down internet service will drop out entirely. In designing the network there are a total of six local repeaters with a maximum distance of 7.3 km., a minimum of 2.4 km. and an average distance of 3.8 km.

The transportation network will be comprised of nine trunk repeaters and locals with a maximum distance of 9.03 km., a minimum of 4 km. and an average distance of 6.84 km.

3 Results and Evaluation

3.1 Network telecommunications design using Radio Mobile software

To calculate radio links there are multiple programs [7], in this case Radio Mobile in version 11.4.4 allows for calculating the links real data obtained from sheets of the shortlisted teams such as transmission power and sensitivity of the radios, and the radiation patterns and antenna gains. Thus the results do not only indicate the possibility of a link, but also with which teams would be this possible.

As can be seen below the results of designed radio links are displayed by set values as azimuth, elevation angle, distance, etc. But the parameters that interest us are really the received signal level (referred to Rx relative on Mobile Radio) which must be between 20 and 26 dB and that the clearance is greater than 60% of the radius of the first Fresnel zone (greater than 0.6F1).

In figure 2 you can verify that it meets the criteria set out above for a particular link.

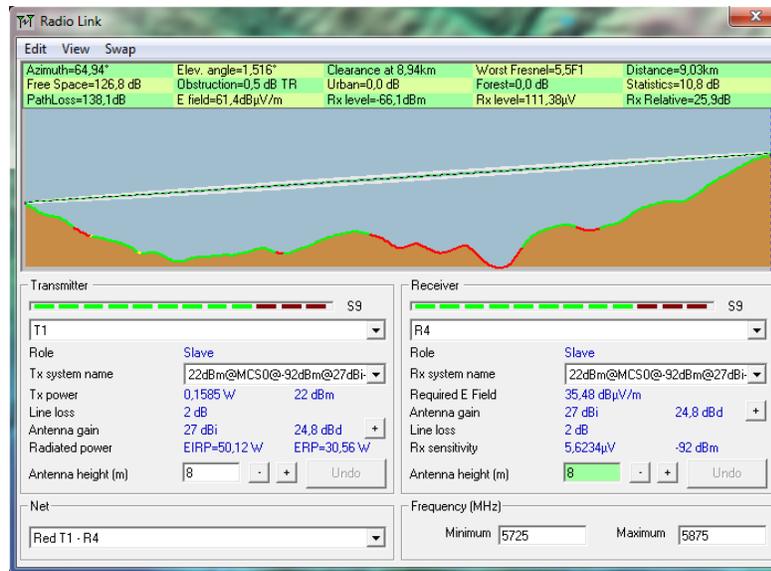


Fig.2. Radio link between two trunk repeaters T1 and R4

Eventually, once all of the calculation procedures listed above are completed, all that remains is to show the final design of the telecommunications network made with Radio Mobile, which can be seen in the following figures:



Fig.3. Final network design seen through Radio Mobile software

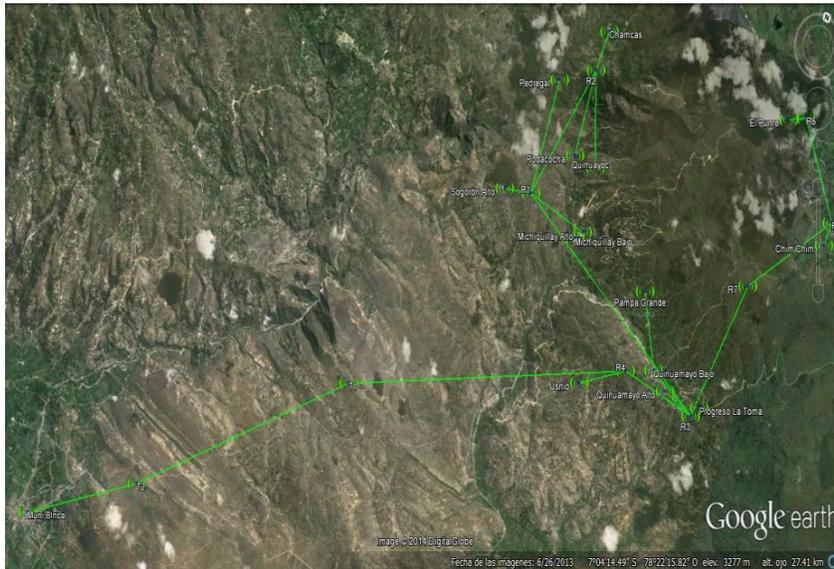


Fig.4. Final network design as seen from Google Earth

3.2 Evaluation of the network

The influence of several parameters in the final network performance has been evaluated. In concrete is evaluated the influence of the sensitivity versus antenna gain.

First, the sensitivity and transmission power of the wireless card used in the final clients systems and the local and trunks repeaters are varied, but antenna gain is constant. Below are the results that have been obtained from the evaluation of the radio link presented in Figure 2:

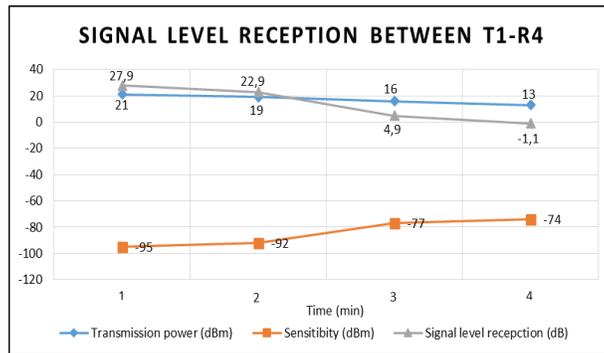


Fig.5. Signal level reception between two trunk repeaters T1 and R4

As seen in the previous figure, if the transmitting power and sensitivity are varied, the signal level at the reception will also be affected. Figure 6 shows the variation of the communication networks links when the sensitivity is changed to -74 dBm and the transmitting power 13 dBm.

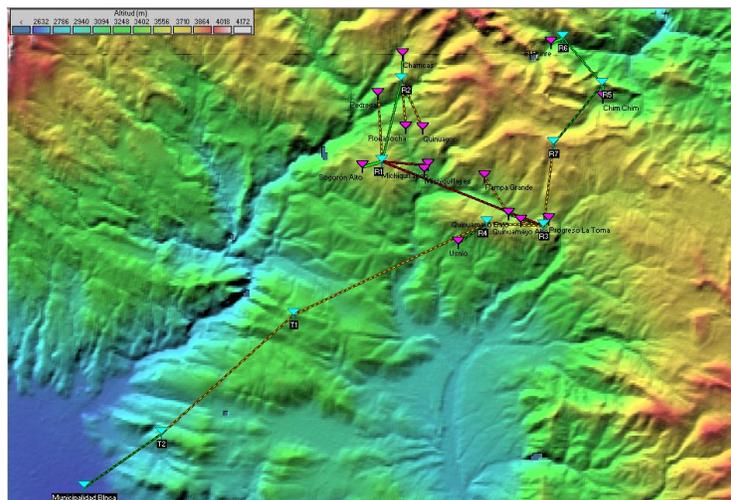


Fig.6. Network design with sensitivity set to -74 dBm and the transmitting power 13 dBm

In the second experiment the gain of the antenna used in the final clients system and the local and trunk repeaters changes, but the sensitivity and transmission power of the wireless card remains constant. Below are the results that have been obtained from the evaluation of the radio link presented in Figure 2:

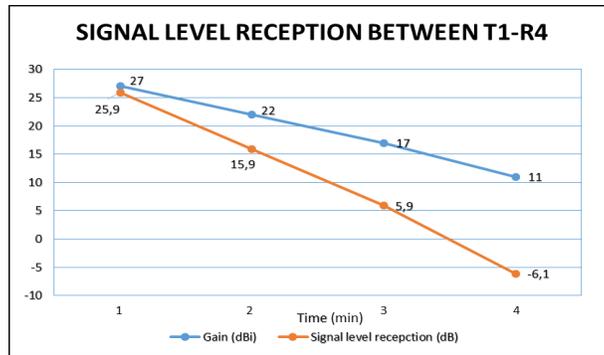


Fig.7. Received signal level between two trunk repeaters T1 and R4

Figure 8 shows the variations of the communications network links by modifying the antenna gain to 11 dBi. As can be seen when varying the gain of the antennas, the network links have a lower signal level as compared to when varying reception sensitivity and transmission power (see Figure 5). Furthermore it should be noted that the red lines represent no connection between various points of the network.

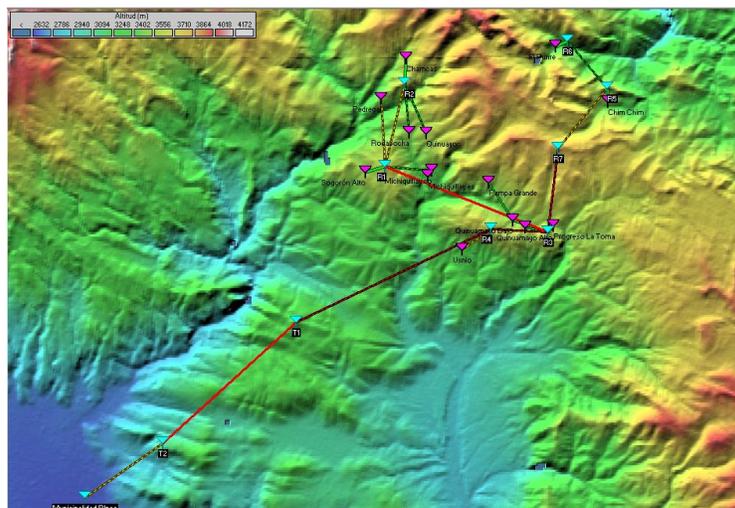


Fig.8. Variation of networking with antenna gain of 11 dBi

5 Conclusions

WIFI In all areas has led to many products on the market, almost all energy efficient, low priced and great flexibility of use, especially in combination with open software developments. WIFI standard was not designed for large networks, their undoubted advantages of cost, use of license free frequencies and high bandwidth, have aroused the interest of several technological agents in developing and undeveloped countries. Even in urban areas of many countries there has been expansion of the distribution of WIFI Internet access to include more outdoor coverage.

With the network communications designed with Radio Mobile, we be able to give Internet access to fourteen communities which are located all along Andes and it may extended to other locations according to the needs of the district's population. Evaluating the communications network it can be noted that through varying the gain of the antennas, the network links have lower signal level in reception than by varying sensitivity and transmission power of the wireless card.

References

1. G. Araujo et al.: Redes Inalámbricas para Zonas Rurales, 2011, pp.11-12
2. Radio Communication Group of the Higher Technical School of Telecommunications Engineering – UPM: Tutorial de Radio Mobile, 2007.
3. R. Coudé.: Radio Mobile Software [on line]. Available in: <http://www.cplus.org/rmw/english1.html> (2014, January 31).
4. RouterBOARD 433AH” [on line]. Available in: <http://i.mt.lv/routerboard/files/rb433ah.pdf>
5. RouterBOARD R52Hn” [on line]. Available in: <http://i.mt.lv/routerboard/files/R52Hn.pdf>
6. HyperLink Wireless 5.8 GHz 802.11a/n Wireless LAN System, 22 dBi Model: HG5822EG [on line]. Available in: http://www.l-com.com/multimedia/datasheets/DS_HG5822EG.PDF
7. HyperLink Wireless 5.8 GHz 802.11a/n Wireless LAN System, 27 dBi Model: HG5827EG [on line]. Available in: http://www.l-com.com/multimedia/datasheets/DS_HG5827EG.PDF
8. Mobius Consulting. Finding the Best Fit - WiFi versus WiMAX [on line]. Available in: http://www.mobiusconsulting.com/papers/WiFi_WiMax_white_paper.pdf (2008, February).

A Visibilimeter Based on a Universal Frequency-to-Digital Converter

R. Barrales-Guadarrama¹, A. Mocholí-Salcedo², M.E. Rodríguez-Rodríguez¹, V.R. Barrales-Guadarrama¹, E.R. Vázquez-Cerón¹

¹Área de Sensores y Procesamiento de Señales, Departamento de Electrónica, Universidad Autónoma Metropolitana—Azcapotzalco, México, D. F., México.

²Grupo de Sistemas de Control de Tráfico, Instituto ITACA, Universitat Politècnica de Valencia, Valencia, España.

Abstract. In this paper, we present a forward-scatter visibilimeter that uses a quasi-digital photodetector and an universal frequency-to-digital converter instead of a conventional analog-to-digital converter as data acquisition system. This feature has allowed the design of a low cost, robust and simple sensor-to-microcontroller interface as demanded by Intelligent Transportation Systems (ITS) applications. An optical system to limit light interference is proposed. The visibilimeter was calibrated from a self-calibrated transmissometer using the same frequency-to-digital technique. This new instrument is capable of 41.24 to 662.5 m visibility range detection and to transmit the information wirelessly to a 100 m distance.

Index Terms— Visibility, Meteorological Optical Range Frequency, frequency-to-digital conversion, intelligent transportation systems.

1 Introduction

Car accidents caused by fog and extreme weather are the worst motorway accidents caused by meteorological conditions. Fog formation occurs when low temperatures over the road surface favour continuous water vapour condensation. Accidents arise when fog density causes a remarkable visibility reduction (less than 40 m) [1]. The severity of the accidents has driven some government departments to implement automatic prevention systems in roads where fog is a traditional problem. For example, in 2009, the California Department of Transportation and the California Highway Patrol started a 'Fog Pilot' project, an awareness fog and reduced-speed warning system that will help protect motorists along a 20 km stretch of freeway on State Route 99 in California's Central Valley. This particular stretch of freeway is notorious for dangerous fog conditions and was the site of a deadly 86-car collision in November 2007. Among other technologies, the 'Fog Pilot' includes visibility sensors. The warnings are communicated to drivers, in under 30 seconds, via the large changeable message signs, which are placed every half mile [2]. Weather monitoring instruments used in ITS applications need to be small, easy to put in operation on multiple measure

sites, which implies a low cost. The design and construction of an ITS visibility sensor must fulfil specific goals according to these considerations.

In this paper, we present a forward-scatter visibility sensor designed to measure a short visibility range of 41.24 to 662.5 m. Besides, a frequency-to-digital conversion technique was introduced in order to make a simplified sensor-to-microcontroller interface and cut costs. These two main features adapt the sensor to an ITS environment.

2 Operating Principles

2.1 The Forward-Scatter Measurement Principle

Several factors make visibility difficult to measure. Weather, sun angle, light intensity, dark adaptation, availability of appropriate visibility targets and individual physical abilities are all factors impacting observers in perceiving conditions in the atmosphere [3]. A standard was adopted to reduce visibility measurement difficulties. The standard quantity related to visibility is the Meteorological Optical Range (MOR). MOR is defined by the World Meteorological Organization (WMO) as: “The length of path through the atmosphere required to reduce the luminous flux in a collimated beam from an incandescent lamp, at a colour temperature of 2700 K, to 0.05 of its original value, the luminous flux being evaluated by means of the photopic luminosity function of the International Commission on Illumination” [4]. When fog is present, visibility sensors determine MOR by measuring the local extinction coefficient, a parameter proportional to the reduction of the luminous flux, which is assumed to be constant around the sensor. MOR is an equivalent observer’s visibility index. In the mid-to-late 70’s, researches began investigating the principle of “forward-scatter” for the measurement of the extinction coefficient, hence the visibility. Few papers report technological results [5], [6]. The visibility is determined by the extinction coefficient σ defined by:

$$\frac{\partial E}{\partial l} = -\sigma E. \quad (1)$$

where E is the intensity of the light at position l . Solving (1), the fraction of light transmitted at length l from the light source is given by:

$$E(l)/E(0) = e^{-\sigma l}. \quad (2)$$

The length at which the light intensity is reduced to 5% can then be found:

$$MOR = 2.996/\sigma. \quad (3)$$

Equation (3) is known too as the Koschmieder’s Law (1925). Visibility can be measured by using optical scatter instruments. Fig. 1 illustrates the basic principle.

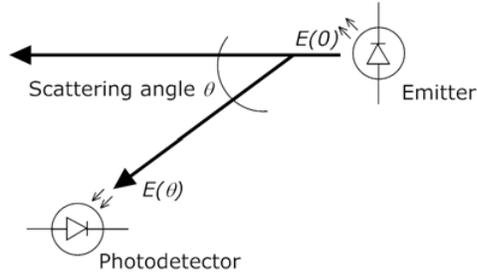


Fig. 1. The measurement principle in optical scatter instruments.

An emitter sends a beam light and the scattered light intensity within the θ angle is received by the detector. A high signal output in the detector corresponds to a high scattered intensity and a low visibility is measured. Different studies have been carried out to estimate visibility from the scattered intensity light [7]-[12]. When $\theta \leq 90^\circ$, the instrument is known as a forward-scattering instrument. Commercial forward-scattering visibility sensors in ITS environments use IR LEDs as light sources. Some studies have theoretically reported the relationship between the extinction coefficient σ and the scattered intensity $E(\theta)$. It has been demonstrated [5], [11] that the extinction coefficient depends on the particle size distribution modelling a fog sample and the scattering angle, but for a range $\theta \in [30^\circ, 40^\circ]$ [5] or $\theta \in [20^\circ, 50^\circ]$ [11], this dependence is minimal and σ is almost proportional to $E(\theta)$. A high lineal correlation between σ and $E(\theta)$ is observed for the angle $\theta = 35^\circ$ [5] or $\theta = 40^\circ$ [11]. Therefore, for a convenient θ range, the following relationship is valid:

$$\sigma = aE(\theta) \quad (4)$$

The PVM instrument [5] and the MIRA instrument [11] use this measurement principle. Both instruments use an IR LED as light transmitter. Visibility is then estimated applying (3).

2.2 Instrument Implementation

The proposed instrument is driven by the need to develop a visibility measurement system that relies on low cost commercially available hardware to simplify the sensor-to-control interface. Besides, the instrument is specified to be used for ITS applications which demands wireless communication capabilities for easy installation. It is important to note that hardware simplification was done by means of a universal frequency-to-digital converter and this approach will lead the design trade-offs in the different instrument stages. Fig. 2 shows the system block diagram.



Fig. 2. Visibilimeter block diagram.

The light source of the visibility sensor is a GaAs IR LED with a total 40 mW radiant flux, emitting light at a 950 nm peak wavelength. As the distance from the lens tip IR LED to the detector is not specified in the radiation pattern curve, a 25 mm diameter plano-convex lens was added as a collimated light lens in order to restrict the radiated fog sample. In order to preserve a low junction temperature, but a high output intensity, the forward current was adjusted to be smaller than the maximum safe continuous value specified in the datasheet (100 mA). Good stable polarization is obtained by means of a low current adjustable positive regulator. As the sensor-to-microcontroller interface is a frequency-to-digital converter, a frequency output sensor should be used. We chose a light-to-frequency converter combining a silicon photodiode and a current-to-frequency converter on single monolithic CMOS integrated circuit. The irradiance range detected by the sensor is 0.001 to 1000 $\mu\text{W}/\text{cm}^2$ with a 940 nm peak response. This detector is manufactured with an integral visible-light cut-off filter and lens. Fig. 3 shows the final optical system design.

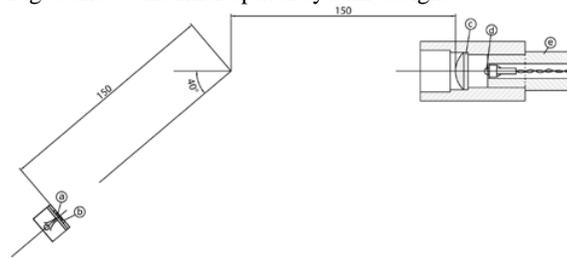


Fig. 3. Mechanical drawing of the final optical sensor design. a. Quasi-digital sensor; b. Protective cover; c. 25 mm plano-convex lens; d. IR LED; e. Adjustable lens mount. Dimensions are in mm.

The quasi-digital photodetector used in this application generates a 50% duty cycle pulse train with frequency directly proportional to light intensity or irradiance. Frequency, as an informative parameter, has many advantages: high noise immunity, high output signal power, wide dynamic range, high accuracy of frequency standards, simplicity of signal switching, interfacing, integration and coding. Therefore the dynamic range of the signal is not limited by supply voltage and noise, as it occurs with systems using analog-to-digital converters (ADCs). Frequency-to-digital converters are capable of high accuracies up to 0.001%. In traditional measuring systems, ADC error is commensurable with the sensor's error [13], [14]. Being a pulse data, the signals of several sensors may be easily multiplexed into one microcontroller. No output standardization for the ADC is necessary as in the case of analog sensors.

Different methods have been designed [13], in order to implement the frequency-to-digital conversion. Microcontrollers offer natural means for implementing such

methods, but these devices require the use of program-oriented conversion methods. These introduce additional error components due to the so-called program-dependent or software related effects, for example, the error due to the delay of reaction to an interruption [15] and the error or shift in time of the response for interruption [16]. All these methods are unable to provide accurate, fast and wide range frequency measurements at the same time when they are programmed into a microcontroller. Actually, the methods of the dependent count are the best methods to implement the frequency-to-digital conversion. These methods combine in themselves the advantages of the classical methods as well as the methods ensuring constant relative error in a broad frequency range and high speed [17]. Taking into account the very wide output frequency range (0.001 kHz to 1000 kHz) of the light-to-frequency converter used in this project, the UFDC-1 [18] universal frequency-to-digital converter based on the methods of dependent count was used to implement the sensor-to-microcontroller interface. Minimum frequency accuracy (1%) was programmed on the UFDC-1 and decimal numbers omitted from the registers. SPI bus is the natural communication link between the UFDC-1 and the microcontroller.

The instrument is not intended to be used as part of a monitoring network in the near future. It is rather to be used as a monitoring instrument wirelessly linked to a large changeable message sign placed along the road. Besides, data monitoring, among other test operations, is made on a regular PC that comes equipped with a Bluetooth wireless link. Because of these considerations and relative integration simplicity, a OEM Class 1 v2.0 + EDR Bluetooth-serial module with a standard 100 m wireless transmit distance, was added to enable wireless communication capability. Fig. 4 shows the block diagram of the main electronic system that allows visibility estimation. System integration meets the dimensions and cost requirements of an ITS device.

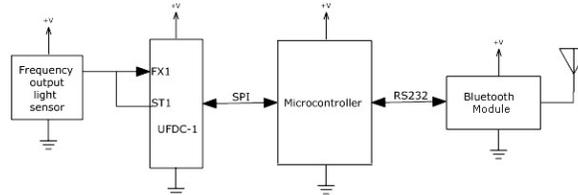


Fig. 4. Block diagram showing the different interconnected parts of the instrument.

2.3 Ambient Light Interference Supresión

Generally, light signals and ambient light interference are deterministically located in the frequency domain and their electrical representations are present at the output of a transimpedance amplifier in photodetection systems. Such signals are relatively easy to separate when they are processed with a first or second order filter. This is not our case. The characteristic transfer function of the quasi-digital photodetector is [19]:

$$f = kE + f_D \quad (5)$$

where f is the output frequency of the sensor, k is the sensor responsivity, E is the irradiance and f_D is the sensor dark frequency. Typically, f_D is very small (0.4 Hz) and can be neglected. Therefore, (5) is rewritten as:

$$f \approx kE \quad (6)$$

The output frequency vs. irradiance curve reported by the photodetector manufacturer [19] is obtained for a particular wavelength $\lambda_p = 940$ nm. It is clear that, when the photodetector is exposed to ambient light, it will generate a particular frequency signal proportional to the irradiance magnitude and the wavelength magnitude present at the moment:

$$\begin{aligned} f_{\lambda_0} &= k_0 E \\ f_{\lambda_1} &= k_1 E \\ f_{\lambda_2} &= k_2 E \\ &\bullet \\ &\bullet \\ &\bullet \\ f_{940} &= k_{940} E \\ &\bullet \\ &\bullet \\ f_{950} &= k_{950} E_S \\ &\bullet \\ &\bullet \end{aligned} \quad (7)$$

The irradiance signal of interest is E_S which is produced by the IR LED at 950 nm. This irradiance must be filtered from the light spectrum received by the photodetector, in order to be further processed by the instrument and obtain the visibility magnitude. The necessary characterization of the optical system in Fig. 3 has been reported elsewhere [20]. The important results obtained from this characterization are: a) the sensor's response to incident irradiances in the interval 800 nm to 1100 nm, Fig. 5, that confirms the sensor's output frequency dependence upon the input wavelength; b) the experimental peak response of the sensor, Fig. 6, located at 920 nm and the normalized response at 950 nm is 0.9; c) the sensor's response when it is provided with a band pass interference filter (BIF) with a central wavelength at 950 nm, and a bandwidth of 10 nm [21], Fig. 7. In the practice, the study in Fig. 7 demonstrates that the only irradiance detected is E_S at 950 nm from the IR LED. Once the BIF is added to the optical system of the visibility sensor, the instrument characterization may be carried on.

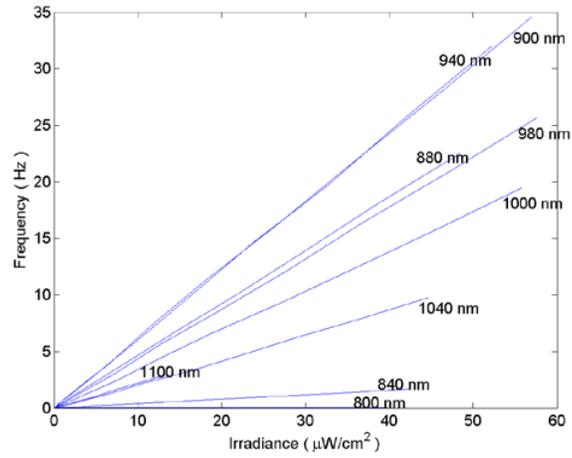


Fig. 5. Sensor's output frequency dependence upon the input wavelength.

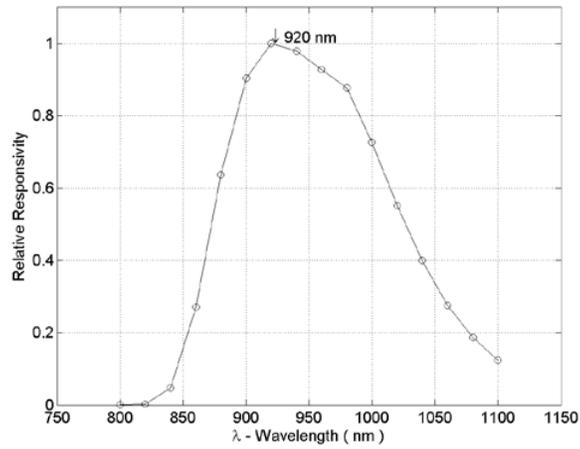


Fig. 6. Sensor's relative responsivity dependence upon the input wavelength.

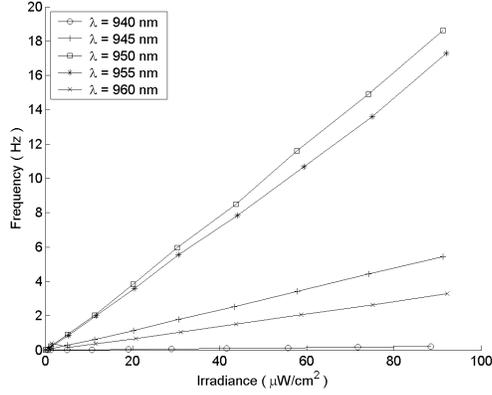


Fig. 7. Response of the sensor provided with the band pass interference filter with central wavelength at 950 nm.

3 Experimental Setups

3.1 Frequency Characteristic

Output frequency response is one of the key characteristics of the instrument in order to obtain reliable visibility measurements with enough accuracy. Therefore, an experimental setup allowing the analysis of the relation $f_{OUT} = g(f_0)$, where f_{OUT} is the frequency measured by the UFDC-1 and f_0 is an input reference frequency, was designed. The experimental setup used to assess the output frequency response of the instrument is shown on Fig. 8.

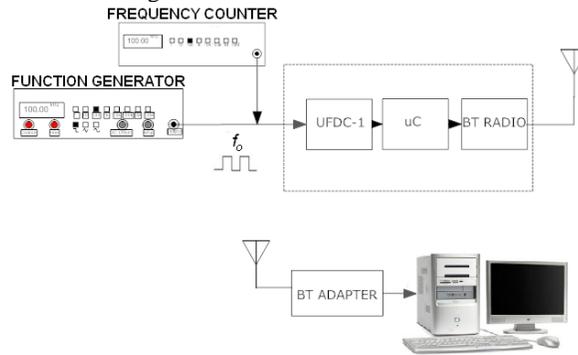


Fig. 8. Experimental setup to assess the front-end output frequency linearity. A TTL square wave of frequency f_0 was generated by a HP 33120A function generator as test signal for the instrument. The input reference frequency was measured with

an ESCORT EGC-3230 frequency counter. The output frequency, measured by the UFDC-1, was wirelessly transmitted to a computer and displayed in a virtual terminal.

3.2 Experimental Photodetector Responsivity

Experimental photodetector responsivity value must be taken into account in the signal processing program that allows the ambient light interference suppression. As shown in Fig. 5, photodetector responsivity depends on the wavelength used and must be experimentally obtained since we work with a 950 nm wavelength and not the 920 nm peak detector response. Fig. 9 shows the setup to obtain this parameter.

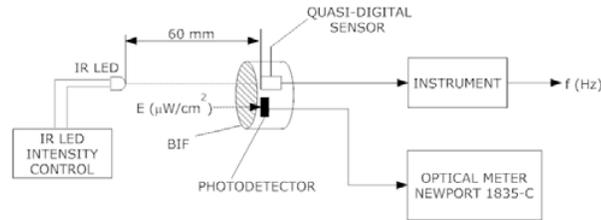


Fig. 9. Experimental setup to measure the photodetector responsivity at $\lambda = 950$ nm.

The IR LED intensity control provides a variable beam whose irradiance E is measured by the Newport 1835-C optical meter. The proportional frequency associated to this value and generated by the quasi-digital sensor is directly measured by the UFDC-1 in the instrument. The distance between the IR LED and the optical meter photodetector was chosen in a way to obtain high irradiance values, but to prevent photodetector saturation. The experiment was conducted under dark room conditions with the BIF as ambient light control measure.

4 Experimental Results

4.1 Instrument Frequency Linearity

An ascending and descending data exploration, using the experimental setup of Fig. 8, was done in order to apply a linear regression for the relation $f_{OUT} = g(f_0)$. Fig. 10 shows the graphical result. The slope is $m = 1.0041$ Hz/Hz and the y intercept is $b = 4.7123$ Hz. These parameters were computed within the range [45 Hz, 800300 Hz], where the instrument sent valid lectures. The line defined by the equation $f_{OUT} = 1.0041f_0 + 4.7123$ Hz is the best straight line from which the maximum non-linearity error in % FSS is $(2608.6245/(800300-45)) \times 100 \approx 0.326\%$.

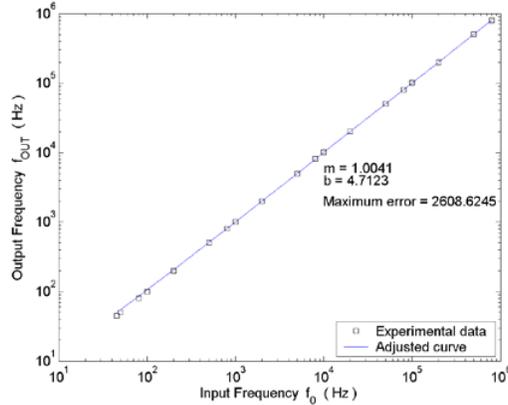


Fig. 10. Frequency response of the instrument.

4.2 Photodetector Responsivity Measurement

A lineal regression upon the data obtained using the experimental setup in Fig. 9 was performed in order to estimate the practical quasi-digital photodetector responsivity k . Fig. 11 shows the experimental data distribution and the related adjusted curve when the experiment is conducted under lab conditions. In this case: $k_1 = 0.036$ kHz/ μ W/cm². Fig. 11 also shows the issues when the same experimental setup is used to obtain the quasi-digital sensor responsivity under field conditions (sunlight at 12:00 pm, sensor oriented to have the sun behind and the IR beam line oriented in the North line). In this case: $k_2 = 0.04$ kHz/ μ W/cm². Data analysis shows that relative error between the straight line slopes k_1 and k_2 is 10%. Responsivity $k = k_2$ was chosen to be 0.04 kHz/ μ W/cm² in the instrument.

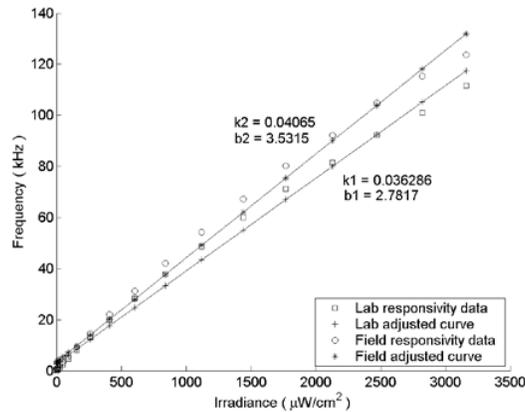


Fig. 11. First order adjusted curve for the measurement of the sensor responsivity under laboratory conditions and field conditions.

5 Calibration

As the BIF keeps a low relative error between the responsivity measured in laboratory and the responsivity measured in the field and non intense sunlight is present when ground fog appears because this type of fog is caused by the radiation cooling of the Earth's surface (radiation fog) which is caused after sunset, when the Earth receives no heat from the Sun [22], we decided to perform the calibration of the instrument under laboratory conditions. If the ambient light is sufficiently reduced, the instrument can make an estimation of the extinction coefficient from (4):

$$\sigma = aE(40^\circ) = \frac{a}{k_2} f. \quad (8)$$

The Koschmieder's Law allows the instrument to estimate visibility from (3):

$$MOR = \frac{2.996}{\sigma} = \frac{2.996}{(a/k_2)f} \quad (9)$$

k_2 is the sensor's responsivity. a is a constant that depends upon light beam intensity, the scattering geometry and the detector sensitivity. The only practical method of determining the constant a is to compare the forward-scatter sensor's measurements to those from a transmissometer [23] or a standard visibilimeter. Due to the impossibility of getting a visibilimeter or a transmissometer as standard instruments to proceed to a direct calibration process, an indirect calibration was conducted. Fig. 12 shows the calibration setup of our own. A second quasi-digital sensor was added to the optical system of the visibilimeter in the $\theta = 0^\circ$ direction, separated by a distance $l = 300$ mm from the light source emitting a constant irradiance E_0 in order to form a small transmissometer.

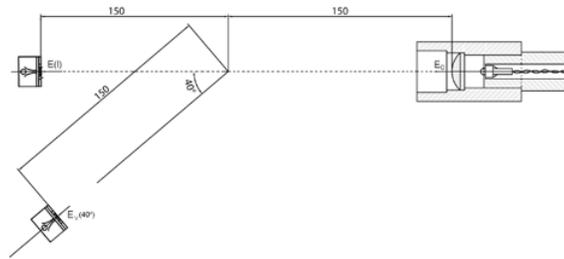


Fig. 12. Mechanical drawing of the calibration setup.

This setup was then placed in a Weiss Technik salt spray test chamber SC450 programmed to run a salt fog test (DIN ES ISO 9227) and frequency registers from the two sensors were taken once fog saturation was attained and the chamber deactivated. The chamber was isolated from ambient light. The average value of the frequency f_0 associated with the light source and the frequency generated by the transmissometer

sensor f_T are arranged in the Bouguer-Lambert Law to compute the extinction coefficient within the chamber:

$$\sigma_T = \frac{\ln\left[\frac{E_0}{E_l}\right]}{l} = \frac{\ln\left[\frac{f_0/k_2}{f_T/k_2}\right]}{l} = \frac{\ln[f_0/f_T]}{l}. \quad (10)$$

Assuming a homogeneous atmosphere, the visibilimeter extinction coefficient will be the same as for the small transmissometer. From (8):

$$\sigma_T = \sigma_V = aE_V(40^\circ) = a \frac{f_V(40^\circ)}{k_2}. \quad (11)$$

Frequency f_V measurements from the visibilimeter sensor generate indirect measurements of the irradiance E_V received by the instrument at $\theta = 40^\circ$ and frequency f_T measurements from the small transmissometer generate indirect measurements of the extinction coefficient σ_T . To obtain an estimation of a , a first order regression was performed for these indirect registers. Fig. 13 shows the result for two register sets.

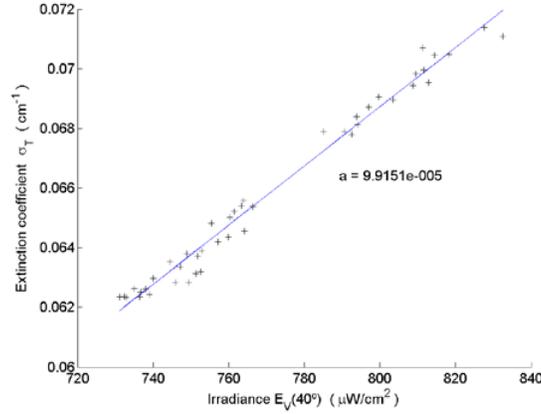


Fig. 13. a estimation by first order regression.

The spray chamber allows a visibility sensor frequency range register $f_V(40^\circ)$ register from 1854 to 29788 Hz. For this range, we have a corresponding extinction coefficient range σ_V from 0.0045 to 0.0726 cm^{-1} where this figure has been computed using the relation: $\sigma_V = aE_V(40^\circ) = a\left(\frac{f_V(40^\circ)}{k_2}\right) = 9.9151 \times 10^{-5} \left(\frac{f_V(40^\circ)}{40.65}\right)$ according to the calibration process. Applying the Koschmieder's Law (3) to this range the response of the visibility sensor may be estimated. This result is shown in Fig. 14.

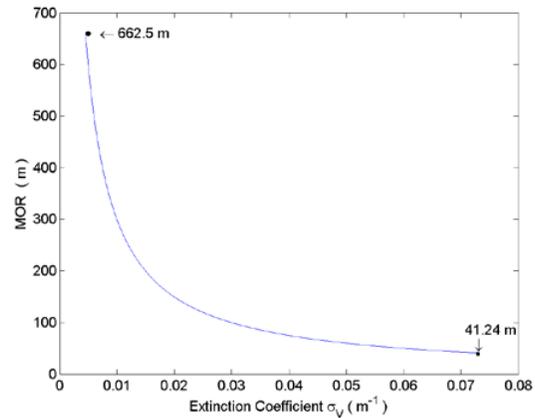


Fig. 14. Estimated MOR response of the forward-scatter visibility sensor based on universal frequency-to-digital converter.

Fig. 15 shows the visibilimeter when calibration was performed with the spray test chamber.



Fig. 15. Photograph of the visibilimeter during calibration procedure.

6 Conclusions and Future Work

Very good linearity is observed between the UFDC-1 input and output frequencies. The maximum non-linearity error in % FSS is 0.326%. Nevertheless, the claimed frequency measurement range, 0.001 kHz to 1000 kHz, is not entirely swept by the instrument. The lower frequency range limit is 45 Hz and this could be the result of an involuntary zero introduction generated by the lag low pass filter connected at the UFDC frequency input, recommended by the manufacturer. Fortunately, this condition does not have an impact on the visibilimeter response because the MOR lower limit

detected occurs at 1854 Hz. Accurate visibility measurements in ITS applications are not necessary because warning systems alert drivers before they reach a fog area and they would have sufficient time to reduce speed. Therefore, minimum frequency accuracy (1%) was programmed on the UFDC-1. A positive point with regard to the calibration procedure may be mentioned: the use of a small transmissometer to estimate extinction coefficient is valid because the transmissometer is self calibrating, according to the Bouguer-Lambert Law. Main instrument characteristics are: a estimated visibility range of 41.24 to 662.5 m, good UFDC linearity, reduced size, telemetric operation within 100 m and low cost. Although uncertainty is not meaningful, a comparison with a standard visibility sensor must be done to establish real uncertainty. Though the first measurement instruments using UFDCs are developed to monitor only one measurand with no inherent interferences, studies must be carried on to understand the differences of using an UFDC when inherent noise or other kind of interferences are present in the signal variation.

Actually, the design of the visibility reported in this paper has been submitted as an invention before the Instituto Mexicano de la Propiedad Industrial, October 24 2012, in order to obtain a patent grant.

7 Acknowledgements

Authors would like to thank the Universitat Politècnica de Valencia (Spain) and the Universidad Autónoma Metropolitana—Azcapotzalco (México) for their support provided in this work.

References

1. K. R. MacHutchon, A. Ryan, "Fog Detection and Warning, A Novel Approach to Sensor Location", 1999 *Proc. IEEE AFRICON*, Vol. 1, ISBN 0-7803-5546-6, pp. 43-50.
2. News Section, p. 8, (April-March 2009), *ITS International Electronic Magazine*, [Online]. Available: www.itsinternational.com.
3. J.T. Bradley, K. Kraus, T. Townshend, "Federal Citing Criteria for Automated Surface Weather Observations", *7th Symp. on Meteorological Observations and Instrumentation*, New Orleans, LA, 1991. pp. 207-210.
4. Organización Meteorológica Mundial, *Guía de Instrumentos y Métodos de Observación Meteorológicos*, Ginebra, 1990, Serie: OMM 8, ISBN: 9263300089.
5. J. V. Winstanley, M. J. Adams, "Point Visibility Meter: A Forward Scatter Instrument for the Measurement of Aerosol Extinction Coefficient", *Applied Optics*, Vol. 14, No. 9, September 1975.
6. J. L. Horner, "Analog Visibility Computer", *Appl. Opt.*, Vol. 15, 1976, pp. 999-1002.
7. S. Twomey, H. B. Howell, "The Relative Merit of White and Monochromatic Light for the Determination of Visibility by Backscattering Measurements", *Applied Optics*, Vol. 4, No. 4, April 1965, pp. 501-506.

8. Robert W. Fenn, "Correlation Between Atmospheric Backscattering and Meteorological Visual Range", *Applied Optics*, Vol. 5, No. 2, February 1966.
9. H. Vogt, "Visibility Measurement Using Backscattered Light", *Journal of the Atmospheric Sciences*, Vol. 25, 1968, ISSN: 0022-4928, pp. 912-918.
10. Robert A. Maffione, David R. Dana, "Instruments and Methods for Measuring the Backward-Scattering Coefficient of Ocean Waters", *Applied Optics*, Vol. 36, No. 24, 20 August 1997.
11. S. A. Tjugum, J. S. Vaagen, T. Jakobsen, B. Hamre, "Use of Optical Scatter Sensors for Measurement of Visibility", *J. Environ. Monit.*, 2005, Vol. 7, pp. 608-611.
12. Roberto Nebuloni, "Empirical Relationships between Extinction Coefficient and Visibility in Fog", *Applied Optics*, Vol. 44, No. 18, 20 June 2005, pp. 3795-3804.
13. N. V. Kirianaki, S. Y. Yurish, N. O. Shpak, V. P. Deynega, *Data Acquisition and Signal Processing for Smart Sensors*, John Wiley & Sons, Chichester, UK, 2002, ISBN 978-0-470-84317-8, pp. 89-126.
14. N. Hlupic, J. Butorac, M. Kresic, "Improved Frequency Measurement by Means of DMM and Verification of its Specifications", *IEEE Trans. Instrum. Meas.*, Vol. 54, No. 5, pp. 1957 – 1963, 2005.
15. M. Prokin, "[Double Buffered Wide-range Frequency Measurement Method for Digital Tachometers](#)", *IEEE Trans. Instrum. Meas.*, Vol. 40, No. 3, pp. 606 – 610, 1991.
16. S.Y. Yurish, F. Reverter, R. Pallàs Areny, "Measurement Error Analysis and Uncertainty Reduction for Period and Time Interval-to-digital Converters Based on Microcontrollers", *Meas. Sci. Technol.* Vol. 16, No. 8, 2005, pp. 1660-1666.
17. N.V. Kirianaki, S. Y. Yurish, N. O. Shpack, "Methods of Dependent Count for Frequency Measurements", *Measurement* 29, 2001, pp. 31-50.
18. (2004), Universal Frequency-to-Digital Converter (UFDC-1), Specification and Application Note, [Online], Available: http://www.sensorsportal.com/DOWNLOADS/UFDC_1.pdf.
19. (2005) Infrared Light-to-Frequency Converter (TSL245R), Datasheet, [Online], Available: <http://www.taosinc.com/productdetail.aspx?product=46>.
20. R. Barrales-Guadarrama, A. Mocholí-Salcedo, M. E. Rodríguez-Rodríguez, V. R. Barrales-Guadarrama, E. R. Vázquez Cerón, *Validación del método de medición de un visibilímetro provisto de un convertidor universal de frecuencia a código digital*, Reporte de investigación. Universidad Autónoma Metropolitana, México, 2013.
21. (2011) 700 – 999 nm Bandpass Interference Filters, [Online], Available: <http://www.edmundoptics.com/products/displayproduct.cfm?productid=3198&PageNum=6&Sort=displayOrder&Order=asc#products>
22. AeroGrapher's Mate, Module 05—Basic Meteorology, Chapter 5, Atmospheric Phenomena, Radiation Fog, [Online]. Available: http://www.tpub.com/content/aerographer/14312/css/14312_143.htm
23. D. C. Burnham, "Fog, Snow, and Rain Calibrations for Forward-scatter Visibility Sensors", Scientific and Engineering Solutions, Inc., Orleans, MA, 1993.

Levels of induced voltage between rectangular magnetic loops

J. H. Arroyo-Núñez^{1,2}, A. Mocholí-Salcedo¹, I. Rivas-Camero², A. Arroyo-Núñez¹, H. Jiménez Mimila².

¹ The Institute for the Applications of Advanced Information and Communication Technologies, Traffic Control Systems, Universidad Politécnica de Valencia, Edificio 8G Camino de Vera s/n 46022 Valencia, Spain
<http://www.itaca.upv.es/view.php/Investigacion/SCT>

² Universidad Politécnica de Tulancingo, División de Ingenierías, Calle Ingenierías No. 100, Col. Huapalcalco, Tulancingo de Bravo 43629, Hidalgo, México.
<http://www.upt.edu.mx>

Abstract. A comparison of induced voltage in magnetic rectangular loops by effect of the magnetic coupling between them is presented. This comparison is made between the theoretical values and experimental values, which are obtained by calculation and by direct measurement, respectively. These results will be used to design a system of short-range communication for the exchange of information between road infrastructure and vehicles.

1 Introduction

Magnetic loops are devices used for decades in vehicle detection [1]. Therefore it is easy to find it located in many paths used for this purpose. The magnetic loops have a great potential for to be used in other applications. As an example, there is the use of some loops for obtain the magnetic profile of the vehicles [2]. In this case the loop works like a sensor for detect the amount of metal that passes over it, allowing to identify the type of vehicle.

The short-range communication using magnetic induction is used in systems of identification by radio frequency known as RFID. The most common applications of these systems are oriented, to access control or product detection. This last application is widely used in the shopping centers to detect the output of unauthorized products.

Therefore, to design the appropriate electronic devices that allow short-range communication, using magnetic loops that are on the highways, it is necessary to know the voltage levels which can be induced. These voltage levels will be based on the size of the loop, number of turns, intensity of current, distance between loops and the frequency of the signal.

This paper presents the values obtained using a mathematical model that allows calculating the induced voltage levels and the results are compared with the experimental values, to determine the margin of error between the two measures.

These results will be used to design a system of short-range communication for the exchange of information between road infrastructure and vehicles.

2 Theoretical analysis

The work is based on Faraday's law which established [3]:

$$\text{FEM} = \oint \vec{E} \cdot d\vec{l} = -\frac{d\phi}{dt} = -\frac{d}{dt} \int_s \vec{B} \cdot d\vec{s} \quad (1)$$

EQ. (1) shows that it is possible to induce an electromotive force (EMF), in a conductor in presence of a time variable magnetic field. If the receiver is a rectangular loop located in the XY-plane, parallel with the loop transmitter, induced EMF will depend on area that cover the loop receiver, the magnetic field that crosses perpendicular (B_z), the frequency and the distance between both loops. These four aspects are determinants in the magnitude of the induced EMF. See Fig. 1.

The component of the magnetic field perpendicular to the plane of the loop is represented by B_z . This value depend on the spatial location in the coordinate plane (x , y , z), the size of the rectangular loop ($2a \times 2b$) and electric current (I). Figure 2 shows the location of the loop. Therefore, B_z is represented by the EQ. 2 [4] [5].

$$B_z = \frac{\mu_0 I}{4\pi} \left(\left[-\frac{(x+a)}{R_1(R_1+d_1)} + \frac{(x-a)}{R_2(R_2+d_1)} + \frac{(x+a)}{R_3(R_3+d_2)} - \frac{(x-a)}{R_4(R_4+d_2)} \right] - \left[\frac{(y+b)}{R_1(R_1+c_1)} - \frac{(y+b)}{R_2(R_2+c_2)} - \frac{(y-b)}{R_3(R_3+c_1)} + \frac{(y-b)}{R_4(R_4+c_2)} \right] \right) \quad (2)$$

Where:

$$\begin{aligned} R_1 &= \sqrt{(x+a)^2 + (y+b)^2 + z^2} \\ R_2 &= \sqrt{(x-a)^2 + (y+b)^2 + z^2} \\ R_3 &= \sqrt{(x+a)^2 + (y-b)^2 + z^2} \\ R_4 &= \sqrt{(x-a)^2 + (y-b)^2 + z^2} \\ c_1 &= -a - x \\ c_2 &= a - x \\ d_1 &= -b - y \\ d_2 &= b - y \end{aligned} \quad (3)$$

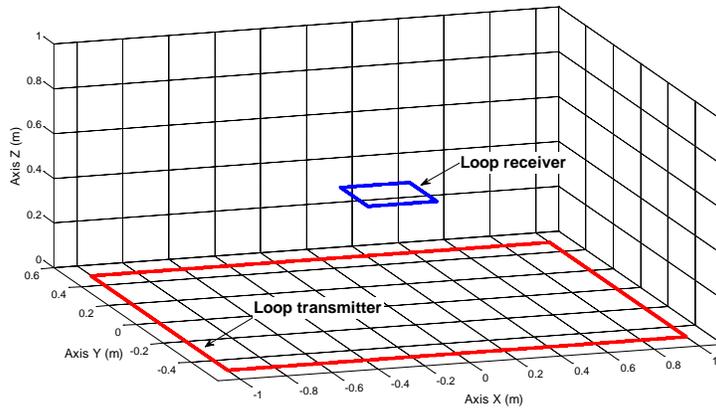


Fig. 1. Transmitter and receiver magnetic loop

The objective of this study is to determine the voltage that is induced in a receiver loop from the current flowing through a loop transmitter.

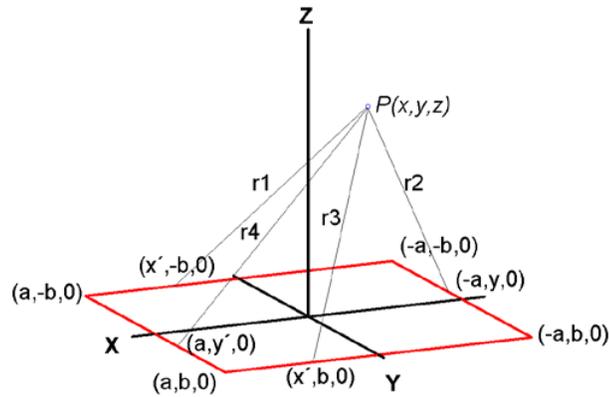


Fig. 2. Magnetic loop and some point P of analysis of the magnetic field

3 Simulation

To carry out the simulation of the voltage induced in the loop receiver, the following parameters for each of the loops were established.

transmitter loop :

dimensions of the loop: 2 X 1 meters

applied current: sinus wave with 100 mA rms at 100 KHz
 number of turns: 6
 receiver loop:
 dimensions of the loop: 0,30 X 0,20 meters
 number of turns: 5

The distance between both loops located in parallel (see Fig. 1) was 0,5 m. Estimates of induced voltages, are presented in the Table1. These values correspond when the loop transmitter is located on the XY-plane as shown in figure 2.

For the calculation of the voltage induced in the loop receiver at various points, was carried out a sweep along the axes X and Y each 0,25 m with a constant height, on the Z axis (0,5 m). In Fig. 3 it can see the points corresponding to the center of the loop receiver, where the calculations were made. The view displayed corresponds to the XY-plane. The simulation was realized using EQ. (1) implemented in MatLab by numerical integration.

Fig. 4 show in graphic form the simulation results of the voltage induced. In this graph it can see that the maximum voltage levels are in the center of the transmitter loop. This behavior is due to the way presenting magnetic field at this point [3].

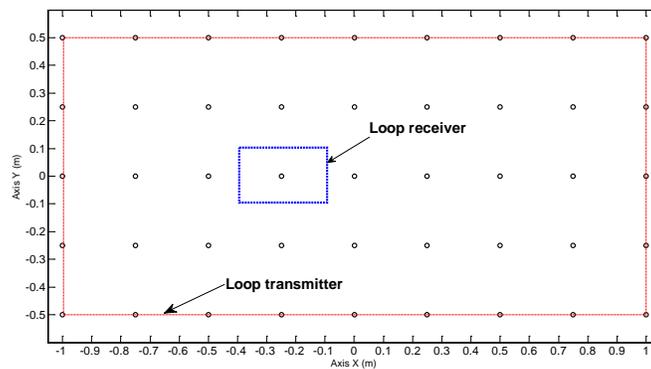


Fig. 3. Location of the loop receiver centers

Table 1. Theoretical calculated values of the voltage induced in the loop receiver (mV)

		Axis X (m)								
		-1	-0,75	-0,5	-0,25	0	0,25	0,5	0,75	1
Axis Y (m)	0,5	10,2288	19,9197	23,6445	24,1753	24,1434	24,1753	23,6445	19,9197	10,2288
	0,25	20,4763	37,251	43,7692	44,8628	44,8957	44,8628	43,7692	37,251	20,4763
	0	23,6862	42,6033	50,0458	51,3473	51,4062	51,3473	50,0458	42,6033	23,6862
	-0,25	20,4763	37,251	43,7692	44,8628	44,8957	44,8628	43,7692	37,251	20,4763
	-0,5	10,2288	19,9197	23,6445	24,1753	24,1434	24,1753	23,6445	19,9197	10,2288

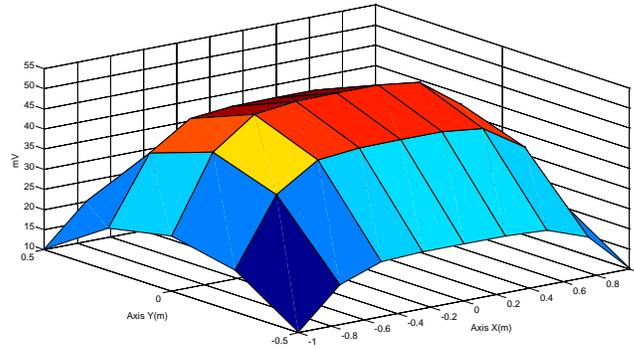


Fig. 4. Levels of voltage induced in the loop receiver

4 Experimentation

To compare the calculated results versus real data, the physical implementation of the loops was conducted according to the specifications of the simulation, see fig. 5.

Table 2 shows the values of the voltage induced in the loop receiver, these values were measured with an oscilloscope. Table 3 shows, in absolute value, the relative percentage difference between the experimental and theoretical values at each test point.

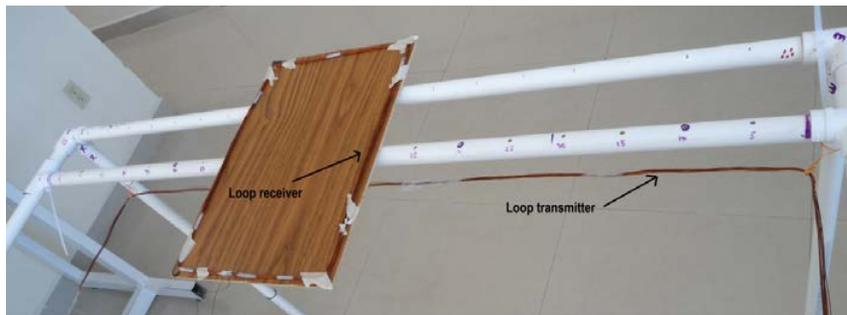


Fig. 5. Implementation of the transmitter and receiver loops

Table 2. Experimental values of the voltage induced in mV

		Axis X (m)								
		-1	-0,75	-0,5	-0,25	0	0,25	0,5	0,75	1
Axis Y (m)	0,5	11,3	21,2	25,4	26,3	26,3	26,2	25,6	20,9	11,4
	0,25	22,4	39,2	46,2	47,8	48,1	47,9	47,5	40,1	22,2
	0	25,2	45,2	52,5	52,1	52,5	52,3	52,1	45,1	25,2
	-0,25	22,2	39,5	46,3	47,2	47,9	47,5	47,1	38,9	22,3
	-0,5	11,2	21,3	25,1	26,1	26,2	26,1	25,4	21	11,3

Table 3. Relative percentage difference between the experimental and theoretical values

		Axis X (m)								
		-1	-0,75	-0,5	-0,25	0	0,25	0,5	0,75	1
Axis Y (m)	0,5	10,47%	6,43%	7,42%	8,79%	8,93%	8,38%	8,27%	4,92%	11,45%
	0,25	9,39%	5,23%	5,55%	6,55%	7,14%	6,77%	8,52%	7,65%	8,42%
	0	6,39%	6,10%	4,90%	1,47%	2,13%	1,86%	4,10%	5,86%	6,39%
	-0,25	8,42%	6,04%	5,78%	5,21%	6,69%	5,88%	7,61%	4,43%	8,91%
	-0,5	9,49%	6,93%	6,16%	7,96%	8,52%	7,96%	7,42%	5,42%	10,47%

5 Conclusions

From these results it is possible to affirm that: a) on average, the relative deviation between the experimental and theoretical values is less than 10% and b) The values that exceed this average are located at points where there are a discontinuity of the magnetic field generated by the loop transmitter, which corresponds at the edges of this [4].

With the results obtained, it is possible to predict the voltage induced between rectangular magnetic loops with a margin of error less than 10% in the center of the loop transmitter.

Tests carried out at different heights of the receiving loop have shown that if the transmitter and receiver distance between the loops is increased the margin of error can be reduced.

If the loops are placed at shorter distances may increase the margin of error. This is due to the magnetic field near of the loop transmitter presents steep slopes, so any minimal variation in the location of the loop receiver on the XY-plane gives rise to induced voltages very different.

Therefore, the work described in this document allow you to set parameters for the design of loops used for short-range communication. In the mathematical model it is possible to set the following parameters: the supply voltage, the distance of coverage,

the size of the loops, the number of turns and the frequency of the signal. With these parameters, you can meet the voltage level received, which adhere to the reality with a margin of error not exceeding 10%.

References

1. Traffic Detector Handbook: Third Edition. U.S. Department of Transportation Federal Highway Administration. 2006.
2. Arroyo, A., Mocholí, A. Sistemas Sensores Empleados en ITS. VII Congreso Español Sistemas Inteligentes de Transporte. Valencia, España. Septiembre 2007.
3. Johnk, Carl T. A. Engineering Electromagnetic Fields and Waves. Second Edition. John Wiley & Sons. 1988.
4. Arroyo Núñez, J.H., Mocholí Salcedo A. Communication Between Magnetic Loops. 16th World Road Meeting. Lisbon, Portugal. Mayo 2010.
5. Misakian, Martin. Equations for the Magnetic Field Produced by One or More Rectangular Loops of Wire in the Same Plane. Journal of Research of the National Institute of Standards and Technology. Volumen 105, Number 4, July-August 2000.

Acoustic Communication System for Underwater Wireless Sensor Networks

Underwater communication

Raúl Sáez, Antonio Sánchez, Ricardo Mercado, Àngel Perles, Pedro Yuste,
Juan Vicente Capella, Sara Blanc, Vicente Busquets, Juan José Serrano, and
Rafa Ors

Instituto ITACA, Universitat Politècnica de València, Camino de Vera s/n, 46022
Valencia, Spain;
e-mail: ucert@upv.es;
website: <http://ucert.upv.es>

Abstract. Underwater wireless sensor Networks (UWSN) provide a variety of attractive working fields such as aquaculture, offshore exploitation and underwater monitoring. These applications require spreading an important number of nodes which requires to develop simple, reliable and energy efficient modem architectures to reduce the economical cost, the developing time and the power consumption, still being efficient and robust.

This paper overviews research efforts of our research team conducted during the last year to develop an acoustic communication system for UWSN. To address the cost and power consumption limitations, the design and the implementation of a low cost and energy efficient underwater modem and a new acoustic-triggered wake-up system were researched. The result of combining both researches resulted in the development of ITACA-S1000 modem, a solution with a very low power consumption and with a very advanced remote wake-up system. Since this modem was the first coherent FSK solution presented to date, it was necessary to reach a new optimal multipath and Doppler-shift correction algorithm for coherent FSK modem architectures with resources and power limitations to archive an efficient and robust communication.

Keywords: underwater wireless sensor networks; acoustic modems; uwsn; underwater sensor networks; wake-up

1 Introduction

Underwater Wireless Sensor Network (UWSN) development is of current interest due to the potential variety of attractive working fields such as aquaculture, offshore exploitation, biological monitoring as well as water and seafloor pollution, seismic activity and ocean currents. Applications of these fields may require one or more of the following: short range underwater communications between collaborative devices; short/medium range communications between sea bottom

and surface; and longer range communications between the site and land. In these applications, it is often desirable to eliminate or reduce the number of wires and connectors to a minimum to reduce cost and maintenance and increase reliability. Practical solutions for these applications mainly use floating buoys that recollect data from a large number of submerged sensor nodes via wireless acoustic communication[1] and communicate them to a base station using ISM frequency bands.

Since sensor networks are formed by a large number of nodes, sensor nodes must be low cost. Moreover, nodes must be self-powered because replacement is difficult once deployed. So, they must be equipped with a battery which discharge determines the node lifetime. In order to prolong nodes' life, it is recommendable to halt node activity when no data are needed to be transmitted or received. During these non-activity periods a node can operate in a low-power mode, but can't receive any message. The solution to this issue is asynchronously to wake up a slept node through an acoustic signal[2]. For this, an acoustic-triggered wake-up system must be include within its architecture. It is desirable that the wake up system can recognize predefined patterns in order to individually wake up network nodes. Thus, the solution is more flexible and can be adapted to many scenarios.

This paper describes research efforts of our research team conducted during the last year to develop an acoustic communication system for UWSN. In 2011, a flexible, energy efficient and low-cost underwater modem was designed[3] due to difficulties in finding an adequate underwater acoustic modem to address the cost and power consumption limitations of UWSN. Besides, a new acoustic-triggered wake-up system for low power nodes with predefined pattern recognition was designed[4] because of problems to find a wake-up solution useful for low-power nodes. In 2012, the combination of both researches and the addition of several improves resulted in a flexible, low cost and energy efficient acoustic modem architecture[5]. The result of this architecture was born the ITACA-S1000 modem, the solution with the less power consumption and with the remote wake-up system most advanced to date. However, since this new modem was the first coherent FSK solution presented so far, there were no efficient correction algorithms designed to improve the efficiency and the robustness of modem communications. Hence, a new algorithm for both Doppler-shift and multipath corrections for coherent FSK modem was researched[6].

This paper is organized as follows. Section 2 describes the research on an acoustic-triggered wake-up system. Section 3 describes the research on the low cost and energy efficient acoustic modem architecture. Section 4 describes the optimal multipath and Doppler-shift correction algorithm for coherent FSK modem. Finally, section 5 concludes this paper.

2 An acoustic-triggered wake-up system

In 2011, a new flexible, energy efficient and low-cost underwater modem was designed and implemented for UWSN[3]. With the purpose of prolonging nodes'

life, a search was conducted in order to find an underwater wake up system with a power consumption as low as possible and with an order of magnitude similar to the power consumption of the new modem. However, previous solutions[7][2] resulted not adequate. Therefore, a new wake-up solution adapted to these requirements was researched. The result was a new acoustic-triggered wake-up system for low power nodes presented in [4].

The new acoustic-triggered wake-up system waits listening to the channel to detect possible incoming wake-up signals. If a signal is received, the system processes it, waking up the modem MCU just in case of a positive recognition. The wake-up signal is On-Off-Keying modulated in order to save energy[8].

On the receiver side, a specific hardware is required to always listen the channel, waiting for a wake-up signal. The design contains a commercial IC and a specific matching net between piezoelectric and IC input designed to match both impedances and suitably couples the acoustic incoming signal. The commercial IC is the AS3933 of Austria Microsystems[9] which includes false wake-up detection and has a low power consumption suitable for modem requirements. The matching net is formed by a 3-stage T-structure band-pass filter with passive inductors and capacitors. This structure was found the most suitable since capacitors avoid any circuit bias modification and inductors mimic coil-antenna magnetic coupling.

On the transmitter side, since the modem transmits data with FSK modulation, only a software adjustment on modulation is required. Hence, no hardware modifications are needed to produce the OOK modulation. The modem simply switches power amplifier output on and off according to transmitted symbol and the bit synchronism is archived by Manchester coding.

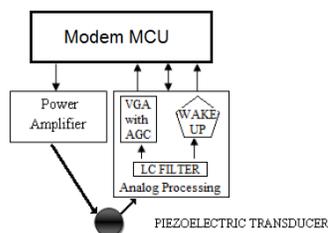


Fig. 1. Asynchronous wake-up system

In 2012, an improve to the acoustic-triggered wake-up system and a simulink model to acoustic-triggered wake-up systems were presented in [10]. The simulink model is designed to accurately and quickly evaluate network performance with acoustic-triggered wake-up capability in a predefined scenario. With this model, it is expected evaluate future extensions of the system. The improve of the wake-up system consisted in an extension of pattern recognition for recognizing multiple patterns. The multiple pattern recognition is based in extending the

wake-up signal and performing a detection algorithm. The wake-up signal is extended with an m additional bits pattern. This new pattern is OOK modulated too.

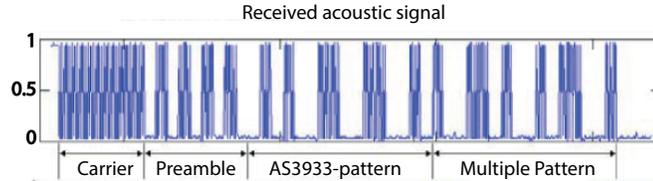


Fig. 2. Wake-up signal extension

The detection algorithm uses this additional pattern to be detected by the modem MCU. First, the modem remains in sleep mode while the wake-up peripheral listens to the channel. After detecting a valid AS3933-pattern, the wake-up peripheral wakes up the modem MCU. Then, the MCU checks the m additional bits. If the additional bits aren't a valid pattern, the MCU sleeps again. To minimize the power consumption, the MCU works at a frequency 60 times slower than the frequency in the transmission and reception modes while checking the additional pattern.

3 A low cost and energy efficient acoustic modem architecture

The first description of the ITACA modem was given in [3]. In 2012, the design was improved with enhanced features such as a triggered wake-up system, CCA, and RSSI measurement. The result of this improve was a low cost and energy efficient acoustic modem architecture presented in [5].

The acoustic modem architecture is built around a low power microcontroller (MCU) and includes several blocks designed under cost, complexity and consumption constraints. The figure 3 shows the modem architecture block diagram.

The transducer chosen for this architecture is a commercial piezoelectric-based transducer, the HUMMINBIRD XP 9 20[11]. The use of this element is a key factor that must be considered when a low cost underwater node is wanted since its price is significantly lower than hydrophones. The transducer has two identified bands around 85 kHz and 200 kHz. However, due to water attenuation, it is more worthwhile to use the lowest possible frequency band: 85 KHz.

is based on an XOR gate and filtering on a lead-lag low-pass filter while the controlled oscillator is implemented by the MCU. It is an energy-efficient HW-SW solution because both the external phase detector and the low pass filter relieve the MCU from the most consuming tasks in demodulation.

The triggered wake-up system is designed to work both a synchronous and asynchronous manner. The synchronous wake-up is implemented through a real time clock (RTC) with a 32.768 Khz crystal which can be programmed by MCU via serial interface. The RTC has several outputs that are activated when any alarm condition are met. These outputs are used to wake-up the MCU from its deeper low-power modes. Although this implementation does not any real innovation, it remains very useful and was included in the modem architecture to enable the implementation of upper layer protocols. The asynchronous wake-up is implemented through the asynchronous wake-up system described previously in section 2.

The CCA feature is implemented as a physical-layer mechanism and an MCU software routine. Physically, CCA requires carrier frequency (CF) detection or RSSI observation, both of which are supported by the already implemented wake-up circuit without additional hardware. The wake-up peripheral integrates a frequency detector based on a zero crossing counter. After CF detection, the gain of the amplifier is set to the maximum and the AGC reduces this level according to the received signal input level. After 1.06 ms, the AGC algorithm is complete and a stable RSSI value can be read from the corresponding register of the peripheral. Then, the MCU routine can read the RSSI value for eight symbols to obtain an average value above 2 dB.

As a result of this architecture was born the ITACA-S1000 modem, the solution with a very low power consumption and with a very advanced remote wake-up system. Figure 3 shows a comparative analysis between commercial acoustic modems with modem ITACA-S1000 using the concept of normalized power for each working state: idle, reception and transmission. This parameter, shown in (1), describes the relationship between the average power consumption (\bar{P}_{modem}), the maximum distance reached (D_{max}) and the spectral efficiency in a concise and quantitative way. The lower the value of this parameter, the higher the efficiency of the modem in each state.

$$P_{normalized} = \frac{\bar{P}_{modem}(W)}{D_{max}(m) \cdot Spectral_{efficiency}(bps/Hz)} \quad (1)$$

The ITACA-S1000 presents the lowest normalized power in idle state because of its revolutionary idle consumption, only $11\mu W$. In transmission, the efficiency is the highest due to the power stage designed specifically. Finally, in reception, its efficiency is similar to the best commercial options, ATM modems. Therefore, considering the normalized power, ITACA-S1000 is the most optimal solution in its operating range.

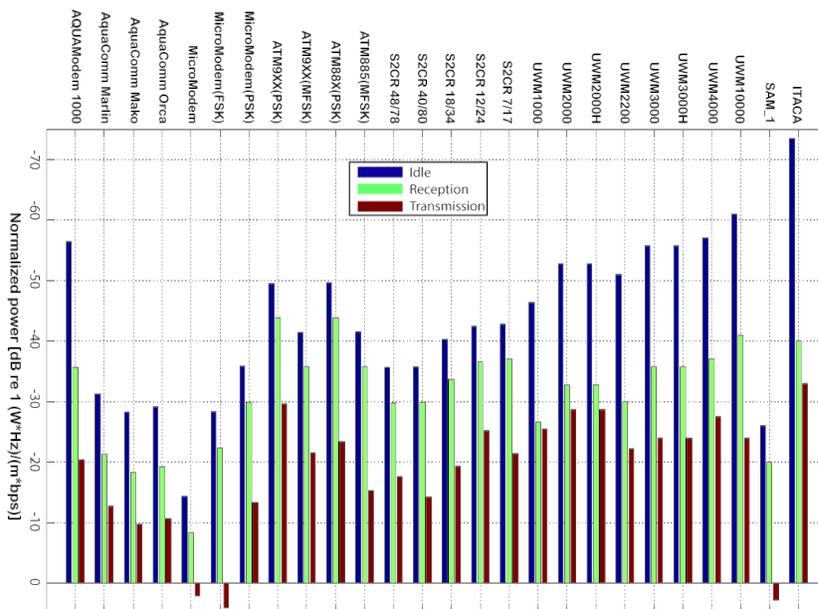


Fig. 4. Comparison between commercial acoustics modem with modem ITACA-S1000

4 An optimal multipath and Doppler-shift correction algorithm for coherent FSK modem

To reach a reliable acoustic communication, the multipath and Doppler effects must be addressed in the communication[14], specially in shallow water. On the one hand, multipath distortion is produced by echoes from the sea-bottom, the surface or submerged structures. On the other hand, Doppler-shift is produced as an effect of the relative movement between transmitter and receiver, and also can be produced by sound speed changes. In 2011, to improve the reliability and robustness of modem communications and to adapt the ITACA-S1000 modem to tough underwater environments, the incorporation in the modem architecture of correction algorithms to address both effects was considered. Unfortunately, neither optimal multipath nor Doppler-shift correction algorithms were found for coherent-FSK modulation in the literature so far. Therefore, a new algorithm for both Doppler-shift and multipath corrections for coherent FSK modem was researched and was presented in [6].

This new algorithm was optimized to be run in low resources and low power modem architectures. The algorithm is based in a closed feedback loop and reuses the existing modem FSK decoder. Two main blocks, a Doppler estimator and an adaptive filter, are used to compensate both Doppler and multipath effects with an initial training step.

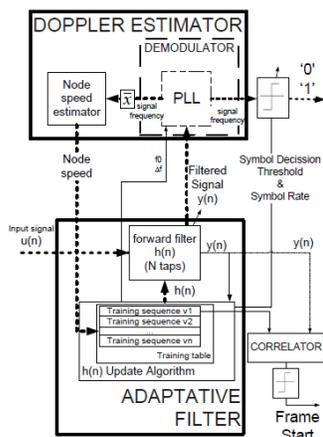


Fig. 5. Block diagram of training algorithm

The algorithm needs that same extra information is added at start of frames in order to perform the training step. This extra information is a training sequence with specific characteristics which can be obtained using primitive polynomials[15] to generate Pseudo Noise (PN) sequences.



Fig. 6. Frame format with extra training information

An experimental testbed was deployed in a small craft marina where frames were sent to a 10 meter distance with a relative movement ranging from -1 to +1 m/s. On the one hand, the error probability was analyzed. Table 1 shows the error probability both with and without corrections. Results revealed that the more positive the speed is, the higher error probability is obtained. However, correction algorithms were able to correct all errors. On the other hand, algorithms convergence and training process were also analyzed. Results showed that the Doppler correction algorithm converges with 25 symbols - @ 1kbps - in the worst case, the multipath compensation algorithm converges in only 4 symbols even in the worst multipath scenario and a 31 bit PN training sequence is enough to train the receiver.

Table 1. Experimental mean error rate probability

	-1 m/s	0 m/s	+1 m/s
No correction	6.5%	15%	29%
Multipath & Doppler correction	0%	0%	0%

5 Conclusions

This paper describes our contributions for underwater wireless sensor networks through the development of a low cost, energy efficient and reliable acoustic communication system.

The developed acoustic modem is an ultra-low power and low-cost modem architecture based in piezoelectric transducers, that combines a typical ultra-low power microcontroller-based core with energy-efficient mechanisms. This architecture uses coherent-FSK modulation and the demodulation has been cleverly divided into a dual hardware-software solution, which reduces processing requirements as well as the consumed power. Data rate reached with this architecture is 1 kbps, enough for UWSN deployment and comparable to other scientific research solutions.

The developed acoustic-triggered wake-up system is capable of detect pre-defined patterns and perform selective wake-up. With this feature, modems remain in sleep mode as long as possible reducing the overall power consumption. Besides, the wake-up signal is generated without additional hardware, being the modem data interface also used to both generate and receive the acoustic wake-up signal without being necessary a double acoustic transducer. Finally, since the wake-up system is based on a commercial IC, it can operate in different frequency bands and it can be easily integrated in several platforms.

To improve the efficiency and the robustness of modem communications, a novel algorithm for multipath and Doppler-shift effects correction for coherent-FSK acoustic modem was developed. The algorithm is simple and needs very low extra resources. This enables its applicability in low power modems to build long life underwater networks.

In conclusion, our contributions would help in the definition and evaluation of energy-efficient network communication policies based on asynchronous wake up medium access control, enable the deployment of wide underwater sensor networks to perform various applications and improve the efficiency and the robustness of underwater communications based in coherent-FSK acoustic platforms.

Acknowledgments

This work was supported by the Technical University of Valencia Research Project SP20120889 and Spanish government under projects CTM2011-29691-C02-01 and TIN2011-28435-C03-0.

References

1. Yang Xiao. *Underwater Acoustic Sensor Networks*. Auerbach Publications, Boston, MA, USA, 1 edition, 2009.
2. Jack Wills, Wei Ye, and John Heidemann. Low-power acoustic modem for dense underwater sensor networks. In *Proceedings of the 1st ACM International Workshop on Underwater Networks*, WUWNet '06, pages 79–85, New York, NY, USA, 2006. ACM.
3. A. Sanchez, S. Blanc, P. Yuste, and J.J. Serrano. A low cost and high efficient acoustic modem for underwater sensor networks. In *OCEANS, 2011 IEEE - Spain*, pages 1–10, June 2011.
4. A. Sanchez, S. Blanc, P. Yuste, and J.J. Serrano. Rfid based acoustic wake-up system for underwater sensor networks. In *Mobile Adhoc and Sensor Systems (MASS), 2011 IEEE 8th International Conference on*, pages 873–878, Oct 2011.
5. Antonio Sánchez, Sara Blanc, Pedro Yuste, Angel Perles, and Juan José Serrano. An ultra-low power and flexible acoustic modem design to develop energy-efficient underwater sensor networks. *Sensors*, 12(6):6837–6856, 2012.
6. A. Sanchez, S. Climent, P. Yuste, A. Perles-Ivars, and J. J. Serrano. Low complexity multipath and doppler-shift correction algorithm for reliable underwater coherent-fsk acoustic modems: Short paper. In *Proceedings of the Seventh ACM International Conference on Underwater Networks and Systems*, WUWNet '12, pages 21:1–21:5, New York, NY, USA, 2012. ACM.
7. Albert F. Harris, III, Milica Stojanovic, and Michele Zorzi. When underwater acoustic nodes should sleep with one eye open: Idle-time power management in underwater sensor networks. In *Proceedings of the 1st ACM International Workshop on Underwater Networks*, WUWNet '06, pages 105–108, New York, NY, USA, 2006. ACM.
8. M.S. Durante and S. Mahlke. An ultra low power wakeup receiver for wireless sensor nodes. In *Sensor Technologies and Applications, 2009. SENSORCOMM '09. Third International Conference on*, pages 167–170, June 2009.
9. Data sheet, “as3933 – 3d low frequency wakeup receiver”. Austria Microsystems. v1.0 Jul 01, 2010.
10. A. Sanchez, S. Blanc, P. Yuste, I. Piqueras, and J. J. Serrano. Advanced acoustic wake-up system for underwater sensor networks. *Communications in Information Science and Management Engineering(CISME)*, 2(2):1–10, 2012.
11. Humminbird: Fishfinder, depth finder, gps chartplotters, side imaging sonar technology and marine fish finders. Available online: <http://ww.humminbird.com> (accessed on 3 April 2012).
12. R. Jurdak, C.V. Lopes, and P. Baldi. Software acoustic modems for short range mote-based underwater sensor networks. In *OCEANS 2006 - Asia Pacific*, pages 1–7, May 2006.
13. Saunvit Pandya, J. Engel, J. Chen, Zhifang Fan, and Chang Liu. Coral: miniature acoustic communication subsystem architecture for underwater wireless sensor networks. In *Sensors, 2005 IEEE*, pages 4 pp.–, Oct 2005.
14. Jim Partan, Jim Kurose, and Brian Neil Levine. A survey of practical issues in underwater networks. In *Proceedings of the 1st ACM International Workshop on Underwater Networks*, WUWNet '06, pages 17–24, New York, NY, USA, 2006. ACM.
15. Frank Ehlers, Warren Fox, Dirk Maiwald, Martin Ulmke, and Gary Wood. Advances in signal processing for maritime applications. *EURASIP Journal on Advances in Signal Processing*, 2010(1):512767, 2010.

Modeling and Simulation of Underwater Wireless Sensor Networks

Underwater communication

Raúl Sáez, Salvador Climent, Ricardo Mercado, Àngel Perles, Pedro Yuste,
Juan Vicente Capella, Sara Blanc, Vicente Busquets, Juan José Serrano, and
Rafa Ors

Instituto ITACA, Universitat Politècnica de València, Camino de Vera s/n, 46022
Valencia, Spain;
e-mail: ucert@upv.es;
website: <http://ucert.upv.es>

Abstract. Many research areas have benefited from simulations to acquire reliable data before conducting real experiments. In underwater wireless sensor networks (UWSN), the deployment, especially in deep waters, can be difficult and expensive. On the other hand, the cost of underwater wireless nodes are expensive partially because of their complex acoustic transceivers. Therefore, simulations are an essential tool to test and tune new features and algorithms before deploying a UWSN to conduct a real experiment.

This paper overviews research efforts of our research team conducted during the last year to contribute on the design of UWSN by means of modeling and simulations. The first contribution was the adaptation of an energy-efficient and robust architecture for WSN to the subaquatic acoustic medium by means of an ns-3 simulator model. Then, a low-power underwater wake-up model for the ns-3 simulator was developed to test and improve wake-up systems. Finally, as a result of all acquired knowledge, a complete underwater wireless sensor network ecosystem of models for the ns-3 simulator was researched.

Keywords: uwsn; wake-up; modeling; simulation; ns-3

1 Introduction

Underwater Wireless Sensor Networks (UWSN) have become an important area of research due to its wide range of applications, ranging from submarine surveillance to monitoring of the marine environment. In these networks the transmission is done by means of acoustic waves, since electromagnetic waves are heavily attenuated underwater. However, acoustic signal propagation is five orders of magnitude lower than its radio-frequency counterpart. On radio-frequency networks, this propagation delay is negligible but it has to be considered on underwater acoustic networks. Besides, acoustic waves present a signal spreading proportional to the distance and different signal attenuations depending on distance and frequency.

Moreover, the deployment and maintenance costs of these networks are usually high and the cost of the nodes is expensive partially because of their more complex acoustic transceivers. Hence, there is a need for low-cost, low power modems capable of extending the nodes' battery life as much as possible. Nowadays acoustic modem uses wake-up systems which enables it to listen and recognize certain stimuli sent prior to the actual data packet and wake-up the main circuitry to receive it. So, the modem can remain in low power mode for long time periods. Another option to compensate for the energy requirements is the use of energy harvesting mechanisms that can obtain energy from the environment to extend the nodes' battery life[1][2].

Many research areas have benefited from simulations to acquire reliable data before conducting real experiments. Since the deployment costs of these networks are high, simulations are an essential tool to test and tune new features and algorithms before their implementation with real hardware to avoid node or even network failures. However, one has to be certain that the simulation results are as accurate as possible.

This paper describes research efforts of our research team conducted during the last year to contribute on the design of low cost and energy-efficient UWSN by means of modeling and simulations. In 2011, an energy-efficient routing protocol for wireless sensor networks, EDETA[3], was adapted to the subaquatic acoustic medium. The result of this adaptation was EDETA-e[4], a power-aware routing protocol for UWSN which minimizes the energy consumption. Then, a model of the protocol for the ns-3 simulator[5] was developed to study the behavior of the protocol in various underwater scenarios. In 2012, using this model, a study of different scheduling and retransmission techniques applied to EDETA-e[6] was conducted to analyze their performance in terms of energy consumption, delays, packet lost rate and duplicate packets. At the same time, a low-power underwater wake-up model for the ns-3 simulator[7] was developed to evaluate and compare acoustic platforms with wake-up system. Thereafter, this model was used to evaluate the adaptation of several MAC protocols to make use of the wake-up capabilities[8]. In 2013, as a result of all acquired knowledge, a UWSN ecosystem for the ns-3 simulator with energy-harvesting and wake-up capabilities[9] was developed to accurately simulate underwater applications before their development.

This paper is organized as follows. Section 2 describes the routing protocol model obtained after adapting the EDETA protocol to the subaquatic acoustic medium. Section 3 describes the low-power underwater wake-up model for the ns-3 simulator. Section 4 describes the complete underwater wireless sensor network ecosystem of models for the ns-3 simulator. Finally, section 5 concludes this paper.

2 The EDETA-e routing protocol model

In underwater acoustic networks, where current transceivers have more power consumption compared with their radio-frequency counterpart and the propa-

gation speed might vary over time, having a planned and fixed infrastructure can help the deployment of underwater networks. Moreover, since GPS signals are heavily attenuated underwater, the localization of sensor nodes and AUVs (Autonomous Underwater Vehicles) is not trivial and the fixed nodes can be used for localization proposes.

In 2011, the Energy-efficient aDaptive hiErarchical and robusT Architecture (EDETA) for WSN[3] was adapted to underwater environment. The result was EDETA-e, a power-aware routing protocol for UWSN which minimizes the energy consumption. EDETA-e is a subset of EDETA that allows the engineer to assume control over the network formation and the delays. Since cluster heads are fixed at design time, EDETA-e only considers one initialization phase. After that, all nodes will always remain in the normal operation phase.

In order to evaluate the protocol, a model for the ns-3 simulator was developed in [4]. The model uses a modified version of underwater acoustic network (UAN) model which is aware of the received packets signal level and has support to transmit in different radio channels. Figure 1 depicts a simplified class diagram of the modified classes of the UAN model.

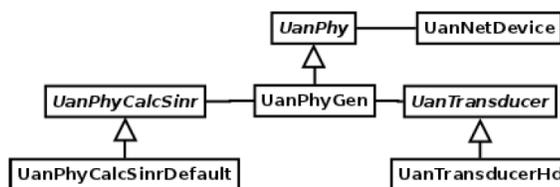


Fig. 1. Simplified class diagram of UAN modified methods

The received signal strength is passed to the upper layers attaching a tag to the received packet at the StartRxPacket method of the UAN phy layer. To implement the different radio channels functionality, the current channel is attached to the packet with another tag. Besides, it was also necessary to modify the CalcSinrDb, SendPacket and Receive methods from the UanPhyCalcSinrDefault, UanPhyGen and UanTransducerHd classes to correctly handle send and receive packets and SNR calculations. Moreover, a data link layer was implemented in order to make the EDETA implementation independent of the MAC layer. This layer extracts sender and destination addresses and the received signal strength of the packet and sends them to the EDETA model along with the rest of the EDETA packet.

Simulation results using this model in three different square scenarios concluded that EDETA-e is a very suitable protocol for subaquatic sensor networks, presenting, in addition, new features in this field.

In 2012, to optimize the protocol performance in terms of energy consumption, packet delay, number of duplicate packets and packet loss, simulations

using the EDETA-e model were conducted in [6] to analyze impacts of different delay-aware and non-delay-aware scheduling and retransmission techniques when applied underwater. The techniques under study were TAck, TnoAck, DAck, DnoAck, DnoAck2 and DCsma. TAck is a TDMA schedule with acknowledgment and data packet loss. TnoAck is a TDMA schedule without acknowledgment. DAck is a delay-aware schedule with acknowledgment and data packet loss. DnoAck is a delay-aware schedule without acknowledgment. DnoAck2 is a delay-aware schedule without acknowledgment. DCsma is a delay-aware schedule with acknowledgment and CSMA retransmission period.

Finally, a comparative study between EDETA-e using different scheduling and retransmission parameters and DBR[10] was conducted. The study consisted in simulations of both protocols using the same MAC layer (ALOHA protocol) and the same modem parameters. Results showed that the average end-to-end delay for EDETA-e was higher than the delay for DBR. However, in terms of lost data due to collisions, EDETA-e experienced no collisions since it is a fully scheduled algorithm while DBR presented high packet loss. On the other hand, results showed that using EDETA-e the last node dies more than 100 times later. Hence, the study demonstrated that EDETA-e is very energy efficient compared with the DBR protocol.

3 A low-power underwater wake-up model

In 2012, an ultra-low power and flexible acoustic modem with wake-up system was designed to develop energy-efficient underwater sensor networks in [11]. To evaluate possible future features and compare this platform with another acoustic platforms with wake-up system, a low-power underwater wake-up model for the ns-3 simulator was developed in [7]. A simplified component diagram of the model is depicted in figure 2.

In the model, the UanPhy component models the underwater physical layer and it includes different Signal-to-Noise-Interference Ratio (SNIR) and Packet Error Rate (PER) models. The UanMac component models the medium access protocol used by the nodes in the configured network. The UanNetDevice component models a Network Interface Card (NIC) and is used by the upper layers to send and receive packets to/from the network. The UanChannel module models an underwater channel[12] and delivers all the packets to the UanPhy components connected or listening to this specific underwater channel. The UanPhy component is also connected to a DeviceEnergyModel component which models the energy consumption of the different radio states and, in turn, is connected to an EnergySource component which models different energy containers like batteries. Finally, the UanMacWU module is responsible for doing the actual channel assessment and sending the wake-up packet (UanPacketWU) before the actual packet that the UanMac module intends to send. As figure 2 shows, two UanPhy and two UanMac are used to model the regular modem and the wake-up system. The two UanPhy modules are equal in functionality and the only dif-

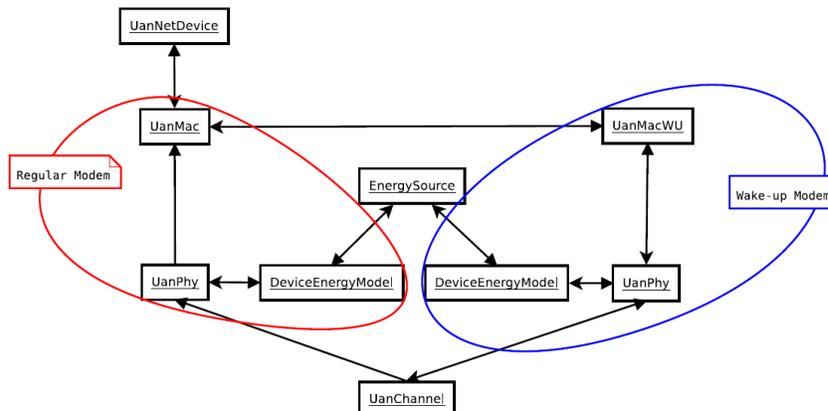


Fig. 2. Simplified component diagram of the proposed underwater wake-up model

ference between the two is the consumption parameters, which are set to match the consumption parameters of the regular modem and the wake-up system.

The model works through the interaction of the regular modem model and the wake-up system model. In reception, a packet will always be preceded by an `UanPacketWU` when it is received from another node, hence the `UanMacWU` will receive it and see if the packet destination is for the node. If it is not, it will discard the packet. If it is, it will wake-up the `UanPhy` in charge of the regular radio so it can receive the packet and send it to the `UanMac` directly. In transmission, the `UanMac` module asks the `UanMacWU` for the current channel status and, if the channel is free, it sends its `UanPacket` with the data from the upper layers or its own control packets to the `UanMacWU`. After that, the `UanMacWU` sends the wake-up packet and the packet from the `UanMac` immediately after it.

In order to evaluate the model, a study of the behavior of different medium access techniques adapted to make use of the wake-up capabilities was conducted in [8]. The MAC protocols under study were MACA, FAMA and ALOHA-CS and the studied parameters were energy consumption, packet delay and the impact of transmission speed on the performance of protocols.

4 A complete underwater wireless sensor network ecosystem of models

In 2013, as a result of all acquired knowledge, a UWSN ecosystem of models for the ns-3 simulator with energy-harvesting and wake-up capabilities was developed in [9]. This ecosystem enables to accurately simulate underwater ap-

plications using real data (such as weather conditions, sound speed profile or bathymetry) from the location where the system will be deployed.

The ecosystem extends the currently available energy and underwater frameworks of the ns-3 simulator with two models, a model of an energy-harvesting module and the low-power underwater wake-up model described in the previous section. The energy-harvesting module is included in the physical layer, since it is a real energy-harvesting hardware model. The low-power wake-up model implementation is divided between physical and MAC layers. The figure 3 shows a diagram of this ecosystem.

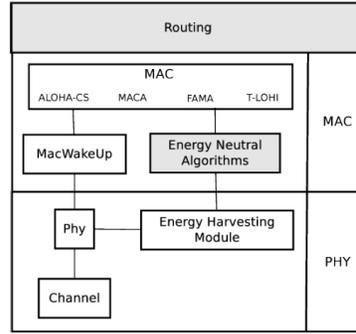


Fig. 3. UWSN simulation ecosystem

The implementation of the energy-harvesting model was presented in [13]. This model is divided into four sub-models, each represented by one class in the ns-3 simulator. A simplified class diagram is depicted in figure 4. The CapacitorEnergySource class models capacitor charge and discharge curves. The SensorEnergyModel class describes the sensor energy consumption (temperature, humidity, etc.). The RadioEnergyModel class represents the radio energy consumption at different modes, and the SolarPanelEnergyModel class models the energy obtained from solar panels.

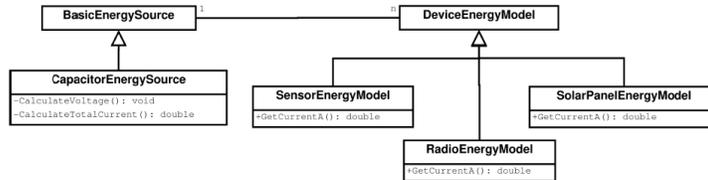


Fig. 4. Simplified class diagram of the ns-3 energy-harvesting model

To demonstrate the utility of this ecosystem of models, a study of an example scenario of a real monitoring installation in an offshore fish farming facility was conducted. The scenario architecture is depicted in figure 5, where different sensor nodes are placed near the fish nets and the sea bottom. These nodes are capable of measuring various environmental variables on demand and sending the data to the sink. Sink nodes are placed on buoys on the sea surface and equipped with solar energy-harvesting capabilities. They also include a radio modem to communicate with an onshore installation. Finally, all nodes are equipped with a low-power, low-cost underwater acoustic modem with integrated wake-up capabilities, as presented in [11].

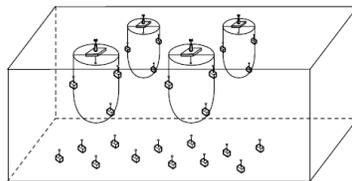


Fig. 5. Simulated scenario diagram

In this scenario, a sink might require a group of nodes to periodically send some environmental variables. To keep the architecture as flexible as possible, there is no need for these nodes to be known a priori. For example, the sink can send a message asking for certain information and the desired sample frequency. Besides, the nodes that are capable of providing this information have to compete to acquire the channel and send the information.

The study consisted of two different parts, an analysis of the energy consumption of nodes using the energy-harvesting module and a study of several MAC protocols using the low-power underwater wake-up model. The analysis of the energy consumption of nodes was performed varying different node parameters and their contribution to system behavior. For the study of MAC protocols, ALOHA-CS, MACA, FAMA and T-Lohi were adapted (when possible) to take advantage of the modem wake-up capabilities. Finally, to ensure accuracy, all simulations of the study were carried out employing the Bellhop propagation model and extracting data from the location where the application will be deployed.

5 Conclusions

Simulations have proven to greatly facilitate the research and development of new algorithms and features for underwater sensor networks. This paper describes our contributions for underwater wireless sensor networks through the development of accurate models that enable the performance of accurate simulations by providing real weather and marine conditions.

The adaptation of the EDETA protocol to underwater behavior resulted in EDETA-e, a power-aware routing protocol which minimizes the energy consumption. The developed ns-3 model of EDETA-e enabled to analyze the behavior of the protocol in the subaquatic medium by means of simulations. Simulations were performed applying different scheduling and retransmission techniques to the protocol. Results showed that EDETA-e is a very suitable protocol for subaquatic sensor networks, presenting, in addition, new desired features not present in other approaches in this field.

The developed low-power underwater wake-up model can be used to test and compare the design of new underwater wake-up systems in terms of energy consumption. Besides, the study about the adaptation of several medium access control (MAC) protocols to make use of the wake-up capabilities has contributed with new interesting and usable data for the MAC layer of UWSN. Therefore, this study proves that the wake-up model can be used to study, through simulations, new protocols based on wake-up features.

Finally, the complete ecosystem of models for the ns-3 simulator enables the simulation of complex scenarios using real meteorological data and marine conditions from the location of the UWSN. Thus, simulations performed with this platform are accurate and provide useful data before the deployment.

In conclusion, our contributions cover from the physical layer to routing layer of UWSN. The developed models can be used to improve the development time of new algorithms and features for UWSN, perform the analysis and the design of new underwater modems with wake-up systems and improve deployments of real UWSN application in different scenarios.

Acknowledgments

This work was supported by the Technical University of Valencia Research Project SP20120889 and Spanish government under projects CTM2011-29691-C02-01 and TIN2011-28435-C03-0.

References

1. Geoff V. Merrett, N.M. White, N.R. Harris, and B.M. Al-Hashimi. Energy-aware simulation for wireless sensor networks. In *Sensor, Mesh and Ad Hoc Communications and Networks, 2009. SECON '09. 6th Annual IEEE Communications Society Conference on*, pages 1–8, June 2009.
2. V. Sharma, U. Mukherji, and V. Joseph. Efficient energy management policies for networks with energy harvesting sensor nodes. In *Communication, Control, and Computing, 2008 46th Annual Allerton Conference on*, pages 375–383, Sept 2008.
3. J.V. Capella, A. Bonastre, R. Ors, and S. Climent. A new energy-efficient, scalable and robust architecture for wireless sensor networks. In *New Technologies, Mobility and Security (NTMS), 2009 3rd International Conference on*, pages 1–6, Dec 2009.
4. S. Climent, J. Capella, A. Bonastre, and R. Ors. A new model for the ns-3 simulator of a novel routing protocol applied to underwater wsn. In *Proceedings of the 2011 International Conference on Wireless Networks*, page 18–21, 2011.

5. Network simulator 3 (ns-3), 2012. Available online: <http://www.nsnam.org> (accessed on 5 May 2014).
6. Salvador Climent, Juan Vicente Capella, Nirvana Meratnia, and Juan Jose Serrano. Underwater sensor networks: A new energy efficient and robust architecture. *Sensors*, 12(1):704–731, 2012.
7. Salvador Climent, Antonio Sanchez, JuanVicente Capella, Sara Blanc, and Juan-José Serrano. Modelling and simulation of underwater low-power wake-up systems. In Francisco Martins, Luís Lopes, and Hervé Paulino, editors, *Sensor Systems and Software*, volume 102 of *Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering*, pages 75–88. Springer Berlin Heidelberg, 2012.
8. Salvador Climent, Antonio Sanchez, J. V. Capella, and J. J. Serrano. Simulating mac protocols under real underwater sensor networks assumptions. In *Proceedings of the Seventh ACM International Conference on Underwater Networks and Systems*, WUWNet '12, pages 34:1–34:2, New York, NY, USA, 2012. ACM.
9. Salvador Climent, Juan Vicente Capella, Sara Blanc, Angel Perles, and Juan José Serrano. A proposal for modeling real hardware, weather and marine conditions for underwater sensor networks. *Sensors*, 13(6):7454–7471, 2013.
10. Hai Yan, Zhijie Jerry Shi, and Jun-Hong Cui. Dbr: Depth-based routing for underwater sensor networks. In *Proceedings of the 7th International IFIP-TC6 Networking Conference on AdHoc and Sensor Networks, Wireless Networks, Next Generation Internet*, NETWORKING'08, pages 72–86, Berlin, Heidelberg, 2008. Springer-Verlag.
11. Antonio Sánchez, Sara Blanc, Pedro Yuste, Angel Perles, and Juan José Serrano. An ultra-low power and flexible acoustic modem design to develop energy-efficient underwater sensor networks. *Sensors*, 12(6):6837–6856, 2012.
12. Federico Guerra, Paolo Casari, and Michele Zorzi. World ocean simulation system (woss): A simulation tool for underwater networks with realistic propagation modeling. In *Proceedings of the Fourth ACM International Workshop on UnderWater Networks*, WUWNet '09, pages 4:1–4:8, New York, NY, USA, 2009. ACM.
13. Antonio Sánchez, Salvador Climent, Sara Blanc, Juan Vicente Capella, and Ignacio Piqueras. Wsn with energy-harvesting: Modeling and simulation based on a practical architecture using real radiation levels. In *Proceedings of the 6th ACM Workshop on Performance Monitoring and Measurement of Heterogeneous Wireless and Wired Networks*, PM2HW2N '11, pages 17–24, New York, NY, USA, 2011. ACM.

Simulation of All-IP Access Network with Configurable Traffic Sources

David Gomez-Cuadrado¹, Victor Sempere-Paya², Miguel A. Rodriguez-Hernandez³, Angel Gomez-Sacristan¹

¹ Universitat Politècnica de València, Valencia, Spain
dagmecua@teleco.upv.es, agomez@dcom.upv.es

² ITI, Universitat Politècnica de València, Valencia, Spain
vsempere@dcom.upv.es

³ ITACA, Universitat Politècnica de València, Valencia, Spain
marodrig@upvnet.upv.es

Abstract. This paper describes a tool for the evaluation of capabilities in a convergent all-IP access network where IP traffic sources generated by the user at the UNI (User Network Interface) are configurable in number and in behavior. This tool, based on the simulator OMNeT++, allows the user to define different traffic types and to analyze the main Quality of Service (QoS) parameters: bandwidth use, delay and packet loss. This paper shows the simulation results for a typical corporative scenario, where voice, video and data services share the same access. Also this tool could be a valuable help for capacity planning for All-IP network design.

1 Introduction **AOI: Wide Area Networks*

All-IP networks [1] represent the evolution of existing networks of voice, data and multimedia. IP protocol is becoming the standard in business and industrial environments due to its characteristics of open system, ubiquity, greater flexibility and lower costs. Traditional services including voice, data and multimedia, in addition to other devices for control and monitoring, such as sensors and industrial networks, converge toward this architecture.

Unlike traditional telecommunications networks, in All-IP networks, services and applications of differing natures share resources in a common infrastructure. It is therefore important to introduce mechanisms in the network traffic management to ensure that every service meets the quality objectives listed in the international recommendations [2] [3].

In a real-world scenario, multiple sources and heterogeneous traffic (isochronous for voice traffic, variable for video services, asynchronous traffic for alarms, best-effort, etc.) with different needs of latency, packet loss, and jitter are competing for a scarce bandwidth. Using Diffserv [4], mechanisms to prioritize some traffic over others can be used, but analytically describing the behavior of this scenario is a complex task, especially in a dynamic environment. However, simulating corporative scenarios

with configurable traffic sources and evaluating their response to several variations of relevant parameters can help saving time and resources.

The optimization of bandwidth to maintain the QoS requirements, or even allowing a certain degree of degradation in less critical services may represent an important source of savings for the companies, given that there is a direct relationship between bandwidth use and costs.

The traffic sources have been designed to be completely configurable in number and behavior. Three categories of traffic sources have been defined: voice, video and data. Each category has specific configuration parameters and behavioral attributes: number of sources, arrival rate, average duration, etc. Those attributes will be described in the upcoming sections.

There has been a special focus on facilitating a flexible management of the simulation environment and the comparison of results in different scenarios. For this reason, the parameters that govern the behavior of the sources and the operations of the simulator are stored in XML (Extended Markup Language) files, which is a format that can also be generated by external applications.

The structure of this article is as follows: Section 2 shows the simulation environment with the description of the traffic sources. Section 3 describes the QoS and queuing mechanisms. Section 4 includes the simulation of an arbitrary scenario, and finally, Section 5 contains the conclusions.

2 Traffic Sources

For the implementation of the simulator, OMNeT++ [5] and some modified elements of the INET library [6] have been used. The OMNeT++ software has been selected because its modularity, ease of integration and previous experience of use in other research tasks.

OMNeT++ is a discrete event simulation tool designed to simulate computer networks, multi-processors and other distributed systems. A simulation model in OMNeT++ consists of hierarchically nested modules that communicate between themselves with messages. The active modules are called simple modules and their behavior is defined in files that are written in C++. These simple modules can be grouped, forming compound modules.

The simulator allows the creation of new functional elements, like the traffic sources, the QoS mechanisms (marking, policing and managing queues) [4], or a module to measure the bandwidth.

Three different categories of traffic have been defined: voice, video and data. Within each category, different types can be defined as a function of a set of specific parameters. There is no limit to the number of sources to simulate.

Traffic parameters can be fixed values or random variables defined using statistical distributions. Specifically, these parameters may follow a uniform distribution, an exponential distribution, or a normal distribution. In the case of using a normal distribution or an exponential distribution, the simulation allows defining a minimum and maximum value.

2.1 Voice Sources

For the implementation of a voice source, an On-Off model [7] has been used. In this model there is a state in which packets are sent and a state in which no traffic is generated (silence). The parameters that define the behavior of a VoIP source are the codec used, the average call duration, and the average time of inactivity (see Figure 1).

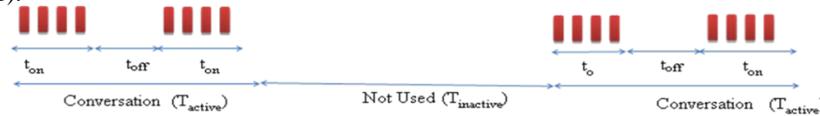


Fig. 1 VoIP Source

Both the time for the On state and for the Off state are random variables that follow a Weibull distribution [8] with the parameters showed in Table 1

Table 1 Weibull distribution parameters

Parameter	Form Factor	Scale Factor
Ton	1.423	0.824
Toff	0.899	1.089

Fig. 2 shows the bandwidth use produced by a voice source with codec G711 (size of the packets of 172 bytes with a separation between packets of 20 milliseconds) [9], and the distribution of the periods On and Off as defined in Table 1.

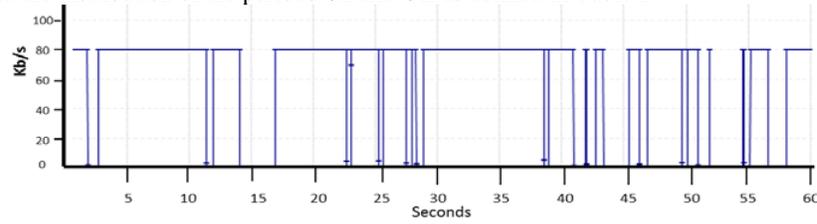


Fig. 2 VoIP bandwidth (On-Off model)

In the case of 50 concurrent sources equal to the previous case sharing the same class of service in the interface, we can see a statistical gain of approximately 40%, because the model includes the times of silence Toff. (Fig.3)

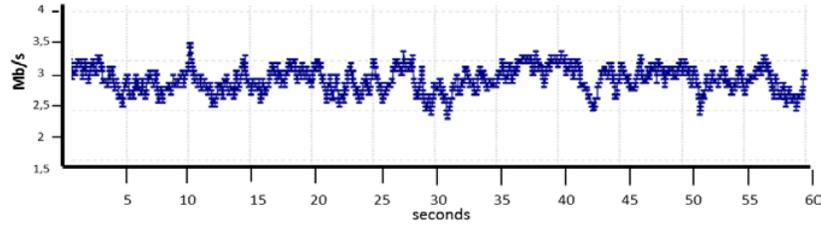


Fig. 3 VoIP bandwidth for 50 concurrent sources

2.2 Video sources

Several bursts are generated with a variable bit rate to simulate the video sources. The video quality is simulated using different bit rates and different number of frames per second which determines the spacing between frames. By default it uses H. 264 to encode the video traffic, with four available qualities: SD, HQ, HD and HD+.

It is possible to estimate H. 264 bitrate as a combination of parameters: the resolution (number of pixels in each frame), the amount of frames per second, and a factor that represents the amount of motion in the image, being 1 low motion (for example a person speaking without moving much), 2 for medium motion (some degree of movement, but predictable and orderly), and 4 for a high motion [10].

We can assume variations between 75% and 150% of the calculated optimum value. Table 2 shows the bitrate calculated for the case of 25 frames per second and low motion.

Table 2 Video Quality Parameters

Quality	Resolution	Bitrate
SD	180p (320x180)	75~150 kbps
HQ	360p (640x360)	300~600 kbps
HD	720p (1280x720)	1200~2500 kbps
HD+	1080p (1920x1080)	2750~5500 kbps

The duration of the video and the time when the source does not generate any traffic (free channel) can also be defined (see Fig. 4).



Fig. 4 Video traffic distribution

As an example, the bandwidth generated by an HQ quality video traffic source with 25 frames per second, low amount of motion, and sending traffic during 60 seconds, is shown in Fig 5. Bandwidth varies from 300 to 600 Kb/s that matches with the HQ video Quality model.

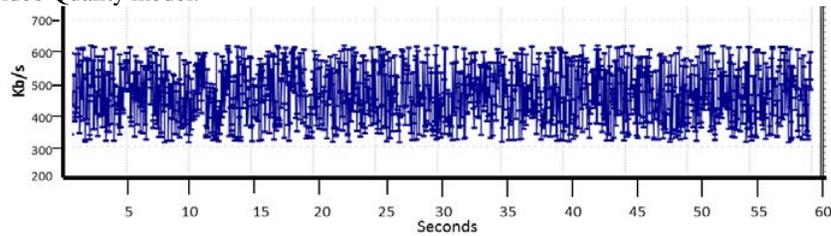


Fig. 5 Video Bandwidth

In the case of concurrence of 30 traffic sources like the previous one, sharing the same class of service in the interface, a small statistic multiplexing gain can be achieved because the small bit rate variations. Now bandwidth varies between 12.5 Mb/s and 15.4 Mb/s.

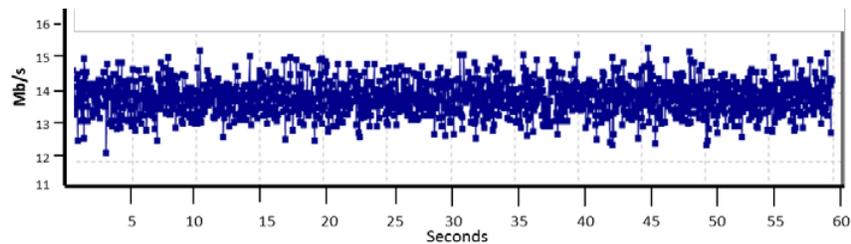


Fig. 6 Video Bandwidth for 30 concurrent sources

2.3 Data sources

The data traffic includes files that can contain data and/or images. In turn, these files can be sent in bursts. The average size of each data resource and each image is configurable, as well as the number of those resources of data and images. It is possible to define the number of files that are sent in each burst, the separation between these files, the number of sessions that we are going to generate, and the average time of separation between sessions (see Fig. 7).

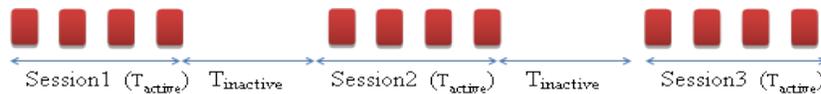


Fig. 7 Data traffic distribution

Fig.8 shows an example where a data source sends 1 burst every 2 seconds composed by a random number of files with an exponential distribution of average 4, with minimum value of 1 and maximum value of 8. These files are composed of a random number of images and text resources. Text resources have an average size of 10KB, while images have an average size of 500KB.

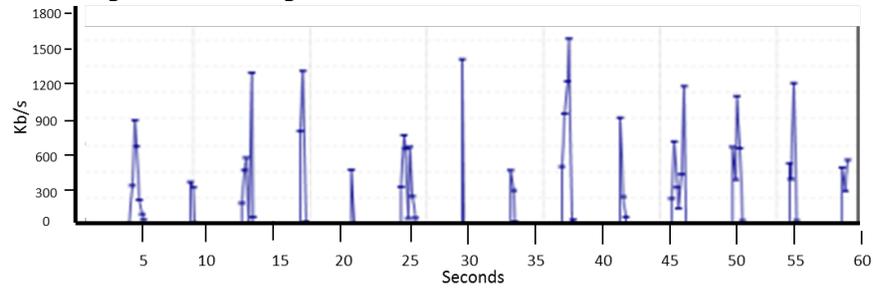


Fig. 8 File & Data transfer bandwidth

Fig 9 shows the effect of multiplexing 80 sources like the previous one, sharing the same class of service in the interface. Each source has been activated between 0 and 30 seconds interval using a uniform random variable.

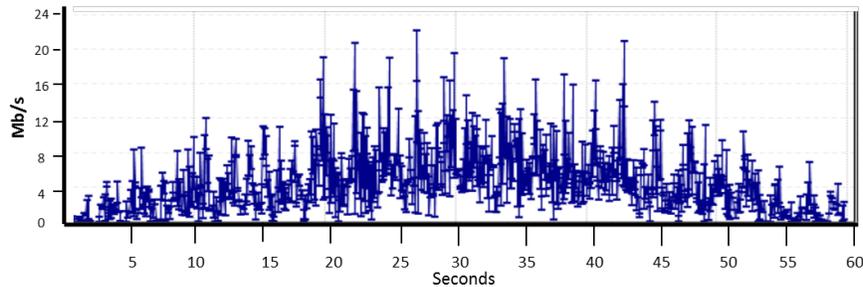


Fig. 9 File & data transfer bandwidth of for 80 concurrent sources

3 QoS and Queues management.

In the simulation environment, the basic QoS functions have been programmed. These functions include packet classification, marking and policing, as well as the queues management, with the aim of ensuring a minimum standard of quality to every type of traffic.

In order to get a simulation as similar as possible to a real-world environment, a Telefonica's service called Macrolan [13] is used as a reference. This service defines three traffic categories: Multimedia, Gold and Silver. The multimedia traffic includes

VoIP traffic, Gold traffic includes video traffic and real time data, while silver includes best-effort traffic.

A basic simulation model has been developed for the queues management, using the Cisco Catalyst 3650 [14] Switch-Router as reference, which is also used in the Macrolan Service. It has 4 queues: one for voice traffic, one for critical traffic (which includes the traffic management, control, etc.), one for the real-time traffic (video and gold), and a last one for the silver traffic.

The queues used for real time traffic (the first 3) are small (10-50 packets) in order to minimize the generation of delays and jitter, while the last one, dedicated to best effort traffic, is much larger (500 packets) to minimize packet loss.

Traffic is classified and filtered in the router. Filter rules depend on various criteria, for example: IP address (source/destination), UDP port (source/destination), etc. Once the packet is classified, the router will mark the DSCP field on the IP packet.

Once the packets are marked, they will go to a QoS block where quality mechanisms associated with each type of traffic are implemented. This check is performed by the module "meter" and its behavior depends on the values of the parameters configured (CIR-Committed Input Rate, CBS-Committed Burst Size, EIR-Excess Information Rate and EBS-Excess Burst Size[4]). In the case of CIR non-conformity, the packet is remarked with a greater chance of discarding in case of congestion, or discarded in the case of surpassing the EIR.

The output of this module is connected to the four queues explained above. The queue for voice traffic will have the most priority (priority queue), and the other 3 queues will be scheduled by a WRR (weighted round robin) algorithm, and its weights can be customizable by the user.

4. Simulation of an arbitrary scenario

A simulation has been made using as a reference a corporative access to an All-IP network with different services: voice, video, access to databases, access to servers and transfer of files, all of them sharing the available bandwidth of 100 Mb/s. Tables 3 to 5 describe the characteristics of the sources used.

The volume of traffic exceeds the maximum available bandwidth in the interface so there will be congestion. QoS mechanisms must act on the incoming traffic, generating delays and packet losses for services of lower priority.

Simulation time is configurable but to show better the effect of congestion, the simulation will last for 30 minutes, although it could last for longer.

Table 3 Voice sources

Service	Activity	Sources	Códec	Average Duration	Inactive time
Corporative Telephony	Low	85	G.711	2 minutes	40 minutes
Call Center	High	15	G.711	3 minutes	2 minutes

Table 4 Video sources

Service	Sources	Quality	FPS	Motion	Average Duration	Inactive time
Videoconference	30	HD	25	Medium	30 minutes	20 minutes
Video surveillance	20	HQ	25	Medium	5 minutes	55 minutes
Telepresence	1	HD	25	High	60 minutes	0 minutes

Table 5 Data sources

Service	Sources	Type of traffic	Average Burst Size	N° of Bursts
Database Access	100	Real-Time	10 files x (100B-5KB)	60
Server Access	100	Best effort	5 files x (1KB-100KB)	70
Big files transfers	20	Best effort	1 file (100KB-100MB)	30

The following QoS parameters have been chosen (see Table 6)

Table 6 QoS Parameters

Traffic	CIR	EIR	Queue	Queue Size	Scheduler
Voice	2 Mbps	4 Mbps	Droptail Queue	10	Priority
Gold	50 Mbps	100 Mbps	FIFO Queue	100	WRR (weight 2)
Silver	10 Mbps	100 Mbps	Droptail Queue	500	WRR (weight 1)

With this simulation scenario, the traffic generated is higher than 100Mbps, the maximum available bandwidth. Fig. 10 shows the evolution of bandwidth vs. time until saturation

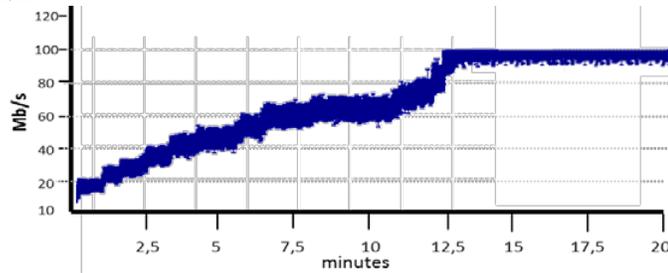


Fig. 10 Bandwidth usage at UNI interface

The program allows simulating different QoS configurations, selecting which type of traffic is prioritized and the amount of bandwidth available for it. For voice traffic, as seen in Fig. 11 there is no packet loss, and delay is in the worst case less than 1 millisecond.

The CIR assigned to voice services, 2 Mb/s, is overprovisioned. Peak value for voice traffic is 1.5 Mb/s in the worst case in this scenario.

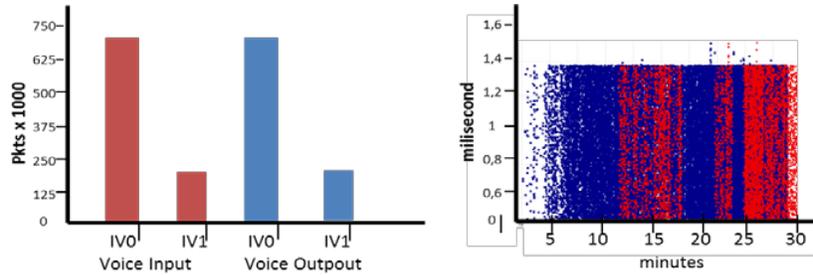


Fig. 11 Traffic Voice Packets and Voice Delay

In the case of real-time traffic (video and real time database access) more traffic than the guaranteed is generated, so packet loss and delay will have a greater effect in this kind of services.

Fig. 12 shows that there is packet loss for every real-time application (from left to right, video conferencing, video surveillance, telepresence, and database access). Around 30% of packets are lost in 15 minutes of simulation. Packet loss is clearly unacceptable for video services [8].

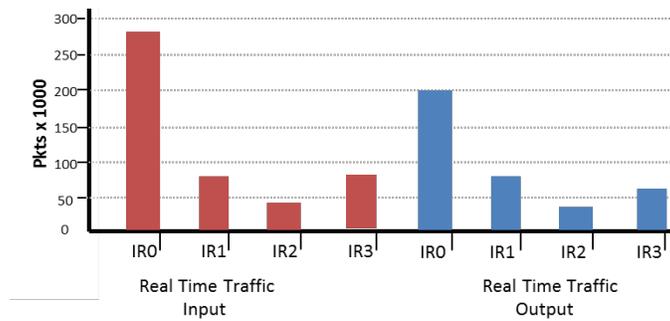


Fig. 12 Packets sent (red) and received (blue) in real time traffic

Fig. 13 shows that delay for video conferencing reach 14 milliseconds because the small size of the queue.

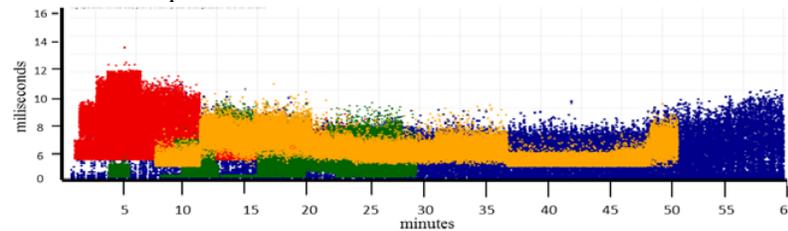


Fig. 13 Delay in real time traffic. In red video conferencing, in blue video surveillance, in green telepresence, and in yellow databases access traffic

5. Conclusions

The analytical characterization of performance in an All-IP environment is a complex task due to heterogeneous traffic sources compete for a limited bandwidth. To solve this problem a simulation tool has been developed. The tool allows the user to define different traffic sources connected to a virtual UNI to analyze the behavior of an All-IP access network.

The parameters of traffic sources can be selected for different traffic types: voice, video and data. The virtual UNI interface can also be configured in terms of bandwidth access, queues length and behavior, as well as QoS parameters (CIR, EIR.)

The presented tool can be very useful during All-IP access networks design to help optimizing different telecommunications scenarios in aspects as architecture or behavior of the sources of traffic, as well as the QoS configuration to be applied.

In this paper we have used the simulator to analyse and optimize a specific scenario. During analysis we have detected that the best effort traffic is well provisioned, but there was overprovisioning for voice traffic and under provisioning for real-time traffic. The optimization after network analysis provisioned in a better way the three traffic types, improving the performance of the All-IP access network.

References

1. ITU-T Recommendation Y. 2001- Next generation Networks General Overview
2. ITU-T Recommendation Y.2112: A QoS control architecture for Ethernet-based IP access networks (2007)
3. ITU-T Recommendation Y.2113: Ethernet QoS control for next generation networks (2009)
4. Evans M. Filfilsh C.: Deploying IP and MPLS QoS for Multiservice Networks: Theory and Practice, Morgan Kaufmann (2007)
5. OMNeT++, <http://www.omnetpp.org/>
6. INET, <http://inet.omnetpp.org/>
7. Shaikh, A.D., Blow, K.J., Eberhard, M.A., Fowler, S.A.: Language independent on-off voice over IP source model with lognormal transitions, IET Communications 7 (2013) 1449 - 1455
8. Hazewinkel, M., Weibull Distribution, Encyclopedia of Mathematics, Springer (2001)
9. ITU-T Recommendation G.711 Pulse code modulation (PCM) of voice frequencies (1990)
10. Kush A.: H.264 for the rest of us, <http://www.images.adobe.com> (2009)
11. Telefonica, Macrolan Service (<http://www.movistar.es/grandes-empresas/soluciones>) (2014)
12. Cisco Systems Inc., Cisco Catalyst 3650 Configuration Guide (<http://www.cisco.com/c/en/us/support/switches/catalyst-3650-series-switches/products-installation-and-configuration-guides-list.html>) (2014)

Low Intrusion Active Hybrid Monitor for Nodes of Sensor Networks

Marlon Navia, José Carlos Campelo, Alberto Bonastre, Juan José Serrano.

IUI ITACA, Sensor Networks Lab, Universitat Politècnica de València.
Camino de Vera, s/n, 46022, Valencia, Spain.
marnamen@teleco.upv.es; {jcampelo,bonastre,jserrano}@itaca.upv.es

Abstract. Despite the growth of sensor networks development at last years, this kind of networks suffers problems in transmission due various factors. Several systems to monitor these networks have been proposed. In this paper a low intrusion active monitor node, focused on obtaining a log of sensor node operation, is presented. For this purpose, parallel transmission is used instead of serial transmission proposed by others similar approaches. Each event in the node is coded with few bits, adding timestamp information. This document starts with a brief about proposed monitors for sensor networks, after that the connection and operation scheme of the monitor node is described, and finally preliminary operation tests are presented.

Keywords. Low intrusion, hybrid monitor, sensor networks.

1 Introduction.

Sensors networks usually experience problems or errors in operation or data transmission. Several causes can be identified, such as interferences in the transmission medium, security attacks (specially in the wireless sensor networks - WSN -[1]), adverse environmental conditions, malfunctioning nodes, and others. Although during design or implementation of this type of network, debugging and operation testing is usually made, mainly in the development phase, the conditions during deployment can be very different and usually unanticipated events arise.

To evaluate the performance of a sensor network, even in controlled conditions or in a real environment, the so called *monitoring systems* (or simply *monitors*) are used. Monitors can focus on many performance parameters, such as throughput, jitter, response time or reliability, and even sometimes, they are focused to security and intrusion detection in the network, as described in [2].

When a monitor measures any of these parameters, two approaches may be used. *Active monitors* involve hardware and/or software in the sensor nodes, interacting with it. This way, active monitors require the modification of the sensor nodes to be monitored, and thus interfering with its normal operation. Measured parameters may vary from unmonitored system, being greater the difference as the cost in the monitored node increases.

On the other hand, *passive monitors* rely on the observation of the external behavior of the system without any interference with its normal operation. No incidence on monitored system performance is caused, but only externally observable measures can be done.

A *software monitor* is usually implemented by means of a specific applications or plug-ins to the operating system, which access to the system status and report the relevant information to the user. Usually, a software monitor has deep information about the system functioning, but it may interfere with the monitored system.

A *hardware monitor* consists on electronic devices which connect to the monitored system, recollecting data from interesting system points. Hardware monitors use to be less intrusive than software monitors.

A monitor can also combine both approaches, in order to obtain a full vision of the system keeping the interference to the minimum. These are the so called *Hybrid monitors* [3].

In this paper a proposal for an active hybrid monitor, with very low intrusion is presented. This monitor records the events occurring in a node of a sensor (or control) network in a non-volatile memory (Secure Digital flash) for later analysis. Furthermore it is able to be incorporated as a part of a more complete monitoring platform, such as described in [4].

2 Related Works

Control networks or sensors, both wired and wireless, have had great development and deployment at the last years. Not only the scope is diverse (agriculture, industry and medicine among the main), but also the technology used for deployment. The case of WSN is where you see a wider range of options [5]. The monitoring systems and techniques proposed for the evaluation of such networks are not too diverse, although there are several that stand out and are detailed below:

SNMS (Sensor Network Management System) [6] is one of the best known monitoring systems. It is a complete management system, focused on working with any type of sensor network. This built on TinyOS, and allows making a review of the state of a node, or save information locally in order to avoid unnecessary network traffic.

Sympathy [7] is a tool that works as a passive monitor, and can detect and debug pre-and-post deployment errors. It operates by analyzing the data arriving at the sink of a sensor network, applying metrics, and inferring where a fault or failure can be produced in the network. The implementation of this mechanism depends on the knowledge of the network behavior. Sympathy also considers the aggregation of a small overhead on the network to increase its accuracy.

SNIF [8] and Pimoto [9] are examples of passive monitors, whose approach is to deploy a network of sniffers with two radio interfaces: one to capture all transmissions from the nodes; and another to transmit, via Bluetooth, the captured information to a device. In the first case the device works as a sink and analyzes information. In the second the device that receives the data is a computer that forwards it, via TCP/IP,

to a central server for analysis, and can also work over more than a sensor network simultaneously.

Memento [10] is an example of active monitor, which adds its code protocol to the node you want to monitor. It uses bitmaps to encode the event of a node and transmit, and can detect problems in a node based on the information provided by their neighbors in the network.

A lightweight tracing, as an active monitor, is proposed in [11]. In this proposal node events are saved by using a very light coding (3 bytes in performed test) in non-volatile memory for further reconstruction and debugging of node and network behavior.

Passive diagnosis for WSN PAD [12] is a monitor system with little intrusion, based in a probabilistic diagnosis approach, based on a Belief (or Bayesian) Network, to infer the root causes of abnormal WSN function. This adds a probe in each node that marks packets with very little overhead (2 bytes in tests).

Finally, in [13] a new three-layer reference model for WSN Monitor Platform (WSN-MP) is presented. It is based on the division of the monitoring issue in three independent layers. *Monitoring Layer* focuses on capturing, analyzing and visualizing information; *Information Layer* deals with information coherence and finally *Interchange Layer* supports information management.

3 The Proposed Monitor

A Monitoring Platform for the evaluation of the operation and performance of a Sensor Network is currently being developed. This Monitoring Platform can be easily developed by following the methodology presented in [13].

In this first approach, the internal state of the node in each moment will be measured by the monitor, obtaining a detailed trace of the states reached by the mote. The main characteristics of the monitoring platform should be:

- It must have the greatest possible independence between the application / network and monitor.
- The intrusion in both software and hardware should be minimal.
- The system must be able to record information of the nodes.

3.1 Monitoring Platform Architecture.

The first step in the design of the new WSN-MP, according to the purpose of [13], consists on the specification of the *Monitoring Layer*, where the MDP's (Monitoring Data Points) which are relevant for the desired observation have to be identified. For each MDP, a *probe* must be considered, in order to capture the relevant information at the correct time.

In this part of the Monitoring Platform, only the internal state of the mote is considered. This state is obtained from the so-called *State_MDP*.

To measure the data from *State_MDP*, a composed probe is proposed. It will be also a hybrid monitor, as far as it combines both hardware and software elements.

Software elements consist on traps located through the code of the monitored mote. Each change of state, the appropriate numeric code is sent through the physical interface. Hardware elements consist on a monitor node attached to the mote, which receives the code from the software traps by means of a physical interface. The physical communication between both elements takes place through a parallel interface as shown in fig 1.

The data transmission from sensor node is made in parallel to the monitor using 8 or 16 lines, plus a line that sends an interrupt signal whenever an event occurs. Of the 8 transported bits 4 indicate the event and 4 may be used for additional information. Additionally, if the node supports it and can provide more information, others 8 bits could be used to transmit another parameter. The connection scheme between monitor node and sensor node is shown in Fig. 1.

As previously it was mentioned, the *Information Layer* deals with information coherence. Issues related to data format and synchronicity must be solved at this point. Data format is based on ASCII text, forming a record for each trap transmitted, which includes a Time Stamp field, the Trap Code and the Trap Sub-code (optional). These fields are divided by commas, and the record is finished by a carriage Return. On the other hand, the synchronism is obtained from a Real Time Clock (RTC) located at monitoring node.

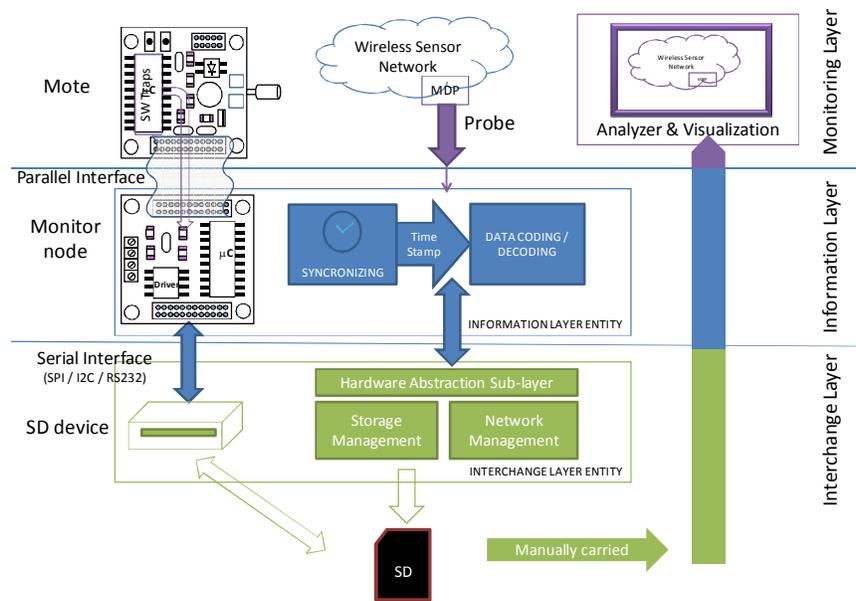


Fig. 1. Architecture of the proposed Monitoring Platform.

Finally, *Interchange Layer* supports information management. This task includes data transmission and data storage. The storage of generated log data can be local or remote. In local mode it can use a non-volatile memory, e.g. a SD memory. In remote mode several ways can be used to implement it. We can use another network, wired or wireless, to transmit the information to a sink created for this objective; or we can connect another device, e.g. a computer, and transmit the information directly to it.

In this Monitoring Platform a storage system is considered, implemented by means of a Secure Digital (SD) card. These SD will be manually carried and downloaded to the Analysis subsystem, where the appropriate algorithms will join and process all the traces obtained.

3.2 Monitor Implementation.

Our purpose is based on the left side of Fig. 1, this is the mote events capture, formatting and data storage. To test the proposed hybrid monitor node we have built an experimental node. For this we have used a sensor mote to be monitored, and the monitor node was implemented in a STM32F0Discovery board¹.

The hardware monitor is based on the STM32F051 microcontroller, a low cost and consumption ARM Cortex-M0 with several interfaces and functions, which has enough capacity to act as a monitor. The GPIO ports in the mote which are connected to the monitor node must be set up at output mode. The connection to the SD module is done through a SPI port. The connections of the mote with its transducers and RF communications devices are not relevant to the monitor node.

Table 1. Defined codes.

Code	Meaning
#define Log_Reset 0x00	//Node Reset/Initialization
#define Log_Sense0 0x01	//Read sensor 0 (first/unique)
#define Log_Sense1 0x02	//Read sensor 1 (second if it's)
#define Log_Sense2 0x03	//Read sensor 2 (third if it's)
#define Log_Wakeup 0x04	//Wake up from sleep/stop
#define Log_RxData 0x05	//Node receives data
#define Log_TxData 0x06	//Node sends data
#define Log_RxACK 0x07	//Node receives ACK
#define Log_RRoute 0x08	//Node reroutes data (if apply)
#define Log_Sleep 0x09	//Node goes to sleep mode
#define Log_Stop 0x0A	//Node goes to stop mode
#define Log_LowBat 0x0B	//Low battery indication
#define Log_SinkRx 0x0C	//Sink receives data
#define Log_SinkTx 0x0D	//Sink sends data
#define Log_SinkEr 0x0E	//Error in sink
#define Log_Error 0x0F	//Error in node

¹ <http://www.st.com/web/catalog/tools/FM116/SC959/SS1532/PF253215>

The list of possible events and their coding, defined in a header file (monitor.h), is shown in Table I. It should be noted that not all events apply to all types of nodes, and the attachment of a monitor node to the network gateway has been also considered. These events codes can be transmitted alone (one code for message) or combine several codes in the same event log. For instance, a wakeup event (Log_Wakeup) due to ACK reception (Log_RxACK) can be reflected in just one event log.

For this implementation, the additional code in the mote is minimal. We have considered only a function to send the code and an additional value if required. An example of the code, where GPIOB is used to transmit the log code, and pin 0 of GPIOA is used to activate the interruption in the monitor node, is shown below.

```

/** Pass log code and value */
void WriteLog_wVal(uint16_t cod, uint16_t val){
    uint16_t value=cod | (val<<4);
    GPIO_Write(GPIOB, value);
    GPIO_WriteBit(GPIOA, GPIO_Pin_0, Bit_SET);
    GPIO_WriteBit(GPIOA, GPIO_Pin_0, Bit_RESET);
}

```

Event logs are generated during working time, when sensor node is up, as the state of the mote evolves. In Fig. 2 an example of these events is shown. Every time an external interrupt gets to the monitor node, the associated code is added to the registry log along with a timestamp, reflecting the exact time of arrival (including fractions of a second) and any attached value, if exists. As previously mentioned, data fields are separated by a semicolon and the record finishes with a carriage return. These records are stored in a SD (Secure Digital) memory as a text file.

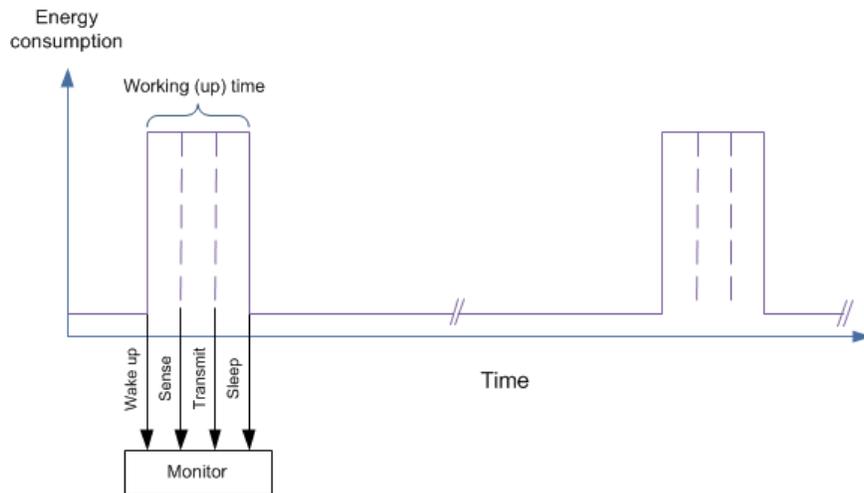


Fig. 2. Events logged in a typical working cycle.

The SD module is connected to the monitor, unlike [11] where it connects non-volatile memory directly to the sensor node. To record the information on the SD memory the open source libraries FatFs, available in [14], are used. These libraries, which are generic to work with the FAT file system microcontroller, must be adapted to the characteristics of the used hardware.

Figure 2 shows an example of events captured in a typical working cycle. During working time of the mote, several events are recorded: wake up, data capture (any sensor), data transmission, and change to sleep mode.

4 Preliminary Test

Initial tests were made to verify the correct operation of the proposed monitor node. A wireless sensor node that measures temperature each 20 seconds and sends the measurement to its gateway was considered as the mote to monitor. The Fig. 3 shows a photograph of both mote and monitor node, and their connection for the tests. The lines between them include data lines (events and additional data), and an interruption line.

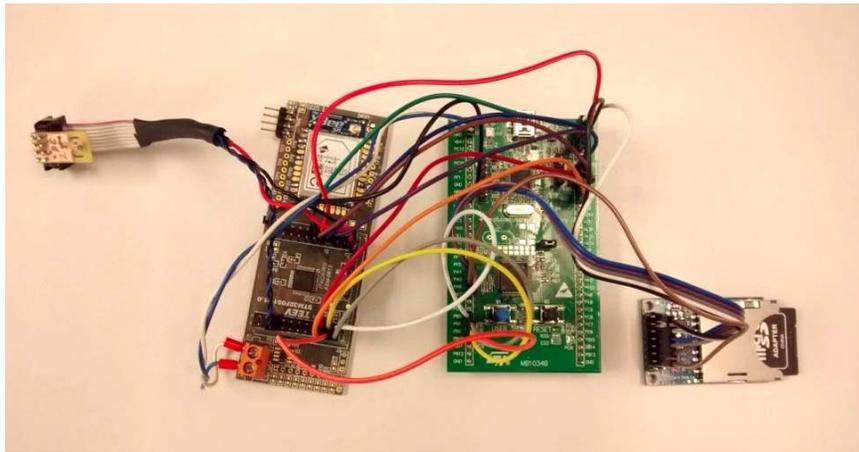
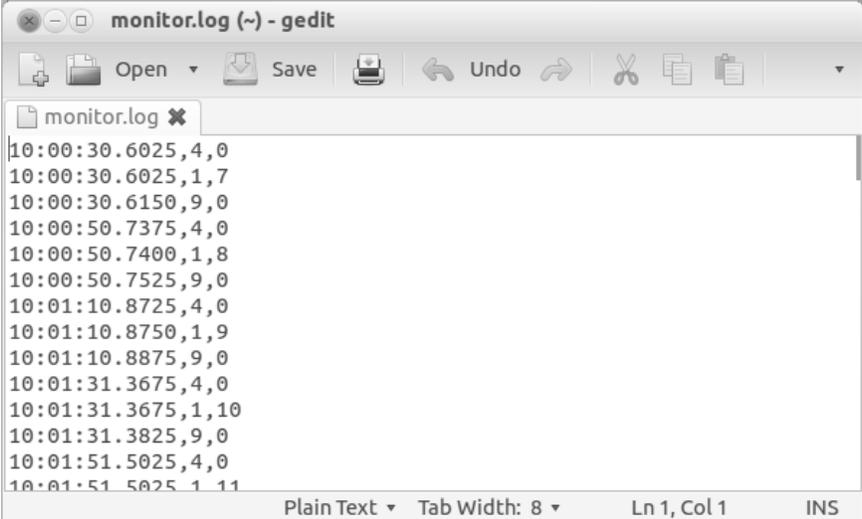


Fig. 3. Mote and developed monitor node.

The Fig. 4 shows a screen capture of content from log file generated by monitor node. Each element is separated by a semicolon. First column is the event timestamp with format hh:mm:ss.subss. The second column is the event code; and third column is an additional value (optional).

It can be appreciated that the working cycle (when node is working) is very short, just a few milliseconds.

The monitoring node is also battery powered. To optimize the monitor battery, monitor node turns into sleep mode whenever it identifies the STOP or SLEEP event in the mote, and it wakes up when the corresponding external interrupt is detected.



```
monitor.log x
10:00:30.6025,4,0
10:00:30.6025,1,7
10:00:30.6150,9,0
10:00:50.7375,4,0
10:00:50.7400,1,8
10:00:50.7525,9,0
10:01:10.8725,4,0
10:01:10.8750,1,9
10:01:10.8875,9,0
10:01:31.3675,4,0
10:01:31.3675,1,10
10:01:31.3825,9,0
10:01:51.5025,4,0
10:01:51.5025,1,11
Plain Text Tab Width: 8 Ln 1, Col 1 INS
```

Fig. 4. Content of a log file generated by monitor node.

In this prototype the synchronization of the monitor node is achieved by direct communication from a time source. This time programming may be done by means of a direct connection via USART port, in order to set up the date and time in monitor node, from a PC. Next development can include the possibility of synchronization by means of wireless communication protocol.

The overhead caused by the monitor node is minimum compared with the normal operation time of mote, as tests results show.

The stored information in the SD memory can be analyzed later in an application developed for this purpose, or by any existing program with a specific plug-in designed for this data. Also it is able to compare the log events generated with other nodes or the sink.

5 Conclusions and Future Work.

As a first step for a Wireless Sensor Network Monitoring Platform, a hybrid monitor has been implemented. This hybrid monitor is composed by software traps and additional hardware. Traps are inserted along the code of the monitored mote, and consist on a few instructions that send a predefined code through a parallel interface. A monitoring node, receiving through this parallel interface, constitutes the additional hardware. Received data is stored in a Secure Digital flash memory for later analysis,

by means of an application developed for this purpose. It is also possible to compare the log events generated with other nodes or the gateway.

This approach is a very interesting option for monitoring sensor networks, due to the very small interference caused in network operation.

This monitor is being incorporated into a more complex monitoring platform, which may complement the captured information, using several sniffer nodes, in order to obtain a complete view of network behavior.

Acknowledgements

This work was supported by the Technical University of Valencia Research Project SP20120889 and Spanish government under projects CTM2011-29691-C02-01 and TIN2011-28435-C03-0.

References

1. M. Patel y A. Aggarwal. "Security Attacks in Wireless Sensor Networks: A Survey". In IEEE International Conference on Intelligent Systems and Signal Processing (ISSP) 2013, pp. 329-333.
2. A. Rachedi. "Monitoring Mechanism for Wireless Sensor Networks: Challenges and Solutions". In I. El Emary and S. Ramakrishnan, editors, *Wireless Sensor Networks*, 2013, pp. 595-621. CRC Press.
3. A. Schoofs, G.M.P. O'Hare, and A.G. Ruzzelli. Debugging Low-Power and Lossy Wireless Networks: A Survey. *IEEE Communications Surveys & Tutorials*, Vol. 14, No. 2, 2012, pp. 311-321.
4. I. Piqueras, J. C. Campelo, R. Ors, J. J. Serrano. "Hybrid monitoring of wireless sensor networks". In IEEE International Conference on Wireless Information Technology and Systems (ICWITS), 2012, pp. 1-4.
5. S. Atanasov. "An overview of wireless communication technologies used in wireless sensor networks". In International Scientific Conference eRA-8, 2013.
6. G. Tolle and D. Culler. "Design of an Application-Cooperative Management System for Wireless Sensor Networks". In 2nd European Workshop on Wireless Sensor Networks (EWSN), 2005, pp. 121-132.
7. N. Ramanathan, K. Chang, R. Kapur, L. Girod, E. Kohler, and D. Estrin. "Sympathy for the Sensor Network Debugger". In 3rd ACM Conference on Embedded Networked Sensor Systems (ACM SenSys 2005), 2005, pp. 255-267.
8. M. Ringwald, K. Römer, and A. Vialetti. Snif: Sensor Network Inspection Framework. Department of Computer Science, ETH Zurich, Technical Report, 535, 2006.
9. A. Awad, R. Nebel, R. German, and F. Dressler. "On the Need for Passive Monitoring in Sensor Networks". In 2008 11th EUROMICRO Conference on Digital System Design Architectures, Methods and Tools, 2008, pp. 693-699.
10. S. Rost, and H. Balakrishnan. "Memento: A Health Monitoring System for Wireless Sensor Networks". In 3rd Annual IEEE Communications Society on Sensor and Ad Hoc Communications and Networks, 2006, pp. 575-584.

11. V. Sundaram, P. Eugster, and X. Zhang. "Lightweight tracing for wireless sensor networks debugging." In Proceedings of the 4th International Workshop on Middleware Tools, Services and Run-Time Support for Sensor Networks, MidSens 09, 2009, pp. 13-18.
12. Y. Liu, K. Liu, and M. Li. "Passive Diagnostis for Wireless Sensor Networks." In IEEE/ACM Transactions on Networking, Vol. 18, No. 4, 2010, pp. 1132-1144.
13. J. Capella, J. Campelo, A. Bonastre, R. Ors. "Proposal of a reference model for WSN's monitoring platforms". Sensors. In press. (2014)
14. ChaN. "FatFs generic FAT file system module". Available in http://elm-chan.org/fsw/ff/00index_e.html, 2011 . Accessed on May 2014.

Spring forces based Interactive Automatic Timed Parallel Automaton Layout Algorithm for Process Mining output Representation

Carlos Fernández-Llatas¹, Antonio Martínez-Millana¹, Jose Miguel Benedí²,
and Vicente Traver¹

¹ ITACA Institute, Universitat Politècnica de València, Spain

² PRLHT, Universitat Politècnica de València, Spain

Abstract The current standardization culture has motivated a critical increase of the normalization processes in all kind of enterprises. The design of those processes to deploy them in an integrated way in the system require an important effort that can be unaffordable manually. The use of Process Mining technologies not only can help to process experts to design those process from scratch, but also, permit a continuous evaluation of the deployment of the processes. However, to take profit of these algorithms it is crucial an automatic layout system that allows an easy understanding of the process presented.

In this paper an interactive automatic layout for Timed Parallel Automata algorithm for recalculating the graphical representation workflows automatically for better human understanding of Process Mining Algorithms is presented.

1 Introduction

Currently, the normalization of processes is one of the most common issues in all the enterprises. The processes formalizing philosophy allows, not only create efficient and effective ways to adequately coordinate all the activities of the organization, but also allows to know with precision the actual status, enabling the traceability and allowing to anticipate the problems on the processes.

However the design of process is not an easy task. The processes designed are usually an approach to the most adequate process that tends to minimize the bureaucratic impact to the process in itself in order to be effective. One of the most accepted solutions for limit the bureaucratic impact of the normalization of processes is the integration of the processes with the ICT infrastructure of the organization using Workflows [6]. However, the formalization of processes for automating systems by using ICT require a high knowledge of all the stages of the organization in order to cover all the coordination among stakeholders that is not always able to achieve.

A relatively new approach to support the design of processes is Process Mining. Process Mining [1][4] is a research field that follows the Pattern Recognition approach for supporting process designers. In this way, this methodology use the

set of events occurred in a process as the training corpus for algorithms that are able to infer Workflow based models identifying the process that can be understood by experts. In fact, Process Mining algorithms usually sacrifice accuracy in order to have a higher understandability. In addition, This methodology not only allows infer the process in an understandable way, but also allows perform a continuous analysis of the adherence of stakeholders to the process evaluating the stability, efficiency and efficacy of the process. This can be used in an Interactive Pattern Recognition approach[5] to perform a iterative design of the process.

However, to allow a easier understanding of the process, it is crucial provide an adequate visualization tool. For that, is necessary to create an automatic Workflow layout that present the process in an understandable way. PROM [9] is the most classic Process Mining tool, which provides graph automatic layout based of dot files. But, the result of those layouts are static and are not interactive with the process expert. That means that the user is not able to adequate the model to his understandability and it is the user who should habituate to the proposed layout.

In this paper, an interactive spring based force automatic layout for Workflows based on Timed Parallel Automaton (TPA) [3] is presented. This algorithm is able to interact with user by stabilizing the modifications proposed by process expert.

This paper is organized as follows. First in next section the basis of the algorithm are described. Then, in Results section the algorithm and the application developed are presented and, finally, the last section conclude the paper.

2 Spring Forces and Directed Graphs

The automatic drawing of Workflows is an open issue. Algorithms like GLEE [7] (also known as Microsoft Automatic Graph Layout (MSAGL)) have been used in literature to draw Workflows in an hierarchical way. However these kind of approaches have problems to represent large Workflows with a high density of states or transitions. One of the most common problems is the *Spaghetti effect* [1] that makes the process unreadable. This effect is due to the long transitions that produce Workflows that have a low hierarchical model.

Other research fields like chemistry or biology use graphs to represent aesthetically big structures, like molecules or multidimensional structures. There is a big amount of drawing graphs algorithms that can be used in this way [8]. Although those structures are not specifically workflows, it is possible to take the basis of these ideas to perform algorithms that build aesthetical structures that can be more easily understood by users. In addition, the use of evolutive Algorithms for modeling graphs[10] can be a solution for creating layout algorithms able to correct the modifications in the users layout making the design in an iterative way.

Our hypothesis is that the creation of an algorithm that achieves a convergence with the minimum length of the transitions can limit the *Spaghetti effect*.

To allow that, we can treat the transitions among nodes as springs. The concept of spring forces applied to graph drawing has been first presented by Eades [2]. A Spring is an artifact that generates a repulsion force when is compressed, and a attraction force when is elongated. Figure 1 represent the typical Spring Force Curve depending on the compression or elongation of the artifact as is defined by the Hookes Law.

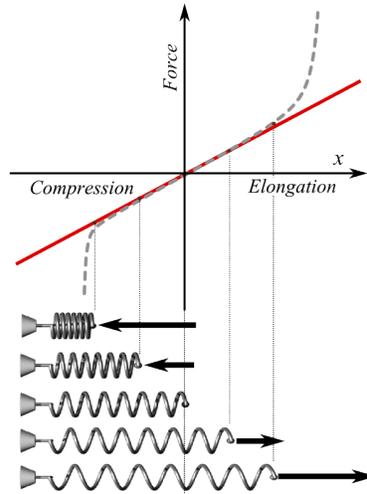


Figure 1. Spring Forces

Mathematically, the equation of the spring can be represented as:

$$F_s = C_1 * \log\left(\frac{d}{C_2}\right) \quad (1)$$

where d is the distance among the nodes and C_1 and C_2 are constants that define the stability model of the Spring.

The objective of this paper is to present and deploy an iterative algorithm in a process mining tool and test the readability in a real high density system. As Workflow representation model we have selected Timed Parallel Automaton (TPA) [3] due to their high expressivity and the availability of a web application that allow process experts to customize the layouts.

3 Results

In this section, the Interactive Spring Force layout algorithm and its mathematical basis is presented. We have selected TPA as the workflow representation model. TPA is a automaton based formalism that is able to represent parallel

situations. This is the main difference with graph, that means that $N \times N$ transitions are allowed, being N the set of nodes of a Workflow. For the algorithm, we have assumed that a transition is each combination $n_0 \times n_1$ in $N \times N$ in TPA node transitions.

Based on equation 2, the equation 2 represents the base force (SF^{ab}) of a transition among the nodes a and b :

$$SF^{ab} = K_1 * \log\left(\frac{M_{ab}}{K_2}\right) \quad (2)$$

$$M_{ab} = \sqrt{dx_{ab}^2 + dy_{ab}^2} \quad (3)$$

where K_1 and K_2 are the constant that define the behavior of the spring and M_{ab} the euclidean distance. For the algorithm we have fixed the constants to $K_1 = 2$ and $K_2 = 10 * NodeRadius$ Where $NodeRadius$ is the radius of the node of the structure.

In Workflows, even in high density problems, is very usual that two nodes have not transitions between them. This makes that, only with spring forces, is highly probable that several nodes could be placed in the same space. For that, it is necessary to add a base repulsion force among all the nodes. This repulsion force between two nodes a and b (RF^{ab}) is presented in the equation 4

$$RF^{ab} = -1 * \frac{K_3}{M_{ab}^2} \quad (4)$$

where K_3 is the offset of the decay of the force with the distance. In our problem, we have fixed K_3 to $100 * NodeRadius$. This force decay dramatically with the distance so it have a negligible effect over long distance nodes and an strong effect over very near nodes.

Both forces are applied over the distance between two points. Then it is possible to calculate a vector with the same direction than the straight between the two points. In this way, the vector $\overrightarrow{SF_{ab}}$ represents force than affects to node a from node b .

$$\overrightarrow{SF_{ab}} = \overrightarrow{d_{ab}} * \frac{SF^{ab}}{M_{ab}} \quad (5)$$

where $\overrightarrow{d_{ab}}$ is the distance vector between a and b .

In an analog way, we represent the repulsion force:

$$\overrightarrow{RF_{ab}} = \overrightarrow{d_{ab}} * \frac{RF^{ab}}{M_{ab}} \quad (6)$$

Using both forces, we can calculate the correction ($\overrightarrow{FC_a}$) that should be applied to node a in each iteration:

$$\overrightarrow{FC_a} = K_4 * \left(\sum_{i=0}^n \overrightarrow{SF_{ai}} + \sum_{j=0}^n \overrightarrow{RF_{aj}} \right) \quad (7)$$

Where K_4 is the constant that define the affectation of the force in each iteration. We have fixed K_4 to 10. The total force is calculated in each iteration using the summation of forces of all nodes.

Using all equations the algorithm is defined as follows:

Data: TPA, Number of Iterations

Result: TPA with Automatic Layout

```

while not last iteration do
  | forall the Node  $a$  in TPA do
  | | forall the Node  $b$  in TPA do
  | | |  $SF_{ab} = CalculateFAForces(a,b);$ 
  | | |  $RF_{ab} = CalculateFNForces(a,b);$ 
  | | end
  | end
  | forall the Node  $a$  in TPA do
  | | CalculateCorrection( $a,SF,RF$ );
  | end
end

```

Algorithm 1: Spring Forces Layout Algorithm

We have implemented the algorithm in a C# portable library ensuring its portability to Windows , Web, Silverlight, Windows Phone and XBox Applications. We have tested the algorithm using a localization corpus in the surgery area of one real hospital. The corpus has the indoor positioning of 521 patients that was in a surgery process between March and June of 2013.

On one hand, the Figure 2 represent the inferred flow of the process followed by the 541 patients ordered in a hierarchical way. On the other hand, the Figure 3 represent the flow once the algorithm has converged.

4 Discussion and conclusions

The presented algorithm is able to correct the positioning of the nodes over a inferred Workflow. This not only allow the user have a better first view of the process, but also, helps in the repositioning once the user has modified the workflow. This allows an interactive support to process expert, helping him in the graphic understanding of the flows. The algorithm is simple and quick and can be implemented easily. The comparison between Figure 2 and Figure 3 shows how the *Spaguetti effect* has disappeared and the workflow is much more legible.

In future work, we are planning the use of this algorithm with medium and low density Process Mining corpus in order to test the behavior of the algorithm in the distribution of different hierarchical sequences in the same workflow. Also we will perform usability validation with experts.

References

1. Wil M. P. van der Aalst. *Process mining: discovery, conformance and enhancement of business processes*. Springer, Berlin [u.a.], 2011.

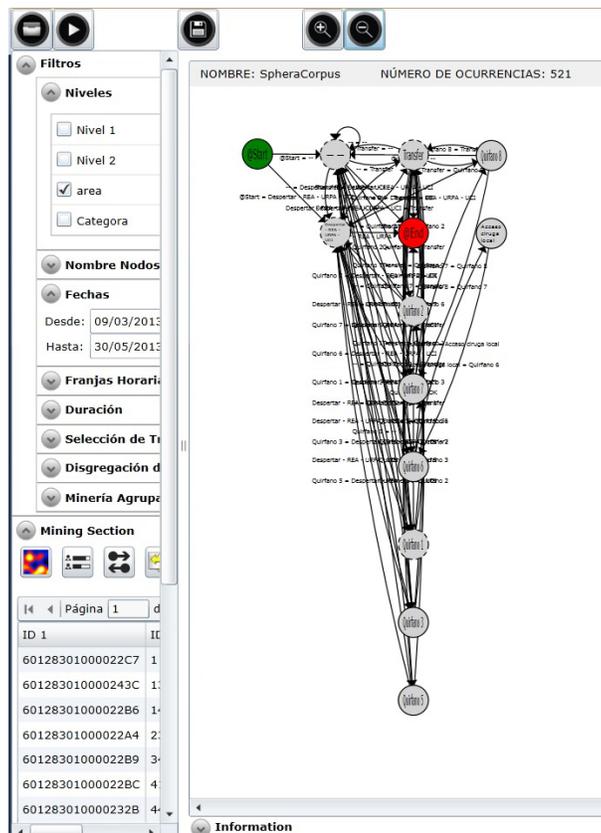


Figure 2. Workflow Inferred with Hierarchical Flow

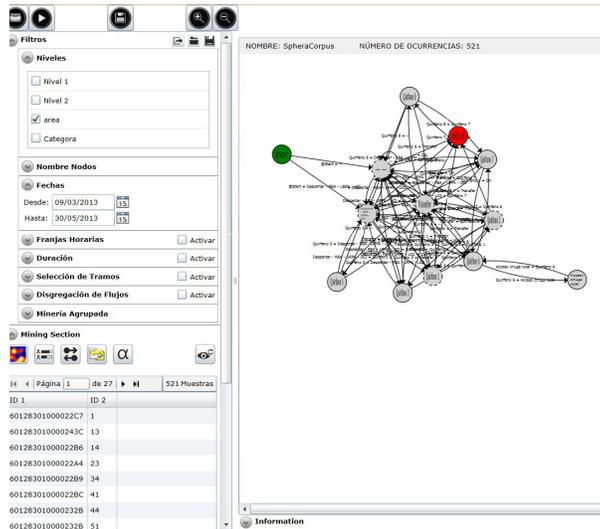


Figure 3. Workflow after iterative corrections

2. P. Eades and Lin Xuemin. How to draw a directed graph. In , *IEEE Workshop on Visual Languages, 1989*, pages 13–17, October 1989.
3. C Fernandez-Llatas, S F Pileggi, V Traver, and J M Benedi. Timed parallel automaton: A mathematical tool for defining highly expressive formal workflows. In *Modelling Symposium (AMS), 2011 Fifth Asia*, page 56â61. IEEE, May 2011.
4. Carlos Fernandez-Llatas, Teresa Meneu, Jose Miguel Benedi, and Vicente Traver. Activity-based process mining for clinical pathways computer aided design. *Conference proceedings: ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference*, 2010:6178–6181, 2010.
5. Carlos Fernandez-Llatas, Teresa Meneu, Vicente Traver, and Jose-Miguel Benedi. Applying evidence-based medicine in telehealth: An interactive pattern recognition approximation. *International Journal of Environmental Research and Public Health*, 10(11):5671–5682, October 2013.
6. Dimitrios Georgakopoulos, Mark F Hornick, and Amit P Sheth. An overview of workflow management: From process modeling to workflow automation infrastructure. *Distributed and Parallel Databases*, 3(2):119â153, 1995.
7. Lev Nachmanson, George Robertson, and Bongshin Lee. Drawing graphs with GLEE. In *Graph Drawing*, page 389â394. Springer, 2008.
8. Roberto Tamassia. *Handbook of Graph Drawing and Visualization*. Chapman and Hall/CRC, Boca Raton, edicion: new. edition, September 2013.
9. Boudewijn F van Dongen, Ana Karla A de Medeiros, H M W Verbeek, A J M M Weijters, and Wil M P van der Aalst. The ProM framework: A new era in process mining tool support. In *ICATPN*, pages 444â454, 2005.
10. Qing-Guo Zhang, Hua-Yong Liu, Wei Zhang, and Ya-Jun Guo. Drawing undirected graphs with genetic algorithms. In Lipo Wang, Ke Chen, and Yew Soon

Ong, editors, *Advances in Natural Computation*, number 3612 in Lecture Notes in Computer Science, pages 28–36. Springer Berlin Heidelberg, January 2005.

Selectable Error Detection and Correction Levels Error Control Codes

Joaquín Gracia-Morán¹, Luis-J. Saiz-Adalid¹, Pedro-J. Gil-Vicente¹,
Daniel Gil-Tomás¹, and J.-Carlos Baraza¹

¹ Instituto ITACA, Universitat Politècnica de Valencia,
Camino de Vera s/n, 46022 Valencia, Spain
{jgracia, ljsaiz, pgil, dgil, jcbaraza}@disca.upv.es

Area of Interest: Fault Tolerant Systems

Abstract. Traditionally, Error Correction Codes (ECC) works with codeword digits exposed to the same error rates. Nevertheless, with the actual height of intermittent faults, it would be interesting to divide a codeword according to possible different error rates. This work summarizes Flexible Unequal Error Control (FUEC) codes. This new codes family divides codewords into any required number of areas, establishing for each one the adequate error detection and/or correction levels. We propose an algorithm that automates the code generation process (at design time). We can apply different selection criteria in order to choose one of the codes meeting the requirements. The code generated allows fast encoding and decoding, as it is implemented by using simple logic operations.

1 Introduction

Multiple applications in today's computer systems use Error Correction Codes (ECC), such as providing reliable delivery and storage of digital data over unreliable communication channels and memories. Normally, these ECC consider that all bits in a codeword require the same error control level. Nevertheless, in different application domains, the bit error rate (BER) does not homogeneously affect to all codeword bits [1]. For example, nowadays, 6.2% of errors in memory subsystems [2], and 39% of hardware errors in microprocessors reported to operating systems have an intermittent nature [3].

In this way, the interest in asymmetric error control codes has been growing. Unequal Error Control (UEC) codes [1] establish different error control levels in diverse codeword areas. These codes split codewords in two parts: one area is strongly controlled (it can apply full error correction or burst error correction), but the second area is weakly controlled, applying only single error correction (and sometimes double error detection).

As multiple errors are becoming more and more frequent in today's systems [4], single error correction may be insufficient for the weakly controlled area. On the other

hand, full error correction in the strongly controlled area requires a high level of redundancy. Thus, the main drawback of existing UEC codes is their lack of flexibility. Designers are not able to use different error control functions on each area, according to the specifications of each system, application or context of use.

This work summarizes Flexible Unequal Error Control (FUEC) codes, a new type of UEC codes enhanced for flexibility [5]. FUEC codes establish any desired number of control areas in a codeword, deploying the adequate error control strategy in each part.

The rest of this paper is organized as follows. Section 2 overviews Flexible Unequal Error Control (FUEC) codes, while Section 3 sums up the methodology to generate FUEC codes. Section 4 presents implementation details about a FUEC code example. Finally, Section 5 provides some conclusions and ideas for future work.

2. Overview of Flexible Unequal Error Control (FUEC) Codes

Let us consider an application with variable Bit Error Rate (vBER). For instance, Volume Holographic Memories (VHM), where the BER of readout data from the edge is much higher than that from the center of the media [12]. In this case, as the BER is proportional to the distance from the center of the media, more than two areas should be considered.

We can illustrate how FUEC works with a simple example. For the sake of simplicity, we will consider 12-bit words, divided in three 4-bit areas (see Fig. 1(a)).

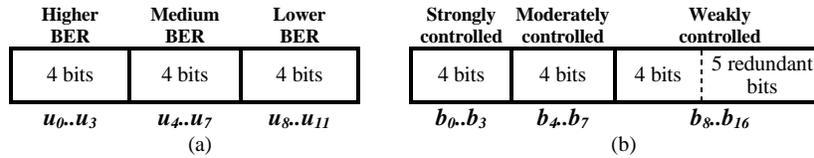


Fig. 1. Layout of the input word (a) and the encoded word (b).

We will consider next fault hypothesis:

- i. Strongly controlled area (higher BER): all 1-bit errors and all 2-bit burst errors should be corrected, and all 2-bit errors and 3-bit burst errors should be detected;
- ii. Moderately controlled area: all 1-bit errors and all 2-bit burst errors must be corrected, with no additional detection;
- iii. Weakly controlled area (lower BER): 1-bit error correction and 2-bit burst error detection will be implemented.

With these conditions, we will need 5 redundant bits (see Fig. 1(b)) as explained later [5]. With this value, we can find an H matrix to solve the error control requirements. Next, the methodology used to generate FUEC codes is applied to this example.

3. Methodology Description

Different parameters must be set when designing a code:

- i. the data length (k);
- ii. the encoded word length (n) and
- iii. the set of error vectors to be corrected (E_+) and detected (E_Δ).

Also, other parameters have to be considered to define E_+ and E_Δ when FUEC codes are to be generated:

- i. the number of control areas of the codeword;
- ii. the boundaries (that is, the first and last bits) of each area and
- iii. the error control level to apply to each area.

With these parameters, and if it exists, it is possible to obtain an $((n-k) \times n)$ \mathbf{H} matrix able to correct and detect the selected errors. The methodology proposed considers all possible matrices, taking into account that \mathbf{H} must satisfy condition (1), that is, there must be a different syndrome for each correctable error. The condition for additional error detection is defined in (2). In other words, each detectable error must generate a syndrome which is different to all the syndromes generated by the correctable errors. However, and as they have to be detected, several detectable errors may present the same syndrome.

$$\mathbf{H} \cdot \mathbf{e}_i^T \neq \mathbf{H} \cdot \mathbf{e}_j^T; \forall \mathbf{e}_i, \mathbf{e}_j \in E_+ \mid \mathbf{e}_i \neq \mathbf{e}_j \quad (1)$$

$$\mathbf{H} \cdot \mathbf{e}_i^T \neq \mathbf{H} \cdot \mathbf{e}_j^T; \forall \mathbf{e}_i \in E_\Delta, \mathbf{e}_j \in E_+ \quad (2)$$

where $\mathbf{e} = (e_0, e_1, \dots, e_{n-1})$ models the error induced by the channel. If no error has occurred in the i th bit, $e_i=0$; otherwise, $e_i=1$.

Searching \mathbf{H} can be considered a Boolean satisfiability (SAT) problem. Previous proposals to solve this problem [6][7] are focused on specific applications. Our proposal is more general: in three successive steps, our algorithm is able to find any binary linear block code, if it exists, just selecting the set of error vectors to be corrected. So, the first step is to determine E_+ and E_Δ . Then, the algorithm tries to find an \mathbf{H} matrix able to solve conditions (1) and (2). Finally, as several solutions can be found, one of them can be selected using different criteria.

Determining E_+ and E_Δ Error Vector Sets. According to the example summarized in Fig. 1, the set of correctable errors is $E_+ = E_0 \cup E_1 \cup E_{B2}^{0..7}$. It includes the no-error vector (E_0), all single bit errors (E_1) and all 2-bit burst errors in the areas with higher and medium BER ($E_{B2}^{0..7}$). To find an error correcting code, the condition $|E_+| \leq 2^{n-k}$ must be satisfied ($|E_+| = |E_0| + |E_1| + |E_{B2}^{0..7}| = 1 + n + 7 \leq 2^{n-12}$). From this expression, $n \geq 17$, that is, at least 5 redundant bits are required. If $n = 17$, $|E_+| = 25$

and $2^{n-k} = 32$. The layout of the encoded word is shown in Fig. 1(b), and the vectors representing the errors to be corrected are included in Table 1.

Table 1. Vectors representing the errors to be corrected in the considered example.

Correctable errors (E_+)	Error subset
(0000000000000000)	No error (E_0)
(1000000000000000) (0100000000000000) (0010000000000000) (0001000000000000) (0000100000000000) (0000010000000000) (0000001000000000) (0000000100000000) (0000000010000000) (0000000001000000) (0000000000100000) (0000000000010000) (0000000000001000) (0000000000000100) (0000000000000010) (0000000000000010) (0000000000000001)	Single bit errors (E_1)
(1100000000000000) (0110000000000000) (0011000000000000) (0001100000000000) (0000110000000000) (0000011000000000) (0000001100000000)	2-bit burst errors in the areas with higher and medium BER ($E_{B_2}^{0..7}$)

Let us define now E_Δ . As decided in the example design requirements, $E_\Delta = (E_2^{0..3} - E_{B_2}^{0..3}) \cup E_{B_3}^{0..3} \cup E_{B_2}^{7..16}$. The seven (32–25) syndromes not used for correction are employed for the detection of these error vectors, grouped in Table 2. Depending on the requirements of the code to be designed, other sets of error vectors can be generated.

Table 2. Vectors representing the errors to be detected in the considered example.

Detectable errors (E_Δ)	Error subset
(1010000000000000) (1001000000000000) (0101000000000000)	2-bit random errors in the area with higher BER, excluding correctable errors ($E_2^{0..3} - E_{B_2}^{0..3}$)
(1x10000000000000) (01x10000000000000)	3-bit burst errors in the area with higher BER ($E_{B_3}^{0..3}$)
(0000000110000000) (0000000011000000) (0000000001100000) (0000000000110000) (0000000000011000) (0000000000001100) (0000000000000110) (0000000000000011) (0000000000000001)	2-bit burst errors in the areas with lower BER ($E_{B_2}^{7..16}$)

Computing the Parity-check Matrix (H). With the previously selected set of error vectors (E_+ and E_Δ), the algorithm has to find an H matrix that satisfies the conditions (1) and (2). A huge computational effort is required as all the matrices are considered. To lighten this process, a recursive backtracking algorithm has been developed. It checks partial matrices and adds a new column only if the previous matrix satisfies the requirements, discarding those matrices that do not meet the requirements.

The algorithm starts with a **partial_H** matrix formed with $n-k$ rows and only one column. If this matrix accomplishes the requirements, new columns are added recursively. Both the initial and the added columns must be non-zero, so there are $2^{n-k} - 1$ combinations for each column. The pseudo code of the partial matrix checker procedure is presented in Fig. 2.

The Hamming weight is used to check whether the error vectors have all their 1s in the first n_{cols} bits. This is needed because it is impossible to calculate the syndrome for an error vector with 1s in columns not included in the partial matrix.

```

Procedure CheckPartialMatrix partial_H  $(n-k) \times n_{cols}$  /*  $n_{cols} \in [1..n]$  */
  SyndromeSet = {}
  For each error vector e in  $E_+$ 
    partial_e =  $(e_1, e_2, \dots, e_{n_{cols}})$ 
    If HammingWeight(e) = HammingWeight(partial_e)
      newSyndrome = CalculateSyndrome(partial_H  $\times$  Transpose(partial_e))
      If newSyndrome in SyndromeSet then Return /* Not valid partial matrix */
      Else Add newSyndrome to SyndromeSet
    End if
  End for
  For each error vector e in  $E_\Delta$ 
    partial_e =  $(e_1, e_2, \dots, e_{n_{cols}})$ 
    If HammingWeight(e) = HammingWeight(partial_e)
      newSyndrome = CalculateSyndrome(partial_H  $\times$  Transpose(partial_e))
      If newSyndrome in SyndromeSet then Return /* Not valid partial matrix */
      Else Do Nothing /* newSyndrome not stored in this case */
    End if
  End for
  If  $n_{cols} = n$ 
    Add partial_H to SolutionsSet
    Return
  Else
    For each possible new_column /*  $n-k$  bits, excluding the all 0 combination:  $2^{n-k}-1$  possible values */
      CheckPartialMatrix [partial_H | new_column]  $(n-k) \times (n_{cols}+1)$ 
    End for
  End if
End procedure

```

Fig. 2. Partial matrix checker procedure.

The first loop adds syndromes for all selected error vectors to **SyndromeSet**, verifying also if the new syndrome generated has been previously added. If this is the case, two different error vectors generate the same syndrome, so **partial_H** cannot be part of a valid solution. If the algorithm arrives at the end of the first loop, all selected error vectors have different syndromes.

Then, a second loop calculates syndromes for the detectable errors. If these syndromes have been previously added to **SyndromeSet** (i.e. the syndrome is associated to a correctable error), **partial_H** cannot be part of a valid solution. Unlike the first loop, the calculated syndromes are not added to **SyndromeSet**, as these syndromes are associated to more than one detectable error. If the algorithm arrives at the end of the second loop, all selected correctable errors have non-equal syndromes, and the detectable errors have distinct syndromes from those used for correction.

At this moment, **partial_H** has n columns, so a solution has been found, adding it to **SolutionsSet**. If **partial_H** presents fewer columns, the third loop will add the necessary columns.

Even if the condition $|E_+| \leq 2^{n-k}$ is satisfied, it is not guaranteed the existence of a FUEC code with the selected requirements. If no solution is found, **SolutionsSet** is empty (the searched code does not exist). The solution is to increase the redundancy or to reduce the number of error vectors to be corrected or detected. Once **SolutionsSet** is obtained, a solution can be chosen according to different criteria:

- First matrix found. This option reduces the H generation time.

- Smallest Hamming weight of H. This solution commonly reduces the number of logic gates in a hardware implementation, etc.

Next paragraph presents the code obtained for our case study, applying the smallest Hamming weight of H criterion.

4 Code Implementation

Attending to the error vectors shown in Table 1 and Table 2, and selecting a solution with smallest Hamming weight in H, one possible solution is shown in Table 3, where \mathbf{u} is the input word $\mathbf{u}=(u_0, u_1, \dots, u_{k-1})$ that represents the original data, \mathbf{b} is a vector of n bits $\mathbf{b}=(b_0, b_1, \dots, b_{n-1})$ where the code adds the required redundancy, and \mathbf{r} is the received word $\mathbf{r}=(r_0, r_1, \dots, r_{n-1})$.

In this case, \mathbf{u} is part of \mathbf{b} , and the parity bits are located in the columns with only one 1. For example, bit b_{13} of the codeword is a parity bit, because the corresponding column in \mathbf{H} has only one 1, in the second row. As it can be seen in (3), each parity bit is calculated by XORing the bits with a 1 in its row. Similarly, \mathbf{s} is calculated using \mathbf{r} .

Table 3. FUEC code for the considered example.

b/r_0	b/r_1	b/r_2	b/r_3	b/r_4	b/r_5	b/r_6	b/r_7	b/r_8	b/r_9	b/r_{10}	b/r_{11}	b/r_{12}	b/r_{13}	b/r_{14}	b/r_{15}	b/r_{16}
u_0	u_1	u_2	u_3	u_4	u_5	u_6	u_7	u_8	u_9	u_{10}	u_{11}	See (3)				
1	0	1	0	0	1	0	0	1	1	1	1	1	0	0	0	0
0	1	0	1	0	0	1	0	1	1	0	0	0	1	0	0	0
0	0	1	0	1	0	1	1	1	0	1	0	0	0	1	0	0
1	0	0	1	1	0	1	0	0	1	0	1	0	0	0	1	0
0	1	0	0	1	1	0	1	0	0	1	1	0	0	0	0	1

The formulas for the proposed code are (being s_i the calculated syndrome):

$$\begin{aligned}
 b_i &= u_i, \forall i \in \mathbb{N} : 0 \leq i \leq 11 \\
 b_{12} &= u_0 \oplus u_2 \oplus u_5 \oplus u_8 \oplus u_9 \oplus u_{10} \oplus u_{11} & s_0 &= r_0 \oplus r_2 \oplus r_5 \oplus r_8 \oplus r_9 \oplus r_{10} \oplus r_{11} \oplus r_{12} \\
 b_{13} &= u_1 \oplus u_3 \oplus u_6 \oplus u_8 \oplus u_9 & s_1 &= r_1 \oplus r_3 \oplus r_6 \oplus r_8 \oplus r_9 \oplus r_{13} \\
 b_{14} &= u_2 \oplus u_4 \oplus u_6 \oplus u_7 \oplus u_8 \oplus u_{10} & s_2 &= r_2 \oplus r_4 \oplus r_6 \oplus r_7 \oplus r_8 \oplus r_{10} \oplus r_{14} \\
 b_{15} &= u_0 \oplus u_3 \oplus u_4 \oplus u_6 \oplus u_9 \oplus u_{11} & s_3 &= r_0 \oplus r_3 \oplus r_4 \oplus r_6 \oplus r_9 \oplus r_{11} \oplus r_{15} \\
 b_{16} &= u_1 \oplus u_4 \oplus u_5 \oplus u_7 \oplus u_{10} \oplus u_{11} & s_4 &= r_1 \oplus r_4 \oplus r_5 \oplus r_7 \oplus r_{10} \oplus r_{11} \oplus r_{16}
 \end{aligned} \tag{3}$$

According to the value of the syndrome bits, the error can be corrected or detected, as shown in Table 4.

Formulas proposed in (3) can be easily implementable. The encoder is as simple as a XOR tree or equivalent circuitry or algorithm per parity bit. Syndrome bits are calculated in the same way. Correction can be implemented using a 5-to-32 binary decoder and some logic gates. It is important to note that no correction is performed in case of "error detection".

Table 4. Syndrome lookup table (estimated errors) for the proposed code.

$s_4 s_3 s_2 s_1 s_0$	Error in...	$s_4 s_3 s_2 s_1 s_0$	Error in...	$s_4 s_3 s_2 s_1 s_0$	Error in...	$s_4 s_3 s_2 s_1 s_0$	Error in...
0 0 0 0 0	No error	0 1 0 0 0	bit r_{15}	1 0 0 0 0	bit r_{16}	1 1 0 0 0	Detection
0 0 0 0 1	bit r_{12}	0 1 0 0 1	bit r_9	1 0 0 0 1	bit r_5	1 1 0 0 1	bit r_{11}
0 0 0 1 0	bit r_{13}	0 1 0 1 0	bit r_3	1 0 0 1 0	bit r_1	1 1 0 1 0	bits r_6, r_7
0 0 0 1 1	Detection	0 1 0 1 1	bit r_9	1 0 0 1 1	Detection	1 1 0 1 1	bits r_6, r_1
0 0 1 0 0	bit r_{14}	0 1 1 0 0	Detection	1 0 1 0 0	bit r_7	1 1 1 0 0	bit r_4
0 0 1 0 1	bit r_2	0 1 1 0 1	bits r_4, r_5	1 0 1 0 1	bit r_{10}	1 1 1 0 1	Detection
0 0 1 1 0	Detection	0 1 1 1 0	bit r_6	1 0 1 1 0	bits r_3, r_4	1 1 1 1 0	Detection
0 0 1 1 1	bit r_8	0 1 1 1 1	bits r_2, r_3	1 0 1 1 1	bits r_1, r_2	1 1 1 1 1	bits r_5, r_6

This methodology can be used not only to design FUEC codes. As an example of its versatility, in [8] we have generated different codes that improve codes presented previously [9][10][11]. By using the methodology presented in Section 2, we have generated a Hamming code modified to detect 2-bit and 3-bit burst errors, maintaining the Single Error Correction feature, with no extra redundancy and without increasing the encoder and decoder latencies.

More information about FUEC codes can be found in [5].

5 Conclusions and Future Work

This paper presents the flexibility, feasibility and potentials of Flexible Unequal Error Control (FUEC) codes. These codes are able to establish any desired number of control areas in a codeword, deploying the adequate error control capabilities in each area. This challenge is of great interest when variable bit error rates are present.

With these premises, FUEC codes may result of interest in automotive, aerospace or avionics industries, where different sources of interference, noise or process variations may result in areas with variable bit error rates. For instance, intermittent faults affecting these applications increase the fault rate in the erroneous bits. In addition, the occurrence of multiple faults makes necessary to consider diverse error patterns.

Respect to the computational time required to complete the algorithm, the performance of current version has to be improved. A parallel version of the algorithm is currently under development. However, the FUEC codes included in the paper have been obtained using the sequential version.

Acknowledgement. This work has been partially supported by the Universitat Politècnica de València under the Research Project DesTT (SP20120806), and the Spanish Government under the Research Project ARENES (TIN2012-38308-C02-01).

References

1. E. Fujiwara, Code Design for Dependable Systems, John Wiley & Sons, 2006.
2. C. Constantinescu, "Intermittent faults and effects on reliability of integrated circuits", Procs. Reliability and Maintainability Symposium (RAMS 2008), pp. 370-374, January 2008.

3. E.B. Nightingale, J.R. Douceur, and V. Orgovan, "Cycles, cells and platters: an empirical analysis of hardware failures on a million consumer PCs," *Procs. ECCS*, April 2011.
4. P. Gil et al., "Fault Representativeness", Deliverable ETIE2 of Dependability Benchmarking Project, IST-2000-25245, 2002.
5. L.J. Saiz-Adalid et al., "Flexible Unequal Error Control Codes with Selectable Error Detection and Correction Levels", *Procs. SAFECOMP*, pp. 178-189, September 2013.
6. S. Shamshiri and K.T. Cheng, "Error-Locality-Aware Linear Coding to Correct Multi-bit Upsets in SRAMs", *IEEE International Test Conference*, pp. 1-10, November 2010.
7. A. Dutta and N.A. Touba, "Reliable network-on-chip using a low cost unequal error protection code", *Procs. DFT*, pp. 3-11, September 2007.
8. L.J. Saiz-Adalid et al., "Modified Hamming Codes to Enhance Short Burst Error Detection in Semiconductor Memories", *Procs. EDCC*, pp. 62-65, May 2014.
9. C. L. Chen and M. Y. Hsiao, "Error-correcting codes for semiconductor memory applications: a state-of-the-art review", *IBM Journal of Research and Development*, vol. 28(2), pp. 124-134, 1984.
10. A. Sanchez-Macian, P. Reviriego, and J.A. Maestro, "Enhanced Detection of Double and Triple Adjacent Errors in Hamming Codes through Selective Bit Placement", *IEEE Transactions on Device and Materials Reliability*, vol. 12(2), pp. 357-362, June 2012.
11. A. Sanchez-Macian, P. Reviriego and J.A. Maestro, "Hamming SECD AED and Extended Hamming SEC-DED-TAED codes through selective shortening and bit placement", *IEEE Transactions on Device and Materials Reliability*, vol. 14(1), pp. 574-576, March 2014.
12. W. Chou, and M. Neifeld, "Interleaving and error correction in volume holographic memory systems," *Applied optics*, vol. 37(29), pp. 6951-6968, 1998.

Towards a service-oriented platform for the customized analysis of biomedical data quality

Pablo Miñarro¹, Carlos Sáez¹, Juan Miguel García-Gómez¹

¹Instituto de Aplicaciones de las Tecnologías de la Información y de las Comunicaciones Avanzadas - ITACA. Universitat Politècnica de València.

Abstract. The lack of data quality assurance in biomedical context is a problem that can lead to bad decisions and poor research conclusions. Contributing to solve this problem we present the development of a web service based tool for the measurement of biomedical data quality. It approaches the problem using a combination of different dimensions of data quality and six axes of data. Hence, it is designed as scalable software model based on the strategy design pattern. In addition, it offers a web-based graphical user interface to facilitate the on-line use and the customization of data quality assessments. This work describes the current development of the platform architecture and completes it with the discussion of a use case.

1 Introduction

The lack of quality assurance in biomedical data is a problem that may affect both clinical practice and research which reuse this data, as errors in data can lead to errors in diagnosis or incorrect conclusions. Therefore, it is deemed necessary the creation of a tool that helps to assure a certain degree of data quality (DQ). In this paper we present the software platform to measure DQ we are working into. We will give an overview of the software architecture on which it is based, being designed as a DQ service-oriented platform, and we will discuss a use case as an example.

Many studies have been developed within the biomedical domain about DQ assessment methods [1]. They are mainly based on measuring DQ dimensions, however there is not a general agreement on what to measure, and specially how to do it. In a previous work, we introduced an ongoing work on a DQ model for biomedical domains [6]. Such work identified a set of nine dimensions in which quality can be measured, which were defined according to the literature. These dimensions are: completeness, consistency, duplication, correctness, temporal stability, spatial stability, contextualization, predictive value and reliability

On the other hand, different DQ tools exist outside the biomedical domain, as reviewed in the Gartner Magic Quadrant for DQ Tools [2]. Generally, they offer different set of profiling methods to capture general statistics in order to identify data quality issues. As an example, Talend's "Data Quality" [3] and Microsoft's SQL "Data Quality Services" [4] provide tools to data standardization and elimination of

inconsistencies, as well as they allow for data cleaning and profiling. However, those tools are only focused to completeness, consistency and the checking of duplicate data. Hence, there is little attention on the definition of context-specific measurements, or a lack of a more advanced analysis of changes along different data sources or through time [9]. These topics are present in biomedical data domains, thus while the application of these tools may be profitable, more specific methods are deemed necessary in the context of biomedical data, as this may positively impact in the data analysis if the method is closer to the data characteristic of study.

In this work we present the development of a service oriented platform for the analysis of biomedical data and its architectures as part of the presented DQ framework [6].

2 The service oriented DQ platform

Our solution proposal is a web service offered to third parties for quality control, and more specifically the creation of a web service as a solution, which serves as a start point for the posterior export to libraries, plug-in creation, and the integration with other tools.

The usage of service-oriented-architecture (SOA) provides benefits in terms of simplification in design, decomposing problems into smaller ones by modularization [5]. Also SOA makes usage of services, each one doing a specific task, which can facilitate the scalability and the system maintainability, where services may be oriented to both final users and third parties software tools. SOA tools are often web based services, and this provides an easy to use and a platform independent tool through internet.

To outside, the system must be easy to access and with the most generic inputs and outputs. Thus, we have chosen to use a CSV-formatted input for data as well as different parameters to select the DQ evaluation configuration.

On the other hand, the platform must be sufficiently flexible and maintainable, as it is focused to solve an evolving research field as biomedical DQ. As a solution we have chosen the implementation of a flexible DQ measurement model which permits the measure of quality of data in different dimensions, and how this dimension affects to different axes of our data. This solution relies on the Strategy and Factory Method design patterns [7].

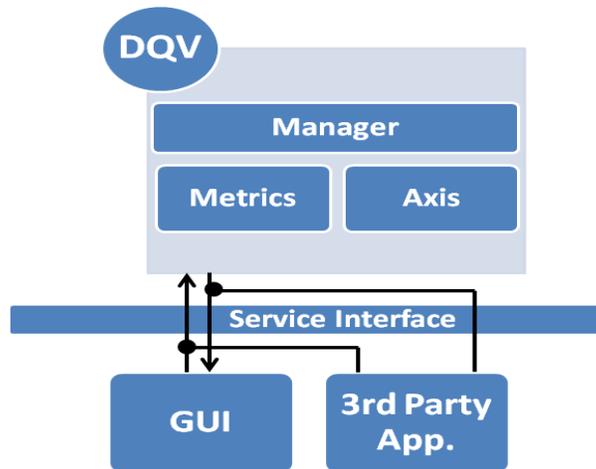


Figure 1. Platform Schema

Fig. 1 shows the schema of our platform. The web service DQV is composed by a Manager, which makes use of metrics and axes. The GUI and other third parties applications make use of this service through a service interface.

3 DQ measurement architecture

Data input in biomedical domains are typically composed of registries and attributes. The values can be observations or context variables. These datasets usually contain different data types and some lost values.

We distinguish between six different applicable axes in the previous defined dimensions in our framework. These axes are: registry, attribute, value, dataset, time and source:

- Registry: The subject of fact, such as patient, episode, contact etc. Conceptually would be rows in a matrix.
- Attribute: The set of the different variables of the dataset. It will be equivalent to the columns of a matrix.
- Value: Represents each data individually, is where the information of data is contained.

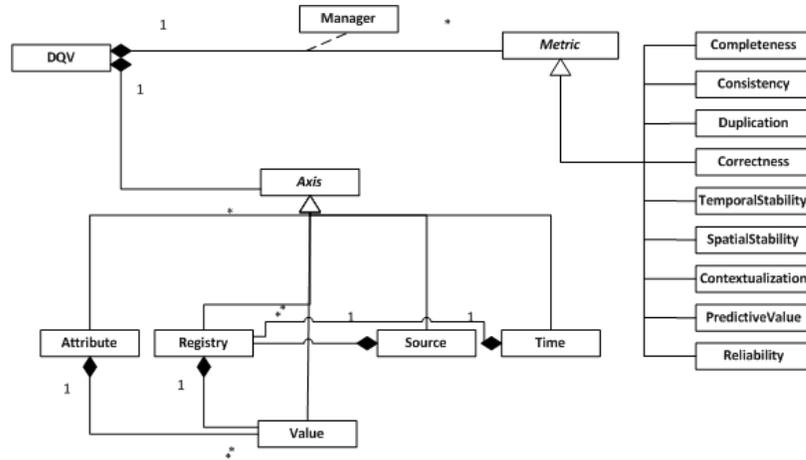


Figure 2. DQ Measurement Architecture

- Dataset: The entire set of values, that is, the whole dataset with all the data.
- Time: It allows comparing data along different periods of time. It can be understood as a special case of a certain attribute variable, the one which tell us when the data has been collected.
- Source: It allows us to compare between data, in this case between different sources where the data has been collected. This axis can also be seen as a special case of an attribute.

Having nine dimensions and six axes provide too many routines to maintain, so we needed to look for a solution that facilitated the scalability and maintenance. To avoid a high number of classes and methods, we have chosen an implementation of the strategy pattern, therefore common abstract classes are inherited into a more specific class, with contains its owns methods.

As a result of studying the functional requirements and the chosen strategy, we propose the architecture showed in the Fig. 2.

The dataset is represented as a set of specific axis classes, such as Attribute, Registry, Source, Time and Value, which inherits form a common “Axis” abstract class. Our Attribute, Registry, Source and Time each contains a set of Value classes, which contains the values of our data.

The DQV class contains all the sets of Attributes, Registries, Source and Time classes and all the Value classes, it also includes the metrics for measuring the data quality in all the dimensions.

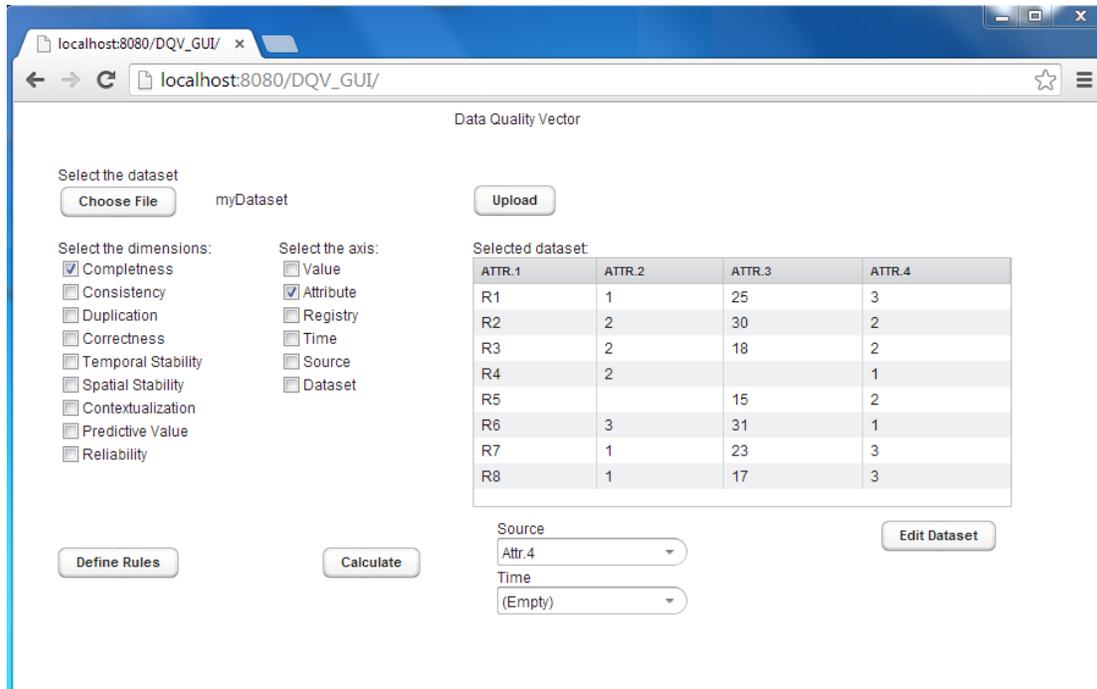


Figure 3. Graphical User Interface

To perform the calculations we have defined a series of classes corresponding to each method to calculate the different dimensions. These metrics, such as Completeness and Consistency in the figure, are inherited from an abstract class “Metric”, such as described in the strategy design pattern.

For dealing with the complexity of having six axes to calculate in each dimension, each one corresponding to the desired axis, we have chosen the usage of an association class “Manager”. This class is in charge of dealing with the different ways of preparing the structure to calculate the dimensions with over the correct axis reusing the metric code.

4 Implementation of DQ dimensions

The complete development of the service will require the combination of advanced techniques and knowledge in biomedical informatics, such as information theory and geometry, machine learning or semantic web. Hence different dimensions require the implementation of specific technology.

Each quality dimension makes reference to different aspects of information; therefore each one uses different methods. Thus, as spatial and temporal stability dimensions are based on statistics and information theory, correctness and predictive value dimensions are based on machine learning; duplication, completeness, consistency and contextualization use rule based systems and semantic web. Finally reliability is based in score generation as accumulative indicators. Thus, the scalable architecture presented before will permit the incorporation of different methods with minor effort as the development of the platform advances.

5 Graphical User Interface

In Fig. 3, we show the first prototype design of our Graphical User Interface (GUI) which is constructed using Vaadin [8]. We have designed the GUI to be easy to use as well as useful for the purpose of our solution. Thus, the user only has to submit the data and select the dimensions and axis to calculate. Furthermore the user has also the options of checking and editing some values of the data if necessary.

The declaration of each attribute type, such as categorical or numeric values, is also possible. Also the option to define rules for the data (to calculate the consistency dimension) is considered.

For temporal and spatial stability, we have two combo boxes to select the corresponding attribute to time and source, which if present, allow the calculation of those dimensions.

6 An use-case of data quality measurement

To show an use-case we will use the synthetic data showed in Fig. 3, defined from current experiments on public databases of breast cancer and heart disease.

For instance, we can calculate the completeness for the axis attribute of the selected dataset file which contains the data, as is showed in Fig. 3.

After uploading the dataset and selecting the completeness dimension along with the axis attribute, we would just have to press over “Calculate” to get the results. As you can notice in Fig. 3, there are two missing values one in the column labelled “Attr.1” and another in the column labelled “Attr.2”.

In case the attribute option is selected for the completeness dimension, the Metric class will use an algorithm to send all the attributes one by one, to get the results of completeness for each attribute. This algorithm will count both the filled and empty values for that attribute and return the results of the percentage of completeness for each attribute as showed in Table 1.

In Table 1 we can observe the selected dimension of our completeness use-case, and the attribute axis where we want to apply the dimension. The results for the four attributes of our dataset labelled as Attr.1, Attr.2, Attr.3 and Attr.4 are shown indicating the percentage of complete data in each attribute.

Table 1. Results of completeness with attribute axis

Completeness	Attribute			
	<i>Attr.1</i>	<i>Attr.2</i>	<i>Attr.3</i>	<i>Attr.4</i>
Result	100%	87.5%	87.5%	100%

7 Conclusions and future work

The current development on a service-oriented platform for biomedical DQ assessment has been introduced. To deal with complexity the implementation of the strategy design pattern has been used, which will facilitate the scalability and maintainability of the service. The DQ platform, as a web service can be accessed by third parties software or by means of the prototype GUI that has been introduced.

Current and further work is including into the platform more functional dimensions. The inclusion of an automatic report system for the presentation of the results is also being studied.

References

- [1] N. G. Weiskopf and C. Weng. Methods and dimensions of electronic health record data quality assessment: enabling reuse for clinical research. *J Am Med Inform Assoc*, 20(1):144–151, Jan 2013.
- [2] <http://www.gartner.com/technology/reprints.do?id=1-1LB9WX9&ct=131007&st=sb> Last accessed 2014/01/15.
- [3] <http://www.talend.com/resource/data-quality.html> Last accessed 2014/01/15.
- [4] <http://msdn.microsoft.com/en-us/library/ff877925.aspx> Last accessed 2014/01/15.
- [5] Prakash M. Nadkarni and Randolph A. Miller, Service-oriented Architecture in Medical Software: Promises and Perils. *J Am Med Inform Assoc*. 2007 Mar-Apr; 14(2): 244–246.
- [6] Sáez C, Martínez-Miranda J, Robles M, García-Gómez JM. Organizing quality assessment of shifting biomedical data. *Stud Health Technol Inform*. 2012;180:721-5.
- [7] Gamma, E. ; Helm, R. ; Johnson, R. ; Vlissides, J. : Design Patterns: Elements of Reusable Object-Oriented Software
- [8] <https://vaadin.com/> Last accessed 2014/01/15.
- [9] [Sáez C, Robles M, García-Gómez JM. Comparative study of probability distribution distances to define a metric for the stability of multi-source biomedical research data. In: Engineering in Medicine and Biology Society (EMBC), 2013 35th Annual International Conference of the IEEE; 2013. p. 3226–3229]

Seasonal Variations of the Earth Rotation Imprint into Noise Processes

V. Milián-Sánchez¹, V.N. Lesnykh², V.A. Kolombet², A. Mocholí-Salcedo³, A. A. Medvedeva², G. Verdú-Martín¹

¹*Institute for Industrial, Radiophysical and Environmental Safety, Universitat Politècnica de València, Camino de Vera, s/n, Valencia, Spain.*

²*Institute of Theoretical and Experimental Biophysics, Russ. Acad. Science, Pushchino, Russia.*

³*Grupo de Sistemas de Control de Tráfico, Instituto ITACA, Universitat Politècnica de Valencia, Camino de Vera, s/n, Valencia, Spain*

Abstract. In this paper we present the current research activity of a group formed by researchers at the Institute of Theoretical and Experimental Biophysics of the Russian Academy of Science (ITEB RAS) in Pushchino, Russia, and at the Polytechnic University of Valencia (UPV), Spain. The research is based on the Shnoll effect which manifests that the scatter in experimental measurements, i.e. the patterns of the corresponding data histograms, is not entirely random. One manifestation is the existence of stellar and solar daily periods in those histograms. We outline the objectives and means of the research group, the current activity and the provisional results, which basically consist in the development of a new method to analyze the experimental results, and the observation of fluctuations in daily periods according to the yearly seasons, which in turn is hypothesized to be related to tidal forces.

Key Words - Macroscopic fluctuations, instrumentation, measurements, sensor network.

1. INTRODUCTION

In the 1960s Simon Shnoll and coworkers suggested that the scatter in experimental measurements is not entirely random [1,2], this effect being caused by some general cosmophysical (cosmogonical) factor(s). This can be seen by comparing the histograms obtained in a series of experiments: whether the processes consist in the rate of a chemical reaction, the fluctuation noise of voltage in a resistor [3], the rate of dark current measured in a photomultiplier [4], the radioactive decay or any other process generating a random signal, a sequence of histograms can be obtained from that signal by dividing the time series in non-overlapping consecutive intervals (Fig.1). The histograms obtained from those intervals are smoothed several times by the running average method to make the comparisons more convenient between each other, and from those comparisons several characteristics described as “macroscopic fluctuations” can be observed [5,6,7,8,9,10].

Altogether, these characteristics constitute the Shnoll effect. A brief description of the Shnoll effect can be found in [11] and a detailed description of the research of Shnoll

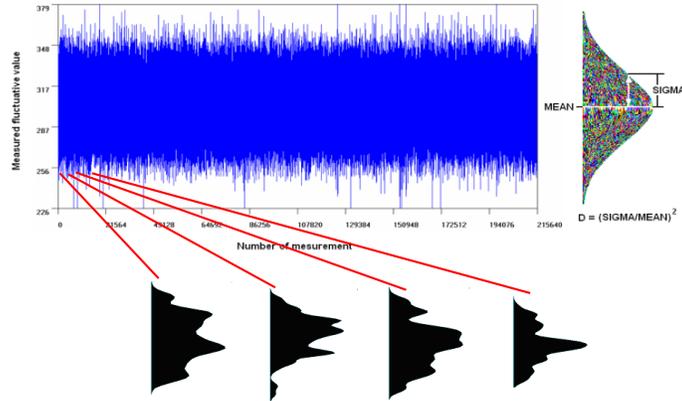


Fig. 1. Noise signal. A sequence of histograms is obtained by dividing the time series in non-overlapping intervals.

and coworkers can be found in his book [12], but in what follows we describe some of their findings:

- a) Similar histograms can be visually found in time series (Fig. 2),

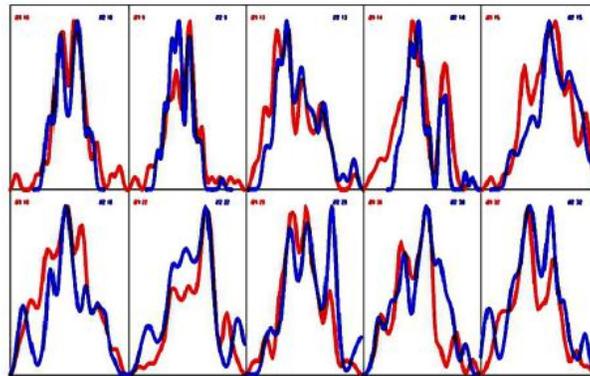


Fig. 2. Superposition of similar (smoothed) histograms.

The smoothed histograms can be more easily compared since the coincidence of peaks (and troughs) and also their height (and depth) can be better analyzed, [13].

- b) Repetition of histogram shapes evolves in accordance with Earth rotation and a constant temporal distance between them can be observed. So, Fig. 3 shows the daily periods, i.e., the probability of finding similar histograms is maximal every 1436 and 1440 min, which coincides with stellar periods referred to fixed stars and to solar period, respectively:

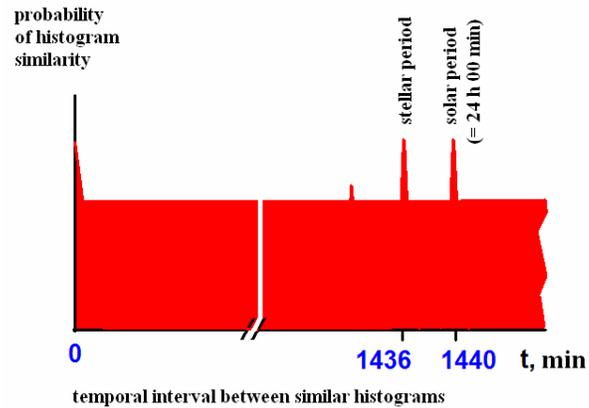


Fig.3. Stellar (every 1436 min) and solar (every 1440 min) daily periods.

Aside, similar periodicities of the sidereal month, lunar evection period, lunar day, calendar year and sidereal year can be found.

- c) Absolute time synchronism, in which the histograms shapes correlate between different experiments, regardless of the distance between the places where they are performed.
- d) Local time synchronism, in which histograms shapes have a high probability to be similar with a time delay given by the difference in longitudes of the experiments locations.
- e) A “near zone” effect can be also observed, which shows that the nearest histograms have a much higher probability to be similar.
- f) A palindrome effect has been presented in [14], the effect consisting in that succession of histogram shapes since 6 am till 6 pm of local time is like a reverse (inverse) histograms succession since 6 pm till 6 am of a following day. This is related to the direction of laboratory rotation corresponding to that of the Earth during its diurnal rotation.
- g) When two radiation beams are West- and East directed, no synchronous similarity of the histograms can be found: instead, similar histograms occur in 12 hours.

h) Moreover, the shape of histograms depends on the direction in space of radiation beam. So the effect of the daily periods is more pronounced in certain situations (see Fig.4, histograms from device No2 obtained in Pushchino), but also the daily periods can almost disappear: this happened to a detector directed upwards in the North Pole [15] (Fig. 4):

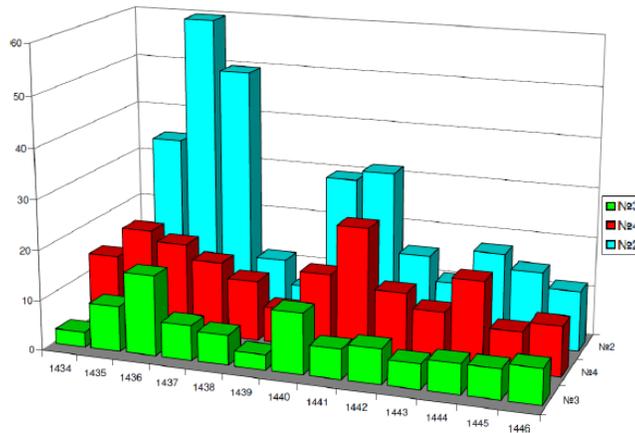


Fig.4. This is Fig. 2 in [15]. Device No.2 in Pushchino with west-ward directed collimator in the laboratory leads to the appearance of similar histograms with the two periods of a sidereal day (1,436 min) and a solar day (1,440 min). Measurements at the North Pole: a) the effect is attenuated for a flat detector (device No.4) directed horizontally sideways, b) the effect almost disappears for a detector directed upwards (device No. 3).

All these results have been obtained by visual comparison, carried out by highly experienced researchers, which is called the expert approach. This means that it is the expert who takes into account the location of the peaks (and troughs) and also their height (and depth), (Fig. 2), and who decides for each pair of histograms, whether they are similar or not. Based on these previous studies, our aim is to continue this research.

2. OBJECTIVES

Our research goals are, at least, to confirm the aforementioned observations by other means, and explore other possible aspects that have not been extensively studied. For instance, it would be of interest to check if the mentioned effects remain unchanged along many previous years and in years to come, and to analyze the histogram evolution depending on the spiral movement of the Earth around the sun on the height of

the center of our Galaxy (and other compact star clusters in our Galaxy) above the horizon [16].

The histogram patterns have been analyzed, so far, mainly by very skilled and trained experts. It is therefore our aim to try to develop a new approach to simplify the work, which could consist in analyzing the histogram pattern by a group of people. This can be helpful as a previous step to develop a computer program to automatize the work.

The signals commonly used are white noise, which follow the Poisson statistics with no preferred frequencies. However, it could also be analyzed the statistics associated to every single segment from where the histograms are drawn. The aim of that study will be to find out if there are some patterns which correlate with histograms behavior.

Most of the results regarding these regular-in-time changes of histogram shapes have been obtained using ^{239}Pu . It could be worth analyzing those effects with other types of isotopes like Tl-204 and others which have not been so extensively used. The works of Jenkins et al. [17] about anomalous behaviors in nuclear decay rates can be an important complementary information source to attempt to find patterns in non-random effects with different nuclides.

The experiments mentioned in the literature with radiation detectors refer mainly to liquid scintillators, solid state devices or Geiger-Müller counters [7, 17]. We can also use a multichannel analyzer to check the intensity of these effects at different energy levels, and also, to try to discriminate those effects in different kinds of radiation.

As pointed out before, aside from ^{239}Pu and other nuclides, other noise signals have been used so far, such as: dark current from photomultipliers [4], thermal noise in a resistor [3], or light fluctuation intensities of a light-diode [12]. But other kind of signals could be analyzed and obtained from other electronic devices, like random noise generators or some specific instruments with the capability to provide a Gaussian noise.

These proposed works can be also complemented with the search for any kind of other factors influencing the measurements results (light, temperature, shielding [18], etc.). For instance, directing the sources beams towards the Centaurus constellation, wherefrom most of the high energy cosmic rays comes into the Earth [19].

Regarding the methodology, instruments for radiation detection must be checked regularly, comparing the obtained measurements with predictions of statistical models.

Old stored records can be compared with newly obtained data, with regular clocks adjustments to ensure synchronicity in measurements held in Pushchino and Valencia.

The new data series will be analyzed by a group of participants.

3. MEASUREMENTS IN PROGRESS

Currently, six radiation detectors are continuously working in ITEB facilities. The data is stored in special data files.

An alpha particles counter (designed by Anatoli Pisakov) has been transferred from ITEB to the nuclear physics lab at the Chemical and Nuclear Engineering Department

in UPV. It is being used in connection with a computing module Arduino Uno with a program developed in ITEB by M.E. Astashev. Data is recorded onto a 1 Gb microSD card. The counter and data acquisition system is able to collect data from different alpha radiation sources. An initial test with a low activity ^{241}Am source has been performed.

Initial tests have been carried out with a random noise generator Hameg HM8035. Another initial test is in progress with an Agilent 33220A with noise generation capability. These tests can be performed with different RC filters to discriminate frequencies and to select specific bandwidths.

Initial tests are in progress with Ludlum 2200 scaler using different radiation sources along with two Geiger-Müller counters. The first step has been performed verifying that the counting equipment follows the predictions of statistical models. These tests will be also used as reference in further measurements.

An initial test has been also performed with a scaler to analyze the noise generated by the instrument. The statistics is good enough to perform comparisons with other noise sources.

4. INSTRUMENTATION

The following instrumentation is available in the laboratories:

In ITEB:

Six radiation detectors.

A rotary device which holds the sources and points the radiation beams at any direction in space. It can be also programmed to keep the radiation beam directed to a certain point in space (Fig. 5):



Fig. 5. Rotary device that points automatically the sources' beams to a fixed point in sky.

In UPV laboratories (nuclear physics and electronics):

One alpha particles counter.

Four scalars, Ludlum M2200

Four Geiger Müller counter tubes ZP1430 (Mullard) and four lead containers.

One multichannel analyzer (Ortec)
 100 MHz Oscilloscope (Hameg 1007)
 20 MHz pulse generator (Hameg HM80359)
 Triple power supply (Hameg HM8040-2).
 Function and arbitrary waveform generator (Agilent 33220A)
 Digital multimeter (Agilent 34401A)

5. Provisional Results and Discussion

First histogram analyses have been performed with the “composite expert” method. In order to do so, a member of the team prepares the data which are forwarded to the other participants, (Fig. 6). After the data is analyzed by the group, the results can be compared with those obtained by the expert. It can be observed that the results are almost the same because the found peaks are shifted about ± 1 min, (Fig. 6). And as an additional result, one can see that the stellar and solar peaks deviate in several minutes from the theoretical 1436 and 1440 min depending on the season of the year (Fig. 6):

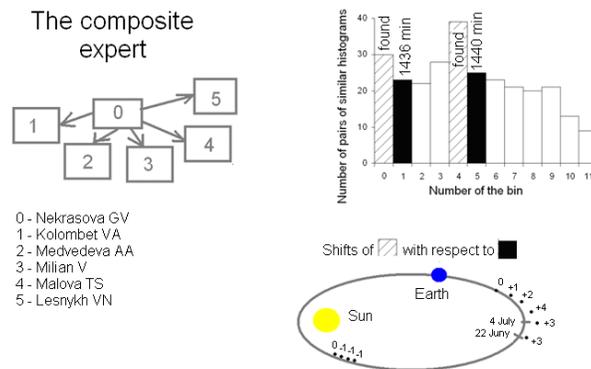
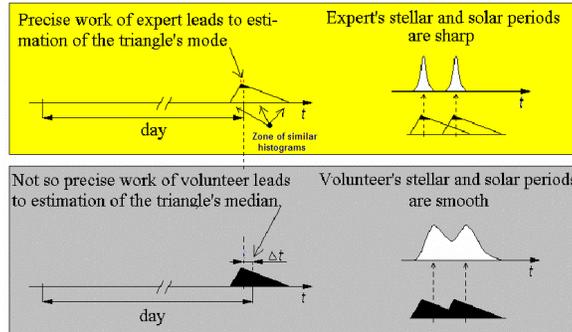


Fig. 6. Peaks obtained by the “composite expert” can slightly deviate from the theoretical ones, but results are good enough. Shift of peaks varies along the year seasons.

This result shows that this new method is good enough to analyze histogram: when the expert method is used, the expert finds a distribution like the one shown in Fig. 4 (counter N°2), where the histograms bins are sharply distributed around the peaks. One can see a kind of triangles (in black) around the sharp peaks (Fig. 7, upper part), and the two corresponding distributions do not overlap. The expert finds the mode of the triangle.

In contrast, the analyses carried out by the group are not so precise (Fig. 7, lower part): the two distributions around the peaks overlap but the result is good enough. The group finds the median of the triangle.



Expert finds the **mode** of the triangle, volunteer - the **median**.

We try to explain trends in the expert-volunteer difference as presence of seasonal variations of the triangle shape.

Fig. 7. Upper part: the expert finds two differentiated peaks corresponding to stellar and solar periods and thus the mode of the triangle associated with the distribution. Lower part: the group finds the triangle's median.

As said before, from this “composite” analysis, the peak's positions changes along the year can be observed. It is hypothesized by one of the authors (VA Kolombet) that daily periods are synchronized with tensions caused by tidal forces in Earth (Fig. 8). Thus, as a first practical application, this method could lead to a new method of measuring tidal forces.

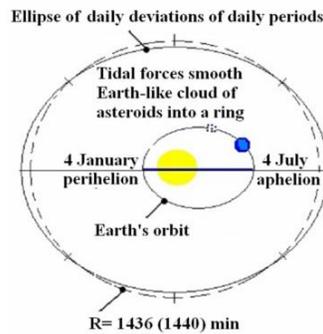


Fig. 8. Tidal forces as a hypothesis about the nature of the observed effect.

Another practical application refers to the question of whether and how cosmophysical factors can affect the measurements obtained with usual instrumentation, especially high precision measurements.

REFERENCES

- [1] S.E. Shnoll, V.A. Kolombet, E.V. Pozharskii, T.A. Zenchenko, I.M. Zvereva, and A.A. Konradov, “Realization of discrete states during fluctuations in macroscopic processes”, *Physics – Uspekhi* 162(10), 1998, pp.1129–1140. <http://ufn.ioc.ac.ru/abstracts/abst98/abst9810.html#d>.
- [2] S.E. Shnoll, V.A. Kolombet, E.V. Pozharskii, T.A. Zenchenko, I.M. Zvereva, and A.A. Konradov, “Cosmophysical origin of ‘macroscopic fluctuations’”, *Biophysics* 43(5), 1998, pp. 864–870 (in russ.).
- [3] A.V. Kaminsky, S.E. Shnoll, “The study of synchronous (by local time) changes of the statistical properties of thermal noise and alpha-activity fluctuations of a ^{239}Pu sample”, arXiv: physics/0605056.
- [4] M.V. Fedorov, L.V. Belousov, V.L. Voeikov, T.A. Zenchenko, K.I. Zenchenko, E.V. Pozharski, A.A. Konradov and S.E. Shnoll, “Synchronous changes in dark current fluctuations in two separate photomultipliers in relation to Earth rotation”, *Astrophysics & Space Science* №1, 2003, pp.105-112.
- [5] S.E. Shnoll, “Macroscopic fluctuations with a discrete distribution of amplitudes in the processes of various nature”. *Advances in Science and Engineering. General Problems of Physico-Chemical Biology*, v. 5, 1985, pp. 130-201.
- [6] S.E. Shnoll, N.V. Udaltzova and N.B. Bodrova, “Macroscopic fluctuations with discrete structure distributions as a result of universal causes including cosmophysical factors In: Proc. First Intern. Congress on Geo-cosmic Relations. Wageningen, Netherlands, 1989, pp.181-188.
- [7] S. E. Shnoll, V. A. Kolombet, E. V. Pozharskii, T. A. Zenchenko, I. M. Zvereva, A. A. Konradov, “Realization of discrete states during fluctuations in macroscopic processes”, *Physics-Uspekhi* 41,(10), 1998, pp. 1025-1035.
- [8] S.E. Shnoll, E.V. Pozharski, T.A. Zenchenko, V.A. Kolombet, I.M. Zvereva, A.A. Konradov, “Fine Structure of Distributions in Measurements of Different Processes as Affected by Geophysical and Cosmophysical Factors» *Phys. Chem. Earth (A)*, Vol.24, No.8, 1999, pp. 711-714.
- [9] S. E. Shnoll, T. A. Zenchenko, K. I. Zenchenko, E. V. Pozharskii, V. A. Kolombet, A. A. Konradov, “Regular variation of the fine structure of statistical distributions as a consequence of cosmophysical agents” *Physics – Uspekhi* 43 (2), 2000, pp. 205-209.
- [10] S. E. Shnoll, I. A. Rubinshtejn, K. I. Zenchenko, V. A. Shlekhtarev, A. V. Kaminsky, A. A. Konradov, N. V. Udaltsova, “Experiments with rotating collimators cutting out pencil of alpha-particles at radioactive decay of ^{239}Pu evidence sharp anisotropy of space”, 2005, <http://arxiv.org/abs/physics/0501004>.
- [11] S. E. Shnoll, “Changes in Fine Structure of Stochastic Distributions as a Consequence of Space-Time Fluctuations”, *Progress in Physics*, V 2, April, 2006, pp. 39-45.
- [12] Shnoll S. E. *Cosmophysical Factors in Stochastic Processes*. American Research Press, Rehoboth (NM, USA), 2012.

-
- [13] S. E. Shnoll Cosmic Physical Factors in Random Processes. Svenska Fiskarkivet, Stockholm, 2009, (in Russian).
- [14] I.A. Rubinshtein, S. E. Shnoll, A.V. Kaminsky, V.A. Kolombet, M. E. Astashev, S.N. Shapovalov, B. I. Bokalenko, A.A. Andreeva, D. P. Kharakoz, “Dependence of Changes of Histogram Shapes from Time and Space Direction is the Same when Fluctuation Intensities of Both Light-Diode Light Flow and ^{239}Pu Alpha-Activity are Measured”, Progress in Physics, V3, July 2012.
- [15] S. E. Shnoll, M. E. Astashev, I.A. Rubinshtein, V.A. Kolombet, S.N. Shapovalov, B. I. Bokalenko, A.A. Andreeva, D. P. Kharakoz, I.A. Melnikov, “Synchronous measurements of alpha-decay of ^{239}Pu carried out at North Pole, Antarctic, and in Puschino confirm that the shapes of the respective histograms depend on the diurnal rotation of the Earth and on the direction of the alpha-particle beam”, Progress in Physics, V3, July 2012.
- [16] Morozov N.A. Influence of the central body of the galaxy and its other dark super-suns on the geophysical and meteorological phenomena that appear around us. Archive of Russian Academy of Sciences, fund 543 (personal fund of N.A.Morozov, Hon. Member of the USSR Academy of Sciences), folder 01, file 72 (188 pages).
- [17] J.H. Jenkins, E.Fischbach, D.Javorsek, R.H.Lee, P.A.Sturrock, “Concerning the time dependence of the decay rate of ^{137}Cs ”, Applied Radiation and Isotopes 74, 2013, pp. 50-55.
- [18] G. Piccardi, The Chemical Basis of Medical Climatology, Charles C. Thomas Publisher, Springfield, Illinois U.S.A.
- [19] J. Abraham, Gregory Snow and Collaboration, Pierre Auger, “Correlation of the Highest-Energy Cosmic Rays with Nearby Extragalactic Objects”, Research Papers in Physics and Astronomy, <http://digitalcommons.unl.edu/physicssnow/64>

An automatic algorithm for the separation of left and right lungs from computed tomography images

Noelia Ruiz¹, Montserrat Robles¹, Roberto Sanz²

¹Instituto de Aplicaciones de las Tecnologías de la Información y de las Comunicaciones Avanzadas (ITACA), Universidad Politécnica de València, Camino de Vera s/n, 46022 Valencia, España

²Hospital Quirón Valencia, R&D Biomedical Engineering, Av. Blasco Ibáñez 14, 46010 Valencia, España

Abstract. In computed tomography (CT) imaging lung segmentation is a precursor to most of imaging pulmonary analysis applications. However, many segmentation algorithms do not include the problem of separating connected lungs, so many quantitative analysis may be affected when this problem appears. For this reason, the following article will present a lung segmentation algorithm which will allow to detect if the lungs are connected (at image level) to each other, and in that case, the algorithm carries out the lung separation in an automatic way. The method consists of three main steps: firstly, the pulmonary parenchyma and the airways are extracted using thresholding and a connected components analysis; then, the airways are removed using a 3D region growing technique; thirdly, a 3D connectivity analysis is done to determine whether there is a connection between the lungs. In case it is true, a 2D connectivity analysis is performed to identify the slices where the connectivity exists; finally, using the active contours technique and a sequence of morphological operations the lung separation is performed. In this way, both lungs are separated in an automatic way, decreasing the user's dedication time and making easier the posterior emphysema and pulmonary volumes measurements. This method has been proved using a set of CT images belonging to a significant case of this problem, obtaining satisfactory results.

Keywords: Image segmentation, X-ray CT, imaging pulmonary, lung separation.

1 Introduction

Computed tomography (CT) is the chosen image modality to obtain lung images in vivo. This imaging technique produces a 3-D volumetric image dataset and it has demonstrated better sensitivity for early detection of lung diseases than other conventional modalities. Advances in CT imaging technology has simultaneously reduced image acquisition time and increased spatial resolution so it is able to offer thin-section CT examinations which allow showing better the anatomic details of lung structures. Nevertheless, all above means a greater data volume and the radiologists have to make a great effort and dedicate more time in order to analyze the whole 3-D volumetric image data. For this reason, systems which allow a high level of automa-

tism in the analysis project are required, especially if quantitative measurements will be done, such as emphysema and pulmonary volumes measurements [1-3].

To be able to detect and quantify abnormalities in certain anatomical structures using CT images, for any methodology, it is required that a first step will be the localization and segmentation of the area of interest. Therefore, the segmentation is a necessary condition for any automatic analysis of medical imaging. In the case of CT lung imaging, it is necessary that lung segmentation is precise, since if there is an error this could bring about an erroneous quantitative analysis. The patient's anatomical variability causes that in some cases the lungs are close together. This means that at image level they appear practically connected due to partial volume effects or owing to the poor resolution of images. There are few methods proposed in literature to solve this problem and as far as we are concerned they do not take into account the different anatomical situations which could be given in the clinic [4-6]. On one hand, in the methodology proposed by Hu et al. [4] different morphological operations are used to perform the lungs separation in an iterative way. The separation implies morphological erosion and a posterior conditional dilatation in order to restore the borders of the lungs without joining the lungs again. Although this method shows suitable results, it is limited when different anatomical situations could be presented in the clinic. On the other hand, De Nunzio et al. [5] propose a completely 3D algorithm which detects the fusions between lungs and by using the region growing technique the corresponding surfaces to the fusions are created. However, this methodology is not robust to those cases in which there are pulmonary alterations caused by diffuse pulmonary diseases, since it has not been tested with pulmonary volumes with several diseases. Finally, in the methodology proposed by Park et al. [6], they use the CT information in a sequential way and a guided dynamic algorithm to achieve the lungs separation. Although their results are promising, the robustness of the algorithm is not fully validated.

The following paper presents a process for the automatic lung segmentation which allows separating the lungs using a methodology which can potentially adapt itself to a great range of anatomical features. Thus, this paper is organized as follows: in Methods, after describing the data set used for development and testing, the algorithm structure will be explained in detail step by step. Then, the results are presented in Results and Discussion. Finally, conclusions and prospects follow in Conclusion.

2 Methods

2.1 Data set description

The proposed method is developed from a set of CT images which are representative of connected lungs problem. The images are obtained with a CT scanner of 64 detectors (Philips Brilliance 64, Philips Healthcare, The Netherlands). The set of images are reconstructed from a standard filter type 'B'. The volume contains 220 slices, each one with a thickness of 3 mm and with a separation between them of 1.5 mm (50% of overlapping). The acquisition matrix off-plan is 512 x 512, with a spatial resolution of 0.98x0.98 mm.

2.2 Methodology

The proposed algorithm to separate lungs in an automatic way when they appear connected consists of several steps. Firstly, the extraction of the lung region will be done using the thresholding method and a connected component analysis. Then, the airways extraction using the 3-D region growing technique will be performed [6-8]. Thirdly, lungs connectivity will be detected doing a 3D connectivity analysis [9]. In case that connectivity will be detected, a 2D connectivity analysis will be established to find the slices which present such connectivity between lungs. Finally, making use of active contours and a sequence of morphological operations, the lungs separation will be done in those slices which have been detected in the previous step.

The following sections will describe with more detail the above steps, which will be processed in a sequential way until it will be achieved to separate totally the lungs in the corresponding volume. All the methodology has been developed in Matlab R2013a (The Mathworks Inc., USA).

Identification of pulmonary region

The first step of the algorithm will have the aim to distinguish the pulmonary parenchyma from the rest of tissues which will be showed in the set of images. In the CT images from healthy patients, pulmonary parenchyma has substantially less attenuation than the rest of tissues because it presents between 80 and 90% of air. This makes that X-rays have minor attenuation when going through such tissue. On account of this, it will be used a thresholding method as a previous step to distinguish the pulmonary parenchyma in the set of images. The alveolus and interstice make up the pulmonary parenchyma, which composes the greatest part of lungs in the CT images. It has a density between -800 and -900 HU (Hounsfield units). Taking into account the density of the parenchyma, it will be established a value of -800 HU as a threshold to realize the thresholding of the CT images. After applying the thresholding step, the next result will be obtained:

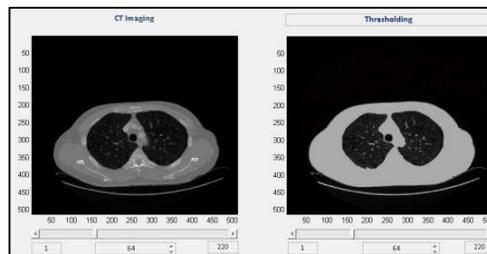


Fig. 1. Thresholding. Left: Original image. Right: Image after thresholding step.

Pulmonary region segmentation

Once the differentiation of pulmonary parenchyma in the CT images is achieved, the pulmonary region segmentation will be done as follows. Given that the pulmonary region is composed of lungs and airways, if we select a point in the trachea and do a

3D connected components analysis, as everything will be connected, the component related with trachea will be the same as the component associated to lungs. Therefore, if this component is represented, both lungs and airways will be shown in the image. The selection of this initial point will be the unique interaction which will be necessary in the whole process.

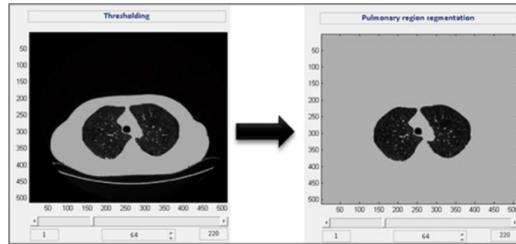


Fig. 2. Pulmonary region segmentation. Left: image after thresholding step. Right: pulmonary region segmentation.

Airways extraction

Inside the pulmonary region, the anatomical structure of interest will be the lungs, so we must extract the airways from the preceding volume as a previous step for its separation. In CT images, the airways appear as a hypointense tubular structure surrounded by a brighter wall, the region growing technique is the common approach to segment this structure which is based on variations of values of grey. In our methodology a 3D region growing technique will be used to manage to extract the airways from the previous volume, as we can see in the following figure:

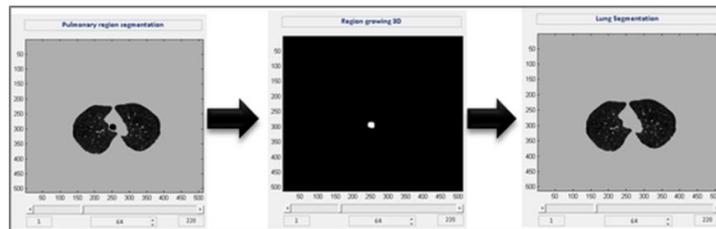


Fig. 3. Airways extraction and obtention of the anatomical structure of interest. Left: pulmonary region. Middle: airways extraction using region growing. Right: lungs segmentation.

Automatic lungs separation

After the above step, the anatomical structure of interest will be available. So it will be the next step to analyze the connectivity between lungs. This connectivity may be caused by partial volume effects or by the poor resolution of the image. This causes an apparent fusion between lungs, leading to interpret in an erroneous way the posterior measurements of pulmonary parenchyma volume or emphysema. The problem has been tackled developing a totally automatic algorithm which will allow to

analyze the whole volume which has been obtained in the previous step. This algorithm will be able to detect the fusion between lungs to carry out their separation. All the steps of the algorithm can be seen in the following figure, each of them will be explained in more detail in the subsequent sections:

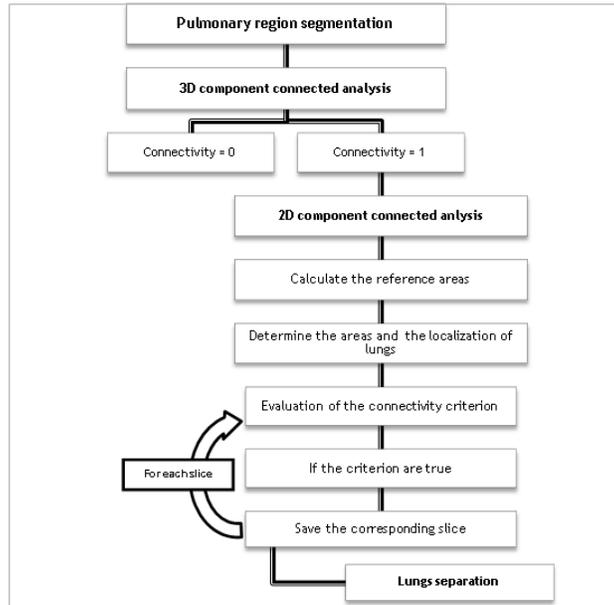


Fig. 4. Scheme of the proposed methodology

3D connected components analysis

Firstly, a 3D connected component analysis will be performed on the anterior volume in order to determine if there is connectivity between lungs. In order to detect such connectivity, the volumes of each component, which are the output of the analysis, are sorted in a descendant way and the two biggest ones are selected. In the case the first area is 100 times bigger than the second area, it will establish that connectivity between lungs exists, so it would mean that it has been found a large, simple and connected component, since a large volume has been labelled with the same label which would correspond with both lungs.

2D connected components analysis.

After checking the connectivity at a 3D level, a 2D connected component analysis will be performed to discover the slices where the connectivity between lungs is present. This analysis will begin from a slice where it is appreciated in a clear way that both lungs are separated. In this case, the slice in which it has been initialized the region growing method is chosen (step 2).

1. Calculation and initial localization of the areas by slice.

From the initial slice to the end of the volume, a 2D connected component analysis will be done slice by slice. It will determine the labels (right lung or left lung) of each of the connected components that have been found in each slice. To achieve this, the areas and the localization of each component will be calculated. The association of the two biggest areas to right lung or left lung will be done based on the localization in the image of the first pixel of each component area. In case of finding a unique area, this will be considered as right lung by default. It will be in the next step, where it will be determined whether this area corresponds to right or left lung or it reflects connectivity between both lungs.

2. Evaluation of the connectivity criteria and identification of the connected slices.

Once the lungs with its corresponding area have been identified in a preliminary way, the following step is to analyze the connectivity in each one of the slices. For it, three criteria are used; the values of two of them will be defined according to the anatomical structure to be analyzed.

The first criterion will be to compare in each slice the areas of both lungs. In this way, if one of the areas is excessively large, it is because both lungs are considered as one, therefore it is a sign of connectivity. On account of this, the value of 100 is fixed to this criterion. The second criterion will be to control the possibility that one lung will disappear (i.e. the slices corresponding to apex or the base of the lungs, at image level) before the other one. If we only would take into account the first criterion to determine connectivity, the extreme slices of the volume in which the above possibility can be given, the connection between lungs would be detected as well. To avoid this, the second criterion will assess the area increment which will be produced between two consecutive slices. So, the value of 30% is fixed to this criterion. Finally, the third criterion will be to check whether there is or not a connection in the previous slice. Therefore, the value of 1 or 0 is fixed to this criterion depending on whether the connection is detected or not in the previous slice, respectively.

To ensure that the connectivity between lungs exists in a determined slice, the following possible cases will be analysed:

- Case 1: none of the first two criteria is fulfilled. This will mean that in the corresponding slice the obtained areas are similar and, according to the previous slice there is not any great increment in the areas. Therefore, it is determined there is no connectivity between lungs in such slice.

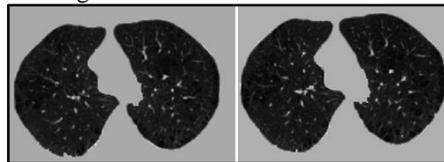


Fig. 5. Case 1. Left: slice in which the lungs are separated. Right: slice in which the lungs are separated as well. In the right slice the first and second criterion are not fulfilled.

- Case 2: the first and second criterion are fulfilled, this implies that in the analyzed slice there is an area which is superior to 100 times the other analyzed area. Moreover, this area increases a 30 percent regarding the area of the previous slice.

Therefore, it would indicate that in the previous slice there was not connectivity between lungs but in the corresponding slice there is.

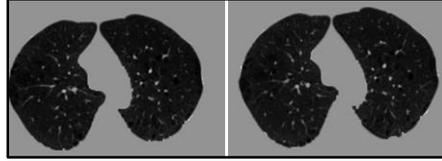


Fig. 6. Case 2. Left: slice in which the lungs are separated. Right: slice in which the lungs are connected. In the slice on the right the first and second criterion are fulfilled.

- Case 3: only the first criterion is fulfilled. This will mean that in the corresponding slice there is an area superior to 100 times the other analyzed area, but with respect to anterior slice, this area does not increase a 30 percent. Therefore, it would suppose that in the previous slice there is connectivity between lungs which remains in the corresponding slice.

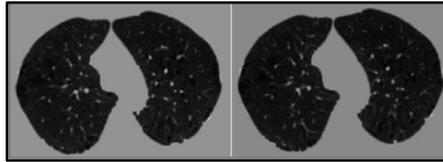


Fig. 7. Case 3. Left: slice in which the lungs are connected. Right: slice in which the lungs are connected as well. In the slice on the right only the first criterion is fulfilled.

Taking into account that case 3 can involve the extreme slices of the volume where it will be possible that one lung disappears before the other one, the third criterion will be analysed to avoid that. In this way, the slices which have connectivity between lungs will be those which fulfil the first and second criterion, or also those which fulfil the first criterion and the third criterion.

Lungs separation

Once the slices where exists connectivity between lungs are known, the next step will be the separation process. To achieve this, the active contours technique is used. This technique requires the binary mask of the initial contour and the image in which the contour has to be adjusted as input parameters. The output parameter will be the adjusted contour. To be able to apply this to our set of CT images and get an accurate contour that ensures the continuity of the pulmonary morphology, the previous array of connected slices will be extended adding five slices above the first connected slice and five slices below the last connected slice. In the event of the slices are nonconsecutive in the previous array, the voids will be filled to ensure, the continuity of the contours.

The proposed methodology also switches between the right lung and the left lung in the calculation of the active contour for each slice. This strategy ensures a robust and accurate cut to separate the lungs, avoiding contour propagation errors (for exam-

ple, the possibility that the active contour covers progressively part of the other lung). Summarizing, the process includes the next steps:

1. The lungs contour is extracted from the previous (non-connected) slice.

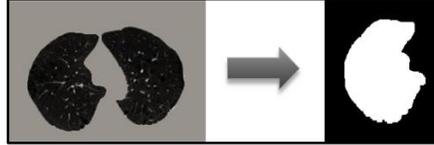


Fig. 8. Lung contour extraction. Left: previous slice. Right: lung contour.

2. The binary mask of the previous contour together with the next slice will be the input parameters to the active contour function. This function will return as an output parameter the contour which has been adjusted to the next slice.
3. Two morphological dilations are done to the binary mask of the contour which has been returned by the active contour function.
4. After morphological dilations, the edge contour is obtained.

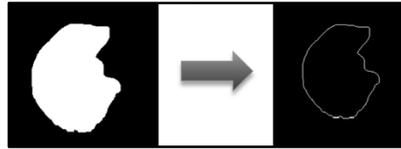


Fig. 9. Obtaining of the edge contour. Left: adjusted contour after the morphological dilations. Right: the edge contour.

5. To achieve an appropriate separation of the lungs, two morphological dilations are done to the previous edge in order to increase its thickness.
6. The inverse of the binary mask of the above dilated edge will be obtained. Then, it is multiplied by the corresponding image of the next slice. Thus, where before in such slice there was union between lungs now there are 0's in the image, achieving the final separation between lungs.

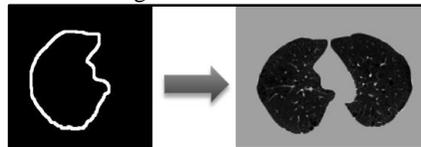


Fig. 10. Lungs separation. Left: edge contour with more thickness. Right: image with separated lungs.

3 Results and Discussion

After applying the methodology to the whole volume of CT images which is composed of 220 slices, the connectivity between lungs is detected in 53 slices. After adding the 10 security slices and completing the array with the corresponding slices

(step 4.3), a total of 65 slices is obtained in which the separation process will be applied (step 4.3).

Once the separation process is applied to the previous array, the most significant results are shown (figure 11). We can observe how the lungs are separated with a certain margin of security in a totally automatic way, avoiding in this way, that in a posterior 3D connectivity analysis they can appear as connected structures.

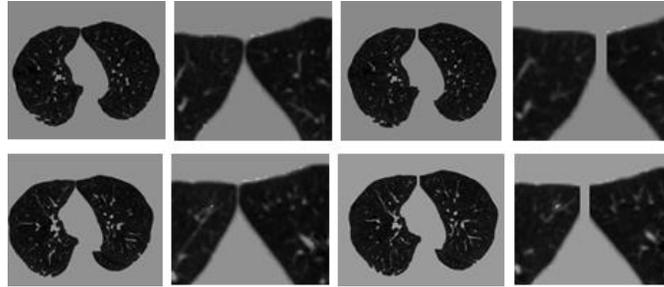


Fig. 11. Obtained results. 1st column: images of the most significant slices with connected lungs. 2nd column: lungs connection. 3rd column: images of the previous slices with the separation between lungs. 4th column: cut between lungs.

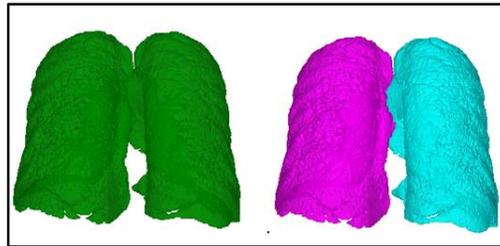


Fig. 12. Volume 3D of lungs. Left: lungs are connected. Right: lungs after separation process.

Our methodology presents some limitations. The separation process entails a slight loss of pulmonary volume, although it is not important concerning the total volume or the own partial volume effects of the original images (table 1). On the other hand, the method has been tested in only one case, although the design has been performed taking into account the knowledge and the experience of a team of radiologists specialised in thorax CT. The method validation will be done in a collaborative way in a wide series of patients.

Initial pulmonary volume	Pulmonary volume after separation process	Difference between volumes	% of lost volume
4.926.283	4.911.143	15.140	≈ 0.3073%

Table 1. Pulmonary volume measurements (measured in pixels)

The proposed methodology offers a new approach to deal with the lung separation problem. The use of active contours has proved to be a robust and fast way to ensure

the integrity of the lung contours during the separation process. The proposed approach is not limited to a specific region of lung connection or lung anatomy, as it would work on lungs presenting anterior or posterior connections, or even both. It uses fully available algorithms, ensuring that it can be implemented on any image processing platform using standard libraries.

4 Conclusions

This paper presents a lung segmentation method which is particularly able to separate the lungs in those cases where they appear as connected in CT images. From a significant case of CT images it has been possible to validate this methodology in a preliminary way.

References

1. Jan-Martin Kuhnigk et al. New Tools for Computer Assistance in Thoracic CT Part 1. Functional Analysis of Lungs, Lung Lobes, and Bronchopulmonary Segments. *RadioGraphics* 2005; 25:525–536.
2. Seth T. Bartel et al. Equating quantitative emphysema measurements on different CT image reconstructions. *Medical Physics* 38, 4894 (2011).
3. David S. Gierada et al. Quantitative CT Assessment of Emphysema and Airways in Relation to Lung Cancer Risk. *Radiology: Volume 261: N° 3-December 2011*.
4. Shiyong Hu et al. Automatic Lung Segmentation for Accurate Quantitation of Volumetric X-Ray CT Images. *IEEE Transactions on medical imaging*, Vol. 20, N° 6, June 2001.
5. Giorgio De Nunzio et al. Automatic Lung Segmentation in CT Images with Accurate Handling of the Hilar Region. *Journal of Digital Imaging*, Vol. 24, N° 1 February, 2011: pp 11-27.
6. Sang Cheol Park et al. Separation of left and right lungs using 3D information of sequential CT images and a guided dynamic programming algorithm. *J Comput Assist Tomogr.* 2011; 35(2): 280–289.
7. Aykac D et al. Segmentation and analysis of the human airway tree from three-dimensional x-ray CT images. *IEEE Trans. Med. Imaging* (2003) 22 940–50.
8. Graham M W et al. Robust system for human airway tree segmentation Proc. (2008) SPIE 6914 69141J.
9. Sedgewick, Robert, *Algorithms in C*, 3rd Ed., Addison-Wesley, 1998, pp. 11-20.

Content Management System: a Review.

José Luis Bayo¹, Álvaro Martínez-Romero¹, Carlos Fernández-Llatas¹,
Vicente Traver¹

¹Instituto ITACA, Universitat Politècnica de València, Camino de Vera s/n, 46022, Valencia,
Spain.
{jobamon, amartinez, cfllatas, vtraver}@itaca.upv.es

Abstract. Usually organizations and research groups have the necessity of exchange and publish information in his projects. The use of Content Management Systems (CMS) can offer an easy way approach to achieve this without the necessity of invest lots of resources in this task. Today there are plenty of products of this type available in the market and it is complicated to decide which the best tool for its necessities is. This review article seeks to create a list of the best tools available to solve the election problem and based on eight points to compare this products and draw some conclusions to support organization managers and technicians in their election.

1 Introduction

When an organization or a research group is working on a project, the amount of information that is generated that has to be exchanged within the consortium or published for dissemination, can suppose in many cases the use of a large quantity of valuable resources. For the creation of portals, and web pages, even 2.0 or 2.5[1] approaches, the majority of development time is spent in replicate again and again the same structure, the same administrator tools and even, often, the same visualization tools. In addition, the security is compromised in each portal due to the lack of a standardized policies of continuous security updates.

One of the adopted solutions to approach this problem is the use of specialized software tools that support Information Technology (IT) professionals in reusing the work. At this point we find the term of Content Management System (CMS) that is being used since the late 1990s. It is difficult to find an exact definition because of the rapid evolution of technologies, but the main concept of all definitions is the same.

A CMS is a system that collects, manages, and publishes information and functionality. At the broadest level, a CMS is responsible for the collection, management, and publishing of chunks of information known as content components [2].

In the market we can find a great number of CMS solutions available, each one offering a wide range of tools such as user account management, contents version trace-

ability, blog, wiki, e-commerce, forums or news. Besides different functionalities provided for a CMS, these are developed in different programming technologies, with different business model and types of license. Considering the variability of these factors, the decision to select one or another is more complex than it seems.

This article offers a review of a subset of CMS found on specialized web tools that analyze usage statistics and ranking list. The CMSs have been selected based on factors like popularity, including products made in several coding technologies and also with different business model. This provides a heterogeneous list, but with more points of view that lists specialized on one of the three mentioned factors.

To compare the CMS capabilities eight filter points have been defined: Open Source license, widely accepted, the existence of a support community, number of features provided natively, extension modules, cost of the platform, easy to use and learning curve to extend or modify the platform.

This review comes from the need to solve the problem of which CMS choose for each project in our research group. With this in mind, two objectives were defined before to start this review. The main goal was to create a help guide to choose between different CMS and the second goal was to disseminate our conclusions to the community, trying to be a start point for other research groups that may need to do a similar task.

2 Materials and Methods

The amount of available CMSs in the market is too great to find all of them and make a pre-selection using only an internet search engine. To accomplish the task of filtering a subset of CMSs, is better to utilize specialized tools publicly available on Internet which collect trends and user feedback. This work selected two complementary tools of this type, as a source of information to constrain the subset of CMSs to evaluate.

The first source used is the statistics generated by the web “BuiltWith”, which provides information on web technology trends on different fields based in its use. In our case using the statistics of the most used CMS technologies on Internet, we can know which CMS are the most widespread. [3]

The second source used is the web “CMS Critic” specialized on CMS reviews, which offers a list of products categorized by aspects like ‘type of license’ or ‘programming language’. But the most valuable data are its ratings in different categories: overall, ease of setup, ease of use, features, help and support, value. [4]

From the different CMS available in this tools, a subset has been selected attending to three factors: a) select the most popular products, b) include at least one product from one of these programming technologies: Java, .Net, Hypertext Preprocessor (PHP); and c) choose products with different types of business model (including the license type: open source, proprietary source, free software, commercial, on premise, hosted).

These three factors provide a heterogeneous list, but offer more points of view that the lists specialized on just one factor. With the final results of our review and taking

into account the requirements of our specific case, we can fine-tune the factors to our desired values.

The main three programming technologies have been selected based on his popularity and web usage statistics, trying to cover more than one community of developers.

To compare each of the selected CMS capabilities, eight filters have been defined to study in each of them.

- **FP-1. Easy to use:** one of the main factors for final users that do not have programming skills. The ease to install, manage the product and publish information are studied at this point.
- **FP-2. Number of features provided natively:** it evaluates the number and the quality of the features that the product offers as core functionalities, without installing third party modules.
- **FP-3. Extension modules:** some products admit extensions and third party modules that provide a vast number of features to enrich the original functionalities.
- **FP-4. Widely accepted:** the use of a widespread product can be a guarantee of perdurability of a product.
- **FP-5. The existence of a community that provides support:** some products are developed or supported only by their communities. Normally, the size and the level of activity of the community behind a product is important to evolve and to solve unusual problems.
- **FP-6. The learning curve to extend and modify the platform:** at some projects the base features provided by CMSs are not enough or some integration between them is needed.
- **FP-7. Open Source:** the reviewed CMS list is not limited on his type of license or business model, however at some research projects it is important to know if the code of some product is available for change or not.
- **FP-8. The cost of the platform:** sometimes a limiting factor in the selection of one product is the cost, not its functionalities. But the features that some products offer can be cost-effective in some projects.

This review includes for each of the selected CMS a brief description of the most remarkable features and a comparison table based on the previous filters that is included at the end.

3 Results

Using the criteria defined to select a subset of candidate CMSs to be reviewed, the list of products pre-selected is, in alphabetical order: Alfresco, Apache Lenya, DNN Platform, Drupal, GateIn, Hippo, Joomla, Liferay, Magnolia, OpenCMS, ProcessWire, SharePoint and WordPress.

A brief description of each one is provided below.

Alfresco: [5]

CMS specialized in document management with services to capture and manage electronic files and automate document-centric business processes. It provides version

control, online previews, workflow management of documents, mobile content management, and many other features to manage digital documents.

The product is open source developed in Java platform and has a cloud-based hosting free license with limited features and different commercial versions of the product for use in internal servers or cloud-based hosting.

This is one of the best Document Management System, but it is highly specialized at this scope of a CMS and the other areas are not well covered or provided. His aim is not to offer generic features to cover different scenarios, the product focused on being good at document management.

Apache Lenya: [6]

CMS hosted by The Apache Software Foundation community. It offers features such as authoring, workflow engine, internationalization, layout, site management and security. The extensibility is provided by using features from other projects like Apache Cocoon project.

Apache Lenya is an Open Source and free CMS, based on Java and Extensible Markup Language (XML). The project is perhaps in one of the best community of developers, but at the moment, the core features are reduced, providing basically features to define articles and publish in a website.

DNN Platform: [7]

DNN formerly known as DotNetNuke is one of the most popular CMS that uses .NET technology. It is an Open Source platform that allows building websites; it has been used by over 800.000 organizations worldwide to build their websites. The platform is available in a free Community Edition and the subscription-based Professional Edition and Enterprise Edition with an Elite Support option.

It provides a quick way to develop and deploy interactive and dynamic websites. It has an easy to use back-end to manage, organize and create content. It has important core modules to manage documents, events, blogs, forms, galleries, repositories, wiki, users and more. There also a module store with lots of free and commercial modules to extend its functionalities.

Drupal: [8]

This software can be used as a CMS but his authors define it as an application framework because of the flexibility it offers in module development and template development. It can be used to develop web applications, from personal blogs to complex enterprise applications.

Drupal is a free and Open Software product that has a big active community and has available more than 14.000 modules to extend it.

Due to his flexible and modular architecture, its learning curve may be higher than other products that are specifically focused on being just a CMS. This is an important point to consider before choose this product.

GateIn: [9]

This software is the evolution of JBoss Portal and uses JBoss as a core technology. It provides an Open Source website framework with the features: Single Sign-On, layout support, user and group management, Lightweight Directory Access Protocol (LDAP) support and implements JBoss Portlet Bridge to extend it with JavaServer Faces (JSF).

JBoss Application Server is a good product to use as a base for a CMS, but attending to the web content, GateIn seems not to be so mature and do not have enough modules at this moment.

Hippo: [10]

Hippo is a CMS in which the main idea is to create contents once, and then publish them to a wide variety of Channels and Devices, like web sites, mobile sites, Facebook, Representational State Transfer (REST) Application Programming Interfaces (APIs), and so on. The Hippo Delivery Tier provides tools for rapid development of websites and channels.

This CMS is an Open Source project developed in Java and there are two versions available: the Community Edition is free but with limited advanced features; the Enterprise Edition contains the full functionality, but you have to purchase it.

The product seems to be a good platform to create static content that can be delivered to multiple devices in an easy way, using graphical tools to do all this work. But it seems not to have another functionalities and they do not have a module repository to extend the core features.

Joomla: [11]

Joomla is a powerful CMS developed in PHP, highly adopted in the Internet. It is easy to use and with a large list of features, but his main capability is its extensibility.

This is a free and Open Software that provides modules like user management, media management, language management, contact management, polls, search, content management, syndication, template management, web services and much more. But the main characteristic is the extension repository with more than 7000 extensions provided by the developer's community.

Liferay: [12]

It is defined as a complete platform for building intuitive, engaging web experiences. Liferay is one of the most popular CMS based on Java programming language, and is recommended by many industry experts.

The platform provides easy and modern interfaces with lots of modules as document library, themes and layout, language support, web publishing, knowledge management, wikis, blogs, messaging, calendar, alerts, security, identity management, performance and scalability.

This is an Open Source software with two different versions: Community Edition is a free distribution with some limited features like support or performance tools. Enterprise Edition is a commercial version and the price depends on the personal requirements.

Taking into account all the above described this seems a very complete CMS, however, it is necessary to consider the learning curve of that type of big systems.

Magnolia: [13]

Magnolia is a product which main goal is to be easy to use, flexible and with a good usability and design. It provides features such as administration, authoring, customization, asset management, internalization, interoperability, mobile access, security and teamwork.

It is an Open Source well-documented platform, developed in Java and with two versions: the Community Edition is free but limited in some advanced functionalities,

installation process and constrained to one single subscriber. The Enterprise Edition is not limited and has official support.

Although support community and extensibility is not so clear on his web site, this product seems to be so easy to use, with good results and enough modules in his base core.

OpenCMS: [14]

Defined as a professional and easy to use website CMS, it provides features to configure, create, edit and structure content in browser based interfaces. Also has a template engine and it is World Wide Web Consortium (W3C) compliant for all contents.

OpenCMS is Open Software and free, it is based on Java and XML technologies. The base core provides an intuitive page editor, site map editor, inline editor, gallery, searching, integrated image processing and more.

Although it seems to be a stable product, it does not have a big community supporting this product and the gallery of extension modules it is not so active and big enough as it would be desirable.

ProcessWire: [15]

ProcessWire is a CMS developed with simplicity in mind, to offer a product that does not require training but without reduced capabilities. This is a lightweight software that allows creating static contents using its core modules. Additionally, it provides a module library to extend its capabilities.

This product is free and Open Source, it is made with PHP. Probably the weakest point is its short life as a product, which makes its use not as widespread as other products.

SharePoint: [16]

This is a product that provides a wide variety of features, from document sharing, social communities and mobile access, to building websites, providing an internal app store and adaptive web site experiences. And of course offers full integration with Microsoft Office products.

This is a non-open source product, but you can extend his features using HTML, JavaScript, PHP or .Net technologies. We can find two commercial licenses, one Online hosted from 3.3 € to 6.5 € plus taxes for each user and another for use in internal servers which price must be consulted being dependent on enterprise necessities.

This can be a good solution for big enterprises, it offers a wide quantity of features and enterprise solutions but it is not Open Source and the ratio cost-functionality is too high for some cases.

WordPress: [17]

Initially WordPress was created as a blogging platform but now is used as a CMS software mainly for his wide acceptance at the web community. It is a user-centered CMS based on be as simple to use as possible.

It provides the possibility to create websites or blogs, it is free Open Source product written in PHP. The development is supported by a vast community of volunteers and it has thousands of extension components available. Over 60 million of users have chosen this product to make his web.

The following table summarizes the main outcomes for the solutions described above when evaluated using the filters described.

Table 1. Summary of the review based on the filter points

CMS	FILTER POINTS							
	FP-1	FP-2	FP-3	FP-4	FP-5	FP-6	FP-7	FP-8
 Alfresco								Free Limited, Commercial
 apache lenya								Free
 DNN								Free, Commercial
 Drupal								Free
 GateIn								Free
 HIPPO								Free Limited, Commercial
 Joomla!								Free
 LIFERAY								Free Limited, Commercial
 magnolia								Free Limited, Commercial
 OpenCMS								Free, Commercial
 PROCESSwire								Free
 SharePoint								Commercial
 WORDPRESS.ORG								Free

4 Discussion and Conclusions

Analyzing all the information collected for each CMS and the outcomes from the comparative table, we think that there are four CMS candidates to be a good solutions to use in research environments: DNN Platform, Liferay, Joomla and WordPress. The specific requirements for each project and the knowledge in programming technologies of each client are the two main factors to drive the selection of a product.

When a project involves the development of a big system and part of the requirements include integration between this software and the CMS or even the system is developed using the CMS core features. We think that it is preferable to focus the at-

tention on the technologies as Java and .NET, technologies that are software frameworks that provides lots of base functionalities, APIs and interfaces to develop standard applications, mobile applications or web applications. Providing performance, security or memory management as part of the framework.

With technologies in mind as a requirement, the expertise and skills of your developer team is the key to decide between the following two CMS.

In our opinion, first of them, the DNN Platform is the election if you have developers with expertise on .NET technologies. This platform has a good extensibility documentation and provides core modules to develop portals that are more than a content authoring site, it provides community modules, file repository module or a module to embed external sites. Another good feature is an easy to use administration interface.

As negative points we can say that, the hosting technology must be Windows with Internet Information Services which tend to be more expensive.

The second CMS Liferay, is developed in Java technology and we think that is the option if your team is formed in this technology. Its main advantages are a good base of core modules and an easy-to-use administrative tool. It provides features to develop a complex enterprise portal, with more capabilities than only generate contents. The main disadvantages are the learning curve to extend it and the cost of web hosting, but normally is below the cost of a Windows hosting.

The other two CMS elected as a good products, are developed in PHP technology, this is a scripting language designed for web development, and it is designed to promote Rapid Application Development (RAD). It has a lot of external libraries or APIs that provide great functionalities to extend its features.

Currently PHP is widely extended for his RAD characteristic and because there are a great offer of free hosting providers that support this technology. Although there are very good platforms and sites developed in PHP, it is not focused on scalability, performance and code maintainability.

If your organization has experience developing projects oriented to web, probably your expertise on PHP is bigger than in technologies as .Net or Java. We believe that this should be a good reason to choose one of the last two products on the list of selected CMS.

At those projects which is needed some web functionalities and none of the CMSs implements your needs, Joomla CMS is a good election, this product is supported by a non-for-profit organization, which provides a product without limited versions. Its main technology PHP, is one of the reasons why it is widely extended on internet.

This CMS has a vast number of modules and great support by the community, what permits to generate a site that has complex functionalities, not only content generation features. The negative points explained by some users are a complex administration tool and that the learning curve is greater than others CMS based on PHP.

Finally if your organization do not have a developers team, the objectives of your project do not require integration with any CMS or there is only a small integration, WordPress is the suggested CMS, this product is the most widespread on Internet of all. Its main features are a very easy-to use interface, not only in administrative tools, also for installation and authoring tools. The most important focus on WordPress is on the generation of content as easy and quickly as possible. Its shortcomings are on the

reduced core features that provides and that have to be covered with extension modules. Luckily it has a large number of extensions already developed by its vast community of users.

References

1. Pileggi, Salvatore F.; Fernandez-Llatas, Carlos; Traver, Vicente. 2012. "When the Social Meets the Semantic: Social Semantic Web or Web 2.5." *Future Internet* 4, no. 3: 852-864.
2. Boiko B.: *Content Management Bible 2nd edn.* Wiley Publishing, Inc., Indianapolis, Indiana (2005)
3. BuiltWith® Pty Ltd.: *CMS technologies Web Usage Statistics*: Available at: <http://trends.builtwith.com/cms> [Accessed 24 jun. 2014].
4. CMS Critic. *CMS Reviews, News, Systems and Articles*. Available at: <http://www.cmscritic.com/> [Accessed 24 jun. 2014].
5. Alfresco Software. *Document Management, Enterprise Content Management*. Available at: <http://www.alfresco.com/> [Accessed 24 jun. 2014].
6. Apache Lenya. The Apache Software Foundation. *Apache Lenya – Open Source Content Management*. Available at: <http://lenya.apache.org/> [Accessed 2 jun. 2014].
7. DNN Corp. *DNN Platform. CMS Platform for Website Management*. Available at: <http://www.dnnsoftware.com/platform> [Accessed 24 jun. 2014].
8. Drupal.org *Drupal – Open Source CMS*. Available at: <https://www.drupal.org/> [Accessed 24 jun. 2014].
9. Red Hat. *JBoss. GateIn Portal – Jboss Community*. Available at: <http://gatein.jboss.org/> [Accessed 24 jun. 2014].
10. Hippo B.V. *Hippo CMS – Enterprise Open Source java CMS*. Available at: <http://www.onehippo.com/> [Accessed 24 jun. 2014].
11. Open Source Matters, Inc. *The Joomla CMS Open Source project*. Available at: <http://www.joomla.org/> [Accessed 24 jun. 2014].
12. Liferay Inc. *Liferay Portal – Enterprise open source portal and collaboration software*. Available at: <https://www.liferay.com/> [Accessed 24 jun. 2014].
13. Magnolia International Ltd. *Magnolia CMS Enterprise Open Source Java Content Management System*. Available at: <http://www.magnolia-cms.com/> [Accessed 24 jun. 2014].
14. Alkacon Software GmbH. *OpenCMS, the Open Source Java Web Content Management System*. Available at: <http://www.opencms.org/> [Accessed 24 jun. 2014].
15. ProcessWire a *Friendly and Powerful Open Source CMS*. Available at: <https://processwire.com/> [Accessed 24 jun. 2014].
16. Microsoft Corporation. *Microsoft SharePoint. Collaboration and management software*. Available at: <http://office.microsoft.com/en-us/sharepoint/> [Accessed 24 jun. 2014].
17. WordPress.org *WordPress, blog tool, publishing platform and CMS*. Available at: <http://wordpress.org/> [Accessed 24 jun. 2014].

Improving reliability in wireless sensor networks: towards intelligent monitoring platforms

Juan Vicente Capella^{1,*}, José Carlos Campelo¹, Alberto Bonastre¹ and Rafael Ors¹

¹ Instituto ITACA, Universitat Politècnica de València,
Ciutat Politècnica de la Innovació. Camino de Vera s/n, 46022 – Valencia (Spain);
{jcapella, jcampelo, bonastre, rors}@itaca.upv.es

Abstract. The applications based on Wireless Sensor Networks (WSN) present a huge growing space. Nevertheless, despite of the numerous laboratory experiments, the real implementations are still scarce. A deeper knowledge of the internal behavior of the WSN in real conditions is necessary to fill the gap between controlled conditions and real environments, mainly in critical applications such as e-health. This way, the so called WSN Monitoring Platforms (WSN-MP) are required in order to characterize, verify and standardize the operation of WSN's. Several WSN-MP approaches have been published, but they have been adapted to measure specific parameters on a specific WSN. In this paper, a generic WSN-MP model is presented, being applicable to any parameter of any WSN. The advantages of the systematization of WSN-MP design and implementations are obvious: reusability, standardization, faster developments.

1 Introduction

A WSN-MP consists on a distributed system which monitors the operation of a WSN, by means of a set of devices and/or modules which acquire information from different elements of the WSN (from motes, wireless network or gateway), collect, analyze and present them to the user providing internal information about WSN operation.

The use of WSN-MP may help in all the states in the life-cycle of a WSN. Researchers on WSN may use a WSN-MP to perform comparative analysis on new proposals. Designers may select the best suitable techniques for a given application requirements. When implementing, the enhanced debugging capabilities added by the WSN-MP are unbeatable. The deployment is much easier when the correct functioning of the motes can be verified in-situ. In operation time, malfunctions may be diagnosed without stopping the system, and the redesign of the system can use more detailed information about current functioning [1]. In addition, these tools can become fundamentals in the standardization and certification of applications based on WSN.

With these goals in mind, WSN-MP's are an increasingly important research line, which is corroborated by the publication of a large number of papers on this area in the scientific literature [2][3][4]. The study of these proposals has shown the great disparity of different WSN-MP's (where each of them uses different ad-hoc solutions), and the lack of methodology for its analysis and design. Additionally, we con-

clude that there is a set of different issues that every WSN-MP must solve to be functional, and the different solutions of these problems can be combined in order to obtain the appropriated characteristics for the desired Monitoring Platform.

From our point of view, this situation is similar to those of computer networks, which began as a set of ad-hoc systems that did not follow any standard. The spectacular advances achieved in this field have been possible, partially, because of the adoption of a universally accepted architecture which promotes the use of standards. Foreseeing the example of Computer Networks, the aim of this paper is to propose a new reference model for WSN-MPs, in order to establish a systematic methodology for the analysis, design, implementation and operation of monitoring platforms on sensor networks. This model allows the comparative analysis of the WSN-MP, and offers a common environment for the design and standardization of new platforms. It is based on decomposing the monitoring process in different levels, identifying its responsibilities and establishing the services, protocols and interfaces needed to accomplish them. This approach offers several new interesting features, such as flexibility (being able to adapt to any type of system or application), simplicity (due to level division), universality (foreseeing any system), and adaptability (being able to follow the future evolution of WSN and its WSN-MP). This model could serve as first approach to achieve a standardization of WSN-MP. It also promotes new research areas in order to complete and develop the issues pointed in this article.

2 Proposed architecture

In this section the proposed model is described, starting with a model overview, followed with a deeper description of each one of its components and concluding remarking the main features that this model offers.

As observed at the bibliography, every monitoring platform has to deal with some issues in order to be functional [5]. Our proposal identifies these problems and classifies them into three categories:

- Data management: First category makes reference to monitoring data. These data must be captured, analyzed and shown to the user in a meaningful way. The semantic meaning of the data is highly related to the application and the needs of the monitoring agent.
- Data coding: Once obtained, data must be expressed in a convenient way, in order to ensure the integration of data provided by heterogeneous sources. Some issues related to this deal with a common time base and capturing conditions.
- Data interchange: As the monitoring platform is distributed in space, all the data obtained must be centralized in order to ensure a global comprehension of the system functioning.

Keeping these different – and usually independent – problems in mind, the proposed architecture is composed by three levels.

Monitoring Level will be located at the upper level of the architecture. This level is in charge of all the issues related and specific to the system under observation. It must

deal with the definition of what should be monitored, how this information must be acquired and the way it has to be processed and shown to the user.

Information Level is located under the Monitoring Level. It must provide the support needed to code the obtained information in a standard way. This level also deals with timing issues, which includes when the information must be captured (*triggering*) and stores this time value into obtained data (*time stamp*).

Finally, *Interchange Level* allows the information captured in different points alongside the WSN to monitor to be transferred to a single point, when must be analyzed and/or visualized by the Monitoring Level.

The model must locate the interfaces between each pair of adjacent levels, in order to keep each level independent of the rest, obtaining in this manner portability and interoperability between components from different manufacturers/developers. It is also necessary the definition of a set of protocols to support the communication needs into each level. The standardization of these protocols will open the possibility of universal reuse of off-the-shelf components. Each of these levels is defined in the following sections.

3.1 Monitoring level

This level deals with the issues related to capture, analysis and interpretation of the information related to the monitored system functioning. It is composed by three subsystems. Capture subsystem acquires data from the observed system. These data are processed by the Analysis subsystem. This Analysis subsystem is in charge of joining and sorting the information of all captures in the monitoring platform, and processing it, applying the appropriate information algorithms (indeed using Artificial Intelligence techniques) in order to extract the required conclusions.

Finally, these results must arrive to the Visualization & Control subsystem. The interface of the user is provided by this module, showing the required values in a useful way. This subsystem may also receive indications, instructions or parameters from the user, with the purpose of modifying the functioning of the Monitoring Platform, in both capture and analyze subsystems. It should be done in a friendly manner and offer a powerful yet simple interface.

3.2 Information level

The information level is responsible from several functions. It is in charge to homogenize the probe's captured data, delivered by upper level in raw mode (without format). It is also concerned with the timing of the capture, applying a time stamp mechanism which makes possible the temporal sorting of events proceeding from different sources, in the same mote or in another mote of the monitored WSN. Finally, it is also responsible for the capture's trigger, providing to the Monitoring Level for the necessary trigger indication.

This level offers its services to the upper level, in such a way that it interchanges information with it (accepts information from capture entity, and delivers it to the

analysis and visualization entities). Similarly, it uses the services offered by lower level (interchange level) to move and/or store the information between Information Level Entities. Virtually, this communication must accomplish the so called Information Level Protocol, which must provide the support for the services of this level.

3.3 Interchange level

Since a monitoring platform is also a distributed system itself, all the data managed by the information level are spatially dispersed, and usually have been captured in different moments. They may be stored in non-volatile memory to be downloaded (or even physically transported) later. It is necessary the guarantee that the information communication between system components is correct. Issues related to data storage and communications (if exist) must be considered at this level. Interchange level must provide the information level from those services required to fulfill its functions.

This level can offer the distributed data storage service to the upper level, making use of one or more communication networks. Monitored data is stored in this level in a structured way, being able to deliver the stored data when it is required by the Analyzer subsystem or by the Visualizer subsystem.

Communication and storage technologies have been developing its own standards in the last decades. The utilization of these standards, and future ones, relies in the definition of a sub-level which can use the services offered by these implementations in an efficient way. In this sense, a Hardware Abstraction Sub-level, which permits the adaptation to any of those existing standards to be applied, is required. Despite of not suggested, a WSN-MP designer may choose to implement its own non-standard network or storage technology, and integrate it with standard upper level elements if the appropriate abstraction sub-level is defined.

An Interchange Level Entity must provide two groups of services:

- Network Management: This element has to deal with communication issues, such as device addressing, device discovering and routing between the different used networks.
- Storage Management: This element has to deal with distributed storage issues such as data coherence, data storing and data querying.

As any other element of this architecture, the definition of standard interfaces makes possible the migration from different implementations, as far as detailed services and functions are provided. In this way, many systems may be applicable, from powerful communication networks (wired or wireless networks, public or private) to the most simple, such as data storage in non-volatile memories carried manually to a data center.

3 Conclusions

The proposed model is based on the division of the monitoring issue in three levels. Monitoring Level focuses on relevant information: data capture, analysis and visualization. Information Level deals with information coherence: data format, synchronization and triggering. Finally Interchange Level supports information management: communication and storage.

Moreover, the WSN-MP designed from this model covers all stages of the life cycle of a WSN. Additionally, this model can be the basis of future certification processes for WSN based applications.

The proposed model is ready to be applied in the analysis, but in order to be applied to the design is necessary to continue with the specification of the interfaces between levels and the research and development of protocols and services of each level. This will allow the development of universally accepted standards. New platforms implemented following this model will maintain the advantages of standardization (reuse and interchangeability of hardware and software, independence between levels, isolated resolution of each task associated with a level, data intercommunication between platforms, etc.) without losing the necessary specialization, keeping the ability to optimize for each WSN and for the appropriated parameters to be monitored depending on the application characteristics. These desired characteristics may include parameters such as dependability, bounded response time, early detection of coverage loss and mobility management.

Acknowledgments

This work was supported by the Technical University of Valencia Research Project SP20120889 and Spanish government under projects CTM2011-29691-C02-01 and TIN2011-28435-C03-0.

References

1. Meier, A.; Motani, M.; Siqun, H.; Künzli, S. DiMo: Distributed node monitoring in wireless sensor networks. In Proceedings of the 11th international symposium on Modeling, analysis and simulation of wireless and mobile systems; MSWiM '08; ACM: New York, NY, USA, 2008; pp. 117–121.
2. Sundaram, V.; Eugster, P.; Zhang, X. Lightweight tracing for wireless sensor networks debugging. In Proceedings of the 4th International Workshop on Middleware Tools, Services and Run-Time Support for Sensor Networks; MidSens '09; ACM: New York, NY, USA, 2009; pp. 13–18.
3. Rost, S.; Balakrishnan, H. Memento: A health monitoring system for wireless sensor networks. In Sensor and Ad Hoc Communications and Networks, 2006. SECON'06. 2006 3rd Annual IEEE Communications Society on; IEEE, 2006; Vol. 2, pp. 575–584.
4. Yunhao, L.; Kebin, L.; Mo, L. Passive Diagnosis for Wireless Sensor Networks. IEEE/ACM Trans. Netw. 2010, 18, 1132–1144.

5. Schoofs, A.; O'Hare, G. M. P.; Ruzzelli, A. G. Debugging Low-Power and Lossy Wireless Networks: A Survey. *IEEE Commun. Surv. Tutorials* 2011, 1–11

Low-Power Sensor Control Manager for Multi-hop Underwater Networks

Paulin FAUVEAU, Radwan FANIDKI, Sara BLANC, Raul SÁEZ, Juan José SERRANO

Instituto ITACA, Universitat Politècnica de València, Cmno. De Vera s/n. Valencia. Spain

E-mails: paulin.fauveau@gmail.com, fanidki@hotmail.fr,
sablaccla@disca.upv.es, rsaez@itaca.upv.es,
jserrano@itaca.upv.es

Abstract. Underwater sensor wireless networks are increasing in interest for seabed and water column monitoring. However, due to the fact of lacking for a predictable energy-harvesting source the life time battery constrain reveals the importance of looking for low-power consumption development platforms to implement underwater sensor networks. This paper presents a study on consumption, wake up timing and performance of an ARM Cortex M4 in different low power mode. The STM32F4 discovery board has been selected as commercial platform for a prototype design of a multihop communication system. The implementation integrates FreeRTOS, SRAM and micro SD storage. The system is intended to the capability of transmitting data from underwater sensor nodes along a deployed lined topology. Communication is done through ARM Cortex UART ports, linked with the S1000 ITACA modem with acoustic capability. Thus, a specific API has been designed to couple both systems. Data are finally received in a pc host sink through Ethernet connection.

Keywords : Low power, Wake up timing, Underwater Network, Wireless Sensor, Sensor control manager;

1 INTRODUCTION

The need to sense the underwater world drives the researcher to develop underwater wireless communication. Wireless information transmission through the ocean with a sensor network is one of the enabling technologies for the development of future application of underwater sensing. Applications of underwater sensing range from oceanographic data collection to Multiple Unmanned or Autonomous Underwater Vehicles, and include instrument monitoring, pollution control, climate recording, and study of marine life. Applications can have very different requirements. Some are fixed or mobile, short or long-lived, best-effort or life-or-death.

These requirements can result in different designs. In our case, we implement an algorithm for a short-range, long-life and low-cost application. The design is oriented to consume as low power as possible in order to extend nodes operational life. One of the advantages of the ARM cortex M4 is that it offers several operating modes which can

switch between different clock sources in order to minimize its power consumption. There are three different modes: sleep mode, stop mode and standby mode. Those modes are very interesting option in case we have a network with low traffic demand.

Previous works on the field leads the RIS group (under UCERT research line) to the development of an ultra-low power consumption acoustic modem [1][2]. This modem has been a great step into point-to-point acoustic communications with best results in ranges of 200-300 meters.

In addition to the modem, sensor processing and multihop wireless network algorithms are in current development. This paper is focused on the implementation of an algorithm that allows nodes sensor measurement and data transference in linear topology. The algorithm has been implemented in ARM Cortex™-M4 Discovery development kits [3].

The paper presents a detailed study of wakeup timing and consumption in three different low power modes. In this study, we discuss several options of communication architecture to improve the performance within a low consumption.

The remainder of this paper is organized as follows. Section II introduces the device, design and the algorithm behavior. Section III analyzes the consumption of the device in three low power modes. The section IV refers to wakeup timing in low power modes. Section V investigates different solutions applicable to multihop transmission in a linear topology; and finally, section VI draws conclusion.

2 UNDERWATER ACOUSTIC LINK

The S1000 modem [2] is the acoustic network link of the system. This modem has been design around 8-bits microcontroller architecture, small is size and with a global low-power consumption thanks to an asynchronous RFID wakeup mechanism [4].

However, S1000 storage is limited. Thus, the modem needs to be connected as to another storage device. To the solution presented in this paper, this device consists of a Cortex M4 core-based design. A Cortex M4 design will remove the storage limit. Next sections will evaluate both, the consumption and the wake up constrains of this solution. Relevant features implemented in the S1000 modem are the following:

- Sleep mode power consumption is under 10 μ W.
- Supports RSSI measurement.
- Allows slave connection by serial transmission (USB, SPI and UART)
- Transmission power is variable to a maximum of 108mW (240 meters range)

The modem supports both data and control frames; control frames allow reconfiguration of: ID wakeup pattern, transmission power (indirectly implies transmission range), and MAC address. These actions are transparent to the Cortex M4 design.

The ITACA S1000 API has been designed to be adaptive, hardware-independent and to cover a range of applications as wide as possible. To archive these qualities, the API provides several configuration commands and a set of static definitions in the API source code.

Commands allow adapting the API to several applications and environments in a

dynamic but limited way. However, major adaptations could require a full change of API behavior which can only be realized before compiling the source code.

The MCU-to-modem communication is performed through a serial communication. The API provides the required interface between the node and the modem to configure the communication channel and transmit/receive messages to/from other nodes.

3 SYSTEM DESIGN

3.1 Network topology

This research is focused on the design of architecture to transmit data from submerged oceanographic sensor nodes to a sink computer in linear topology. The network is built with several nodes composed of stm32F4 microcontroller, including an ARM Cortex M4, linked with an acoustic S1000 modem [2]. Consequently, the ARM Cortex microcontroller will be in charge of mastering the RBR oceanographic sensor and the data communication.

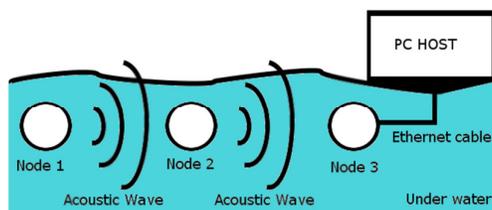


Fig.1. Underwater communication line

3.2 Hardware features

As mentioned above, each node is composed of a Cortex M4, an acoustic modem [2], a sensor system and SD storage. Information from the neighbor node is received by the S1000 modem (assume by the left modem in Fig. 2) and sent via UART connection to the ARM Cortex M4. Moreover, information read from the own node sensors will be store in the μ SD card. The ARM Cortex M4 processes this information and sends both received data and sensor reading to the next remote node through the acoustic modem (assumed by the right modem in Fig. 2).

Sensor data are currently composed of the internal temperature sensor value and RTC (Real Time Clock) value as a time base. To master this program behavior, a multi-tasking management is suitable; this leads to the integration of *FreeRTOS* in the microcontroller. This Real Time Operating System has been chosen for this small and simple design and for this native support for Cortex M4.

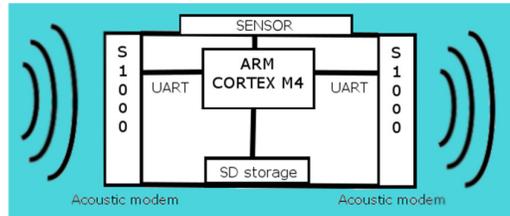


Fig.2. Node architecture

3.3 Network algorithms

Our research conducts us to design algorithms to transmit data from underwater nodes to a computer host. Fig. 3 shows an overview of the system. The ITACA WSN API (IWCA UART) allows the serial connection between the S1000 and the Cortex M4. The API has been integrated in the system, although the analysis of the S1000 link is out of the scope of this paper. Besides, the pc host receives data from the entire nodes through Ethernet. Those data will be processed and stored into a data base.

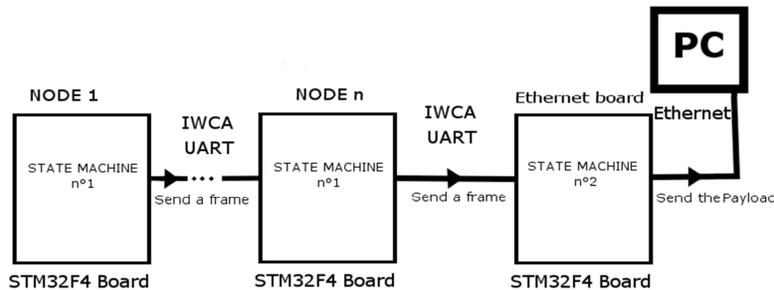


Fig.3. Network concept

Moreover, due to the underwater constraints, the system has to be energy efficient. On this point, Cortex M4 provided 3 state of low power mode:

1. Sleep mode (basic level of sleep)
2. Stop mode (deep sleep)
3. Standby mode (shut down of the system)

Obviously, these modes have different impact on the board consumption and in the system reactivity. This low power management is the focus of the algorithms in Fig.4 and Fig.5.

To wake up the microcontroller, the system uses a very simple additional hardware. Time synchronization is kept by using a Real Time Clock (RTC) interrupt. This hardware is well known and is integrated as on-chip peripheral in this microcontroller. Additionally, the microcontroller can also be waked-up asynchronously by an external interrupt from another device (assumed the S1000 modem). Data received through the serial connection are stored in a SRAM memory (4KBytes). This asynchronous wakeup, to receive data by the S1000 modem serial connection, is coherent with the

S1000 architecture because the modem includes a RFID tag that gives it its asynchronous wakeup feature.

Thus, the RTC interrupt has been selected to start the measurement window in the sensor node. During each measurement window, the node follows four steps:

- Firstly, gets data from sensors.
- Secondly, stores these data into the SD card. An advantage of μ SD storage is to maintain a measurement history that could be rescued from the board at any time.
- Thirdly, the algorithm builds a frame; the payload will contain together data stored in the SRAM and data stored in the SD card.
- Finally, the Cortex M4 board wakes-up the S1000 modem for a new transmission.

State machine n°1: Node sensor

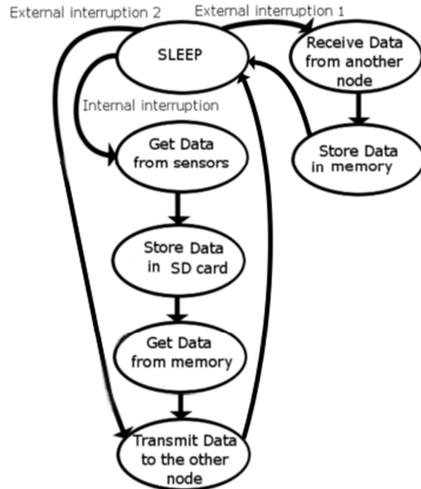


Fig.4. State machine underwater nodes

State machine n°2: Ethernet transmitter

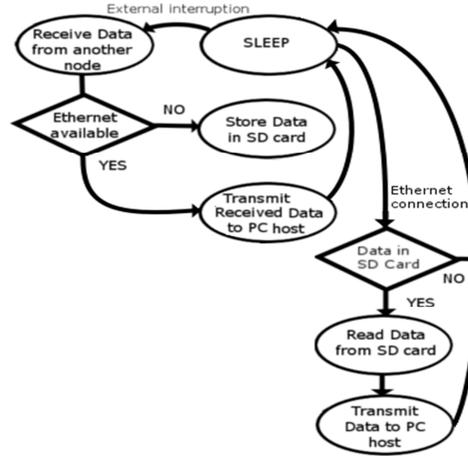


Fig.5 State machine of the Ethernet board

External Interruption 1 represents the link between the Cortex M4 board and the S1000 acoustic modem. Cortex M4 board has been designed as slave. When the S1000 modem receives a new frame, this small modem is not able to store this frame. Thus, the modem wakes-up to the Cortex M4 board who stores the frame in SRAM memory. *External interruption 2* occurred if the next node return a non-acknowledge. In this case, the node re-transmits the frame.

Transmission consists of a point-to-point master-slave asynchronous scheme (Master A and Slave B). That means, when a node A needs to send data to its neighbor, it starts a transmission sending a RFID tag to B. The receiver B wakes up and waits (nothing is sent by B). A sends data immediately and goes to sleep again without waiting for B acknowledge. Just in case a CRC error, B will wakes up A with a RFID tag together with a NON-acknowledge.

In this board, the same reception part of the node sensor board has been designed. The difference consists of the way of storing the incoming data. Actually, it depends on the TCP connection state with the PC host. To avoid a loss, the data are saved into SD card until a connection happens.

The system can be also wake by an Ethernet incoming connection, which triggers the sending of the waiting data in SD card. Once the connection is established, received data are directly transmitting through Ethernet. Due to the Ethernet peripheral device, the consumption of this board will be higher than the node sensor board, and the way of reducing consumption are very limited. Thus, the rest of the paper will be focused on the sensor board program regarding the energy saving.

4 LOW POWER

Embedded system designers are currently facing an urgent need to make a trade-off between power efficiency and high performance, especially with recent high-performance microcontrollers such as the STM32F4 series. Analyze different low power modes allows us to find the best way to reduce the maximum consumption with the best efficiency. Low-power modes are implemented with other features to reduce the average current consumption over the life of the application.

This part gives an overview of the STM32F4 low power mode consumptions. According to the data sheet of the stm32f4xx [3] from ST microelectronics, consumptions of both the Run and the Sleep modes depend on the operating frequency. Table 1 shows the current level for each low-power mode and run mode at the maximum core frequency (168MHz).

Table 1.Current consumption of STM32F4

Low-power Mode (168 MHz CPU clock)	Current consumption Arm Cortex M4
Run mode	70 mA
Sleep mode	12 mA
Stop mode	310 μ A
Standby mode	3,1 μ A

The main way of reducing consumption, is to slowing down the core clock. Consequently, the node sensor board runs with the lowest clock frequency, 8MHz from the on-board crystal. Besides, additional powers saving ways have been implemented to reach the lowest consumption. Like the gating of peripheral clocks when they are unused into the sleep mode or the deactivation of Schmitt Trigger Input for I/O pin providing zero consumption.

However, even if the consumption in *Run* and *Sleep* mode have been reduced significantly as expected, Table 2 highlights extra consumption into *Stop* and *Standby* mode. Actually, on the STM32F4-Discovery board, a pull-up resistor connected to BOOT0 pin of the microcontroller leads to around 500 μ A of current consumption, and other peripherals devices influence the measurement. This reveals the necessity of designing a specific board for the application.

Table 2.Current consumption of node algorithm

Low-power Mode (8 MHz CPU clock)	Current consumption Discovery Board
Run mode	6.56 mA
Sleep mode	1.89 mA
Stop mode	0.86 mA
Standby mode	0.55 mA

5 WAKE UP TIMING

As mentioned above, each low power mode has a direct influence on the system reactivity. This part highlight the wake up time constrains in the different low power modes in the application. There is different ways of configuring the Low power modes; the best compromise between reactivity and consumption leads to this configuration:

- Sleep mode : All peripherals clocks deactivated;
- Stop mode : Voltage regulator in low powered without a flash deep sleep;
- Standby mode: Backup regulator ON (used to maintain backup SRAM content).

The deactivation of peripheral clocks doesn't affect the wake up timing and reduce the current consumption. In the stop mode, enter into a deep sleep brings huge amount of time at wake and doesn't really impact the current level. The backup regulator is used to keep RTC and a backup RAM alive during the standby, this brings tiny consumption but it's absolutely required by the application.

On Table 3, the wakeup times are measured from the wakeup event to the point in which the application code starts the corresponding task. In Stop mode, the system needs to restore CPU clock, and in Standby mode the system reboot when it's waking up.

According to Fig. 6, main sources of latency are caused by the clock restoration and the Operating System mechanism (time to give and take semaphore). After a wake up from standby mode, the program reboots, what explains the increase of the timings. This mode has to be used carefully.

Table 3.Low power mode – Wake up timing

Low-power Mode (8MHz CPU clock)		Wakeup time		
		System wake ⁽¹⁾	External IT	RTC
Sleep mode	<i>Average</i>	5 μ s ⁽⁴⁾		64 μ s
	<i>Total</i>		5 μ s	69 μ s
Stop mode ⁽²⁾	<i>Average</i>	22 μ s	83 μ s	120 μ s
	<i>Total</i>		105 μ s	142 μ s
Standby mode ⁽³⁾	<i>Average</i>	320 μ s	3.37 ms	16 ms
	<i>Total</i>		3.69 ms	16.32 ms

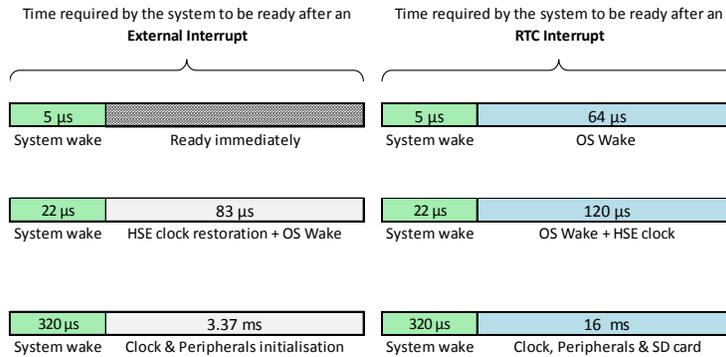


Fig.6. Required time to be ready

6 TEST

To determine the efficiency of the system, tests have to point out the average current consumption in the three low-power mode provided by the Cortex M4, which is also directly linked with the sampling rate. This point raises the question of the minimum time window to keep the system sleeping as far as possible, hence the need of testing the task timings.

In the architecture, the critical point about the time consumption concerns the data transmission time. Indeed, the low UART transfer rate of the acoustic modem (115kbaud) causes trouble, especially as the size of data transmitted grows at each passing of a node. So it is very important to know how much time the system needs to transfer data from a node to another according to the number of sensor data.

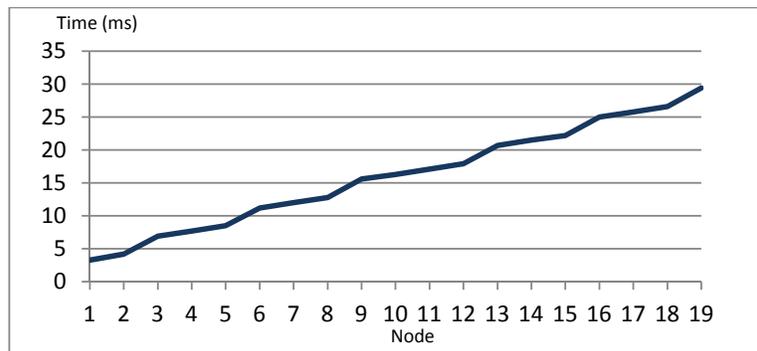


Fig.7. Time to send modem frame via UART depending on the number of nodes

The above test was performed on a sample of 1 to 19 sensor data transmitted, and this illustrates already a large amount of time spent. The non linearity of the transmission time is due to size payload limitation of modem frame. Thus, another frame needs

to be created to send the rest of sensor data. As a result an incremental jump is shown into the time graph of Fig. 7.

By forcing the transmission test to the upper nodes limit (theoretically up to 255), the test highlights memory trouble. Consequently, the application is limited to 100 nodes. At this value, transmission time reach 142ms. A linear trend line of the time graph gives the following formula:

$$t_{\text{uart Rx/Tx}} (ms) = 1.4x + 2$$

- $t_{\text{uart Rx/Tx}}$ corresponding to the reception/transmission time of modem frame sending.
- x corresponding to the number of data sensor to transmit.

UART timing is the only variable components of task time consumption. Table 4 shows the minimal running time spent by the application; it refers to the minimal timing operation for one sensor data transmitted and received.

Table 4. Minimum running time

Mode	Program	Timing
Sleep	Proper data flow	14,42 ms
	CRC error	+7 ms
Stop	Proper data flow	14,60 ms
	CRC error	+7,2 ms
Standby	Proper data flow	37,95 ms
	CRC error	+23,2 ms

The difference was made between a proper sensor data flow and the case of faulty reception. Because the system is waking up and has to transmit again sensor data.

Thanks to the previous current level measurement and the timing test, Fig.8 illustrates the minimum consumption of nodes, according to the sample frequency, the data to transmit in the linear system and the low power mode used.

The best option depends on the final application. In other words; it depends at which frequency we need to obtain information from sensors. In the case of we need to get data quickly, the stop mode is recommended. This mode offers the best compromise between running time and low power consumption.

For lower sampling (less than 1Hz), it's recommended to use the standby mode. It offers a low consumption in regard to the ratio between sleeping time and running time.

7 CONCLUSION

This paper presents an implementation of an algorithm that allows nodes sensor measurement and data transference in linear topology. The algorithm has been implemented in ARM Cortex™-M4 Discovery development kits. One of the advantages of the ARM cortex M4 is that it provides several operating modes which can switch be-

tween different clock sources in order to minimize its power consumption. This work has measured both time and power consumption constrains of this development kits assuming a linear communication topology.

Test had as objectives to determine the efficiency of the system according to the average current level and the minimal time window of sampling, in order to find the best ratio between rapidity and consumption. Test raised issue about time consuming of modem frame transmission and highlighted a memory limitation above 100 nodes; this set the physical maximum value of the underwater sensor network.

Moreover, to be the most power efficient, test leads to recommend the use of *Cortex M4 Stop mode* above 1Hz sampling rate; otherwise the use of *standby mode* is truly better even if this involves a program reboot. However, for these modes, the STM32F4-Discovery board suffers of extra consumption, so it is advisable to design a specific board to fully answer the power savings needs.

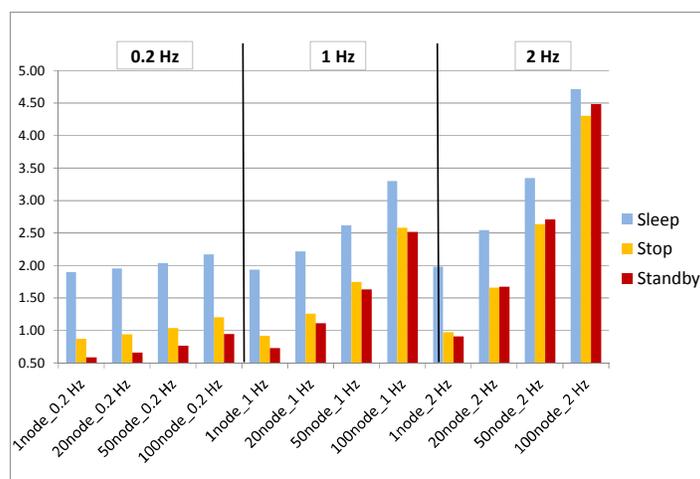


Fig.8. Minimum node consumption

Acknowledge: This work has partially funded by the Project number CTM2011-29691-C02-01 (Sonorización ambiental subacuática para la inspección y monitorización de explotaciones de acuicultura marina – Ministerio de Ciencia e Innovación).

References

1. Sanchez A, Blanc S, Yuste P, Serrano JJ. A low cost and high efficient acoustic modem for underwater sensor networks. OCEANS 2011 IEEE - Spain, IEEE: Santander, 2011; 1–10, doi:10.1109/Oceans-Spain.2011.6003428
2. Sanchez, S. Blanc, P. Yuste, A. Perles, and J. J. Serrano, “An Ultra-Low Power and Flexible Acoustic Modem Design to Develop Energy-Efficient Underwater Sensor Networks,” *Sensors*, vol. 12, no. 6, pp. 6837–6856, May 2012.
3. ST web page; available in : <http://www.st.com/web/catalog/tools/FM116/SC959/SS1532/PF252419>
4. Sanchez A, Blanc S, Yuste P, Serrano J. RFID Based Acoustic Wakeup System for Underwater Sensor Networks. 2011 IEEE 8th International Conference on Mobile Ad-Hoc and Sensor Systems, pp 873–878, doi:10.1109/MASS.2011.103.

Performance Analysis of Telecommunication Systems based on Time-Scale Separation ^{*}

Luis Tello-Oquendo, Vicent Pla and Jorge Martinez-Bauset

ITACA Research Institute
Universitat Politècnica de València, Spain

Abstract. The increasing complexity of the telecommunication systems has made modeling more challenging. A commonly used modeling tool are Markov chains, but the existence of several different components (e.g., user types) often renders its analysis computationally intractable. When these components operate at sufficiently separated time-scales, the quasi-stationary approximation has proven to be accurate and highly efficient. However, while the computational efficiency of this method is maintained no matter what the separation of time-scales is, its precision deteriorates rapidly as the separation narrows. In this paper, we present a new approximate method that extends the outreach of the quasi-stationary approximation by trading-off computational effort in exchange of increased accuracy. The proposed method is iterative in nature and its accuracy can be improved by performing more iterations.

1 Introduction

A large variety of telecommunication systems are inevitably large and complex, mainly due to the interactions among their components as different user types or traffic categories. Rapid progress in technology has also made modeling more challenging. A commonly used modeling tool for performance and dependability analysis of such systems are continuous-time Markov chains (CTMC). One of the main advantages of using Markovian models is that it is general enough to capture the dominant factors of system uncertainty and, in the meantime, it is mathematically tractable [15].

Incorporating all the important factors into the models often results in a large state space of the underlying Markov chain, rendering the analysis problem computationally intractable. Therefore, it is very necessary to develop computationally efficient approximations techniques to compute the performance parameters of the system under study.

In general, it is not possible to divide a large system into its smallest irreducible subsystems completely separable from one another and treat each subsystem independently, we have to deal with situations in which the systems are only nearly decomposable, in the sense that, there are weak links among the irreducible subsystems, which dictate the occasional regime changes of the system. An effective way to treat such near decomposability is the time-scale

^{*} This work has been supported by the Spanish “Ministerio de Economía y Competitividad” through the project TIN2010-21378-C02-02.

separation. That is, we set up the systems as if there were two time-scales, fast vs. slow [15].

When the events of the components of a system occur at sufficiently separated time-scales, the simplest approximation, producing easily computable and accurate results, is the so called quasi-stationary (or, quasi-static) approximation (QSA) [2, 3, 5, 7, 14].

In [4] the authors introduce a new approximation method, also based in time-scale decomposition, called Generalized Quasi-Stationary Approximation (GQSA), that provides a way to trade off computational complexity and accuracy. They apply it to an integrated services system that serve short-lived non-real-time and long-lived real-time traffic. This new approximation aims to improve the accuracy at the price of higher computational cost. In [13] it was shown that, while the new GQSA improves the accuracy in some instances, it does not occur in all of them; and more importantly, it is difficult to predict in which cases accuracy can be enhanced by GQSA.

Taking into account that QSA provide accurate results only when the separation of time-scales is considerable (quasi-stationary regime), the purpose of this paper is to provide an accurate and high efficient approximation method for the performance analysis of telecommunications systems, to improve the accuracy of QSA when the system moves away from the quasi-stationary regime, i.e., enlarge the region of validity of QSA when the separation of time-scales is not enough.

We analyzed two types of telecommunication systems: a cognitive radio system (CRS) with two user types [8], and an integrated services system (ISS) with two traffic categories [11]. These systems present qualitative important differences at the model level, as is described in Sect. 2.

The approximation that we propose is based on the phase-type (PH) distribution which can be characterized as the time to absorption in a finite Markov chain. The two-dimensional Markov process of each system (CRS and ISS) can be represented with a number of transient states and an absorbing state as is detailed in Sect. 3.

We approach the problem with an iterative method. Taking QSA as an initial approximation to the solution vector, we proceed to modify this approximation in such a way that, as the number of iterations increases, the result becomes closer and closer to the exact solution. This is an important advantage with respect to the GQSA approach, because we can use the previous result as the new initial approximation. If indeed there is little change, we should expect to compute the new result in relatively few iterations. Also, an iterative process may be halted once a prespecified tolerance criterion has been satisfied, and this may be relatively lax [12].

The remainder of the paper is organized as follows. Section 2 describes the Markov models, and it details the characteristics of the systems to analyze. Section 3 presents the proposed approximation method based on time-scale separation and PH distribution. Section 4 details the numeric evaluation of the systems and the results of the performance metrics computed with the proposed approximation method at different time-scales to validate its accuracy. Finally, the conclusions are presented in Section 5.

2 Description of Models

In this section, we detail the characteristics of the systems in which we applied the proposed approximation method to evaluate its performance. We describe the two-dimensional CTMC models associated with them.

2.1 Cognitive Radio System

As in [8], we model the PU and SU traffic at the session (connection) level and ignore interactions at the packet level (scheduling, buffer management, etc.). We assume an ideal MAC layer for SUs, which allows a perfect sharing of the allocated channels among the active SUs (all active SUs get the same bandwidth portion), introduce zero delay and whose control mechanisms consume zero resources. In addition, we also assume that an active SU can sense the arrival of a PU in the same channel instantaneously and reliably. In this sense, the performance parameters obtained can be considered as an upper bound.

The Cognitive Radio System has C_1 *primary channels* (PCs) that can be shared by PUs and SUs, and C_2 *secondary channels* (SCs) only for SUs. Let $C = C_1 + C_2$ be the total number of channels in the system. Note that the SCs can be obtained from e.g. unlicensed bands, as proposed in [1]. This assumption is applicable to the *coexistence* deployment scenario for CRNs [10]. Alternatively, as it might be of commercial interest for the primary and secondary networks to *cooperate*, the secondary channels may be obtained based on an agreement with the primary network [10]. A SU in the PCs might be forced to vacate its channel if a PU claims it to initiate a new session. As SUs support *spectrum handover*, a vacated SU can continue with its ongoing communication if a free channel is available. Otherwise, it is *forced to terminate*.

For the sake of mathematical tractability, Poisson arrivals and exponentially distributed service times are assumed. The arrival rate for PU (SU) sessions is λ_1 (λ_2), their service rate is μ_1 (μ_2), and requests consume 1 (1) channel when accepted. We denote by (i, j) the system state, when there are i ongoing PU sessions and j SU sessions. The set of feasible states is $\mathcal{S} := \{(i, j) : 0 \leq i \leq C_1, 0 \leq i + j \leq C\}$ and the cardinality of \mathcal{S} is $|\mathcal{S}| = (\frac{C_1}{2} + C_2 + 1) \cdot (C_1 + 1)$. The state-transition diagram of the system is depicted in Fig. 1. Given the set of feasible states and their transitions in a CTMC, we can construct the global balance equations and the normalization equation. From these, we calculate the steady-state probabilities denoted as $\pi(i, j)$.

The system's performance parameters are determined as follows:

$$P_1 = \sum_{k=0}^{C_2} \pi(C_1, k) \quad , \quad P_2 = \sum_{k=C_2}^C \pi(C - k, k), \quad (1)$$

$$P_{ft} = \frac{\lambda_1(P_2 - \pi(C_1, C_2))}{\lambda_2(1 - P_2)}, \quad (2)$$

$$Th_2 = \sum_{j=1}^C \sum_{i=0}^Z j \mu_2 \cdot \pi(i, j), \quad (3)$$

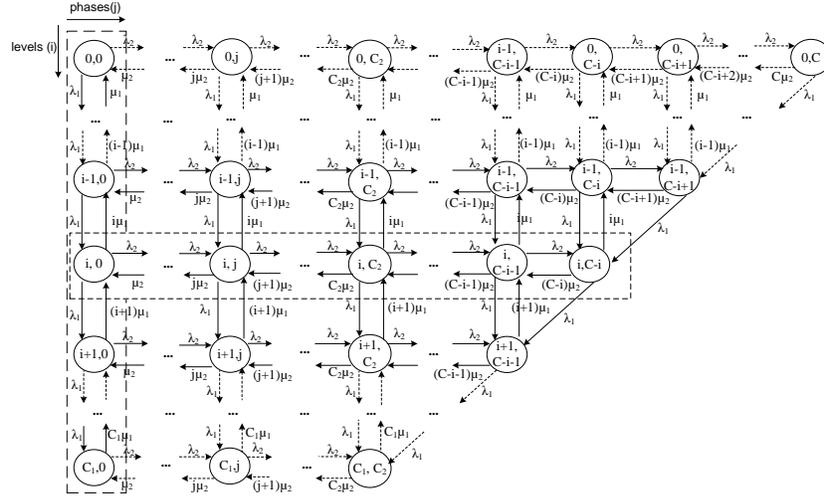


Fig. 1. State-transition diagram, Cognitive Radio System.

where P_1 is the PUs blocking probability, which clearly coincides with the one obtained in an Erlang-B loss model with C_1 servers; P_2 is the SUs blocking probability, i.e. the fraction of SU sessions rejected upon arrival as they find the system full; P_{ft} is the forced termination probability of the SUs, i.e. the rate of SU sessions forced to terminate divided by the rate of accepted SU sessions; Th_2 is the SUs throughput, i.e. the rate of SU sessions successfully completed and $Z = \min(C_1, C - j)$.

2.2 Integrated Services System

We use the same model defined in [4] for an Integrated Services System that serve real-time (RT) and non-real-time (NRT) traffic. We consider a link whose limited resources (C Mbps in total) are shared amongst RT and NRT requests. The RT traffic is given strict priority over the NRT traffic. We initially assume that all RT calls are of the same class each requiring one channel of rate c b/s during its entire service duration to meet its required QoS. We denote by N_{rt} the maximum number of channels for RT calls. When an RT call arrives, it occupies 1 channel if available; otherwise, it is blocked. We set N_{rt} , such that $N_{rt} \cdot c$ is sufficiently smaller than C to avoid starvation of the NRT traffic. Let $n_{rt}(t)$ be the number of RT calls in the system at time t , $t \geq 0$, so $\{n_{rt}(t), t \geq 0\}$ is the RT process. NRT flows are served evenly by the leftover capacity from the RT traffic according to the processor sharing (PS) discipline. Let $n_{nrt}(t)$ be the number of NRT flows in the system at time t , $t \geq 0$. Then, $\{(n_{rt}(t), n_{nrt}(t)), t \geq 0\}$ is the joint RT and NRT process. The capacity available for all the NRT traffic at time t is given by $C_{nrt}(t) = C - n_{rt}(t) \cdot c$. The bit-rate of each admitted NRT flow at time t is $c_{nrt}(t) = C_{nrt}(t)/n_{nrt}(t)$, which is updated with RT or NRT admitted arrivals or departures. To satisfy the QoS of admitted NRT flows, the maximum

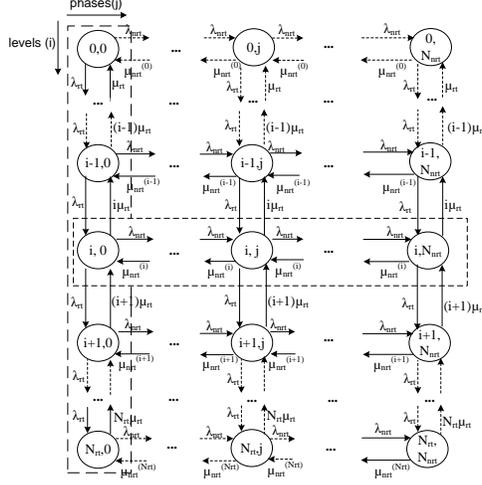


Fig. 2. State-transition diagram, Integrated Services System.

number of concurrent NRT flows is limited to N_{nrt} . Accordingly, an NRT flow arriving at time t is blocked if $n_{nrt}(t) = N_{nrt}$. We assume Poisson arrivals for RT and NRT requests with rates λ_{rt} and λ_{nrt} respectively. The service time of each admitted RT request is exponentially distributed, its service rate is μ_{rt} . On the other hand, as data sessions generate NRT traffic, their sojourn time will depend on the available resources. The size of the flows generated by the data sessions are exponentially distributed with mean L (bits).

We denote by (i, j) the system state, when there are i ongoing RT calls and j NRT flows. Let \mathcal{S} be the set of feasible states as $\mathcal{S} := \{(i, j) : 0 \leq i \leq N_{rt}, 0 \leq i+j \leq N_{rt}+N_{nrt}\}$ and the cardinality of \mathcal{S} is $|\mathcal{S}| = (N_{rt}+1)(N_{nrt}+1)$. The state-transition diagram of the system is depicted in Fig. 2. Given the set of feasible states and their transitions in a CTMC, we can construct the global balance equations and the normalization equation. From these, we calculate the steady-state probabilities denoted as $\pi(i, j)$. We must consider that the service rate of NRT flows varies according to the n_{rt} RT calls in the system as $\mu_{nrt}^{(i)} = \frac{C-i\epsilon}{L}$.

The system's performance parameters can be developed as follows:

$$P_{nrt} = \sum_{k=0}^{N_{rt}} \pi(k, N_{nrt}), \quad (4)$$

$$E[X_{nrt}] = \sum_{j=1}^{N_{nrt}} \sum_{i=0}^{N_{rt}} j \cdot \pi(i, j), \quad (5)$$

$$E[D_{nrt}] = \frac{E[X_{nrt}]}{\lambda_{nrt}(1 - P_{b_{nrt}})}, \quad (6)$$

where P_{nrt} is the blocking probability of NRT flows, $E[X_{nrt}]$ is the mean number of NRT flows in the system and $E[D_{nrt}]$ is the NRT flow average transfer delay.

3 Approximation Method

3.1 PH Distributions

Consider a CTMC on a finite state space $\mathcal{S} = \{0, 1, 2, \dots, m\}$. A PH distribution is the distribution of the time until absorption in a finite Markov chain of dimension $m + 1$, where 1 state is absorbing and the remaining m states are transients. A PH distribution is uniquely given by an m dimensional row vector $\boldsymbol{\alpha}$ and an $m \times m$ matrix \mathbf{T} . The vector $\boldsymbol{\alpha}$ can be interpreted as the initial probability vector among the m transient states (with $\sum_{i=0}^m \alpha_i = 1$), while the matrix T can be interpreted as the infinitesimal generator matrix among the transient states in the continuous case. The random variable that is defined as the time to absorption, is said to have a (continuous) PH distribution [9].

The infinitesimal generator for the CTMC can be written in block-matrix form as $Q = \begin{bmatrix} \mathbf{t} & \mathbf{T} \\ 0 & \mathbf{0} \end{bmatrix}$. Here, $\mathbf{0}$ is a $1 \times m$ vector of zeros. The vector $\mathbf{t} = (t_{10}, t_{20}, \dots, t_{m0})'$ (the prime denoting transpose) where, for $i = 1, 2 \dots m, t_{i0} \geq 0$, with at least one of the t_{i0} s positive, is the absorption rate from state i . The $m \times m$ matrix $\mathbf{T} = [t_{ij}]$ is such that, for $i = 1, 2 \dots m$, with $i \neq j$,

$$t_{ij} \geq 0 \quad \text{and} \quad t_{ii} = - \sum_{j=0, j \neq i}^m t_{ij}, \quad (7)$$

that is, $\mathbf{t} = -\mathbf{T}\mathbf{e}$ where \mathbf{e} is a column vector of ones of appropriate dimension. We call the the pair $(\boldsymbol{\alpha}; \mathbf{T})$ a representation for the PH distribution. The matrix \mathbf{T} is referred to as a *PH generator*.

To ensure absorption in a finite time with probability one, we require that every non-absorbing state is transient. This statement is equivalent to \mathbf{T} being invertible, therefore $-(\mathbf{T}^{-1})_{ij}$ is the expected total time spent in phase j during the time until absorption, given that the initial phase is i [6, Theorem 2.4.3].

For extensive bibliographies and comprehensive theoretical treatment of PH distributions, see [9, Chap.2]. Also [6, Chap.2], is a very readable introduction to the topic.

3.2 Absorbing Markov Chains Method (AbMC)

In the quasi-stationary approximation it is assumed that, when the process enters a level, it takes an infinitely long time to leave this level. Thus, the probability that the process is in a phase j of the level i , given that the process is in the level i , is simply given by the stationary distribution of the level i considered as an isolated CTMC.

In our method, we assume that although the sojourn time in a level will be typically large (consistently with the large separation between time-scales) it is finite. We treat each level as an absorbing Markov chain in which the transitions out the level are transitions to an absorbing state. We now obtain the probability that the process is in a phase j of the level i , given that the process is in the level i , as the fraction of time that the process spends in the phase j before absorption. Note that if we knew the initial probabilities of the phases (i.e., upon entering the level) then the conditional probabilities obtained

by this method would be the exact ones. However, unless the original CTMC has some special structure (for instance, if each level can only be entered by exactly one of its phases), these initial probabilities cannot be obtained without having the stationary distribution of the whole CTMC.

We propose here to use the QSA to estimate the initial probabilities for the phases in each level. Then, we can obtain an estimation of the conditional probabilities for the phases as described above, i.e., as the fraction of time spent in each of them before absorption. Now, from the estimation of the conditional probabilities for the phases in each level, and the probability distribution for the levels, a new approximation for the stationary distribution of the CTMC is obtained. This way, we have obtained a refinement of the initial approximation of the stationary distribution given by the QSA. Moreover, the same process can be repeated iteratively to further improve the approximation.

Based on the basic properties of PH distributions introduced in Sect. 3.1, the iterative procedure described above can be implemented using the following equations:

$$\tilde{\boldsymbol{\pi}}_i^{(k+1)} = \left[\boldsymbol{\alpha}_i^{(k)} (-T_i^{-1}) \mathbf{e} \right]^{-1} \boldsymbol{\alpha}_i^{(k)} (-T_i^{-1}) \quad (8)$$

$$\mathbf{v}_i^{(k)} = \pi_{i-1} \tilde{\boldsymbol{\pi}}_{i-1}^{(k)} \mathbf{u}_{i-1} + \pi_{i+1} \tilde{\boldsymbol{\pi}}_{i+1}^{(k)} \mathbf{d}_{i+1} \quad (9)$$

$$\boldsymbol{\alpha}_i^{(k)} = \frac{1}{\mathbf{v}_i^{(k)} \mathbf{e}} \mathbf{v}_i^{(k)} \quad (10)$$

- Being k the iteration number, $\tilde{\boldsymbol{\pi}}_i^{(k)}$ is the distribution of probabilities obtained by the PH distribution. The initial value is given by

$$\tilde{\boldsymbol{\pi}}_i^{(0)} = \hat{\boldsymbol{\pi}}(j|i), \quad (11)$$

where $\hat{\boldsymbol{\pi}}(j|i)$ is the distribution of probabilities of:

- finding j ongoing sessions in an $M/M/(C-i)/(C-i)$ system with only SUs in CRS.
- finding j NRT flows in an $M/M/1/N-PS$ system with only NRT traffic in ISS.
- $\boldsymbol{\pi}$ is the distribution of probabilities of the elements with strict priority in the system (PUs in CRS, RT traffic in ISS). It is computed using simple recursions since their corresponding CTMC are one-dimensional birth-and-death processes.
- $\mathbf{v}_i^{(k)}$ is a vector that represent the input rates to each state of the level i . The initial value for the iterations is given by:

$$\mathbf{v}_i^{(0)} = \pi_{i-1} \tilde{\boldsymbol{\pi}}_{i-1}^{(0)} \mathbf{u}_{i-1} + \pi_{i+1} \tilde{\boldsymbol{\pi}}_{i+1}^{(0)} \mathbf{d}_{i+1}, \quad (12)$$

where \mathbf{u}_{i-1} is a vector with the transition rates from level $i-1$ to level i and \mathbf{d}_{i+1} is a vector with the transition rates from level $i+1$ to level i .

- $\boldsymbol{\alpha}_i^{(k)}$ is the initial probability vector among the j transient states of the level i .

$$\boldsymbol{\alpha}_i^{(0)} = \frac{1}{\mathbf{v}_i^{(0)} \mathbf{e}} \mathbf{v}_i^{(0)}. \quad (13)$$

Finally, the state probability distribution of the systems can be approximated as $\pi(i, j) \approx \tilde{\pi}(i, j) = \pi_i \cdot \tilde{\pi}_i^{(k)}(j)$. With this distribution of probabilities, we compute the approximate values of the performance parameters, using (1 – 3) for CRS, and (4 – 6) for ISS. Note that, if we were able to know the initial probability vector of the PH distribution (α), just an iteration would be enough to obtain the exact distribution of probabilities $\pi(i, j)$.

4 Numerical Evaluation and results

As a baseline for our study, we implemented the exact solution of the CTMC associated with each system, to calculate the exact values of their performance parameters. For the sake of comparison, we have used system sizes that allowed the exact solution to be computed within a reasonable time. In addition, we implemented both approximation methods: QSA and GQSA to validate and compare the performance of AbMC. We focus on evaluating the relative error (e_r) of each performance parameter. For instance, the relative error of the blocking probability for SUs in a CRS is computed as $e_r(P_2) = \frac{|P_2^E - P_2^{Ab}|}{P_2^E}$, where P_2^E is the exact value of SUs blocking probability and P_2^{Ab} is the approximate value of SUs blocking probability computed by AbMC method.

We evaluate the performance of the systems with different sizes (number of channels available for each type of user or flow) and different load conditions. In CRS, we analyze the blocking probability, forced termination probability and throughput of the SUs. We consider the following values for the number of primary channels: $C_1 = \{70, 80, 90, 100, 120, 140\}$, and for each of them, the following values for the number of secondary channels are considered: $C_2 = \{10, 20, 40\}$.

In ISS, we determine the blocking probability and average transfer delay of the NRT traffic. Keeping c and L constant, we consider the following values for the total link capacity of the system: $C = \{1.92, 7.68\}$ Mbps, that are similar to the ones used in [4].

To establish the load conditions, we set the service rates to 1, and then we adjust the arrival rates to obtain two load conditions: low (L) and high (H), which correspond to blocking probabilities of $1 \cdot 10^{-3}$ and $5 \cdot 10^{-2}$, respectively for each user type or traffic category, given as result four load configurations: *LL*, *LH*, *HL* and *HH*.

To assess the goodness of AbMC, we evaluate the evolution of the relative error as the separation between the time-scales varies. To accomplish this analysis, with the arrival rates adjusted to the specified load condition (*LL*, *LH*, *HL* or *HH*), we use an accelerating factor f , $10^{-5} \leq f \leq 10^5$, to accelerate or decelerate the arrival and departure events of the PUs, in the case of CRS or of the RT traffic, in the case of ISS, while keeping the offered traffic constant. For instance in a CRS, for each value of f , the PU arrival and service rates are obtained as $\lambda_1(f) = f \cdot \lambda_1$ and $\mu_1(f) = f \cdot \mu_1$. Note that the offered traffic $\frac{\lambda_1}{\mu_1} = \frac{\lambda_1(f)}{\mu_1(f)}$ is independent of f .

We varied the accelerating factor f to analyze the behavior of the approximation methods as a function of the separation of time-scales. To obtain the results with the iterative methods, we measured the time to compute *GQSA*

with the respective Radius R and iterated AbMC until that time, e.g., for $AbMC_1$ we iterated the method until the time measured to compute $GQSA_{R=1}$. In the same way for the other representations. The results are shown in Figs. 3 and 4; from them we can make the following observations:

- As expected, when the accelerating factor f decreases ($f \rightarrow 0$), the approximate values of all performance parameters evaluated tend to the QSA and the QSA results are very accurate.
- Note that increasing the radius in GQSA not always ensures a gradual reduction of the relative error [13]. As can be seen in Fig 3, $GQSA_{R=1}$ has a better accuracy in comparison with $GQSA_{R=2}$ and $GQSA_{R=3}$ for $f > 10^{-1}$. Clearly, the trade-off between the accuracy and computational cost will discourage the use of a radius larger than $R = 1$ in GQSA.
- AbMC have a predictable behavior: increasing the number of iterations to find the solution, ensures an improvement of the accuracy. Although they are not represented here due to space limitations, we have observed the same behavior in all performance parameters, for all load conditions and system sizes.

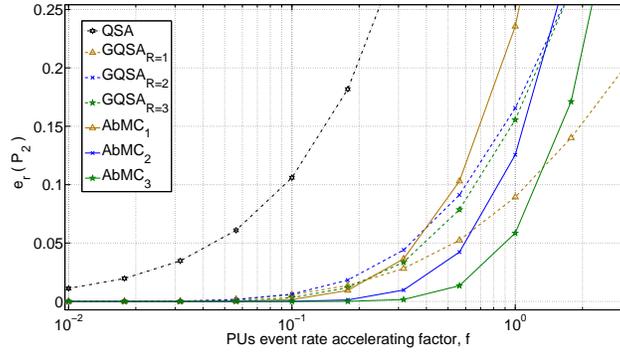


Fig. 3. CRS, Relative Error for the Blocking Probability SUs, LH load condition; $\lambda_1 = 49.239$, $\mu_1 = 1$, $C_1 = 70$; $\lambda_2 = 25.601$, $\mu_2 = 1$, $C_2 = 10$.

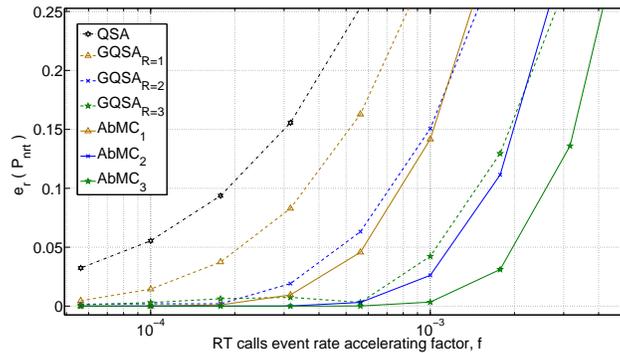


Fig. 4. ISS, Relative Error for the Blocking Probability NRT, HL load condition; $\lambda_{rt} = 38.557$, $\mu_{rt} = 1$, $N_{rt} = 44$; $\lambda_{nrt} = 1.243$, $N_{nrt} = 60$; $C = 7.68$ Mbps, $c = 64$ kbps, $L = 4$ Mb.

5 Conclusions

We provide a new approximate method called Absorbing Markov Chains (AbMC) for the performance analysis of telecommunication systems. We have validated AbMC in comparison with the quasi-stationary approximation and its recently proposed generalization. We have explored the evolution of the accuracy at different time-scales in two telecommunication systems: a cognitive radio system and an integrated services system which, at the model level, present qualitative important differences. The numerical results demonstrate that AbMC extends the outreach of the quasi-stationary approximation by trading-off computational effort in exchange of increased accuracy. AbMC has an iterative approach and as a part of future work, we will study the trade-off between accuracy and computational cost.

References

1. Al-Mahdi, H., Kalil, M.A., Liers, F., Mitschele-Thiel, A.: Increasing spectrum capacity for ad hoc networks using cognitive radios: an analytical model. *IEEE Communications Letters* 13(9), 676–678 (Oct 2009)
2. Alexiades, V., Solomon, A.D.: *Mathematical modeling of melting and freezing processes*. Taylor & Francis (1993)
3. Boxma, O.J., Gabor, A.F., Núñez-Queija, R., Tan, H.P.: Performance analysis of admission control for integrated services with minimum rate guarantees. In: *Proceedings of NGI'06*. pp. 41–47 (2006)
4. Huang, Y., Ko, K., Zukerman, M.: A generalized quasi-stationary approximation for analysis of an integrated service system. *IEEE Communications Letters* 16(11), 1884–1887 (Nov 2012)
5. Hubner, F., Tran-Gia, P.: Quasi-stationary analysis of a finite capacity asynchronous multiplexer with modulated deterministic input. *ITC-13, Copenhagen* (1991)
6. Latouche, G., Ramaswami, V.: *Introduction to Matrix Analytic Methods in Stochastic Modeling*. ASA-SIAM (1999)
7. Liu, S., Virtamo, J.: Performance analysis of wireless data systems with a finite population of mobile users. In: *Proceedings of the 19th International Teletraffic Congress ITC 19*. pp. 1295–1304 (2005)
8. Martinez-Bauset, J., Pla, V., Vidal, J., Guijarro, L.: Approximate analysis of cognitive radio systems using time-scale separation and its accuracy. *IEEE Communications Letters* 17(1), 35–38 (Jan 2013)
9. Neuts, M.: *Matrix-geometric Solutions in Stochastic Models: An Algorithmic Approach*. The Johns Hopkins University Press (1981)
10. Peha, J.: Sharing spectrum through spectrum policy reform and cognitive radio. *Proceedings of the IEEE* 97(4), 708–719 (2009)
11. Roberts, J.W.: Internet traffic, qos, and pricing. *Proceedings of the IEEE* 92(9), 1389–1399 (2004)
12. Stewart, W.J.: *Probability, Markov chains, queues, and simulation: the mathematical basis of performance modeling*. Princeton University Press (2009)
13. Tello-Oquendo, L., Pla, V., Martinez-Bauset, J.: Approximate analysis of wireless systems based on time-scale decomposition. In: *Wireless Days (WD), 2013 IFIP*. pp. 1–6 (Nov 2013)
14. Wong, E., Foh, C.: Analysis of cognitive radio spectrum access with finite user population. *IEEE Communications Letters* 13(5), 294–296 (2009)
15. Yin, G.G., Zhang, Q.: *Discrete-time Markov chains: two-time-scale methods and applications*, vol. 55. Springer (2006)

Multi criteria decision making techniques in dependability benchmarking: How to proceed?

Miquel Martínez, David de Andrés, and Juan-Carlos Ruiz

Instituto de Aplicaciones de las TIC Avanzadas (ITACA)
Universitat Politècnica de València, Campus de Vera s/n, 46022, Spain
Email: {mimarra2, ddandres, jcruizg}@disca.upv.es

Abstract. Dependability benchmarking has become an essential discipline in different domains to perform the assessment of systems and components under faulty conditions by evaluators. However, despite all the efforts made to improve the process of dependability benchmarking, there are still some aspects that require further attention. The analysis of the resultant measures to extract meaningful conclusions from the benchmarks, is a part of the process that has no specific guidelines on how benchmark users should proceed. This negatively affects the essence of dependability benchmarking process, where several attributes must be satisfied, and present some problems in the analysis of results. In this paper, these problems are analysed, and multi-criteria decision-making techniques are proposed as an approach to cope with this problems. In addition to that, this paper explains how these techniques must be introduced in the dependability benchmarking process to satisfy the expected attributes, thus improving the whole process.

1 Introduction

For many years the evaluation of a system's features made reference to the evaluation of those related to its performance. Nevertheless, the need for providing dependable systems in presence of perturbations, has led to a current state of affairs in which many people from both academia and industry evaluate the dependability of systems, in addition to their performance, with comparison and selection purposes.

During all these years, many efforts have been focussed on improving the dependability benchmarking process. The work undertaken in [1] was aimed at defining the guidelines that were used by other works to perform more robust and well defined dependability benchmarks. These guidelines encompass the attributes that must be preserved when developing a dependability benchmark in all the stages of the process. According to this work [1], the dependability benchmarking process can be structured in the three main stages that are depicted in Figure 1.

- A first stage where the requirements of the experiments that will be performed to evaluate the system are defined. Some of these requirements like

the system under benchmarking (SUB) or the benchmark targets (BT) are part of the *configuration setup*. Others more specific to the execution of the benchmark are included in the *execution profile*. Such as the *workload* used to stimulate the system, the type of faults introduced in the system (*faultload and attackload*), the changes affecting the system for some domains (*changeload*), the measurements that will be collected from the BTs during the experimentation, and the final (performance, dependability, security and/or power consumption) measures that will be derived from those measurements.

- The second stage consists in applying these requirements to the experimental platform and execute the experiments to retrieve the expected measurements.
- In the last stage, two phases can be identified: i) The analysis and correlation of the measurements to quantify the measures defined in the first stage, and ii) a posteriori interpretation of these measures to provide conclusions.

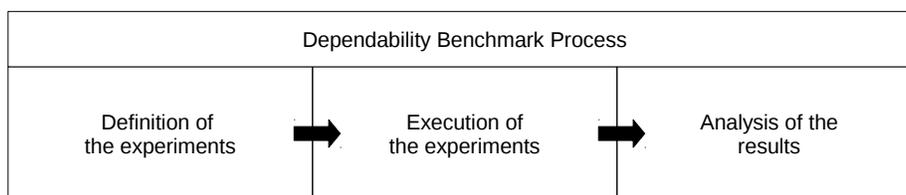


Fig. 1. Main parts of a dependability benchmark

The satisfaction of the dependability benchmarking attributes ensure the quality of its process. Attributes like the **representativeness** is achieved by defining representative requirements for a given application domain, such as measures, workload, changeload, faultload and attackload. The whole dependability benchmark process must be repeatable and reproducible. The **repeatability** is a central property of benchmarking, and it guarantees *”statistically equivalent results when the benchmark is run more than once in the same environment (using the same requirements)*. In the other hand, **reproducibility** grants that *another party obtains statistically equivalent results when the benchmark is implemented from the same specifications and is used to benchmark the same SUB*. The benchmark specifications must be feasible to be applied to various target systems within a particular application domain, thus making the dependability benchmark **portable**. The execution of the benchmark must be as **non-intrusive** as possible, otherwise the measurements gathered from the benchmark will not be representative from a real behaviour. Other attributes like **scalability** or **time and cost efficiency** must be satisfied by a dependability benchmark.

Based on these guidelines, works in different application domains (such as *web servers* [2], *on-line database transactional systems*[3], or *automotive systems*[4],

among others) were devoted to develop dependability benchmarks adapted for each domain. Indeed, the dependability benchmarking process has been studied and improved from different points of view. Works like [5], pointed out that in the dependability community, little attention was paid to the proper characterization of the measurements from a metrological [6] point of view. In fact, works in the field of metrology have been done to improve the quality of dependability measurements in different domains (like [7] and [8]).

However, and despite all these efforts, so far, little or no attention have been paid to rigorously determine how resultant measures should be analysed to provide meaningful conclusions. After analysing many works from the literature, like those presented in [9], the process of how to compute measures from measurements is usually detailed in depth to show the correctness of such process and enabling other researchers to obtain the same measures. However, conclusions are usually based on the evaluator's criteria (which is not a bad thing), but the process followed to provide conclusions from the measures is usually missing or quite vague, making sometimes hard to understand their origin. Considering a dependability benchmarking point of view, this presents a problem of *reproducibility* and *repeatability*. The reproducibility of the analysis is extremely complicated, or impossible, as external evaluators performing the same dependability benchmark are not able to perform the same analysis over their measures, thus their conclusions cannot be compared with those from other works. Repeatability in the other hand is affected because not making explicit the analysis makes evaluators difficult to repeat in new experiments, the same analysis done in previous ones.

Unfortunately, these are not the only problems related to the analysis. The techniques commonly used by the evaluators to analyse the results are usually simplistic or too general. The arithmetic or geometric mean are examples of that, and their application do not allow to perform complex analysis from the measures obtained. Techniques like these ones consider all criteria equally important for the analysis, which is not a realistic approach in real situations. When evaluating a system to assess its performance and dependable features for an specific environment, achieving a good result in some criteria might be considered more important than in others. For this reason, the analysis of dependability benchmarking measures requires from techniques that allow not only the performance of complex analysis, but also that makes the process explicit, thus assuring its repeatability and reproducibility.

The use of *multi-criteria decision-making* (MCDM) techniques have been proposed in previous works to perform the analysis of the results in order to provide meaningful conclusions [10][11]. These techniques, typically used in the field of operational research [12], have been used for many years in different areas (like business [13], education [14], or engineering [15]) to solve problems where conflicting criteria were involved to reach a decision. However, how these techniques should be integrated in the process of a dependability benchmark has not been addressed in detail so far.

This work is aimed at defining how MCDM techniques should be used in the dependability benchmarking domain to perform robust, repeatable and reproducible analysis of the resultant measures. To do so, next section introduces the *multi-criteria decision-making* techniques and gives an overview of their features, and briefly describes a few of them. Section 3 describes how MCDM techniques should be combined within the process of dependability benchmarking, and explain its benefits for the process of analysis. The conclusions of the paper and the in-progress future work are presented in the last section, Section 4.

2 MCDM techniques

MCDM techniques involves evaluation of various alternative solutions upon a set of criteria. The result of multi criteria decision making is the best alternative which secures the highest score with the predefined criteria. These criteria are often conflicting but demanded by the different evaluators. Usually, these criteria are weighted in an order to represent their relevance in the final selection, thus evaluators have to make trade-offs between criteria in order to balance the relevance among them for the analysis.

There are many techniques that can be found in the literature [16] that can be used to perform the interpretation of measures. However, there are certain features that are common to all MCDM techniques.

- Calculate the weight of the criteria for evaluation
- Evaluate each alternative solution separately on each criteria
- Calculate the overall score of each alternative by aggregating the evaluations on each criteria

Among the multiple MCDM techniques available, each of them has a particular procedure to perform the analysis of criteria. These mathematical differences in their application has lead to a wide variety of techniques applied in many domains to solve multi-criteria decision-making problems. According to their particularities, techniques can be grouped into families, in works like [17] a good classification of the different families of available techniques can be found. Depending on the technique, not only the process, but also the criteria from the evaluation can be interpreted in different ways. For example, the analytic hierarchy process (AHP) [18] and its various variants make use of pairwise comparisons using a cardinally deployable verbal scale to fill what is known as *comparison matrices*. From these matrices, the weights that determine the relevance of each criteria to perform the analysis are derived through the eigenvector, and used afterwards.

Other techniques determine the quality of each alternative applying geometrical space distance vectors. The technique for order preferences by similarity to an ideal solution (TOPSIS) [19] is an example of this. From the available alternatives, TOPSIS define an ideal and anti-ideal solution based on the best and worst values for each criteria considering all the alternatives. This way, the best

alternative is considered as the one with the shortest distance in the geometrical space to the ideal solution, and the largest to the anti-ideal one.

Some techniques are more commonly used as tools for screening problems. Potentially all pairwise rankings of all possible alternatives (PAPRIKA) [20] technique defines a set of value levels for each criterion (such as low, medium and high). Undominated pairs (pair of alternatives where one is characterized by higher level of one criterion and another is characterized by higher level of another criterion) are presented to decision makers, which must provide their preference for each pair, either strict preference or indifference. Corollaries are generated from a few of these available preferences and ranks of the alternatives are generated. Yet another family of screening problems are the techniques from the ELECTRE [21] family. ELECTRE I has been improved over time to ELECTRE II, III, IV, Tri. ELECTRE techniques use outranking relationships and create the final ranking of alternatives. They basically analyse the dominance relationships existing between the alternatives. They are particularly good when evaluating large number of alternatives, but can have some issues with contrasting the results.

These are just a few number of techniques among the large amount of available MCDM techniques that can be used to solve multi-criteria decision problems. However, what makes them suitable to be applied in a dependability benchmarking domain is the fact that the process followed in the analysis can remain explicit, thus making the analysis repeatable and reproducible. In fact, their proper integration into the dependability benchmarking process can remove certain uncertainties present in the analysis.

The correct integration of these techniques into the dependability benchmarking process is detailed in next section.

3 MCDM techniques in Dependability Benchmarking

In order to trust the results of a dependability benchmark it is necessary that the attributes mentioned in Section 1 are satisfied to some extent. The problem with the analysis of results, apart from its repeatability and reproducibility is that when the process is over, the evaluator takes part in the analysis of the provided measures to derive some conclusions from them. Performing the analysis of the measures once they are available presents a problem of subjectivity in the analysis. As no restrictions are setted in the technique that must be used to perform the analysis, based on the measures, the evaluator can perform a selfish analysis of the resultant measures to match expected conclusions. This situation, in addition to the fact that not making explicit the process of analysis affects its repeatability and reproducibility, must be understood as a serious problems to consider the conclusions from the dependability benchmark reliable and trustworthy.

Then, it can be stated that not making explicit the process of the analysis may incur situations of error. Different evaluators can analyse the same results and provide conclusions that differ from one another. Or even worse, the results

obtained from similar works cannot be compared due to the impossibility of determining how the analysis was performed. All these, are obstacles for the progress of the research in this domain.

MCDM techniques are a suitable approach not only to make explicit the process followed in the analysis, but also to reduce the subjectivity that can be introduced by the evaluator. Knowing in advance the kind of measures that will be obtained (defined in the first stage of the process), which will be considered as criteria for the analysis, let the evaluator determine the requirements for the analysis *a priori*, before the experiments take place. The technique used to perform the analysis will determine the way these requirements are represented, from now on this representation will be referred as *analysis model*. For example, let us suppose that four measures are defined as outcome measures from the dependability benchmark (M_1, M_2, M_3 and M_4). If AHP were the chosen technique to perform the analysis, the analysis model would be built in a hierarchical way upon those. These measures conform the leaves of the tree and are aggregated into higher level criteria, which at the same time can be aggregated into others until a single criterion (the root of the hierarchy) represents the score of the analysis. This way, the criteria are weighted according to their relevance to achieve their higher criteria, and the analysis can be computed when the values for the lower criteria (measures) are available. Figure 2 shows a possible example of an analysis model of the requirements when using AHP. In this case, M_1 and M_2 would be grouped into criterion C_1 , and M_3 and M_4 into criterion C_2 . The weights (known as priorities in AHP) that determine the relevance of each criterion, are depicted next to the criteria in Figure 2.

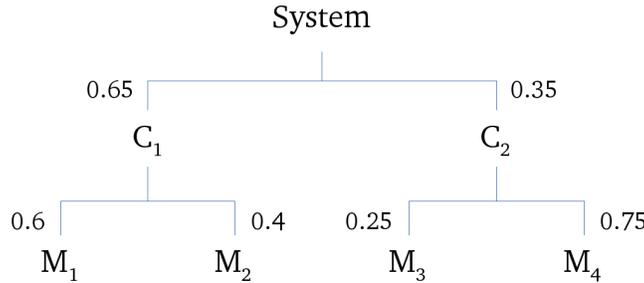


Fig. 2. Example of an analysis model of the requirements using AHP

As mentioned before, in the first stage of the dependability benchmarking process all the aspects regarding the benchmark are settled, ranging from those related to the configuration of the experiments (system under benchmark, benchmark targets, workload, faultload, etc) to the measures that will be obtained from the experimentation. It is in this early stage of the benchmark, where the requirements to perform the analysis must be defined, and so the analysis model build. Building the analysis model at this stage removes the subjectivity from

the process of analysis, as this process is defined before the values for the defined measures are obtained. Thus, evaluators are not tempted to adjust their analysis to provide previously expected conclusions. Figure 3 shows how the dependability benchmarking process should be modified to integrate the use of MCDM techniques to perform the analysis of resultant measures. It can be seen in Figure 3, that in the first stage is where the requirements for the analysis must be defined (thus building the analysis model). Then, the analysis model build in the first stage is applied in the last stage after the values for the measures are obtained, and so the analysis is performed and the conclusions can be provided.

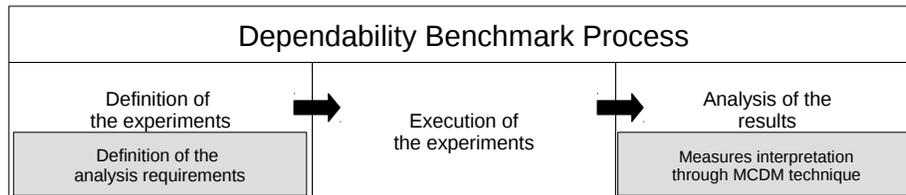


Fig. 3. Integrating MCDM in the dependability benchmark process

4 Conclusions and Future work

Dependability benchmarks have become more robust and accurate during the last years thanks to the efforts done by both the research and the industry communities. However, to date there is no consensus on how the measures provided by these benchmark must be analysed in order to obtain meaningful conclusions for the end users. Currently, the decision of how to proceed with this analysis entirely remains on the evaluator. This fact present some problems for dependability benchmarking that have been pointed out in this paper, like the vague or missing description of the process of analysis performed, the use of techniques that are not prepared to perform complex analysis, and the possibility selfish analysis made by evaluators.

These problems have been tackled in this paper with the use of *multi-criteria decision-making* (MCDM) techniques. These techniques allow the user to perform complex analysis of the measures based on pre-established requirements, where they can determine the relevance of each measure for the analysis. These requirements are the key to build an analysis model that is used to perform the analysis of the measures in order to provide conclusions from the application of the benchmark. Then, as the value of the measures are not required to define how the analysis should be performed, these requirements can be established beforehand, thus removing the *a posteriori* subjective part of the analysis introduced by the evaluators. Also, the representation of the requirements for the analysis through an analysis model makes explicit the reasoning applied by the

evaluator in the analysis. Then, other evaluators can reproduce the same analysis to cross-compare their results, and evaluators can repeat their own analysis later in time if required.

Then, the use of MCDM techniques to analyse the results of a dependability benchmarking assures not only the repeatability and reproducibility of the analysis, but also the possibility to perform a complex analysis more similar to those that would be made in a real situation. Also, their integration in the process as described in this work grants that subjectivity is removed from the whole process of analysis.

However, the number of available MCDM techniques, and the fact that all of them claim to provide the best conclusions, makes it necessary to perform further research to determine ones that should be used for analysis of results in dependability benchmarking. A study of different dependability benchmark results in different domains may be useful to determine the level of suitability of each technique to be applied in a given domain. This is part of ongoing work.

Acknowledgment

Work partially supported by the Spanish project ARENES (TIN2012-38308-C02-01).

References

1. DBench, "Dependability Benchmarking." IST Programme, European Commission, IST 2000-25425, [Online]. Available: <http://www.laas.fr/DBench>, 2013.
2. J. Dures, M. Vieira, and H. Madeira, "Dependability benchmarking of web-servers," in *Computer Safety, Reliability, and Security* (M. Heisel, P. Liggesmeyer, and S. Wittmann, eds.), vol. 3219 of *Lecture Notes in Computer Science*, pp. 297–310, Springer Berlin Heidelberg, 2004.
3. M. Vieira and H. Madeira, "A dependability benchmark for oltp application environments," in *Proceedings of the 29th international conference on Very large data bases - Volume 29*, VLDB '03, pp. 742–753, VLDB Endowment, 2003.
4. J.-C. Ruiz, P. Yuste, P. Gil, and L. Lemus, "On benchmarking the dependability of automotive engine control applications," in *IEEE/IFIP International Conference on Dependable Systems and Networks*, pp. 857–866, 2004.
5. A. Bondavalli, A. Ceccarelli, L. Falai, and M. Vadursi, "A new approach and a related tool for dependability measurements on distributed systems," *Instrumentation and Measurement, IEEE Transactions on*, vol. 59, no. 4, pp. 820–831, 2010.
6. "VIM3: International Vocabulary of Metrology - Basic and General Concepts and Associated Terms." Joint Committee for Guides in Metrology, [Online]. Available: <http://www.bipm.org/en/publications/guides/vim.html>, 2012.
7. L. Angrisani and M. Vadursi, "Cross-layer measurements for a comprehensive characterization of wireless networks in the presence of interference," *Instrumentation and Measurement, IEEE Transactions on*, vol. 56, no. 4, pp. 1148–1156, 2007.
8. M. Choi, N. Park, V. Piuri, and F. Lombardi, "Reliability measurement of mass storage system for onboard instrumentation," *Instrumentation and Measurement, IEEE Transactions on*, vol. 54, no. 6, pp. 2297–2304, 2005.

9. K. Kanoun and L. Spainhower, *Dependability benchmarking for computer systems*, vol. 72. John Wiley & Sons, 2008.
10. M. Martínez, D. de Andrés, J.-C. Ruiz, and J. Frigal, “Analysis of results in dependability benchmarking: Can we do better?,” *M&N 2013, International Workshop on Measurements and Networking*, pp. 127–131, 2013.
11. M. Martínez, D. D. Andres, and J.-C. Ruiz, “Gaining confidence on dependability benchmarks’ conclusions through “back-to-back” testing (practical experience report),” in *Dependable Computing Conference (EDCC), 2014 Tenth European*, pp. 130–137, May 2014.
12. M. Koksalan, J. Wallenius, and S. Zionts, *Multiple Criteria Decision Making: From Early History to the 21st Century*. World Scientific Publishing Company; 1 edition (June 6, 2011), 2012.
13. M.-K. Chen and S.-C. Wang, “The critical factors of success for information service industry in developing international market: Using analytic hierarchy process (ahp) approach,” *Expert Systems with Applications*, vol. 37, no. 1, pp. 694–704, 2010.
14. X. Sun *et al.*, “Construction and operation of analytic hierarchy process about moral education evaluation in colleges and universities.,” *Advances in Information Sciences & Service Sciences*, vol. 3, no. 11, 2011.
15. İ. Ertuğrul and N. Karakaşoğlu, “Performance evaluation of turkish cement firms with fuzzy analytic hierarchy process and topsis methods,” *Expert Systems with Applications*, vol. 36, no. 1, pp. 702–715, 2009.
16. J. Dodgson, M. Spackman, A. Pearman, and L. Phillips, *Multi-criteria analysis: a manual*. Department for Communities and Local Government: London, 2009.
17. V. Skula, *Comprehensive Methodology for Complex Systems’ Requirements Engineering and Decision Making*. PhD thesis, LAAS-CNRS, Toulouse, France, 2014.
18. T. Saaty, “What is the analytic hierarchy process?,” in *Mathematical Models for Decision Support* (G. Mitra, H. Greenberg, F. Lootsma, M. Rijkaert, and H. Zimmermann, eds.), vol. 48 of *NATO ASI Series*, pp. 109–121, Springer Berlin Heidelberg, 1988.
19. Y.-J. Lai, T.-Y. Liu, and C.-L. Hwang, “Topsis for modm,” *European Journal of Operational Research*, vol. 76, no. 3, pp. 486 – 500, 1994. Facility Location Models for Distribution Planning.
20. P. Hansen and F. Ombler, “A new method for scoring additive multi-attribute value models using pairwise rankings of alternatives,” *Journal of Multi-Criteria Decision Analysis*, vol. 15, no. 3-4, pp. 87–107, 2008.
21. B. Roy, “Electre iii: Un algorithme de classements fondé sur une représentation floue des préférences en présence de criteres multiples,” *Cahiers du CERO*, vol. 20, no. 1, pp. 3–24, 1978.

Best Practice in Horizon 2020: From Drawing a Proposal to Final Project Closure

Jose A. Morgado-Migueles¹, M^a Jesús Arnal-Parada²,
M^a José Nodal-López³, and José M. Dahoui-Obon⁴

¹ Instituto Universitario ITACA, Tecnologías para la Salud y el Bienestar,
Universitat Politècnica de València, Camino de Vera, s/n
46022 Valencia, Spain
jmorgado@itaca.upv.es
<http://www.tsb.upv.es>

² Instituto Universitario ITACA, Tecnologías para la Salud y el Bienestar,
Universitat Politècnica de València, Camino de Vera, s/n
46022 Valencia, Spain
mjarnal@itaca.upv.es

³ Instituto Universitario ITACA, Tecnologías para la Salud y el Bienestar,
Universitat Politècnica de València, Camino de Vera, s/n
46022 Valencia, Spain
manodal@itaca.upv.es

⁴ Instituto Universitario ITACA, R&D Management
Universitat Politècnica de València, Camino de Vera, s/n
46022 Valencia, Spain
jodaob@itaca.upv.es

Abstract. To guarantee the success of one initiative within the new EU Framework Programme for Research and Innovation 2014-2020 (HORIZON 2020), it is necessary to implement all kind of best practices since the first steps of the process (defining the Proposal) to the end, when the Project is declared totally closed at the European Commission. According to this, the aim of this paper is to describe these best practices that allow a winner Proposal to become into a competitive and profitable Project for the consortium, trying to avoid possible hazards along this procedure.

1 Introduction

During a random audit done by the European Court of Auditors [1] in order to review how the European Commission perform the appropriate implementation of the Financial Rules defined in the EU Framework Programme for Research and Innovation, (a priori) non relevant issue can be detected (for example, the criteria used by one entity to allocate the equipment depreciation in the projects). Some months later, auditors from the European Commission can visit this entity with one objective: to confirm (in house) the remarks pointed out by the European Court of Auditors in its financial audit report. According to this, these auditors can propose some financial adjustments. In addition, several recommendations can be included in this last report (rec-

ommendations that will be audited again in the future in order to confirm their incorporation to the procedures of the entity).

Far away from being a hypothetical example, the situation described above started in 2008. After the preliminary financial report, the amount to be recovered to the European Commission was up to 450.000 euro. Five years later, the expedient was closed.

To avoid this undesired situation, it is necessary to guarantee the appropriate execution of the Project both technical and financial level. To reach this objective, a wide knowledge of Horizon 2020 have to be the first and most important step in the definition of one specific Proposal.

2 Horizon 2020

Horizon 2020 [2] is the eighth European Union Framework for Research and Innovation with nearly 80 billion euro of funding available since 2014 to 2020. The main goal of this new competitive framework is to promote the synergy between the public and private sectors in order to respond to the actual economic crisis to invest in future jobs and growth looking for the strengthening of the global position of the European Union in Research, Innovation and Technology.

From the previous EU Framework Programme (2008-2013), there are some aspects to consider:

- Horizon 2020, as a single programme, includes several separate programmes and/or initiatives (7th Research Framework Programme – FP7, Competitiveness and Innovation Framework Programme – CIP, Ambient Assisted Living Programme – AAL, European Institute of Innovation and Technologies – EIT, ..),
- Horizon 2020 strongly focuses in social challenges (health, clean energy and transport, among others),
- Simplification of participation rules for all the entities,
- Strong participation by SMEs as principal vehicle to get the main goal of Horizon 2020.

2.1 Priorities

Excellent Science. This priority aim to reinforce and extend the excellence of the science base at the European Union, consolidating the Research and Innovation system on a global scale.

In this priority there are four different funding instruments (18.9 million euro):

- European Research Council (ERC) looking for frontier Research,
- Future and Emerging Technologies (FET) trying to open new fields of innovation by collaborative Research,

- Marie Skłodowska-Curie actions (MSCA) as opportunities for training and career development,
- Research infrastructures (including e-infrastructure) ensuring access to world-class facilities.

Industrial Leadership. The development of the technologies and innovations that will underpin tomorrow's business and help innovate European SMEs to grow into world-leading companies is the goal of this Priority

Industrial Leadership include three different instruments with more than 17.0 million euro budget:

- Leadership in enabling and industrial technologies (LEITs) including the most relevant technologies all over Europe (Information and Communication Technologies, nanotechnologies, materials, biotechnology, manufacturing, and space),
- Innovation in SME, fostering all forms of innovation in all these kind of entities,
- Access to risk finance, leveraging private finance and venture capital for Research and innovation.

Societal challenges. This priority includes 29.7 million euro for funding the main policy priorities all over European Union:

- Health, demographic change and wellbeing,
- Food security, sustainable agriculture, marine and maritime research & the Bioeconomy,
- Secure, clean and efficient energy,
- Smart, green and integrated transport,
- Climate action, resource efficiency and raw materials,
- Inclusive and reflective societies,
- Secure societies.

Other partnering instruments. 4.6 million funding euro for the next items:

- Public-Private Partnerships,
- Public-Public Partnerships,
- European Innovation Partnerships.

2.2 Rules for Participation

Horizon 2020 has as a main identity the simplification in the rules for participation for all the beneficiaries thanks to a single set of rules adapted for the whole research and innovation cycle, covering all research programmes and funding bodies. According to these rules for participation, the evaluation criteria is also simple reaching three items (excellence, impact and implementation).

Regarding to financial aspects, the funding scheme is also reduced to maximum of 100% of the total budget. There is only one limitation for innovation projects, limited

to 70% of the total eligible costs (except non-profit entities, where there is no limitation). By other hand, the indirect eligible costs define a flat rate of 25% of direct eligible costs.

In addition, it is important to point out the next three characteristic aspects:

- Horizon 2020 introduce new forms of funding aimed at innovation, such as pre-commercial procurement, inducement prizes, dedicated loan and equity instruments,
- Horizon 2020 also improve the rules applied on intellectual property in order to keep the balance between legal security and flexibility, allowing open access publications, and
- The European Commission considers fewer and better targeted financial controls and audits following an audit strategy focused on risk and fraud prevention.

3. Implementation of best practices

After introducing Horizon 2020, this is de moment to work on the proposal, the future project to be chosen by the European Commission to grant a financial contribution for the final implementation of this project.

3.1 Drawing the Proposal

According to the European Commission Decision C(2013)8631, 10th December 2013, Horizon 2020 officially started with the publication of the Work Programme 2014-2015 [3]. In this Work Programme, the European Commission define the expected objectives (calls) to be covered by the different Proposals submitted by competitive way.

To propose a Project within one of the calls included in this Work Programme, it is necessary to fill at maximum the expectatives of that call. In addition, taking into account the evaluation criteria in order to reach a successful Proposal.

For this reason, a high technological and innovate Project can be rejected by the European Comission because of the low quality of the impact and/or excellence.

Many times, the best way to receive feedback about the fortress of a Proposal is to meet the National Contact Point [4] in order to share the vision of this Proposal. In addition, it can be recommended to go directly to the European Commission by asking any already know Project Officer.

3.2 Making of the Proposal

Once a Proposal is drawn as a competitive and winner initiative, it is time to write it according to the Guide for Participants [5] and the templates available in the Participant Portal [6].

In this phase, it is very important to reach an efficient and productive consortium, including partners from different countries with clear tasks in the Project. In addition,

the writing of the Proposal has to be clear and detailed, from the State of the Art where this Project has to be framed to the solution proposed by the Consortium in order to solve the problem described within the Project.

During the writing of the Proposal, several items have to be taken into account:

- The objectives of the Project have to be carefully defined, considering that these objectives should be reachable at the end of the Project.
- The solution proposed by the Consortium has to be clearly innovated and technological realizable,
- The impact of the results development within the Project has to be easily measurable at the end of the Project,
- The budget for the whole Project and the use of resources necessary for the final implementation of the Project have to be appropriated dimensioned according to the tasks to be carried out,
- The management, both at Consortium level and Project level, has to be clearly defined, detailing all the possible situations during the life of the Project.

3.3 Submitting the Proposal

The next necessary step is the submission of the Proposal by using the telematics tools included in the Participant Portal [6].

In this moment, the activities of lobbying are very important in order to present the Proposal in different forums and also to get the support of different relevant people involved in the decision process.

All the partners from different countries usually meet their National Contact Point trying to release the Proposal and get the compromise for defending it in the European Commission.

Sometimes, the most relevant key persons from the consortium lead this initiative contacting other relevant and well positioned people directly placed in Brussels, where the European Commission has its headquarters.

3.4 Formalizing the Project

When the European Commission decided to finance a Proposal, the consortium is invited to negotiate the final conditions of the Project with the Project Officer. The next step is to formalize the relationship between the consortium and the European Commission.

The coordinator of the consortium and the European Commission have to sign the Grant Agreement (GA). By other hand, all the partners of the consortium have to sign the Consortium Agreement (CA).

In this CA, it is very important to define the responsibilities of each partner and/or coordinator within the consortium. In addition, the management bodies have to be clearly defined in order to avoid ambiguity in the decision making process in the Project, including the payments and/or recovery of the funds received from the European Commission.

By other hand, the protection of the background and sideground in the Project have to be included in the CA in order to protect the interests of every partner. This is a very complicated issue just because the nature of all the entities part of the consortium (companies, universities, non-profit associations, Public Bodies...).

3.5 Implementing the Project

From the technical point of view, the implementation of the Project is specified in Annex I of the GA (also named Description of Work), defined by the consortium according to the recommendations received from the Project Officer. In this sense, the rhythm is marked by the consortium itself.

From the financial point of view, unlike the technical implementation of the Project, the rules are defined by the European Commission. In Horizon 2020, the financial rules are included in the Annotated Model Grant Agreement [7]

Some relevant aspects to take into account are listed below:

- Due to the burden on proof for eligibility on the beneficiaries, it is important to keep sufficient supporting documents to show the declared costs are eligible,
- The eligible costs have to meet the following items:
 - a) they must be actually incurred by the beneficiary;
 - b) they must be incurred in the period;
 - c) they must be indicated in the estimated budget set out in Annex 2 of GA;
 - d) they must be incurred in connection with the action as described in Annex 1 and necessary for its implementation;
 - e) they must be identifiable and verifiable, in particular recorded in the beneficiary's accounts in concordance with the accounting standards applicable in the country where the beneficiary is established and with the beneficiary's usual cost accounting practices;
 - f) they must comply with the applicable national law on taxes, labour and social security; and
 - g) they must be reasonable, justified and must comply with the principle of sound financial management, in particular regarding economy and efficiency,
- For personnel costs, it is mandatory to record the effort from every person involved in the Project by using timesheets in order to have the evidence of the productive hours declared.
- For the calculation of the personnel ratio, it is necessary to have a clear idea of the payments done by the beneficiary to their workers and the nature of those payments. In addition, this ratio should be calculated using the data from the last financial natural year calculated in the entity (in FP7, it was possible to calculate the personnel ratio according to the duration of the period, many times different from one natural year),
- For travel costs, it is very important to get supporting documents in order to justify the necessity of these travels and the people involved. Not all the travels can be included in Horizon 2020,

- For equipment costs, the allocation of every equipment has to be carefully calculated considering the life of this equipment (depreciation) and the percentage of use within the Project. It is not possible to declare (and claim) the entire cost of one equipment,
- For subcontracting cost, the beneficiaries have to collect three different offers from three different services providers in order to justify the objective election of one of them. In addition, any subcontracting has to be specifically included in Annex 1 of GA,
- For other costs, it is highly advisable to support any cost with documents that justify the necessity of that cost.

3.6 Closing the Project

At the end of the Project, it is essential to remember the requirements from the European Commission, including the commitment of keep alive during almost 3 years after the last payment of the European Commission to the coordinator:

During this time defined by Horizon 2020, it is possible to be chosen to receive external auditors from any European Body in order to review in house the appropriate implementation of the Project at financial level (not forgetting the relevance of the technical aspects of this same Project).

4 Conclusions

One common mistake during this process is to write a high quality Proposal with overstated budget/funding. As a consequence, if the Proposal is accepted by the European Commission, the beneficiary receive all the funding but is not able to justify it according to the actual rules.

For this reason, it is important to be coherent during all the process. One minor maladjustment at the beginning can be suffered along the Project. For example, a low increase (5%) of the real cost of one person/month during the Proposal means a higher declaration of effort (15%). Technically, it can be perfectly reasonable; but financially, it supposes an inappropriate distribution of the resources. In fact, this could be a reason enough to be audited by the European Commission.

In conclusion, as soon as one entity incorporates the guidelines and recommendations from the European Commission according to Horizon 2020, the success rate will increase exponentially, minimizing futures threats when the Project is already closed.

References

1. <http://www.eca.europa.eu/en/Pages/ecadefault.aspx>

2. <http://ec.europa.eu/programmes/horizon2020/en>
3. http://ec.europa.eu/research/participants/data/ref/h2020/wp/2014_2015/main/h2020-wp1415-intro_en.pdf
4. http://ec.europa.eu/research/participants/portal/desktop/en/support/national_contact_points.html#c.contact=country/sbg/Spain/1/1/0&function_details.function_abbr/sbg//1/1/0&contact_name/asc
5. http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/pse/h2020-guide-pse_en.pdf
6. http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/pse/h2020-guide-pse_en.pdf
7. http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/amga/h2020-amga_en.pdf

Proposal of a semiautomatic conversion algorithm from archetypes to relational databases

Antonio Félix¹, Monsterrat Robles², Diego Boscá²

¹ Servicio de Emergencias Sanitarias, Conselleria de Sanitat de la Comunitat Valenciana,
Spain

Tonyete@gmail.com

² Instituto de Aplicaciones de las Tecnologías de la Información y de las Comunicaciones
Avanzadas (ITACA), Universitat Politècnica de València, Camino de Vera s/n,
46022 Valencia, España
{Mrobles, Diebosto}@upv.es

Abstract. The purpose of this study is to investigate the feasibility of applying a conversion algorithm between archetypes and relational databases. The paper applies an algorithm to an existing archetype of extrahospitalary emergencies. The result is evaluated with a relational database from the same organization for the same data. This algorithm shows some problems that can be easily fixed by an operator who performs an analysis of the results using a complete understanding of the data. Therefore, semiautomatic conversion algorithm is the best choice to perform this task.

1 Introduction

Modeling the data through clinical archetypes ensures interoperability across and within Electronic Health Records [1]. Interoperability between systems can be considered from different approaches. Technical interoperability takes into account the physical and logical interfaces that allow the exchange of information. The syntactic interoperability alludes to the document exchange, not giving importance to the content in the documents. The organizational interoperability is established as the cooperation between systems, thanks to sharing a common context and workflow. This article has its topic on semantic interoperability. Semantic interoperability is the state between two entities/applications when, attending to a specific task, one entity accepts information from the other and this task can be done successfully, without needing to intervene an external operator [2].

The ISO 13606 EHR interoperability standard was created by the European Committee for Standardization to define the common characteristics of EHR and incorporate them to information models and messages between systems [2]. ISO 13606 establishes a dual model of the information: the Reference Model and the Archetype Model. The Reference Model has the global characteristics from EHR, how they gather and which is the context of each one. It reflexes the steady characteristics from EHR. The Archetype Model is the formal definition of a specific combination of basic components

from the Reference Model applied to a specificity within a field of knowledge or clinical organization.

Hierarchical components of the Reference Model are the EXTRACT of EHR, the highest level container of part or wide EHR from a single patient; the FOLDER, the highest level component of organization of EHR, that divides it in compartments relative to provided assistance for just one condition; the COMPOSITION, the set of information introduced in EHR by a health worker, as a result of a clinical meeting; the SECTION, data of the EHR inside a COMPOSITION that belong to a clinical header, and reflex the information flow in a structured way which conveys the meaning; the ENTRY, the information introduced in EHR as a result of a single clinical action, observation, clinical interpretation or clinical purpose; the CLUSTER, the way to represent information which is equivalent to columns or tables; the ELEMENT, the last hierarchical node, which has a single data value [3].

There are different possibilities for the definition of a Database Management System (DBMS). In 1970, Edgar Frank Codd defined the Relational Model for database management [4]. This model establishes that all data are group in relations. Each relation has attributes, which are an ordered pair of attribute name and type name. A tuple is an ordered set of attributes values. Each relation can be explained as a table, attribute as a column inside a table and a tuple as a row inside a table. Each table can have one or more attributes which value identify unambiguously a tuple, which is called primary key. Also, a relation can have one or more attributes that allow the combination of data from different relations, which is called foreign key. It can also define constraints over data. All this structure can be viewed in the database scheme.

Conversion between archetypes and relational databases can be useful. A clinical organization can be required to perform a DBMS to respond the need of data storage. In this case, health professionals can build an archetype of their field of knowledge, and IT staff will implement a scheme of database. This is especially important when dealing with big archetypes. A conversion algorithm will save time of this task to them.

The purpose of this paper is to evaluate a proposal of a semiautomatic conversion algorithm from archetypes to relational databases.

2 Methods

An archetype built for extrahospitalary emergencies will be used as a basis for the application. The archetype for extrahospitalary emergencies can be seen in figure 1. The result of this process will be compared with a scheme from a relational database implemented for that clinical situation. The relational schema for the implemented database can be seen in figure 2. Both documents were generated from the same system, but in different ways. Thus, they respond to same problem but were built in separate procedures.

The following items are evaluated in the validation process of conversion from archetype to relational database:

- Number of tables.

- Number of attributes.
- Integrity of data, reflected in primary keys and foreign keys.

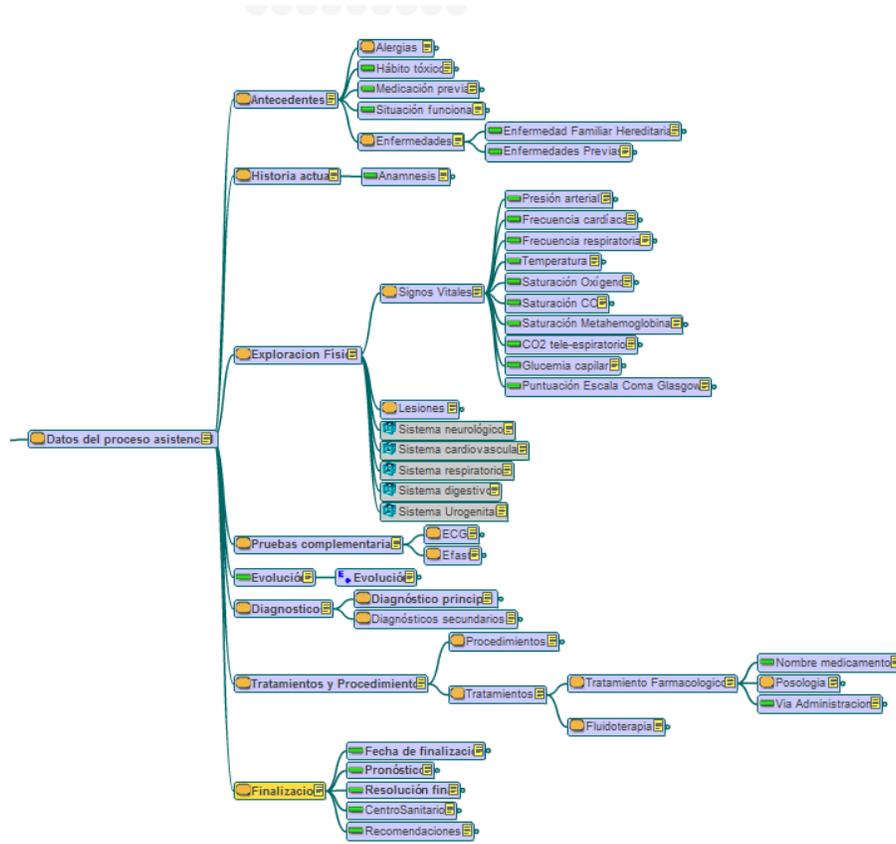


Fig 1. Extrahospitalary Emergencies Archetype

The algorithm is described as a sequence. We traverse the archetype and find the different SECTIONS. Each SECTION is processed in the same way:

1. If SECTION has other SECTIONS, each one has to be analyzed until a SECTION contains just lower elements of the Reference Model (i.e. ENTRY, CLUSTER, or ELEMENT).
2. Analyze maximum occurrences of the ENTRIES. We apply step 4 to ENTRIES with maximum occurrence of 1. For every other value of maximum occurrences we apply step 3.
3. Group selected ENTRIES by maximum occurrence, creating “i” groups. Then we analyze if there is a relationship between the maximum occurrences of the entries (i.e. if they are supposed to exist all at the same time). If there

is a relation, all ENTRIES are treated as a group in step 4. If there isn't a relation, each ENTRY from the “i” group is treated on their own in step 4.

4. Create a table for each individual ENTRY from step 2 and for each individual ENTRY and for each “i” group from step 3. We assign an automatically generated table name following the rule:
SECTION_(SECTION)_ENTRY

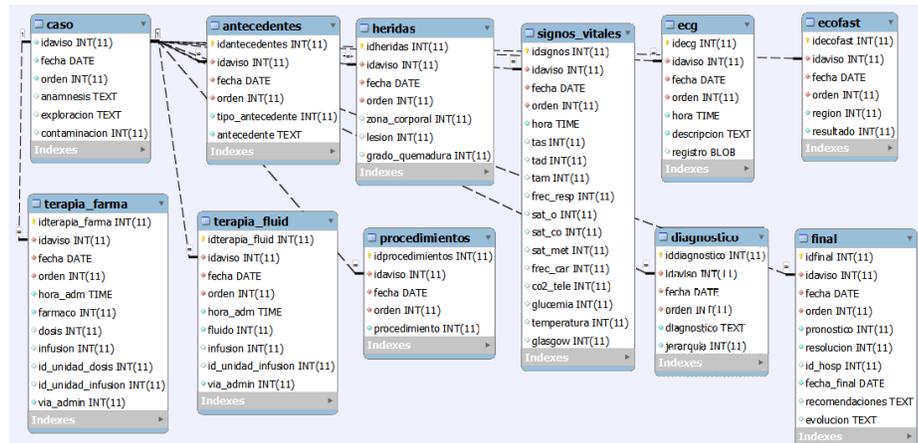


Fig. 2. Extrahospitalary Emergencies Relational Database

The assignment of attributes will be according to these rules:

- Each ELEMENT from the ENTRY generates an attribute.
- The primary key for the table will be an ID attribute for the table generated for this purpose.
- Each table will have a foreign key pointing to the root node.

3 Results

The original database scheme, as shown in figure 2, shows 11 tables with 90 attributes: 14 primary keys, 33 foreign keys and 43 clinical attributes. Figure 3 shows the result of applying the algorithm. The algorithm creates 19 tables with 123 attributes: 21 primary keys, 54 foreign keys and 48 clinical attributes.

First discussion is the fact that the algorithm increases the number of tables, and consequently, primary and foreign keys.

One of the transformed SECTION classes that contributes to the increase of tables is “Antecedentes” SECTION. Figure 2 shows a single table conjugating the five types of “Antecedentes”. Proposed algorithm generates five different tables. First modification to the algorithm could bring closer both solutions: inside a SECTION, all

ENTRIES and SECTIONS generate a single table, keeping the occurrence maximum rule. In this case, we can obtain a table with a primary key with foreign keys pointing “Caso”, and one attribute for each ELEMENT inside the ENTRIES of the SECTIONS in “Antecedentes” SECTION. But this solution will not be complete, because each tuple will have one occurrence for each ELEMENT, although it will be NULL value. The solution of figure 2 shows a single table with two clinical attributes: one attribute with different types of “Antecedentes” as a list of coded texts, and an attribute describing the record. This is the exactly the same kind of problem with the transformation of “Diagnosticos” SECTION, to which we have applied the same solution.

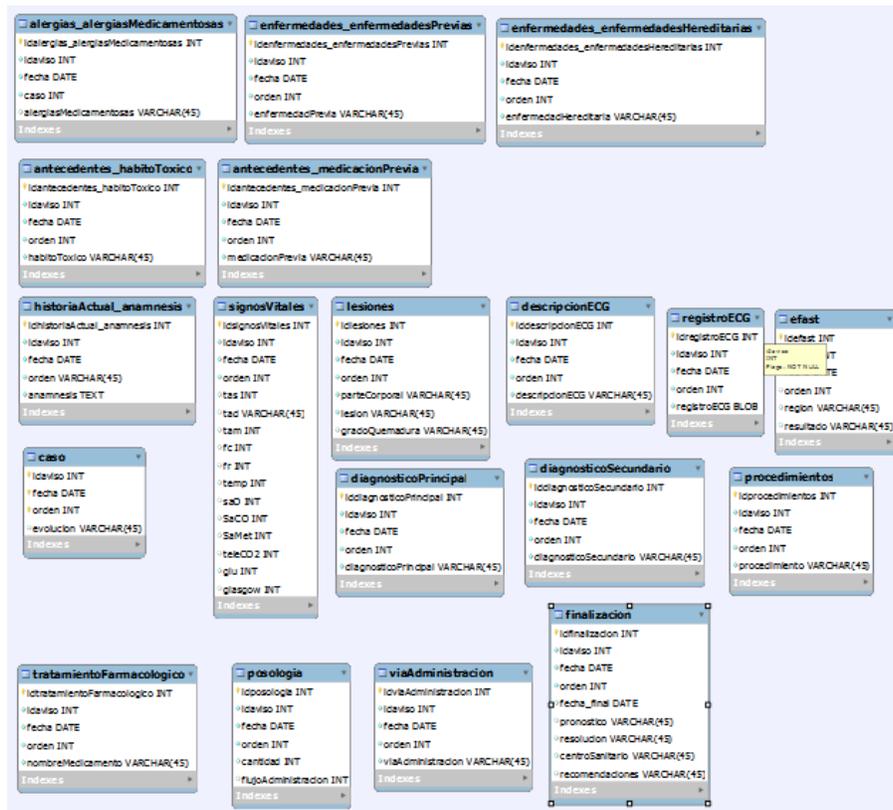


Fig. 3. Relational Database Generated by Proposed Algorithm

A different problem appears with “Tratamiento Farmacológico” SECTION. Figure 4 shows the result from the algorithm at this SECTION. The algorithm generated three tables and they have lost their relationship with the original meaning, which is about a pharmacological treatment. First table in figure 4 has the name of the medicine. The second table has the amount and the flux of administration. The third table has the

route of administration. These last two tables have the foreign keys pointing to the “Caso” SECTION, or clinical case. But this is a wrong meaning: the amount, flux and route of administration refer to medicine, which is the first table in figure 4. A solution can be performed by changing the foreign key and point to medicine. But the inclusion of this change in the algorithm will result in more mistakes with other SECTIONS with similar structure. SECTION “Tratamiento Farmacológico” has an ENTRY and two SECTIONS. SECTIONS inside “Tratamiento Farmacológico” could have foreign keys pointing to the ENTRY. But same structure in SECTION “Antecedentes” would make no sense. SECTION “Antecedentes” is about previous medical records. It has two SECTIONS, one for medical allergies, and another for previous diseases. Also has three ENTRIES. There are two reasons that show why is wrong to take this approach. First reason is that we could not know which one of the three ENTRIES will be the chosen to be pointed by the foreign keys. The second reason is that it would make no sense that a foreign key from allergies points to previous functional situation of the patient.



Fig. 4. Algorithm result over "Tratamiento Farmacológico" Section

4 Conclusion

An automatic conversion algorithm from archetypes to relational databases is extremely difficult to implement. The hierarchical components of the Reference Model from ISO 13606 can be combined in similar ways, as “Antecedentes” and “Tratamiento Farmacológico”, but the relationship between these entities can be very different. Therefore, an algorithm applied to these two similar SECTIONS will have different kinds of mistakes in some cases.

For the moment, the semiautomatic process will be a great solution. The automatic part will save time to IT team. But it is necessary a complete understanding of both the clinical information structure and archetype structure to adjust the results of the algorithm.

In the future, we will explore further this research, trying to develop a complete automatic conversion algorithm from archetypes to relational databases.

References

1. Tapuria A, Kalra D, Kobayashi S, “Contribution of Clinical Archetypes, and the Challenges, towards Achieving Semantic Interoperability for EHRs”, *Healthcare Informatics Research*, vol. 19, no.4, p. 286-92, 2013.
2. Muñoz Carrero, A.; Romero Gutiérrez, A.; Marco Cuenca, G.; Abad Acebedo, A.; Cáceres Tello, J.; Sánchez de Madariaga, R.; Serrano Balazote, P.; Moner Cano, D.; Maldonado Segura, J.A., “Manual práctico de interoperabilidad semántica para entornos sanitarios basada en arquetipos”. Madrid: Unidad de investigación en Telemedicina y e-Salud – Instituto de Salud Carlos III. Mensor, julio de 2013
3. EN 13606-1:2007, Health Electronics - Electronic health record communication - Part I: Reference Model. CEN- European Committee for Standardization.
4. Codd, E. F., “A relational model of data for large shared data banks”- *Communications of the ACM*, vol. 13, no.6, p.377-387 1970.

Fast parameter estimation based on B-Splines to model enzymatic activities: a pesticide degradation application

Area of interest: “Biomedical Informatics”

Pedro Mayorga^{1,*}, Fernando Ramos, Edgar Dantán-González,
Laura Ortiz-Hernández, Claudia Díaz-Camino, Alfonso Estudillo¹

¹Universidad Politécnica de Chiapas, Ingeniería Biomédica, Calle Eduardo J. Selvas
S/N, Col Magisterial, Tuxtla Gutiérrez, Chiapas, C.P. 29082, México

*mayorga@ib.upchiapas.edu.mx

Abstract. Molecular modeling and simulation methods provide important skills to gain an insight into the function of biological process. In particular, such methods have been applied to the characterization of enzyme activities. In this report, we modeled enzymatic activities by using a B-Spline approach and tested it against experimental data.

We have established a compromise that satisfies both the enzymatic reactions and the observed data by formulating a linear least square problem whose optimal values represent the control points of the B-Splines that best fit the enzymatic model.

The proposed model yielded better results, with respect to the adjustment to the observed data, than the Runge-Kutta method of fourth order. We also found that the Fast B-Spline Method proposed in this work provides: (i) better local control at the observed points; (ii) soft adjustment to the observations; and (iii) the global optimum is the solution of a linear system whose matrix is banded, symmetric and positive-definite; therefore, fast and efficient algorithms exist for its solution in $\mathcal{O}(n)$ time.

Keywords: B-Splines, pesticide degradation application, enzymatic activities

1 Introduction

One relevant aspect of performing modeling and simulations in molecular biology is that such computational approaches can provide an understanding of the fundamental mechanisms of enzyme catalysis, such as the analysis of catalytic interactions [13].

Molecular modeling and simulation methods are increasingly making important contributions to the study of the structure and function of biological macromolecules, wherein one of the most important applications is the modeling and analysis of enzyme mechanisms, highlighting that the biological catalysis is an important interface between chemistry and biology [18].

Kinetics and the regulation of biomolecular binding reactions are the basic blocks of all complex systems, such as macromolecules [10].

In this work, we were interested in modeling the interaction of enzymes with certain pesticides. On the one side, some investigations have been carried out to confirm that pesticides may affect the activities of enzymes in soils [6, 7]. Pesticides may enter the soil through agricultural practices. In addition, pesticides can also enter the soil through accidental spillage, leaks at pesticide dump sites, industrial wastes or urban pollution. In soil, enzymes contribute to the total biological activity of the soil-plant environment, wherein their catalytic efficiency may be influenced by the composition of the surrounding environment. In [23] the influence of four pesticides, glyphosate, paraquat, atrazine and carbaryl, characterized by different physicochemical properties, on the activities of invertase, urease and phosphatase isolated from a number of soils was investigated. This study confirmed that pesticides affect the behavior of these enzymes.

Conversely, certain enzymes have been reported to reduce considerably the toxicity of pesticides. For instance, several enzymes called phosphotriesterase (PTE) hydrolyze particular organophosphate (OP) compounds that have been widely used as synthetic pesticides [28, 20]; these compounds irreversibly inhibit acetylcholinesterase (AChE), a key enzyme in the central nervous systems of insects and non-target organisms, including humans. Inhibition of AChE causes the accumulation of acetylcholine in the synaptic junction and disrupts the normal transmission of nerve impulses. Excessive levels of acetylcholine ultimately cause neuromuscular paralysis leading to death by asphyxiation.

Several PTEs have been isolated from different microorganisms and it has been found that these enzymes hydrolyze and detoxify OPs, which reduces their toxicity by decreasing their ability to inactivate AChE. Important studies have revealed how the hydrolysis effectiveness varies with the kind of substrate with which the enzyme interacts. For example, the degradation of methyl-parathion (MP) in the metabolic pathway of some bacteria starts with PTE enzymatic activity, which hydrolyzes MP into dimethyl-tiophosphoric acid and p-nitrophenol (PNP). We have already reported [6] the characterization of one *Enterobacter* sp. bacteria strain capable of hydrolyzing MP, and other pesticides, through the *opdE* gene that codes for a hydrolase. As a result of this hydrolysis a quantity of PNP appears, which can be used to supply data to a dynamic model.

Modeling enzymatic reactions is usually carried-out by dynamic models represented by a system of differential-algebraic equations [11, 8, 14, 16, 4, 1, 15, 17, 2, 3, 25] and is referred as Parameter Estimation Problem. The classical parameter estimation problem is described as follows:

$$\frac{dy}{dt} = f(\mathbf{x}, \mathbf{y}, \mathbf{p}, \mathbf{v}, t), \quad (1)$$

subject to

$$\begin{aligned} \mathbf{x}(t_0) &= \mathbf{x}_0, \\ \mathbf{h}(\mathbf{x}, \mathbf{y}, \mathbf{p}) &= \mathbf{0}, \\ \mathbf{g}(\mathbf{x}, \mathbf{y}, \mathbf{p}) &\leq \mathbf{0}, \\ \mathbf{p}^L &\leq \mathbf{p} \leq \mathbf{p}^U, \end{aligned} \quad (2)$$

where \mathbf{p} is the vector of control parameters, \mathbf{x} is the vector of control variables (concentrations or states), \mathbf{v} is the vector of constant parameters, \mathbf{f} , \mathbf{g} are possible equality and inequality constraints of the system, finally, \mathbf{p}^L and \mathbf{p}^U are (optional) lower and upper bound of the control parameter \mathbf{p} .

The optimization problem described by (1) and (2) generally has many local optima and the gradient based methods do not converge to the global optimum. On the one hand, some meta-heuristic optimization algorithms such as Simulated Annealing, Genetic Algorithm, Evolutionary Strategy, Genetic Programming, Differential Evolution and Multiobjective Approaches have been used to solve particular problems with success [2, 3, 25]. Although meta-heuristics approaches solve the optimization problem, they are not efficient. On the other hand, the B-Splines are piecewise polynomial functions and are widely used for data interpolation, even with noise. Some applications of B-Splines recently developed are to model growth of bacteria strains [24], clustering time-series gene expression [26], the exploration of phenotype microarray kinetics [26], dengue forecast [12] and to estimate the mutual information with application to gene expression data [7].

Zhan and Yeung [27] use a spline approximation to solve (15) by transforming the parameter estimation problem in a Linear Programming Problem (linear optimization with linear constraints) which allows the use of fast and efficient algorithms to solve the optimization problem. The algorithms used for the optimization of a Linear Programming Problem have polynomial complexity [19], then, we can improve this complexity time through an optimization problem of order $\mathcal{O}(n)$, as we will see later, the B-Spline Method of degree 2 proposed in this work has this property.

This paper proposes a Fast B-Spline Method (FBSM) for modeling enzymatic activities associated with the degradation of pesticides using real experimental data. We propose a Linear Least Squares Problem which balances the adjustment of the enzymatic reaction described by a SODE (see equation (4)) and the adjustment to the observations. Moreover, the linear system related with the solution is banded, therefore, the solution can be computed with fast and efficient algorithms. An additional contribution of this work is that we test our approach with real wet-lab experiments related with pesticides.

The objective of this work is twofold: 1) to improve the time complexity of the parameter estimation problem by formulating a FBSM based on an algorithm of order $\mathcal{O}(n)$; 2) To apply the proposed method for modeling the enzymatic activities involved in the pesticides degradation using real experimental data.

The remaining sections of this paper are organized as follows: In the results section, we describe the enzymatic reaction, the problem to be solved, the FBSM for parameter estimation and its formulation. At the end of the section the results provided by this model is compared with the classical Runge Kutta Method which serves as reference; in the conclusion section, we summarize the main properties of the proposed method by outlining its advantages and limitations. Finally, future work related with this investigation is described.

2 Materials and Methods

One of the main focuses of this work is to test FBSM with real experimental data. Consequently, we briefly describe below the materials and methods used to obtain the experimental results.

The Cons002 strain was cultured overnight in 3 ml of minimal medium (MM) with a composition (g/l): 2.92 KH_2PO_4 , 2.74 K_2HPO_4 , 0.20 $MgSO_4 \cdot 7H_2O$, 2.00 KNO_3 , and 0.99 $(NH_4)_2SO_4$ containing 50 mg/l glucose and 20 mg/l MP with an initial OD600 of 0.05 at 28°C and agitated at 100 rpm. Samples (1 ml) were collected at 0, 4, 8, 12, 16, 20, 24 and 48 h. After being harvested, the cells were suspended in 2 ml of distilled water and sonicated at 25% efficiency for 75 s (3 pulses of 25 s each) in a sonicator (Branson Sonifier 450). After sonication, the solution was centrifuged (17,000g for 10 min), and the intracellular fractions were tested for hydrolase activity. Hydrolysis of MP was measured spectrophotometrically by monitoring the production of PNP at 405 nm ($\epsilon_{\max} = 17000 \text{ M}^{-1} \text{ cm}^{-1}$) for 6 h at 28°C; assays were performed in 20 mM Tris-*HCl* (pH 9). Enzyme activity was defined as the amount of enzyme required to release 1 mol of PNP per min at 37°C. The data are reported as the specific activity (mU/mg of protein).

3 Enzymatic reaction

A general enzymatic reaction takes the form of



where E , S and P are the enzyme, substrate and the product respectively, and k is the proportionality (rate) constant (coefficient) that may be determined either experimentally or estimated theoretically [14].

The rate equation of (3) is given by a system of differential equations cited in [11] as:

$$\begin{aligned} \frac{d}{dt}[E] &= v_E = -k[E][S], \\ \frac{d}{dt}[S] &= v_S = -k[E][S], \\ \frac{d}{dt}[P] &= v_P = k[E][S], \end{aligned} \quad (4)$$

where $[E]$, $[S]$ and $[P]$ represent the function concentration of the enzyme E , substrate S and the product P with respect to time, respectively; v_E , v_S and v_P are the velocity rate functions for E , S and P , respectively.

Our purpose is to find functions $E(t)$, $S(t)$ and $P(t)$ that must satisfy two conditions, the interpolation and the dynamic model. The first condition is satisfied when the functions pass through a set of observations, in our case, the observations were obtained from laboratory experiments. The second condition satisfies the system of differential equations (4) representing the kinetic model of an enzymatic activity (3).

4 The B-Spline interpolation method

The B-Spline interpolation, is widely used to build a continuous function that passes through a set of given points or observations [9, 29, 5, 21]. A relevant property of the B-spline interpolation is the local control, which means that in case of modifications around the i th observation the function is just affected locally around such observation. This property is very important for our purposes in this paper because, although we are finding a smooth and continuous function, it should be adapted to disturbances caused either by the nature of the experiment or by noise, without changing its global behavior.

Definition The B-Spline of degree j is defined in terms of m points called knots and $m - j - 1$ basis functions. Consider the m knots as $u_1 \leq u_2 \leq \dots \leq u_m$, and the basis B-splines $B_{i,j}$ of degree j defined using the Cox de Boor recursion formula

$$B_{i,0}(t) = \begin{cases} 1 & \text{if } u_i \leq t < u_{i+1} \\ 0 & \text{otherwise} \end{cases} \quad i = 1, \dots, m-1; \quad (5)$$

and

$$B_{i,j}(t) = \frac{t - u_i}{u_{i+j} - u_i} B_{i,j-1}(t) + \frac{u_{i+j+1} - t}{u_{i+j+1} - u_{i+1}} B_{i+1,j-1}(t) \quad i = 1, \dots, m-j-1; \quad (6)$$

then, the B-Spline function $y(t)$ is defined as

$$y(t) = \sum_{i=1}^{m-j-1} B_{i,j}(t) p_i, \quad (7)$$

where the points $p_i \in \mathbb{R}^d$ are called control points or de Boor points. By definition, if $u_{i+1} - u_i = u_i - u_{i-1}$ for $i = 2, \dots, m-1$, the B-spline is uniform, otherwise the B-spline is non-uniform.

The $B_{i,j}$, $i = 1, \dots, m-j-1$ are the basis functions and are also called blending functions, because of their use in interpolation.

The B-Spline function of degree j is $j-1$ times continuously differentiable and the derivative of a B-Spline function is another B-Spline function. For example, the derivative of the $B_{i,j}(t)$ is defined in equation (8) and the derivative of the B-Spline in equation (9), where $y'(t)$ has one degree less and one more control point with respect to $y(t)$.

$$B'_{i,j}(t) = \frac{j}{u_{i+j} - u_i} B_{i,j-1}(t) - \frac{j}{u_{i+j+1} - u_{i+1}} B_{i+1,j-1}(t) \quad (8)$$

$$y'(t) = \sum_{i=1}^{m-j} B_{i,j-1}(t) q_i; \quad q_i = \frac{j}{u_{i+j} - u_i} (p_i - p_{i-1}) \quad (9)$$

Interpolation with B-Splines Interpolation of n observed data (t_k, y_k) , for $k = 1, \dots, n$ using B-Splines is equivalent to solve a system of linear equations for the unknown control points p_i . We can observe that the errors of the observation k is $\epsilon_k = y_k - y(t_k) = y_k - \sum_{i=1}^{m-j-1} B_{ij}(t_k)p_i$, which is a linear function of p ; then, we can minimize the sum of squared errors of the optimization problem (10)

$$\min \xi(p) = \frac{1}{2} \sum_{k=1}^n \left(y_k - \sum_{i=1}^{m-j-1} B_{ij}(t_k)p_i \right)^2 = \frac{1}{2} \|\mathbf{y} - \mathbf{B}\mathbf{p}\|^2, \quad (10)$$

where the optimal solution of the equation (10) is the solution of the following linear system

$$\mathbf{B}^T \mathbf{B}\mathbf{p} = \mathbf{B}^T \mathbf{y} \quad (11)$$

where $\mathbf{p} \in \mathbb{R}^{m \times d}$ is the matrix whose entries are the control points, $\mathbf{B} \in \mathbb{R}^{n \times m}$ is the matrix where an element belonging to row r and column s is defined as $\mathbf{B}_{rs} = B_{s,j}(t_r)$ and $\mathbf{y} \in \mathbb{R}^{n \times d}$ is the matrix of the n observations \hat{y}_k , $k = 1, \dots, n$. Then, the interpolation function evaluated on the knots is $\hat{\mathbf{y}} = \mathbf{B}\mathbf{p}$, see equation (7).

For noisy data, the smoothing B-Spline function can be achieved by minimizing the following function:

$$\min \xi(p) = \frac{1}{2} (1 - \gamma) \sum_{k=1}^n (y_k - \hat{y}(t_k))^2 + \frac{1}{2} \gamma \int_{t_1}^{t_n} (\hat{y}''(v))^2 dv, \quad (12)$$

where $\gamma \geq 0$ controls the smoothness of the solution. The solution is close to the observations when γ tends to zero, and becomes to straight line when γ tends to 1. The roughness penalty has the form $\int_{t_1}^{t_n} (\hat{y}''(v))^2 dv = \mathbf{p}^T \mathbf{\Psi}\mathbf{p}$, where the element of row i and column j of $\mathbf{\Psi}$ is $\int_{t_1}^{t_n} B_i''(v) B_j''(v) dv$. Taking the matrix notation, the equation (12) can be represented as follows:

$$\min \xi(p) = \frac{1}{2} (1 - \gamma) \|\mathbf{y} - \mathbf{B}\mathbf{p}\|^2 + \frac{1}{2} \gamma \mathbf{p}^T \mathbf{\Psi}\mathbf{p}, \quad (13)$$

whose the optimal \mathbf{p}^* is the solution of the following linear system:

$$((1 - \gamma) \mathbf{B}^T \mathbf{B} + \gamma \mathbf{\Psi}) \mathbf{p} = (1 - \gamma) \mathbf{B}^T \mathbf{y}. \quad (14)$$

5 The proposed model: The FBSM for enzymatic activities

As we mentioned previously, we have a set of observed data $\{\tilde{t}_j, \tilde{E}_j, \tilde{S}_j, \tilde{P}_j\}$, $j = 1, \dots, n$ of an enzymatic reaction as described in equation (3) and we need to find the B-Spline functions $E(t)$, $S(t)$ and $P(t)$ such that $E(t)$, $S(t)$ and $P(t)$ satisfy the equations (4) while remaining close to the observed data and control the roughness of the functions due to the noise.

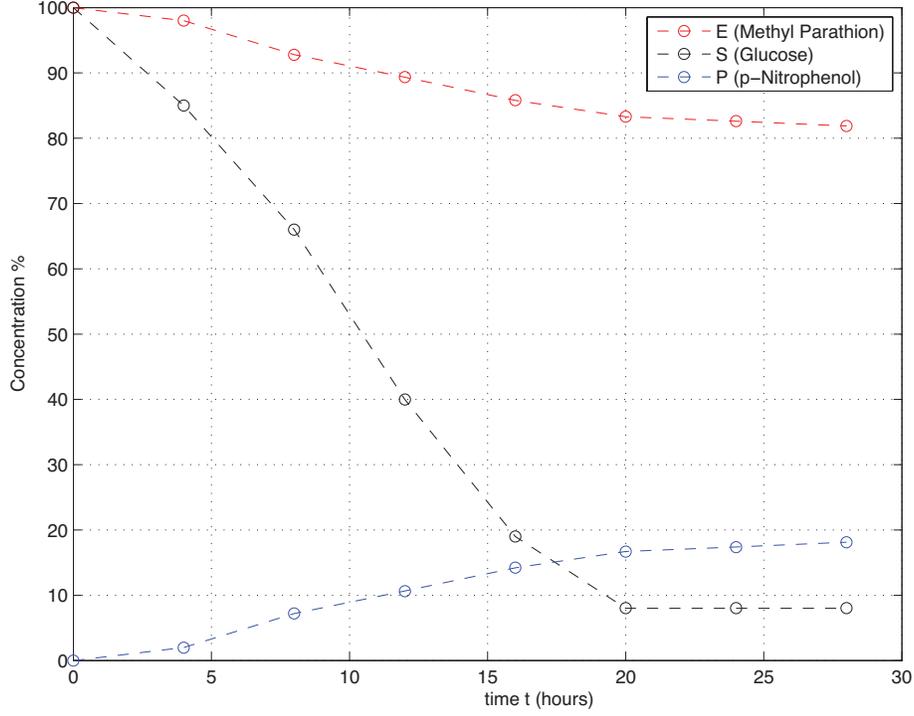


Fig. 1. Observations of the degradation of methyl parathion by a suspension of *Enterobacter* sp. cells obtained from wet lab experiments.

We propose a LSP containing the balance between the dynamic part, the adjustment of the observations and the smoothness of the solutions:

$$\arg \min_{x \in \{E, S, P\}} \xi(\mathbf{p}) = \sum_{x \in \{E, S, P\}} \frac{1}{2} \delta \left(\lambda \|\tilde{v}_x - \dot{x}\|^2 + \mu \|\tilde{x}_i - x\|^2 \right) + \frac{1}{2} \gamma \int_{t_1}^{t_n} (\ddot{x}(v))^2 dv, \quad (15)$$

where \mathbf{p} is the set of control points of the B-Spline functions $E(t)$, $S(t)$ and $P(t)$; $0 \leq \lambda \leq 1$ ($\mu \equiv 1 - \lambda$) keeps the balance between satisfying the SODE and the interpolation criteria; $0 \leq \gamma \leq 1$ ($\delta \equiv 1 - \gamma$) controls the smoothness of the proposed solution $x \in \{E, S, P\}$. The \dot{x} and \ddot{x} are the first and second derivative of x with respect to the time t . The reaction velocity of the compounds \tilde{v}_x is defined in equation (4).

Given that $E(t)$, $S(t)$ and $P(t)$ are B-Spline functions, the solution of (15) is the solution of the following system of linear equations:

$$\left(\delta\left(\lambda\dot{\mathbf{B}}^T\dot{\mathbf{B}}+\mu\mathbf{B}^T\mathbf{B}\right)+\gamma\Psi\right)\mathbf{p}_x=\delta\left(\lambda\dot{\mathbf{B}}^T\tilde{v}_x+\mu\mathbf{B}^T\tilde{x}\right), \quad x\in\{E,S,P\}, \quad (16)$$

where \mathbf{B} is the matrix of Blending functions, $\dot{\mathbf{B}}$ represents its first derivative and Ψ is the smoothness matrix defined in the previous section.

The matrix $\mathbf{A}=\delta\left(\lambda\dot{\mathbf{B}}^T\dot{\mathbf{B}}+\mu\mathbf{B}^T\mathbf{B}\right)+\gamma\Psi$ is banded, symmetric and positive definite, then, we can solve the linear system (16) using a fast and efficient algorithm in $\mathcal{O}(j^2n)$ time [22], where j represents the degree of the Blending functions. If the B-Spline degree is much smaller than the dimension of A , then the solution can be computed in $\mathcal{O}(n)$. In our experiments we use B-Splines of degree 2, to take advantage of this feature.

With this formulation, we do not need a powerful exploration algorithm, as in the case of the genetic algorithm used in the literature, for finding the optimal set of parameters; thus, we gain in accuracy and speed of convergence by the use of efficient algorithms for linear systems.

As we mentioned above, the B-Spline approach described in this work produces B-Spline functions capable of adjusting the dynamic model (4) of the enzymatic reaction (3) while staying close to the observed experimental data.

In this paper, we propose that $E(t)$, $S(t)$ and $P(t)$ are modeled with B-Splines of degree 2, because the resultant curves are smooth and continuous in the first derivative and the parameter estimation problem can be solved with fast and efficient algorithms in $\mathcal{O}(n)$.

6 Analysis and discussion of results

In this section we discuss the simulation results of the biodegradation of methyl-parathion by a suspension of *Enterobacter* sp. strain cells.

As a comparison purpose, we simulate the enzymatic activity with the traditional Runge-Kutta Method of fourth order (which is the most common procedure when we have only the initial values) as in [4, 8, 11].

In order to simulate the enzymatic activity we tested several values of the rate constant k and we have found that the value $k=2.1917\times 10^{-4}$ minimizes the objective function using the traditional Runge-Kutta method of order 4.

In Figure 1, we present the observed experimental data describing the biodegradation of methyl-parathion by *Enterobacter* sp. cells. This data was used to compute \tilde{x} , v_x for the optimization function (15).

We simulated the enzymatic reaction using the RK4. The RK4 simulation is shown in Figure 2.

Comparing the plots in Figure 2 against Figure 3, we note a better performance in favor of the B-Spline model. Moreover, the measurements show that the B-Spline model is about 50 times better than the RK4, because the average error of the RK4 is 42.11 and the average error of all the simulations is 0.88 which represents that in average the B-Spline model is 47.85 times better than the classical RK4.

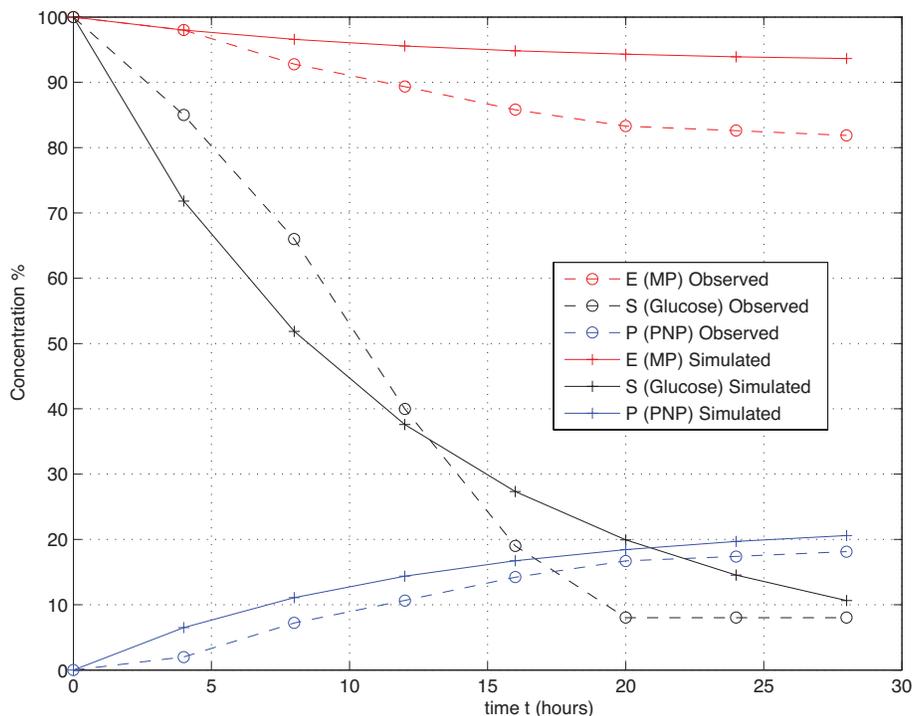


Fig. 2. Simulation result of the degradation of methyl parathion by a suspension of *Enterobacter* sp. cells obtained from wet lab experiments using the RK4 method.

We solve the optimization problem for five thousands values of $\gamma \in (0, 1)$, the mean and the standard deviation are plotted in Figure 4; we see that the solutions belong to a well defined shape around the observations, specially around the PNP and the MP that are the enzymes we need to measure in this study.

In Figure 5, we plotted several simulations trying to compare how the solutions change with respect to the two control variables λ and γ related with the adjustment to the dynamic model and the smoothness of the curve, respectively. We can see that the solutions get out of a close neighborhood of the observations as λ increases, because the counterpart parameter $\mu \equiv 1 - \lambda$, which controls the adjustment to the observations, decreases. If γ grows the solutions get away from a close neighborhood of the observations because the simulation results are smoother and tend to a straight line.

We obtain two important advantages of the FBSM: 1) it is continuous and smooth; 2) the solution can be computed with fast and efficient algorithms due to the fact that the matrix of the linear system is banded, symmetric and definite-positive.

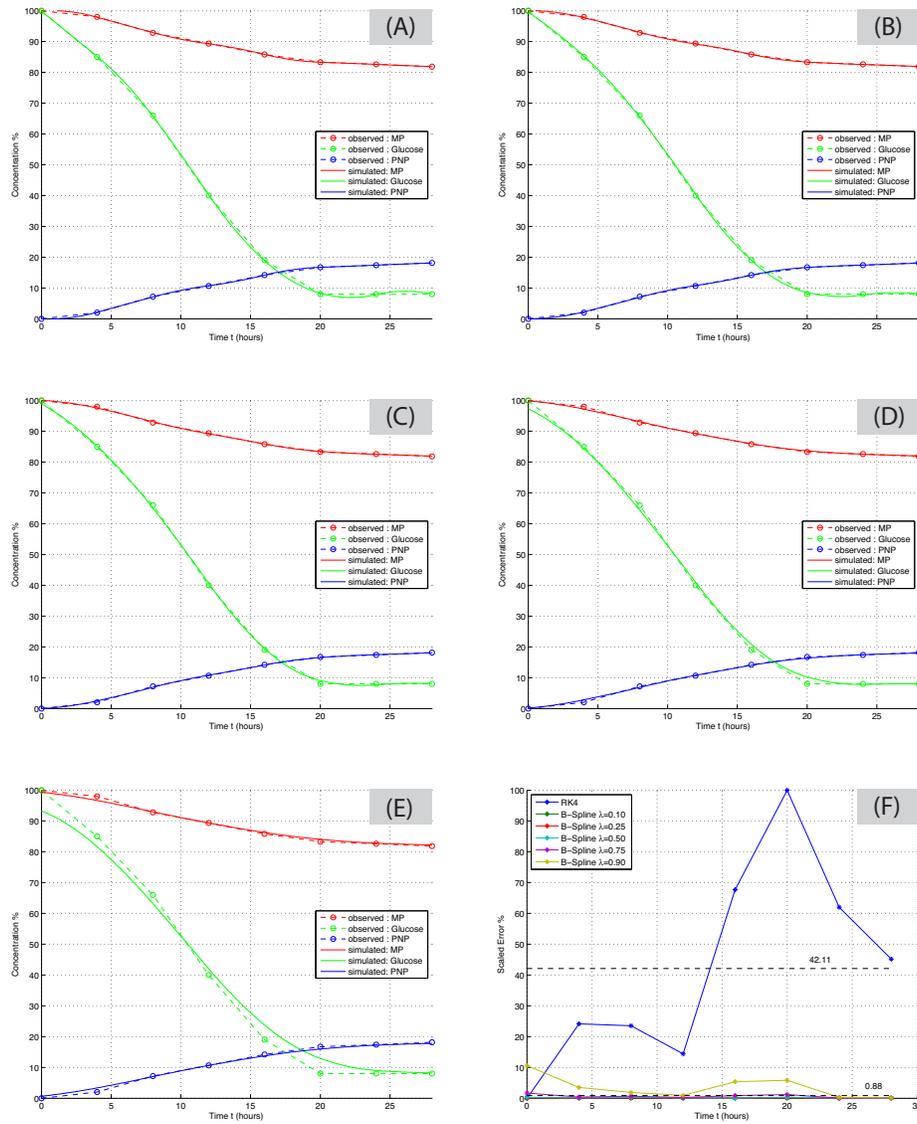


Fig. 3. B-Spline Simulation of experiment showed in Figure 1, using the B-Spline dynamic model without smoothing (i.e. $\gamma = 0$). (a)-(e) Simulations with $\lambda \in \{0.10, 0.25, 0.50, 0.75, 0.90\}$. (f) shows the error plots for the RK4 and the different simulations.

Different solutions of the linear system (16) for the LSP (15) are presented in Figure 3.

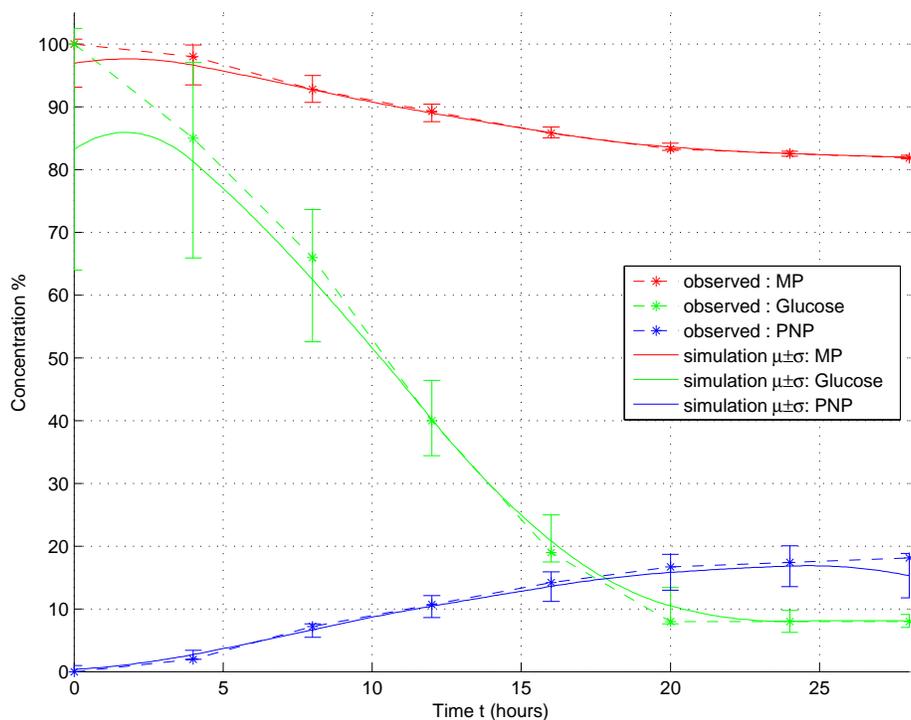


Fig.4. B-Spline Simulation of experiment showed in Figure 1, using the B-Spline dynamic model using 5 thousands values of γ equally spaced within the open interval (0,1). The dotted lines are the observations, the solid line is the average value of all the 5 thousands simulations and the error bars show the standard deviation around the average value.

7 Conclusions

In this work we have highlighted the importance of modeling enzymatic activities related to pesticides, because (i) pesticides affect the enzymatic activities of soils and (ii) the degradation of pesticides through hydrolysis is mediated by particular proteins isolated from soils. We have focused on the second point.

Using systems of differential equations to model the enzymatic activities, we have focused on developing models capable of satisfying both the system of differential equations representing kinetic laws and the observed experimental data. Thereby, the main objective of the present work was to improve the adjustment to the observed experimental data, which will improve the comparison between the results provided by experimental results and those yielded by computational modeling.

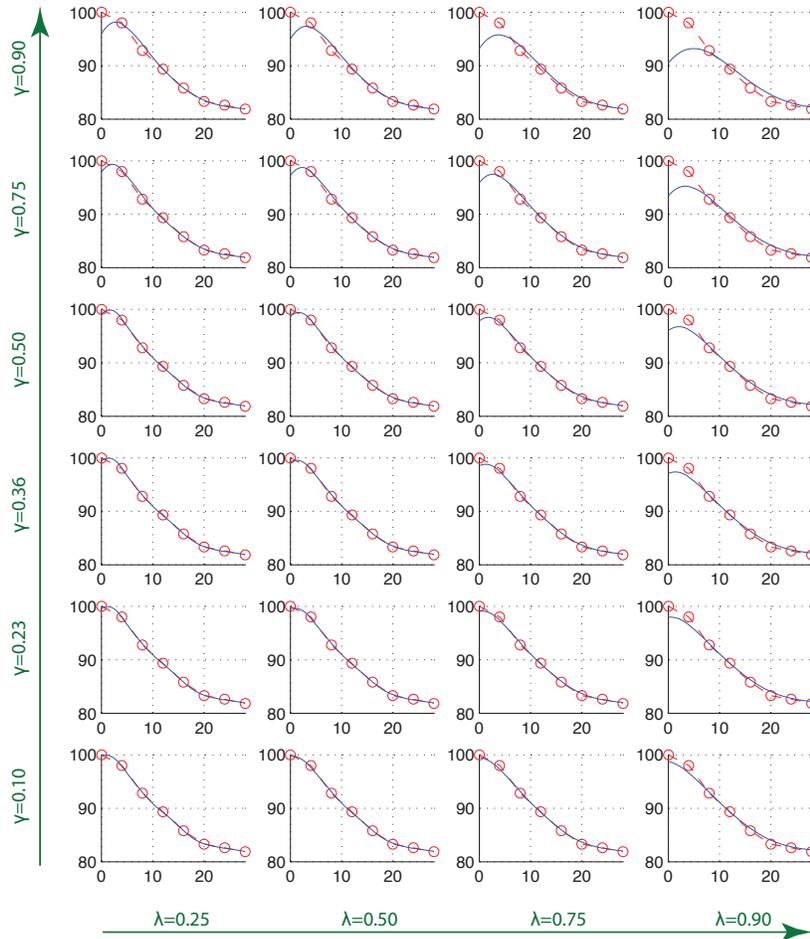


Fig. 5. B-Spline Simulations of the degradation of MP, using the B-Spline dynamic model. A red dotted line with circles represents the observations of the degradation of MP, a blue solid line represents the solution of the B-Spline dynamic model. By Columns from left to right, the plots have values of $\lambda = \{0.25, 0.50, 0.75, 0.90\}$. By Rows from down to top, the plots have values of $\gamma = \{0.1, 0.23, 0.36, 0.50, 0.75, 0.90\}$.

The accuracy concerning the adjustment to the observed data is as follows: the adjustment of the FBSM is about 50 times better than the Runge Kutta of order 4.

An important advantage of the FBSM is that it is possible to obtain the optimal functions with fast and efficient techniques for a Linear System whose matrix is banded, symmetric and positive-definite ($\mathcal{O}(n)$).

Future work will focus on testing the model with others experimental data to verify whether the applicability of the FBSM over different biological reactions is feasible.

8 Acknowledgements

The authors express their gratitude to the Mexican National Council for Science and Technology (CONACYT) for financing this work through the project grant: 84205, *Research of computational formal models for gene regulation*.

References

1. Ashyraliyev, M., Fomekong-Nanfack, Y., Kaandorp, J.A., Blom, J.G.: Systems biology: parameter estimation for biochemical models. *FEBS Journal* 276(4), 886–902 (2009)
2. Ashyraliyev, M., Jaeger, J., Blom, J.G.: Parameter estimation and determinability analysis applied to *Drosophila* gap gene circuits. *BMC Syst Biol* 2, 83 (2008)
3. Banga, J.: Optimization in computational systems biology. *BMC Systems Biology* 2(1), 47+ (2008)
4. Beier, S.P., Hede, P.D.: *Chemistry*. Ventus Publishing ApS (2010)
5. Caglar, N., Caglar, H.: B-spline method for solving linear system of second-order boundary value problems. *Comput. Math. Appl.* 57(5), 757–762 (Mar 2009)
6. Chino-Flores, C., Dantán-González, E., Vázquez Ramos, A., Tinoco-Valencia, R., Díaz-Méndez, R., Sánchez-Salinas, E., ML, C.G., Ramos-Quintana, F., ML, O.H.: Isolation of the opdE gene that encodes for a new hydrolase of *Enterobacter* sp. capable of degrading organophosphorus pesticides. *Biodegradation* 23, 387–397 (2012), 10.1007/s10532-011-9517-6
7. Daub, C., Steuer, R., Selbig, J., Kloska, S.: Estimating mutual information using B-spline functions - an improved similarity measure for analysing gene expression data. *BMC Bioinformatics* 5(1), 118+ (2004)
8. Érdi, P., Tóth, J.: *Mathematical models of chemical reactions: theory and applications of deterministic and stochastic models*. Nonlinear science, Manchester University Press (1989)
9. Eubank, R.: *Nonparametric Regression and Spline Smoothing*. Statistics, Textbooks and Monographs, Marcel Dekker (1999)
10. Goodrich, J., Kugel, J.: *Binding and Kinetics for Molecular Biologists*. Cold Spring Harbor Laboratory Press (2007)
11. Higham, D.J.: Modeling and Simulating Chemical Reactions. *SIAM Rev.* 50(2), 347–368 (May 2008)
12. Hii, Y., Rocklöv, J., Wall, S., Ng, L., Tang, C., Ng, N.: Optimal lead time for dengue forecast. *PLoS Negl Trop Dis* 6(10), e1848 (2012)
13. van der Kamp, M.W., Mulholland, A.J.: Computational enzymology: insight into biological catalysts from modelling. *Natural product reports* 25(6), 1001–1014 (Dec 2008)

14. Klappa, P.: Kinetics for Bioscientist. Ventus Publishing ApS (2009)
15. Kutalik, Z., Tucker, W., Moulton, V.: S-system parameter estimation for noisy metabolic profiles using newton-flow analysis. *IET Syst Biol* 1(3), 174–80 (May 2007)
16. Linford, J.C., Michalakes, J., Vachharajani, M., Sandu, A.: Multi-core acceleration of chemical kinetics for simulation and prediction. In: Proceedings of the Conference on High Performance Computing Networking, Storage and Analysis. pp. 7:1–7:11. SC '09, ACM, New York, NY, USA (2009)
17. Matsubara, Y., Kikuchi, S., Sugimoto, M., Tomita, M.: Parameter estimation for stiff equations of biosystems using radial basis function networks. *BMC Bioinformatics* 7(1), 230+ (2006)
18. Mulholland, A.J.: Computational enzymology: modelling the mechanisms of biological catalysts. *Biochemical Society transactions* 36(Pt 1), 22–26 (Feb 2008), <http://dx.doi.org/10.1042/BST0360022>
19. Nocedal, J., Wright, S.J.: Numerical Optimization. Springer (Aug 2000)
20. Ortiz-Hernández, M., Quintero-Ramírez, R., Nava-Ocampo, A., Bello-Ramírez, A.: Study of the mechanism of *Flavobacterium* sp. for hydrolyzing organophosphate pesticides. *Fundamental Clinical Pharmacology* 17(6), 717–723 (2003)
21. Pan, M.S., Tang, J.T., Yang, X.L.: An adaptive median filter algorithm based on B-spline function. *Int. J. Autom. Comput.* 8(1), 92–99 (Feb 2011)
22. Press, W.H., Teukolsky, S.A., Vetterling, W.T., Flannery, B.P.: Numerical recipes in C (2nd ed.): the art of scientific computing. Cambridge University Press, New York, NY, USA (1992)
23. Sannino, F., Gianfreda, L.: Pesticide influence on soil enzymatic activities. *Chemosphere* 45(4-5), 417–25 (2001), <http://www.biomedsearch.com/nih/Pesticide-influence-soil-enzymatic-activities/11680737.html>
24. Sekse, C., Bohlin, J., Skjerve, E., Vegarud, G.E.: Growth comparison of several *Escherichia coli* strains exposed to various concentrations of lactoferrin using linear spline regression. *Microb Inform Exp* 2, 5 (2012)
25. Sun, J., Garibaldi, J.M., Hodgman, C.: Parameter Estimation Using Metaheuristics in Systems Biology: A Comprehensive Review. *IEEE/ACM Trans. Comput. Biol. Bioinformatics* 9(1), 185–202 (Jan 2012)
26. Vaas, L.A.I., Sikorski, J., Michael, V., Göker, M., Klenk, H.P.: Visualization and curve-parameter estimation strategies for efficient exploration of phenotype microarray kinetics. *PLoS One* 7(4), e34846 (2012)
27. Zhan, C., Yeung, L.: Parameter estimation in systems biology models using spline approximation. *BMC Systems Biology* 5(1), 14+ (2011)
28. Zhang, R., Cui, Z., Zhang, X., Jiang, J., Gu, J.D., Li, S.: Cloning of the organophosphorus pesticide hydrolase gene clusters of seven degradative bacteria isolated from a methyl parathion contaminated site and evidence of their horizontal gene transfer. *Biodegradation* 17, 465–472 (2006), 10.1007/s10532-005-9018-6
29. Zheng, W., Bo, P., Liu, Y., Wang, W.: Fast B-spline curve fitting by L-BFGS. *Comput. Aided Geom. Des.* 29(7), 448–462 (Oct 2012)

The ITACA-WIICT'14 is a meeting forum for scientifics, technicians and other professionals who are dedicated to Information and communication technologies study and research. Its fundamental scope is to promote the contact among scientific and professionals, improving the cooperation as well as the technological transfer among professionals.

