New Trends in Electronics Technology

ISBN: 978-970-773-345-9

innovación editorial lagares
M E X I C O

New Trends in Electronics Technology

Gerardo Romero, Aldo Méndez, Marco Panduro y René Domínguez

New Trends in Electronics Technology

Edited by:
Gerardo Romero
Aldo Méndez
Marco Panduro
René Domínguez

NEW TRENDS IN ELECTRONICS TECHNOLOGY

Edited by

Gerardo Romero

Aldo Méndez

Marco Panduro

René Domínguez

innovación editorial lagares
M E X I C O

New Trends in Electronics Technology

D.R. © 2007 por Innovación Editorial Lagares de México, S.A. de C.V.
Av. Álamo Plateado No. 1 – 402
Fracc. Los Álamos
Naucalpan, Estado de México
C.P. 53230
Teléfono: (55) 5240-1295 al 98
email: editor@lagares.com.mx

ISBN: **978-970-773-345-9**

Diseño de Portada: Enrique Ibarra Vicente

Cuidado Editorial: Alicia Benet Vélez

Primera edición septiembre, 2007

IMPRESO EN MÉXICO / PRINTED IN MEXICO

Preface

This book presents some applications of electronics technology in the areas of telecommunications, control, evolutionary computation and photonic technology. More detail about the content of each area of interest is presented in the following sections.

Telecommunications

The wireless technology is a truly revolutionary paradigm shift, enabling multimedia communications between people and devices from any location. In addition, it has become a ubiquitous part of modern life, from global cellular telephone systems to local and even personal-area networks. Future wireless networks will be integrated into every aspect of daily life, and therefore could affect our life in a magnitude similar to that of the Internet and cellular phones. However, the emerging applications and directions require fundamental understanding on how to design and control wireless networks that lies far beyond what the currently existing theory can provide. Therefore, the first section, Wireless Networks, consists of 5 chapters. The first chapter, "Performance Analysis of Medium Access Control Protocol in Wireless Mobile Communications" by A. Méndez *et al.* presents the performance evaluation of the S-ALOHA, used as random access channel, and it analyzes the aspects that are highly sensitive on the performance of cellular systems. The second chapter, "Transmitter Precoding for MAI/ISI Rejection in a Wireless TDD/DS-CDMA System" by M. Luna-Rivera and D. Campos-Delgado proposes a transmit pre-filtering technique for downlink time-division-duplex (TDD) CDMA communications which employs the conventional matched filter detector at the mobile station. Analytical and simulation results are provided to illustrate the advantage obtained by using the precoding scheme at the transmitter. The third chapter, "Review of Frequency Selectivity Parameters for Broadband Wireless Signals" by V. Hinostroza-Zubía presents a system for indoor environment channel characterization. The Results demonstrate that the fading is within specific limits, these results could help to the designers of adaptive receivers to estimate the channel more accurately. The fourth chapter, "Quadruple Play: On the Wireless Communications Convergence in México" by A. Andrade presents a vision on the wireless communications convergence. In addition, a comprehensive introduction of the mainstream wireless mobile technologies and their evolution paths to the future in terms of the expected voice and data spectral efficiency. In the last chapter of this section, "Copying the Human Eye Strategies to Design Antenna Arrays", by D. Betancourt and C. del Rio, the behavior of the human eye is analyzed obtaining different solutions for the main common trade-offs of antenna array systems, in particular regarding the angular resolution and the signal/noise ratio.

Control

In recent years, the development of control systems technology has increased significantly; we can see this in the fact that almost every technological innovation considers a control dispositive for its operation. The second section of this book contains both theoretic and applied results of the main topics related to recent control theory. In the chapter 6 the author, C. Elizondo-González, presents some results based in root bounding to guarantee the robust D-stability property for linear time invariant systems. Such bounds are applied to obtain a theorem that determines the conditions to be achieved in the polynomial as to its roots have the real part bounded in the left side of the complex plane. Also, an example utilizing the results and applying a recent stability theorem for the linear time invariant system is presented. In chapter 7 E. Alcorta-García, discusses different model-based approaches and stresses one of the methods in particular the so called observer-based approach. This approach uses a dynamic model of the transformer to be supervised in order to reduce the rate of false alarms. The main advantage of observer-based methods is that the required stability is achieved. Next, in chapter 8 G. Sanahuja *et al.* show a comparative analysis of a linear and nonlinear control laws to stabilize a VTOL aircraft. A linear control law using LQR method is obtained and compared with respect to three nonlinear control strategies obtained using the well-known backstepping technique and the saturation functions. The performance of the controllers is compared by simulation but also in real-time experiences. The robustness of the control algorithms with respect to aggressive perturbations is illustrated with real-time experiments. In chapter 9, E. Gorrostieta *et al.*, present a methodology for the development of projects in mechatronics field appears which has been applied in several projects where it has valued the development and its behavior. The iteration of the different disciplines appears on projects development and obtaining in this way one a better integration, also this methodology has been used in the development of new industrial machinery. A

mechatronics design method is proposed as a part of the research and engineering interaction activities, but also the manufacture aspects and complex mechanical adjustments are considered. The methodology has been applied to the neural networks and fuzzy logic control of a pneumatic valve used in a flexible manipulator robot. In chapter 10, G. Romero *et al.* present sufficient conditions to verify the robust stability property of a class of time delay systems which are described as an interval plant with uncertain time delay. The main result is obtained on the basis of two polynomials that can be easily computed using Kharitonov's polynomials. In chapter 11, H. Sira-Ramírez and E. Barrios-Cruz present a high gain reduced order observer with integral estimation error injections is used for the fast determination of the unknown constant mechanical load in a DC motor. The motor is controlled by an exact tracking error dynamics passive output feedback control scheme which demands knowledge of the load parameter in the feed-forward terms alone and the on-line availability of the armature current. The determination of the feed-forward control is based on a certainty equivalence differential parametrization of the motor armature current and voltage input variables in terms of the angular velocity desired trajectory. The scheme results in a sensor-less, certainty equivalence, adaptive trajectory tracking feedback control scheme jointly exploiting the energy dissipation structure of the system and its flatness. The results are illustrated by means of digital computer simulations.

Evolutionary Computation

Taking a page from Darwin's 'On the origin of the species', computer scientists have found ways to evolve solutions to complex problems. Harnessing the evolutionary process within a computer provides a means for addressing complex engineering problems-ones involving chaotic disturbances, randomness, and complex nonlinear dynamics that traditional algorithms have been unable to conquer. Indeed, the field of evolutionary computation is one of the fastest growing areas of computer science and engineering for just this reason; it is addressing many problems that were previously beyond reach, such as rapid design of medicines, flexible solutions to supply-chain management problems, and rapid analysis of battlefield tactics for defense. Potentially, the field may fulfill the dream of artificial intelligence: a computer that can learn on its own and become an expert in any chosen area.

This section consists of two chapters. The first chapter, "Evolutionary Computation Techniques for Two Computational Biology Problems" by C. Brizuela-Rodríguez et al. deals with the design of evolutionary algorithms for two well known computational biology problems: whole-genome shotgun assembly and a simplified model of the protein folding. The folding problem is one of most challenging open problems in biology, and any clue on how to solve it or its approximations will be very valuable. The idea to solve the problems is to use a genetic algorithm tailored to specific problem knowledge. The proposed method has proven to be an effective alternative to solve both problems. The second chapter, "Design of Non-uniform Phased Linear Arrays using a Multi-objective Genetic Algorithm" by M. Panduro deals with the design of non-uniform phased linear arrays for smart antenna systems. The design problem is modeled as a multi-objective optimization problem with nonlinear constraints. A multi-objective genetic algorithm denominated NSGA-II is employed as the methodology to solve the resulting optimization problem. The addressed problem considers a driving-point impedance restriction and the design of non-uniform arrays to have a steerable radiation pattern. Experimental results show the effectiveness of the NSGA-II for the design of non-uniform phased linear arrays.

Photonic Technology

In view of recent advances in photonics devices and optical fiber communications systems, it is clear that electrical engineering student and higher should have considerable exposure to optoelectronics and fiber optics in their undergraduate and graduate educations, respectively, even if their areas of specialization are not photonics. Since photonics compromises many different disciplines and it has a relevant roll in several areas of the human activity, is important provide a brief overview of main topics of subjects selected in lightwave technology. With this in mind, the material of the Optoelectronics engineering section is organized in 3 chapters. The chapter "Specialty Optical Fibers in Laser and Sensing Applications" by R. Selvas-Aguilar *et al.*, give an overview of the investigation carried out on specialty optical fiber during the last 40 years and what has been done in Mexico. Rare-earth-doped optical fibers and photonic crystal fibers are well described in this article as those kinds of fibers have important applications in laser and sensor systems. The chapter, "Integrated InP Photonic Switches" by D. May-Arrioja and P. LiKamWa, present an integrated 1x3 optical switch that operates using the principle of carrier-induced refractive index change in InGaAsP multiple quantum wells. Finally, the last chapter "Non-linear optical effects in liquid crystals" by R. Dominguez-Cruz *et al.*, present the optical characterization of Kerr optical nonlinearity in liquid crystals through Z-scan technique.

Finally, we would like to thank every one who collaborated, one way or the other, in the writting of this book. Special thanks to M.B.A. Irma Pérez-Vargas and her crew, the PC UAT staff: Diego, Iván, Oliver, Charly, Reyna, Gloria, Adriana,

Tony, Luz, and Diana. Also, our gratefulness goes to the administrative Staff of the Universidad Autónoma de Tamaulipas represented by the Rector, M.E.S. José Ma. Leal Gutiérrez, and the UAM Reynosa Rodhe represented by the Director, Jaime Alberto Arredondo Lucio. And last, but not least, thanks to the greatest Scientist of all for His invaluable collaboration all along the way: Thank you God!.

The editors,
Gerardo Romero
Aldo Méndez
Marco Panduro
René Domínguez

Contents

PART I

TELECOMMUNICATIONS

Chapter 1

Performance Analysis of Medium Access Control Protocol in Wireless Mobile Communications

Aldo Mendez1, David Covarrubias2, and Cesar Vargas3

1 UAT Autonomous University of Tamaulipas, UAT-UAMRR; Reynosa, Tamaulipas, México; Apdo. Postal 88779. almendez@uat.edu.mx
2 CICESE Research Center; Ensenada, Baja California, México; Apdo. Postal 22860. dacoro@cicese.mx
3 ITESM-CET Campus Monterrey; Monterrey, Nuevo León, México; Apdo. Postal 64849 M. cvargas@itesm.mx

Abstract

In this chapter the performance of the Slotted Aloha protocol on mobile radio networks is evaluated, when Rayleigh fading, shadowing and the spatial distribution of the mobile terminals are taken into account. A new expression for the backlog steady-sate probabilities is given, and the capture probability is found to be more enhanced under the influence of shadowing environments – especially when combined with the other two effects. In addition, we present a performance evaluation of Slotted Aloha (S-Aloha) of those parameters, which most influence the performance of stability and efficiency.

Keywords: Capture effect, retransmission algorithm, S-Aloha, stability, spatial distribution, shadow, Rayleigh effect.

1 Introduction

In mobile communication systems, a Medium Control Access (MAC) protocol facilitates communication between the mobile and base terminal beginning with a Mobile Terminal (MT) service request via the random access channel (RACH). Based on this service request, the Base Station (BS) implements a resource assignment method that coordinates MT transmissions on a slot-by-slot basis via a scheduling algorithm that maximizes the utilization of Base Station (BS) resources while minimizing the potential conflict of MTs simultaneously contending for limited BS resources. It is worth emphasizing that given the present randomness characteristics in a mobile communication scenario, it is necessary to count with a MAC technique, that can efficiently cope with the possible interference among MTs which use simultaneously the RACH directed toward the same BS, i.e., the MAC technique should help prevent and solve these problems, as well as optimize the request of the channel in order to have high throughput and low average delay.

It is known from the literature [1] that Slotted ALOHA (S-Aloha), used as a random access channel (RACH), imposes limitations on the maximum throughput (only 36.8% of full channel capacity is achievable), and stability problems can occur as the load in the channel increases. In this chapter an evaluation of S-Aloha (in the uplink communication) is presented, where the aspects that are highly sensitive on the performance of cellular systems are determined. We also offer a different opinion analyzed in [2]-[3] in presenting a alternative version of the state transition probabilities of a cellular system while obtaining the capture probability of a channel exhibiting Rayleigh fading (RF), shadowing (S), and uniform spatial distribution (USD) of the mobile terminals (MTs). In addition, a stabilizing algorithm for the packet retransmissions is presented, which adapts to the instantaneous traffic changes.

The remainder of this work is organized as follows. Section 2 of this chapter offers a different opinion from [2]-[3] in presenting an alternative version of the state transition probabilities of a cellular system while obtaining the capture probability of a channel exhibiting Rayleigh fading, shadowing, and MTs spatial distribution. This section also models the S-Aloha access protocol using Markov chains to provide a detailed insight into the dynamic behavior of the network. Moreover, this section presents an algorithm of retransmission adaptable to the conditions of traffic changes. Section 3 presents the experimental setups and results. Finally, section 4 gives the conclusions of the work.

2 RACH Modeling

The initial stage of the analysis and modeling of the RACH is made on a slot by slot basis, where the slots of the system are numbered sequentially, k=0, 1, 2,…. Moreover, let $\eta_s(k)$ denote the number of backlogged mobile terminals at the beginning of the k-th slot, the random variable $\eta_s(k)$ is referred to as the state of the system. The number of MTs in backlog at the beginning of the (k+1)-th slot, depends on the number of mobile terminals in backlog at the beginning of the k-th slot,

and the number of mobile terminals going from one state to another within the slot. Due to the fact that this is independent of the activities in any previous slot, the process could be represented by a Markov chain modeled by a birth-death process.

According to the analysis made in [4], the process of retransmission and transmission of each mobile terminal for a finite number of mobile terminals, M, is an independent geometric process, in which the probability that i out of the j backlogged mobile terminals program a retransmission in a single slot, represents a binomial distribution, with probability that a mobile terminal retransmits a packet ν, and probability that a mobile terminal generates a new packet φ. By this means, the steady-state transition probabilities, $p_{ij} = \lim_{k\to\infty} \Pr\left(\eta_s(k) = j \mid \eta_s(k-1) = i\right)$ are obtained and the transition matrix P is formed [5]. The steady-state probability vector π, whose elements are π_j is the solution to the finite set of linear equations, [6]-[7]:

$$\boldsymbol{\pi}=\boldsymbol{\pi}\boldsymbol{P}, \text{ and } \sum_{i=0}^{M} \pi_i = 1. \tag{1}$$

It is advisable to analyze the system behavior in a steady-state, since the solution to the distribution in equilibrium of the steady-state of the Markov chain enables the evaluation of the system performance.

Through computer simulations, we have obtained the behavior of the steady-state probabilities considering the previous analysis for a finite population. Figure 1 illustrates the steady-state probability of 20 MTs (minimum quantity of a finite population), with a generation probability φ=0.1, and various retransmission probabilities, ν.

In addition, Figure 1 also illustrates that as the probability of retransmission, ν, increases, there exists a high probability of obtaining all the MTs in backlog, which suggests the design of an algorithm that can handle the probabilities of retransmission adaptably and dynamically. If the probability of generation and retransmission are high as illustrated in Figure 1, all MTs will be in backlog, requiring the application of a Deferred First Transmission (DFT) scheme [8]. Under this scheme, newly arriving data packets join the backlog prior to their first transmission. Each data packet is then independently transmitted within a time-slot with a given probability.

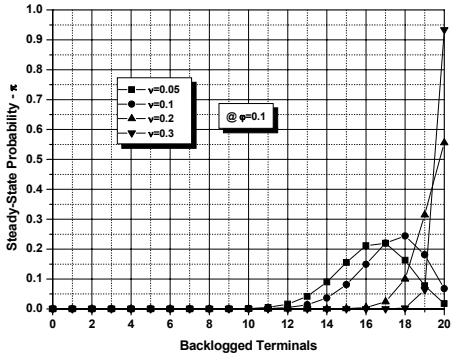


Fig. 1. Steady-state probability of the Random Access Channel.

Based on this scheme, all the packets (new and backlogged), receive equal treatment, facilitating the calculation of the retransmission probability. Recall that for S-Aloha with a high number of MTs in backlog, the average delay increases [1].

2.1 Modeling the Throughput of the RACH

In order to evaluate the throughput of the system, it is considered that all intention of transmission agrees with the beginning of each slot, and that the activity in any given slot is independent of the activity in any previous slot. Taken this into consideration, the fraction of time that a channel transports useful information or throughput, *S*, is equal to the average fraction of slots in any successful transmission.

For a transmission to be successful, there should only be a single transmission in the slot. This indicates that all of the mobile terminals in backlog are in silence and that only a new mobile terminal transmits, or only one mobile terminal in backlog transmits while there is no new packet generated. So the probability of success when *i* mobile terminals are in backlog state, is given by

$$P_{succ}(i) = (1-\nu)^i (M-i)\varphi(1-\varphi)^{M-i-1} + i\nu(1-\nu)^{i-1}(1-\varphi)^{M-i}, \tag{2}$$

and the throughput *S*, is expressed as

$$S = E[P_{succ}(i)] = \sum_{i=0}^{M} P_{succ}(i)\pi_i \,, \tag{3}$$

where the vector of probability of the steady state $\boldsymbol{\pi}$ can be calculated according to what has been discussed in the previous section.

2.2 Modeling the capture effect of the RACH

Up to this point, the modeling of S-Aloha as a RACH, has been considered within a noiseless channel where all data packets arrive at the BS with the same power levels. Under these conditions, when two or more data packets simultaneously arrive at the BS, they will collide and be destroyed. However, since data packets typically arrive at the BS at different power levels, capture effect occurs and reduces the probability of the destruction of the colliding data packets, which results in an increase of system throughput, [1].

Capture probability, $P_{capt}(i)$, $i>0$, is defined as the probability that one of i collided data packets will be successfully received. Capture probability is related to the concept of sensitivity of the receiver (S_{ens}), or in this case, BS sensitivity. According to [9], BS sensitivity is defined as the minimum voltage level at a receiver input required to obtain an output power ratio between 12 and 20 dB. To successfully capture a data packet, its maximum signal level should be greater than the sensitivity of the receiver [10]. In addition, the threshold between the power of the signal and the power of the possible interferers should be greater than a certain margin known this as capture ratio (R), [11]-[12].

According to the previous information, and given the presence of I, where $I\geq 1$, interfering packets in the same cell (each interfering packet with a power of w_{ui}, $i=1, 2, 3, \ldots, I$), the probability of capture is obtained by comparing the power of a useful packet, w_c, with the total power of possible interfering packets according to

$$P_{capt}(I) = \Pr\left(\frac{w_c}{\sum_{i=1}^{I} w_{ui}} \rangle R \quad , \quad \mathrm{w_c} \rangle S_{ens} \right) , \quad \mathrm{I} \rangle 0 \,. \tag{4}$$

Considering this scheme, a new expression for the state transition probabilities has been obtained,

$$p_{ij} = \begin{cases} 0 \,, & j \langle i-1 \,, \\ (1-\varphi)^{M-i} \sum_{c=1}^{i} \binom{i}{c} \nu^c (1-\nu)^{i-c} P_{capt}(c), & j = i-1, \\ 1_{\{M \rangle j\}} \binom{M-i}{j-i+1} \varphi^{j-i+1} (1-\varphi)^{M-j-1} \cdot & \\ \sum_{k=0}^{i} \binom{i}{k} \nu^k (1-\nu)^{i-k} P_{capt}(k+j-i+1) + & \mathrm{M} \geq \mathrm{j} \geq i. \\ \binom{M-i}{j-i} \varphi^{j-i} (1-\varphi)^{M-j} \sum_{k=0}^{i} \binom{i}{k} \nu^k (1-\nu)^{i-k} \cdot & \\ \left[1 - P_{capt}(k+j-i)\right] \,. & \end{cases} \tag{5}$$

Expression (5) is different to that given in [2] and [3], mainly because they no considering the case when the final number of mobile terminals in backlog is equal to M ($j=M$), and this will happen when $M-i$ idle terminals transmit and no packet is captured successfully.

Since the Markov chain is homogeneous, the vector of steady-state probability $\boldsymbol{\pi}$, can be calculated by solving a group of non linear equations $\boldsymbol{\pi}=\boldsymbol{\pi P}$, and $\sum_{i=0}^{M} \pi_i = 1$.

With these probabilities, the throughput performance can be computed by

$$S = E\left[P_{succ}(i)\right] = \sum_{i=0}^{M} P_{succ}(i)\cdot \pi_i \quad , \tag{6}$$

where the probability of success, $P_{succ}(i)$, considering the capture effect, is given by

$$P_{succ}(i) = \sum_{r=0}^{M-i}\binom{M-i}{M-i-r}\varphi^r(1-\varphi)^{M-i-r}\sum_{c=0}^{i}\binom{i}{c}\nu^c(1-\nu)^{i-c}P_{capt}(c+r). \tag{7}$$

According to equations (6) and (7), we need to determine the capture probabilities P_{capt}, while considering the spatial distribution of the MTs, and the presence of radio channel characterized by Rayleigh fading, and shadowing

The probability of capture that considers uniform spatial distribution and the effects of Rayleigh fading, and shadowing is given by

$$P_{capt}(I) = \left[B^2\int_0^{\infty}\int_0^{\infty}\int_0^{1} 2y\cdot\frac{1}{\left(Rzy^4+u\right)z}\cdot\exp\left(-\frac{\left(A-\log u\right)^2}{2\sigma^2}\right)\cdot \exp\left(-\frac{\left(A-\log z\right)^2}{2\sigma^2}\right)dydudz\right]^{I-1}, \tag{8}$$

where $B = \frac{\log e}{\sqrt{2\pi\sigma}}$, $A = \log W_0 + \frac{\sigma^2}{2\log e}$, W_0 is the mean value of the signal power, σ is the standard deviation, y is the distance between the MT and BS, *log* is the logarithm base 10, and R the capture ratio.

2.3 Stability analysis of the RACH

According to the feedback information of a slot (idle, success, and collision), we have modeled the algorithm of retransmission through Markov chains. Utilizing feedback information, the algorithm of retransmission has the capacity to execute channel load control, by performing channel load estimation, $\hat{n}$, at the beginning of each time-slot, [13].

Under these circumstances, each backlogged data packet is retransmitted with a probability $\nu(\hat{n}) = \min\left(1, \frac{1}{\hat{n}}\right)$. In this algorithm, all the MTs observe the feedback channel. Thus, they obtain information about the outcome of each slot. In time-slot k+1, each MT independently performs an update of his estimate $\hat{n}$ and obtains an $\hat{n}_{k+1}$ depending on the outcome of the previous time-slot [14].

$$\hat{n}_{k+1} = \hat{n} + \begin{cases} est\_i \text{ , if slot k was idle,} \\ \text{est\_s , if slot k had one success} \\ \qquad \text{(active state),} \\ \text{est\_c , if slot k had one collision} \\ \qquad \text{(backlog state),} \end{cases} \tag{9}$$

where *est_i*, *est_s*, *est_c* are defined later in this section. Given that all of the MTs have the same estimate $\hat{n}$ for n, the probability of success in slot k+1 is

$$P_{succ_{stab}} = \frac{n}{\hat{n}}\left(1-\frac{1}{\hat{n}}\right)^{n-1}\left(1-p_e\right) \approx \frac{n}{\hat{n}}e^{-\frac{n}{\hat{n}}}\left(1-p_e\right) \quad , \tag{10}$$

where p_e is the probability of the packet not being accepted by the receiver. And the probability that an idle slot is detected, can be obtained as

$$P_{idle} = \left(1 - \frac{1}{\hat{n}}\right)^n \approx e^{-\frac{n}{\hat{n}}} . \tag{11}$$

Finally, a collision may occur with a probability $P_{collision} = 1 - P_{succ_{stab}} - P_{idle}$.

Define the estimate drift, $d(y)$, as the expected value of the difference between $\hat{n}$ and $\hat{n}_{k+1}$; i. e.,

$$d(y) = est\_i \cdot P_{idle} + est\_s \cdot P_{succ_{stab}} + est\_c \cdot P_{collision} \tag{12}$$

where $y = \frac{n}{\hat{n}}$. According to the properties of estimation in equilibrium, $d(1)=0$, symmetric point ($d''(1)=0$), and $d'(1) = \delta$ for an arbitrary $\delta > 0$, the values of *est_i*, *est_s*, and *est_c* are found to be

$$\begin{aligned} est\_i &= 2\delta - e\delta , \\ est\_s &= 2\delta - \frac{e\delta}{1 - p_e}, \\ est\_c &= 2\delta . \end{aligned} \tag{13}$$

With these values, it is possible to calculate in a dynamic form the estimate, Equation (9), obtaining control of the load of the system. Therefore, the objective is to operate with a binary exponential retransmission algorithm [8] that permits us to stabilize the response of S-Aloha, as it will be seen in Section 3.

3 Experimental Setup and Results

In order to evaluate our system, a Real Time Emulator (RTE) [15] was used. To achieve useful results, the RTE must perform a real time emulation of the behavior of the input/output transport chain.

Markov chains, based on the hidden Markov model (HMM), are used to model the functionality transport. That is, for each testing scenario, the Markov chains are properly trained through offline simulations to reproduce the statistical behavior of the channel radio. Once this statistical behavior is understood, the parameters of the HMM within the RTE, are properly tuned to reproduce this behavior with sufficient accuracy from a statistical viewpoint.

This approach allows the emulation of a great number of scenarios under various propagation conditions. Therefore, the main advantage of using a HMM model is the reduction in time, resources, and effort with regard to implementing a real system. In this discrete simulation process the packets of each MT are generated according to a Bernoulli process, with probabilities of generation of 10^{-3} to 1. For each probability of generation, 10,000 data packets are transmitted, where each data packet is contained in a time-slot, and the probability of transmitting a new data packet is 1, independently of the present value of retransmission probability. We also assume that a MT cannot generate a new packet until the present data packet has been transmitted. In addition, we consider w_c and w_{ui} as random variables. Additionally, Table 1 shows parameters utilized in this simulation.

Table 1. *Simulation Parameters.*

Parameters	**Value**
Number of mobile/cellular terminals - MT	80
Standard deviation for outdoors - σ	5 dB
Minimum sensitivity of the base station – S_{ens}	-116 dBm
Power loss factor - α	4
MT's maximum transmission power	125 mW

In our presentation of results, we will be including, in a gradual manner the described parameters of Section 2. The first parameter of S-Aloha as RACH simulated is the throughput.

Figure 2 illustrates the response of S-Aloha in our simulation, considering a uniform spatial distribution (USD) of the MTs and, a capture effect without considering effect of the radio channel (that means there is perfect transmit power

control). The parameter R is the capture ratio which we varied from the almost perfect capture (R=2) up imperfect capture (R=10), [1]. The capture ratio R=1 (perfect capture) wasn't illustrated since this behavior is very unlikely.

Figure 2 illustrates the influence of the capture ratio *R* on throughput system. As indicated, the smallest capture ratios correspond to the highest system throughputs. As the capture ratio increases, the desired signal power requires more power than total interference power so that the packet is captured by the BS and therefore the value of the throughput decreases. Due that with *R*=2 the throughput presents the best response, this value of *R* will be used during the remainder of this chapter. The typical response of S-Aloha (R=∞), will be used as a reference.

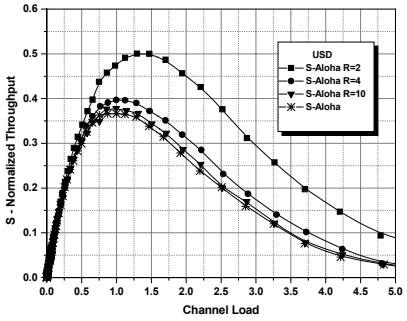


Fig. 2. *Behavior of the Throughput considering Uniform Spatial Distribution (USD) and Capture Ratio.*

The following simulation considers, capture ratio, uniform spatial distribution (USD), Rayleigh fading (RF), and shadowing (S). Figure 3 illustrates these combinations of behaviors.

In Figure 3 we observe that throughput improves substantially when considering the shadow effect. This improvement is due to capture probability increases when the shadow effect and spatial distribution of the MTs are considered. Consequently, the probability that a channel exhibiting Rayleigh fading is sufficiently strong enough to survive a collision is extremely small. Therefore, these results indicate that throughput performance is highly sensitive to shadowing for a given capture ratio.

An additional parameter of interest is the improvement of S-Aloha stability. Figure 6 illustrates a S-Aloha throughput response when utilizing an algorithm of dynamic retransmission, a capture ratio (*R*=2), uniform spatial distribution, and the combined effects of Rayleigh fading and shadowing.

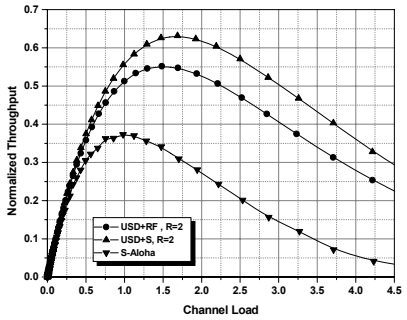


Fig. 3. *Behavior of the Throughput considering Uniform Spatial Distribution (USD), and Radio Channel Effect.*

Figure 4 illustrates that system throughput improves when the capture effect, the radio channel effects, and stabilization through the retransmission probabilities are considered. These effects assist the system to achieve efficiencies greater than 70% and stability in most of the range of channel traffic. The uniform spatial distribution of the MTs and shadowing have a significant effect on channel efficiency, while Rayleigh fading has less of an effect on channel efficiency. Also, there is a better-suited behavior on the handling of the channel traffic, since when the load of the system is increased the throughput value holds constant (approximately 0.65), due to a control load effect. We may say that this behavior is because

it uses a feedback channel which is useful for indicating the state of the channel, and thus controlling the re-transmission probabilities dynamically.

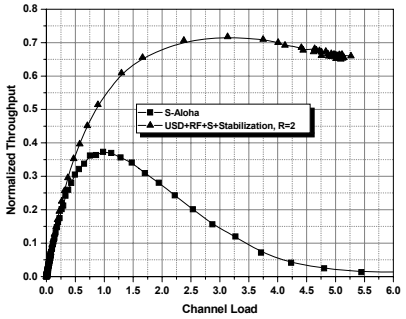


Fig. 4. *Behavior of the Throughput improving the response of stability and efficiency.*

The behavior of the number of backlogged MTs (MTs that collided before and that have packets to retransmit) is shown in Figure 5.

Figure 5 illustrates that the number of backlogged MTs for S-Aloha with ideal channel is approximately 50% of the total MTs at maximum throughput. When applying the dynamic retransmission algorithm, the number of backlogged MTs is becomes less than 7% of the MTs at maximum throughput. By applying a dynamic control to the re-transmission algorithm through the use of a feedback channel, the information channel becomes more efficient by quickly resolving collisions and providing a decreasing the number of low traffic backlogged MTs.

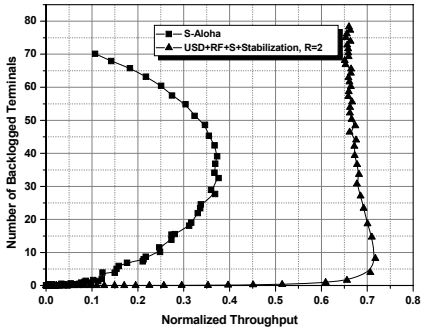


Fig. 5. *Number of mobile terminals in backlogged mode of S-Aloha: typical response and enhanced response.*

According to the statistics obtained in Section 2.2, regarding the number of backlogged MTs and applying Little's Theorem, [4], we can obtain the response of the average delay of S-Aloha. The behavior of this parameter is shown in Figure 6.

According to Figure 6, a near zero slot delay with minimum variation is achieved by utilizing a dynamic stabilization algorithm until the normalized throughput reaches approximately 5 time-slots in the region of maximum throughput. This delay behavior is due to the retransmission algorithm optimizing channel management by dynamically controlling the retransmission probabilities and by quickly resolving data packet collisions.

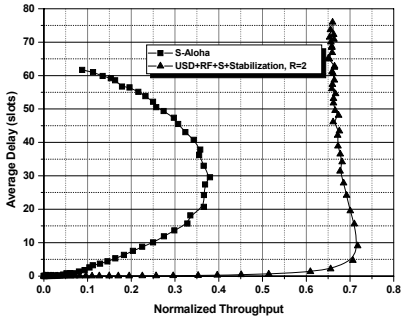


Fig. 6. *Behavior of the Average Delay of S-Aloha: typical response and enhanced response.*

4 Conclusions

We have presented an evaluation of the S-Aloha protocol on a mobile, radio channel, where the shadowing, Rayleigh fading and spatial distribution of the mobile terminals are considered as fundamental aspects that affect the capture effect in the receiver. We have also presented a novel expression for the steady-sate probabilities of the Markov chain that represents the backlog in the system.

We found that shadowing and Rayleigh fading effects actually improve the probability of capture (better in the first case), and that their combined effect, together with a stabilization mechanism and the uniform distribution of mobile terminals, further enhance the result.

5 References

1. Covarrubias, D.: Procedures and Techniques of Dynamic Assignment and Stabilizing of Applicable MAC to Mobile Systems of Third Generation (in Spanish). (1999) Ph.D. Thesis. UPC, Spain.
2. Davis, D. and Gronemeyer, S.: Performance of Slotted ALOHA Random Access with Delay Capture and Randomized Time of Arrival. IEEE Transactions on Communications. 28 (1980) 703-710.
3. Habbab, I., Kavehrad, M. and Sundberg, C.: ALOHA with Capture Over Slow and Fast Fading Radio Channels with Coding and Diversity. IEEE JSAC. 7 (1989) 79-88.
4. Rom, R., and Sidi, M.: Multiple Access Protocol-Performance and Analysis. 1st edn. Springer-Verlag, New York (1990).
5. Bolch, G., et al.: Queueing Networks and Markov Chains: Modeling and Performance Evaluation with Computer Science Applications. 1st edn. John Wiley & Sons, New York (1998).
6. Kleinrock, L.: Queueing Systems Vol I: Theory. 1st edn. John Wiley & Sons, New York (1975).
7. Taylor, H. M. and Karlin, S.: An Introduction to Stochastic Modeling. 1st edn. Academic Press Inc., U.S.A. (1994).
8. Jeong, D. J. and Jeong, W. S.: Performance of an Exponential Backoff Scheme for Slotted-ALOHA Protocol in Local Wireless Environment. IEEE Transactions on Vehicular Technology. 44 (1995) 470-479.
9. Hernando, J. M. and Perez-Fontan, F.: Introduction Mobile Communications Engineering. 1st edn. Artech House Publishers, U.S.A. (1999).
10. Zhang, Z. and Liu, Y-L.: Throughput Analysis of Multichannel Slotted ALOHA Systems in Multiple Log-Normal and Rayleïgh Environment. Proceedings of the IEEE 42nd Vehicular Technology Conference. 1 (1995) 55-58.
11. Zhou, H. and Deng, R. H.: Capture Model for Mobile Radio Slotted ALOHA Systems. IEE Proceedings Communications, 145 (1998) 91-97.
12. Zorzi, M. and Rao, R. R.: Capture and Retransmission Control in Mobile Radio. IEEE JSAC. 12 (1994) 1289-1298.
13. Bertsekas, D. and Gallager, R.: Data Networks. 2nd edn. Prentice-Hall, NJ (1992).
14. Bottcher, A. and Dippold, M.: The Capture Effect in Multiaccess Communications-the Rayleigh and Landmobile Satellite Channels. IEEE Transactions on Communications. 41 (1993) 1364-1372.
15. Covarrubias, D. et al.: An efficient Adaptive Coding Scheme for Data Transmission over a Fading and Nonstationary Mobile Radio Channel. Proceedings of the IEEE 3rd International Symposium on Multi-Dimensional Mobile Communications-MDMC'98. (1998) 168-173.
16. Linnartz, J. P. and Prasad, R.: Near-Effect on Slotted Aloha Channel with Shadowing and Capture. Proceedings of the IEEE Vehicular Technology Conference'89. (1989) 809-813.
17. Linnartz, J. P.: Narrowband Land-Mobile Radio Networks. 1st edn. Artech House, Boston (1993).

Chapter 2

Transmitter Precoding for MAI/ISI Rejection in a Wireless TDD/DS-CDMA System

J.M. Luna-Rivera, and D.U. Campos-Delgado

Departamento de Electronica, Facultad de Ciencias, UASLP, Av. Salvador Nava s/n, C.P. 78290, S.L.P., Mexico. {mlr, ducd}@fciencias.uaslp.mx

Abstract

Transmission of signals through time-varying mobile radio channels destroys orthogonality between the different users' signals of a CDMA system, and thus, causes multiple access interference (MAI) and inter-symbol interference (ISI). Many multi-user detection techniques have been proposed to improve CDMA performance; however they induce higher complexity at the receiver. In this work, we present a transmit pre-filtering technique for downlink time-division-duplex (TDD) CDMA communications which employs the conventional matched filter detector at the mobile station. This precoding technique provides a very simple transmission scheme that combines a cyclic prefix strategy with a signal transformation to reduce the MAI/ISI effects. The main focus is on a constrained minimum mean square error (MSE) pre-filtering that minimizes the MSE together with a transmitter power constraint. Analytical and simulation results are provided to illustrate the advantage obtained by using the precoding scheme at the transmitter.

Keywords: Transmitter Precoding, Cyclic Prefix, Multi-path Channel, CDMA.

1 Introduction

Since the air-interface of the third generation of mobile communications is based on code division multiple access (CDMA) technology, a significant interest exists in enhancing data rates and capacity for this type of systems. In practical conditions, CDMA systems suffer severely from MAI, due to simultaneous usage of the available bandwidth by many users, and ISI, due to the multi-path propagation which gravely reduces the performance of classical CDMA systems with the conventional RAKE receivers. Such interference factors have induced the development of sophisticated signal-processing techniques for data detection. To combat MAI/ISI, multi-user detection (MUD) techniques can be applied at the downlink CDMA receiver. MUD is a well-investigated topic that mitigates MAI/ISI. Current MUD receivers (see [1] and the references therein) offer attractive compromise between performance and complexity, but this has placed a higher computational and cost burden on detectors, demodulators, decoders, etc. Nevertheless, the interest of maintaining low cost and complexity, especially at the mobile units, is as important as ever. This has led to the investigation of alternative signal-processing algorithms. As shown in [2], [3], the signal received at each mobile unit can be improved through transmitter-based processing while keeping the advantage of a simple receiver. This signal pre-processing approach allows significant reductions in complexity by moving computational burden from the mobile units to the base station.

In what follows, a simple bitwise precoding technique with cyclic prefix is described for transmissions over multi-path channels. An optimized precoding matrix is applied in the transmitter using the mean square error (MSE) criterion. It is fair to remark that the MSE minimization at the transmitter is different from the well-known MSE solution at the receiver because transmit processing affects all the received signals before noise is introduced. To maintain the average transmitted power per symbol interval to a desired level, a power constraint condition is incorporated into the optimization problem. The resulting pre-processing technique is performed at the transmitter so that a simple despreading receiver (matched filter) can be utilized at the mobile unit, thereby, eliminating the need for channel estimation and equalization at the receiver. An important requirement for transmitter precoding is, however, the advance knowledge of the wireless channel at the transmitter. In a TDD mode, the downlink and uplink signals are time multiplexed into the same carrier. Hence the downlink channel can be estimated and updated at the base station using the uplink signals continuously [4].

Transmit pre-processing techniques can be performed linearly or nonlinearly. Previous work on linear transmit pre-processing for the downlink of a multi-user CDMA system includes transmit matched filter [5], transmit zero-forcing filter [6] and transmit MMSE [2]. While transmit matched filter maximizes the desired signal portion in the received signal, transmit zero-forcing can completely pre-cancel MAI but with its performance degraded by transmit power scaling. Meanwhile, transmit MMSE finds the optimum precoding transformation that minimizes the mean square error. A comparison of linear precoding techniques based on FIR structures can be found in [5]. A non-linear extension of transmit

precoding is provided by the Tomlinson-Harashima (THP) based precoding techniques [7], [8], and the references within. Further precoding techniques are suggested in [9], [10], [11], [12], [13]. Transmit pre-processing using the MSE criterion has already been applied in different systems [2], [3], [5], [9], [10], [11], [12], [13]. In [2], the optimum precoding transformation in the MMSE sense is considered for CDMA over AWGN and multi-path channels. Noll Barreto et al. proposed in [9] a constrained MMSE transmit filter which minimizes both the MSE together with a transmit power constraint. Different from previous results, the transmit pre-processing scheme presented here is a transmitter precoding strategy which, combined with a DS-CDMA system including cyclic prefix, offers: 1) a multi-user transmitter optimization scheme for combating MAI in multi-path fading channels using a simple matched filter receiver; 2) resilience to multi-path fading by adding a cyclic prefix in the time domain rather than using frequency domain processing as in OFDM systems [14]; 3) simplicity even at the base station. The benefits of this transmission scheme are mainly obtained by incorporating the cyclic prefix strategy from which a multi-user detection problem in multi-path channels is converted to a set of decoupled single user detection problems.

Notations throughout this chapter are defined as follows. Scalars are denoted by *italic*, lowercase **bold** for vectors, uppercase **bold** for matrices, vector 2-norm by $||\cdot||$, matrix Frobenius norm by $||\cdot||_F$ and trace by *tr(·)*.

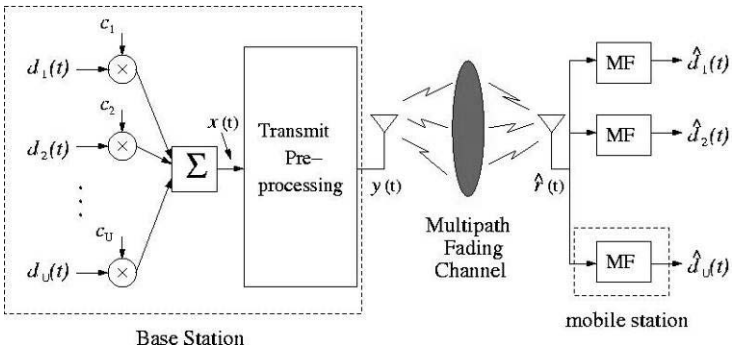


Fig. 1. *The downlink of an U-user TDD/DS-CDMA system with transmit pre-processing.*

2 System Model

The system of interest is the downlink of a synchronous TDD/DS-CDMA system with U active users communicating through a common base-station (BS), see Fig. 1. Considering the same channel impulse response for all users, the received signal at the u-th mobile user can be modeled in matrix notation as

$$\hat{\mathbf{r}}_u(k) = \mathbf{H}_d \mathbf{y}(k) + \mathbf{H}_i \mathbf{y}(k-1) + \hat{\boldsymbol{\eta}}_u(k) \tag{1}$$

where $\hat{\mathbf{r}}_u(k) \in \Re^{(N+\alpha)x1}$ is the received signal of the u-th mobile user at the k-th symbol interval, $\mathbf{y}(k)$ and $\mathbf{y}(k-1)$ are the transmitted signals during the k-th and $(k-1)$-th symbol intervals. The scalars N and α define the system's processing gain and channel delay spread respectively. Following the spreading and transmit pre-processing block, the transmitted signal at symbol interval k is described by

$$\mathbf{y}(k) = \mathbf{TCd}(k) \tag{2}$$

where $\mathbf{d}(k) = [d_1(k), \cdots, d_U(k)]^T$ with $d_u(k) \in \{1, -1\}$ as the k-th data bit from user u; $u \in \{1, \cdots, U\}$. The u-th column vector of matrix $\mathbf{C} \in \Re^{NxU}$, consisting of N chips, denotes the spreading code of user $u, \mathbf{c}_u = [c_{1,u}(k), c_{2,u}(k), \cdots, c_{N,u}(k)]^T$, with $c_{n,u}(k) \in \left\{\pm 1/\sqrt{N}\right\}$; $n \in \{1, \cdots, N\}$. Here transmitter precoding is defined by the linear transformation matrix $\mathbf{T} \in \Re^{(N+\alpha)xN}$ whose structure will be defined in next section. In (1), it is straightforward to show that during the k-th time interval the channel response can be separated into the matrices $\mathbf{H}_d$ and $\mathbf{H}_i$, where the contribution to the transmission of $\mathbf{y}(k)$ is denoted by the Toeplitz matrix $\mathbf{H}_d \in \Re^{(N+\alpha)x(N+\alpha)}$ with first row $\begin{bmatrix} h_{k,0} & 0 & \cdots & 0 \end{bmatrix}$ and first column $\begin{bmatrix} h_{k,0} & \cdots & h_{k,\alpha} & 0 & \cdots & 0 \end{bmatrix}^T$ while the effect of ISI is represented by the Toeplitz

matrix $\mathbf{H}_i \in \Re^{(N+\alpha)x(N+\alpha)}$ with first row $\begin{bmatrix} 0 & \cdots & 0 & h_{k,\alpha} & \cdots & h_{k,1} \end{bmatrix}$. The channel gain corresponding to the *l*-th path during the *k*-th symbol interval is denoted by $h_{k,l}; l = 0, \cdots, \alpha,$ with α as the number of resolvable paths. Finally, the vector $\hat{\boldsymbol{\eta}}_u(k) \in \Re^{(N+\alpha)x1}$ in (1) represents the noise with zero mean and variance σ^2.

3 Transmitter Precoding

The proposed transmit pre-processing process can be summarized as follows. The downlink signal transmitted during the *k*-th symbol interval can be written as given in (2), where $\mathbf{C}$ is the matrix of spreading codes and $\mathbf{T} \in \Re^{(N+\alpha)xN}$ the transmit filter matrix whose structure is defined as

$$\mathbf{T} = \begin{bmatrix} \mathbf{A} \\ \mathbf{B} \end{bmatrix} \tag{3}$$

The transformation matrix $\mathbf{T}$ will prove to be convenient for ISI elimination. The precoding scheme presented here employs a cyclic prefixed strategy to eliminate the ISI effect from the channel. The cyclic prefix is incorporated in (3) with the form of matrix $\mathbf{A} \in \Re^{\alpha xN}$, this cyclic prefix is defined as the last α rows of matrix $\mathbf{B}$. Thus, the transformation in (3) is reduced to choose matrix $\mathbf{B} \in \Re^{NxN}$ according to some optimality criterion. Since ISI occurs only on the first α samples of $\hat{\mathbf{r}}_k(k)$, this effect can be simply eliminated by ignoring the first α samples of $\hat{\mathbf{r}}_k(k)$ at the receiver. After cyclic prefix removal, the received signal for the *u*-th user is then expressed as

$$\mathbf{r}_k(u) = \mathbf{HBCd}(k) + \boldsymbol{\eta}_u(k) \tag{4}$$

where now the channel response is given by the *NxN* circulant Toeplitz matrix

$$\mathbf{H} = \begin{bmatrix} h_{k,0} & 0 & \cdots & 0 & h_{k,\alpha} & h_{k,\alpha-1} & \cdots & h_{k,1} \\ h_{k,1} & h_{k,0} & \cdots & 0 & 0 & h_{k,\alpha} & \cdots & h_{k,2} \\ \vdots & & \ddots & & & & & \vdots \\ h_{k,\alpha} & h_{k,\alpha-1} & \cdots & h_{k,1} & h_{k,0} & 0 & \cdots & 0 \\ \vdots & & \ddots & & & & & \vdots \\ 0 & \cdots & 0 & h_{k,\alpha} & h_{k,\alpha-1} & \cdots & h_{k,1} & h_{k,0} \end{bmatrix} \tag{5}$$

Since the first α samples of matrix $\mathbf{T}$ represents the matrix $\mathbf{A}$ and by neglecting the same number of samples at the receiver, then the received signal can be expressed as given in (4). As a consequence, the cyclic prefix matrix $\mathbf{A}$ does not appear anymore in equation (4). Therefore, the symbols transmitted are estimated by simply multiplying (4) by $\mathbf{C}^T$, that is

$$\hat{d}(k) = \mathbf{C}^T\mathbf{r}(k) = \mathbf{C}^T\mathbf{HBCd}(k) + \mathbf{C}^T\boldsymbol{\eta}_u(k) \tag{6}$$

where the *u*-th element of vector $\hat{d}(k) \in \Re^{Ux1}$, $\hat{d}_u(k)$, represents the estimate of the *k*-th transmitted symbol for user *u*.

3.1 Transmitter Precoding with Power Constraint

Since the aim is to improve the downlink performance of a TDD/DS-CDMA system by pre-processing the transmitting signal, we seek to choose the transmit filter $\mathbf{T}$ in (2) that provides the best performance possible when the mobile unit is constrained to the use of a matched filter receiver. The minimum mean square error (MMSE) is then chosen as the criterion for the design of matrix $\mathbf{T}$ or equivalently $\mathbf{B}$. Be $J = E[\| \mathbf{d}(k) - \hat{\mathbf{d}}(k) \|^2]$ the mean square error with $E[\cdot]$ as the expectation with respect to $\mathbf{d}(k)$ and $\boldsymbol{\eta}_u(k)$. Substituting $\hat{\mathbf{d}}(k)$, as given in (6), then

$$J = \left\| \mathbf{I}_U - \mathbf{C}^T\mathbf{HBC} \right\|_F^2 + U\sigma^2 \tag{7}$$

Where $\mathbf{I}_U$ denotes an UxU identity matrix. It is easy to verify that considering orthogonal spreading codes, *i.e.* $\mathbf{C}^T\mathbf{C} = \mathbf{I}_U$, the optimum transmit pre-processing matrix would be $\mathbf{B} = \mathbf{H}^{-1}$, assuming that $\mathbf{H}$ is positive definite so that $\mathbf{H}^{-1}$ exists. Notice that this optimum solution does not affect the noise variance [2]. A statistical minimization of the MSE considering the receiver noise is also possible but of little practical interest, since the noise power at the mobile stations are not known at the transmitter. However, a crucial disadvantage of this solution is the need of inverting the channel matrix, $\mathbf{H}$, at the transmitter which can cause to require impractical levels of transmitted power by $\mathbf{y}(k)$. Since power is a limited resource, the transmitted signal $\mathbf{y}(k)$ must be subjected to a power constraint condition. Therefore, including the power restriction, the optimization is reformulated as:

$$\min_{B\in\Re^{NxN}} \left\| I_U - C^T HBC \right\|_F^2 \qquad (8)$$

$$\text{subject to: } \left\|\mathbf{BC}\right\|_F^2 = U \ .$$

Where U is the required power of a non-precoding equivalent CDMA system with U users. One way of solving (8) is applying the Lagrange approach [15], resulting in the optimization process with the extended cost function

$$\hat{J} = \left\|\mathbf{I}_U - \mathbf{C}^T\mathbf{HBC}\right\|_F^2 + \lambda\left[\left\|\mathbf{BC}\right\|_F^2 - U\right] \qquad (9)$$

where λ is the Lagrange multiplier. The following propositions assure to find a solution to the minimization of $\hat{J}$.

Proposition 1. The choice of $\mathbf{B}$ that minimizes $\hat{J}$ under the power constraint is

$$\mathbf{B} = \left(\mathbf{H}^T\mathbf{CC}^T\mathbf{H} + \lambda\mathbf{I_U}\right)^{-1}\mathbf{H}^T \qquad (10)$$

with λ selected such that $tr\left(\mathbf{B}^T\mathbf{BCC}^T\right) = U$ is attained.

Proof.: By developing equation (9), we get

$$\hat{J} = tr\left(\mathbf{I}_U\right) - 2tr\left(\mathbf{CC}^T\mathbf{HB}\right) + tr\left(\mathbf{B}^T\mathbf{H}^T\mathbf{CC}^T\mathbf{HBCC}^T\right) + \lambda\left[tr\left(\mathbf{B}^T\mathbf{BCC}^T\right) - U\right]$$

taking the derivative of $\hat{J}$ with respect to $\mathbf{B}$ and λ and matching to zero results in:

$$\frac{\partial\hat{J}}{\partial\mathbf{B}} = -\mathbf{H}^T\mathbf{CC}^T + \mathbf{H}^T\mathbf{CC}^T\mathbf{HBCC}^T + \lambda\mathbf{BCC}^T = 0 \qquad (11)$$

and

$$\frac{\partial\hat{J}}{\partial\lambda} = tr\left(\mathbf{B}^T\mathbf{BCC}^T\right) - U = 0 \qquad (12)$$

From (11), it is straightforward to show that the solution to the constrained optimization problem is given by

$$\mathbf{B} = \left(\mathbf{H}^T\mathbf{CC}^T\mathbf{H} + \lambda\mathbf{I}_U\right)^{-1}\mathbf{H}^T. \qquad (13)$$

After substitution of $\mathbf{B}$ in (9), it can be verified that the optimal cost achieved is

$$\hat{J}_{opt} = \lambda^2\left\|\left(\lambda\mathbf{I}_U - \mathbf{C}^T\mathbf{HH}^T\mathbf{C}\right)^{-1}\right\|_F^2. \qquad (14)$$

Proposition 2. Let $\mathbf{B} = \left(\mathbf{H}^T\mathbf{CC}^T\mathbf{H} + \lambda\mathbf{I}_U\right)^{-1}\mathbf{H}^T \in \Re^{NxN}$. If $\hat{J}$ is the cost function of the constrained optimization problem, then there must exist a $\lambda>0$ such that (12) is always satisfied, provided that $tr\left(\mathbf{H}^T\mathbf{CC}^T\mathbf{H}\right) > 0$.

Proof.*:* Using proposition 1, the parameter λ is chosen subject to the constraint $\left\|\mathbf{BC}\right\|_F^2 = U$, therefore using (12) it is easy to see that

$$f\left(\hat{\lambda}\right) \equiv tr\left(\mathbf{C}^T\mathbf{H}\left(\hat{\lambda}\mathbf{I}_U + \mathbf{H}^T\mathbf{C}\mathbf{C}^T\mathbf{H}\right)^{-2}\mathbf{H}^T\mathbf{C}\right) = U.$$

Note that $f\left(\hat{\lambda}\right)$ is monotonically decreasing, we can conclude that if $f(0) > U$, then there will always exist a $\hat{\lambda} > 0$ which satisfies the power condition since $\lim_{\lambda\to\infty} f\left(\hat{\lambda}\right) = 0.$

Proposition 3. If orthogonal codes are employed, *i.e.* $\mathbf{C}^T\mathbf{C} = \mathbf{I}_U$, a second solution for $\mathbf{B}$ that satisfy (8) is given by

$$\mathbf{B} = \left(\mathbf{H}^T\mathbf{H} + \lambda\mathbf{I}_U\right)^{-1}\mathbf{H}^T \tag{15}$$

where the selection of λ is again carried out according to $tr\left(\mathbf{B}^T\mathbf{B}\mathbf{C}\mathbf{C}^T\right) = U$.

Proof.*:* Multiplying (11) on the left by $\left(\mathbf{H}^T\right)^{-1}$ we get

$$-\mathbf{C}\mathbf{C}^T + \mathbf{C}\mathbf{C}^T\mathbf{H}\mathbf{B}\mathbf{C}\mathbf{C}^T + \lambda\left(\mathbf{H}^{-1}\right)^T\mathbf{B}\mathbf{C}\mathbf{C}^T = 0$$

the above equation can be further simplified when multiplying by $\mathbf{C}^T$ on the left, and then by $\mathbf{C}$ on the right, resulting in the following expression

$$-\mathbf{I}_U + \mathbf{C}^T\mathbf{H}\mathbf{B}\mathbf{C} + \lambda\mathbf{C}^T\left(\mathbf{H}^{-1}\right)^T\mathbf{B}\mathbf{C} = 0$$

after rearranging and simplifying, an alternative solution to $\mathbf{B}$ is

$$\mathbf{B} = \left(\mathbf{H}^T\mathbf{H} + \lambda\mathbf{I}_U\right)^{-1}\mathbf{H}^T \tag{16}$$

The optimum error is given now by

$$\hat{J}_{opt} = \lambda^2\left\|\mathbf{C}^T\left(\mathbf{H}\mathbf{H}^T + \lambda\mathbf{I}_U\right)^{-1}\mathbf{C}\right\|_F^2 \tag{17}$$

Remarks: In general, the optimization problem posted in (8) is non-convex, and there are local solutions that satisfy the necessary conditions for optimality. Hence (10) and (15) are two solutions to (8), but the optimal costs are different, and as it will be shown in the next section, (10) represents the global optimum.

4 Simulation Results

In this section, the BER performance of the transmitter precoding scheme discussed in Section 3 is presented. The analytical and simulated performance of the proposed system, see Fig. 1, are also compared to the performance of a non-precoding DS-CDMA system using a conventional RAKE receiver. The simulated system is a BPSK modulated multi-path CDMA system as described in Section 2. The users data symbols are spread by codes of length $N = 16$, normalized such that $\left\|\mathbf{c}_u\right\| = 1$, and transmitted synchronously over a multi-path fading channel that follows the delays of the vehicular environment as described in [16]. The profile of the stationary channel used in these simulations, characterize as severe multi-path ($\alpha = 10$), is given in Table 1. The scalar α relates the length of the multi-path channel delay spread. It is also assumed that the transmitter has perfect knowledge of the fading coefficients.

In order to find the optimal precoding matrix $\mathbf{B}$ under the power constraint condition, a value of λ must be selected such that $g(\lambda) = tr\left(\mathbf{C}^T\mathbf{B}^T\mathbf{B}\mathbf{C}\right) - U = 0$. Finding the exact roots of the above function is complicated; therefore, we apply the standard Newton-Raphson method to obtain numerically the roots of $g(\lambda)$. The specific root that the process locates depends on the function, its derivative and an initial value, *i.e.* $\lambda_{n+1} = \lambda_n - g(\lambda_n)/g'(\lambda_n)$. To see how the Newton-

Raphson method performs, a graphical representation of the iterative process for a system with *U=10* users and employing Walsh codes is shown in Fig. 2. The iterative method is presented for both choices of $\mathbf{B}$ given in Section 3.1, *i.e.* equations (10) and (15). Considering an initial value of $\lambda = 0.01$, we find one root of $g(\lambda)$, using (10), in $\lambda = 0.1332$. Notice that only three iterations are required to converge to this value. Similarly, an optimal value of $\lambda = 0.1325$ is obtained for the second solution (using (15)). In this second case, a similar number of iterations are required (3 iterations).

After despreading (see (6)), the analytical BER performance for user *u* can be calculated via the *Q*-function as

$$BER = Q\left(\sqrt{\frac{\sigma_u^2}{\left(\rho_u + \sigma^2\right)}}\right) \tag{18}$$

Table 1. *Stationary multi-path fading channel profile.*

Delay in μs	Tap Coefficients
0.00	$h_0 = 0.7142$
0.26	$h_1 = -0.6149$
0.52	$h_2 = 0.0000$
0.78	$h_3 = 0.2337$
1.04	$h_4 = -0.2013$
1.30	$h_5 = 0.0000$
1.56	$h_6 = 0.0000$
1.82	$h_7 = 0.1009$
2.08	$h_8 = 0.0000$
2.34	$h_9 = 0.0000$
2.60	$h_{10} = -0.0801$

where σ_u^2 is the data bit variance at the output of the matched filter for user *u*, the first term in the denominator, ρ_u, is due to residual multiple access interference and the second term is the noise variance. If optimal precoding is considered, *i.e.* $\mathbf{B} = \mathbf{H}^{-1}$ (removing the power constraint conditionat the transmitter), then it is clear to see that $\sigma_u^2 = 1$ and $\rho_u = 0$, therefore we can calculate the BER as

$$BER = Q\left(\sqrt{\frac{1}{\sigma^2}}\right) \tag{19}$$

The application of the power constrained precoding gives the following performance

$$BER = Q\left(\sqrt{\frac{\gamma_{u,u}}{\left(\sum_{i=1,i\neq u}^{U}\left|\gamma_{u,i}\right| + \sigma^2\right)}}\right) \tag{20}$$

with $\gamma_{u,i}; u,i \in \{1,\cdots U\}$ as the (*u,i*)-th element of matrix $\mathbf{\Gamma} \in \Re^{UxU}$ defined by

$$\mathbf{\Gamma} = E\left[\tilde{\mathbf{d}}(k)\tilde{\mathbf{d}}(k)^T\right] = \mathbf{C}^T\mathbf{HBCC}^T\mathbf{B}^T\mathbf{H}^T\mathbf{C}$$

where $\tilde{\mathbf{d}}(k) = \mathbf{C}^T\mathbf{HBCd}(k)$ is noiseless data vector during the *k*-th time interval. In Fig. 3, the analytical and simulated *BER* performance averages over all users are reported as a function of *SNR*, defined as $SNR = 1/2\sigma^2$. Results are presented for a 10 user system employing Walsh codes of length $N = 16$. The analytical results are obtained from the BER expressions in (19) and (20). Expression (19) yields the theoretical single-user system performance, *i.e.* an interference free system, while (20) is used to yield the *BER* performance substituting $\mathbf{B}$ with both equations (10) and (15). Using the results in Fig. 2, the power constrained precoding performance is obtained by taking $\lambda = 0.1332$ and $\lambda = 0.1325$ depending on

whether (10) or (15) is considered. It is evident that there is a good match between the analytical and simulated performance for values of $SNR \le 10$ dB. For higher *SNR* values, an error is introduced between these curves. This level of disagreements is mainly due to the structure of matrix Γ which becomes predominant for high *SNR* values.

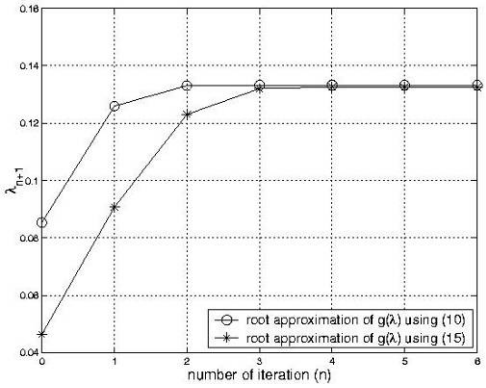


Fig. 2. *Number of iterations needed for Newton's method to converge for the function* $g(\lambda)$ *using (10) and (15) in a system with U=10 users.*

For comparison purposes, the optimal transmitter precoding (without power constraint) performance for *1* and *10* users are also presented. Notice that the case of unconstrained precoding manages to eliminate fully the interferences of the system, resembling the results of a typical AWGN channel. However, this performance is obtained at the expenses of increasing the transmitter power according to the inverse of matrix $\mathbf{H}$. For the previous example, it leads us to increasing the transmitted power to $\|\mathbf{BC}\|_F^2 = \|\mathbf{H}^{-1}\mathbf{C}\|_F^2 \approx 4U$. It is clear, therefore, that a performance penalty is associated with the constrained transmitter precoding relative to the unconstrained precoding. Although constrained precoding degrades its performance, as compare to the optimal precoding, for a *10* users system it gets about *2* dB close for a $BER = 10^{-3}$. Fig. 3 also contains the *BER* performance of a non-precoding DS-CDMA system using a conventional RAKE detector, a system with *U=1* and *5* users are only considered. These results demonstrate the gain achieved by a transmit pre-processing system when comparing to a non-precoding system with RAKE receiver. Recall that a simple receiver is considered in the proposed system, and therefore no channel estimation, adaptive equalizer or feedback from the base station is required.

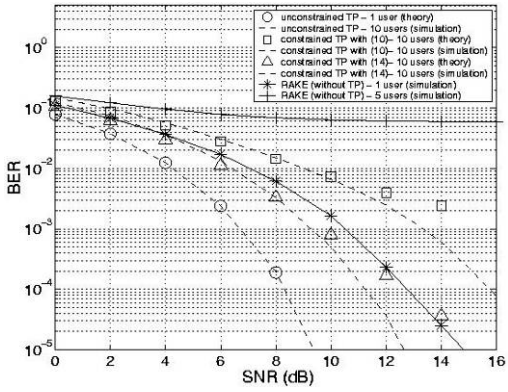


Fig. 3. *Analytical and simulated performance results for power constrained precoding. The non-precoding DS-CDMA system using a RAKE receiver is also presented.*

On the other hand, if we were to incorporate a cyclic prefix for each symbol to combat multi-path, a reduction in the overall data rate is yielded. To mitigate this penalty, a variation to the proposed precoding scheme can be implemented. To

improve the system throughput, a block transmission strategy can then be performed in a similar way as presented in [17] but where it is assumed that the channel is invariant within the block transmission, which may not be true for larger blocks.

5 Conclusions

The main objective of transmit pre-processing scheme presented here was to find the transformation precoding matrix, $\mathbf{T}$, which minimizes in the mean squared error sense the BER at the mobile receivers. Performance evaluations showed that the transmit pre-filtering without power constraint eliminate fully the channel interferences. However, it is achieved by allowing an undesired increase in transmission power. To ensure that the transmitter power with precoding is the same as that without precoding, two precoding solutions were analyzed that satisfy this condition. Simulation results for power constrained precoding were presented and compared with analytical simulation results. These results indicate that the precoding technique is able to mitigate a significant amount of the system's interferences using a simple matched filter at the receiver and under the power restriction at the transmitter. In particular, we observed that for a *10* users system, using the proposed constrained precoding scheme, the performance gets close about 2dB from the optimal case (unconstrained precoding) for a $BER = 10^{-3}$. The fact of using an overhead (cyclic prefix) for each transmitted symbol implies less bandwidth efficiency and is a drawback of this technique, but the simplicity of the Matched Filter receiver and the performance gain with this transmission technique make it desirable for many wireless applications. In addition, to reduce the penalty of incorporating a CP for every symbol transmitted, a block transmission strategy can be performed to improve the system throughput. These results motivate the use of linear precoding techniques for the forward link of TDD/DS-CDMA systems.

Acknowledgments

This research work was supported by Grant PROMEP/103.5/04/1386.

References

1. Moher M.: An Iterative Multi-user Decoder for Near-capacity Communications. IEEE Transactions on Communications, 46(7), 870-880, July 1998.
2. Vojčić B.R., Jang W.M.: Transmitter Precoding in Synchronous Multi-user Communications. IEEE Trans. on Comm., 46(10), 1346-1355, October 1998.
3. Fischer R.H.: Precoding and Signal Shaping for Digital Trans.. Wiley & Sons, NY 2002.
4. Meurer M., Baier P.W., Weber T., Lu Y., Papathanassiou: Joint transmission: advantageous downlink concept for CDMA mobile radio systems using time division duplexing. Electronics Letters, 36(10), 900-901, May 2000.
5. Joham M., Irmer R., Berger S., Fettweis G., and Utschick W.: Linear Precoding Approaches for the TDD DS-CDMA Downlink. The 6th International Symposium on Wireless Personal Multimedia Communications, vol. 3, 323-327, October 2003.
6. Brandt-Pearce M., and Dharap A.: Transmitter-based multiuser interference rejection for the down-link of a wireless CDMA system in a multipath environment. IEEE Journal on Selected Areas on Communications, 18(3), 407-417, March 2000.
7. Harashima H., and Miyakawa H.: Matched-transmission technique for channels with intersymbol interference. IEEE Transactions on Communications, 774-780, August 1972.
8. Tomlinson H. M.: New Automatic Equalizer Employing Module Arithmetic. Electronics Letters, vol. 7, no. 5/6, 38-139, March 1971.
9. Noll Barreto A., and Fettweis G.: Joint Signal Precoding in the Downlink of Spread Spectrum Systems. IEEE Trans. on Wireless Comm., vol. 2, no. 3, 511-518, May 2003.
10. Zerlin B., Joham M., Utschinck W., and Nossek J. A.: Covariance-Based Linear Precoding. IEEE Journal on Selected Areas in Communications, vol. 24, no. 1, January 2006.
11. Choi L., Murch R. D.: Transmit-Preprocessing Techniques with simplified Receivers for the Downlink of MISO TDD-CDMA Systems. IEEE Trans. Veh. Tech., 53(2), March 2004.
12. Dietrich F., Hunger R., Johan M., and Utschick W.: Linear Precoding over Time-Varying Channels in TDD Systems. Proc. ICASSP, vol. V, 117-120, April 2003.
13. Hons H.S., Khandani A.K., and Tong W.: An Optimized Transmitter Precoding Scheme for Synchronous DS-CDMA. IEEE Transactions on Communications, 54(1), January 2006.
14. Wang Z., and Giannakis G.B.: Wireless Multicarrier Communications. IEEE Signal Processing Magazine, 29-48, May 2000.
15. Nocedal J., Wright S.J.. Numerical Optimization. Springer Series in Operations Research, Springer-Verlag 1999.
16. Selection Procedures for the Choice of Radio Transmission Technology of UMTS. (UMTS 30.03 Version 3.2.0).
17. Liu Z. and Giannakis G.B.: Space-Time Block Coded Multiple Access through Frequency-Selective Fading Channels. IEEE Trans. on Comm., 49(6), pp. 1033-1044, June 2001.

Chapter 3

Review Of Frequency Selectivity Parameters For Broadband Wireless Signals

Victor Manuel Hinostroza Zubía

Universidad Autónoma de Ciudad Juárez
Departamento de Ingeniería Eléctrica y Computación
Cd. Juárez, Chihuahua, México C.P. 32310. e-mail: vhinostr@uacj.mx

Abstract

A version of a FMCW (chirp) wideband channel sounder is presented; the transmitted bandwidth can go up to 300 MHz, the repetition frequency of the chirp can go up to 100 and most of the parameters are programmable. The basic block diagram of the sounder is discussed and its important parts are reviewed. Performance specifications of the system are highlighted. The second part of this work is to evaluate the contribution of several related parameters; frequency selective fading, coherence bandwidth and delay spread on the frequency selectivity of the channel. A description of the sounder parameters and the sounded environments are given. The 300 MHz bandwidth is divided in segments as small as 60 kHz to perform the evaluation of frequency selective fading. Sub channels of 20 MHz for OFDM systems and 5 MHz for WCDMA were evaluated. Graphics are provided for a number of bands, parameters and locations in the three different environments. It is also shown the variation of the signal level due to frequency selective fading. The practical assumptions about the coherence bandwidth and delay spread are reviewed and a comparison is made with actual measurements. Statistical analysis was performed over some of the results.

Keywords: Channel characterization, Coherence bandwidth, frequency correlation, frequency selective fading, impulse response and multi-carrier modulation.

1 Introduction

The high data rate that current mobile radio systems use, prompts the need to have a more accurate characterisation of the channel in this kind of environment. The time and frequency dispersion require characterising the environment with a sounder that have distinct features such as high resolution in both time delay and Doppler spread and it should have the ability to match very different testing conditions and it should fit in very different environments regarding space, area and levels of noise. All those requirements cast the specification of the system. These specifications should have the following elements: 1) the time delay resolution should be very fine, in the order of the few nanoseconds. 2) The Doppler spread resolution should be also fine, in the order of a few Hertz. 3) The dynamic range should be high to measure long variation of the channel multipath fading signal. 4) The sensitivity of the system should be high, to match the long variation in signal strength due to the variable architecture of the environments, it should covers several decades in dBs of signal strength deviation.

Taking in consideration the aforementioned specifications a sounder system has been designed, built, tested and several sets of measurements have been carried out with it. The construction of the system emphasise the use of a Direct Digital Frequency Synthesiser (DDFS), this DDFS have a frequency range of up to 400 MHz. To check the specifications of the system, several tests had been carried out: the ambiguity function, the back-to-back test and the two tones test had been conducted and some of these results are shown and explained. From the performed measurements, analysis and data concentration had been carried out, the analysis gave as results the following parameters for each environment: Power delay profile, received signal strength, average delay spread, RMS delay spread, Time variant transfer function, CDFs of RMS delay spread, frequency correlation, coherence bandwidth and CDFs fitting with several probability distributions i.e. Rayleigh, Rician, Lognormal, etc.

The limitations of high data rates in wireless communications are not limited only by noise. As the data rate increases the limitations comes from the Inter Symbol Interference (ISI) due to the dispersive characteristics of the wireless communications channel. The dispersive channel characteristics come from the different propagation paths, multipath, between the receiver and the transmitter. This dispersion can be measured, if we measure the channel impulse response (CIR). As a general rule the effects of ISI on the transmission errors is negligible if the delay spread is significantly shorter than the

duration of the transmitted symbol. Due to the expected increase in demand of higher data rates, wideband multi-carrier systems such as: OFDM and WCDMA are expected to be technologies of choice [1], [11] and [13]. This is because these technologies can provide both, high data rates and an acceptable level of quality of service. However, in order to fulfil those expectations these technologies have to deal with several problems, such as; system scalability, asymmetrical services, high speed data rate coverage, uniform distribution, radio resource management and channel prediction. Of specific interest is the problem of channel prediction or estimation, because this condition, as it was explained before is the main boundary for higher data rates. The study of correlation of the mobile radio channel in frequency and time domains has helped to understand the problem of channel estimation. However, few studies have undertaken the approach of analyzing frequency selective fading as is intended to do in this study. Former works [9], [10], [11], [12] study this problem through analytical approaches (simulations) were some parameters are assumed, in this work actual measured data is used. This work's main interest is frequency selective fading (FSF) estimate and quantification in several environments. This work starts with the result of measurements made with a sounder that uses the chirp technique for sounding. The sounder characteristics are described in separate works. The measurements were made with a 300 MHz chirp signal at a frequency carrier of 2.35 GHz. The sounder was designed and built at UMIST Manchester UK [2] [3]. In part II the theoretical foundations of the channel impulse response are reviewed. Also in this part, the characteristics and dimensions of the three environments sounded are described. In part III, the frequency selective fading evaluation and analysis are reviewed. Plots of the dependency of fading deep and frequency separation of two specific points in the response are studied. At the end conclusions and future work are mentioned.

2 The wideband channel model

The radio propagation channel is normally represented in terms of a time-varying linear filter, with complex low-pass impulse response, h(t, τ). Its time-varying low-pass transfer function is [4] [6]

$$H(t,f)=\int_{-\infty}^{\infty} h(t;\tau)e^{-j2\pi f_\tau}d\tau \tag{1}$$

Where τ represents delay. Using (1) the frequency correlation function for the channel can be written as:

$$E\left\{H\left(t_1;f_1\right)*H\left(t_2;f_2\right)\right\}= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} E\left\{h\left(t_1;\tau_1\right)*h\left(t_2;\tau_2\right)\right\}e^{-j2\pi f_1\tau_1}e^{+j2\pi f_2\tau_2}d\tau_1 d\tau_2 \tag{2}$$

By considering the channel to have uncorrelated scattering (US) and to be wide sense stationary (WSS), the subscript for τ is eliminated and f1 and f2 can be replaced by f + Δf and t1 and t2 replaced by t + Δt, then:

$$R_H(\Delta t;\Delta f)=\int_{-\infty}^{\infty} R_h(\Delta t;\tau)e^{-j2\pi\Delta f\tau}d\tau \tag{3}$$

In (3) RH and Rh represents the correlation of random variations in the channels transfer function and its impulse response respectively. If there are US, then Δt is 0 then:

$$R_h(0;\tau)=E\left\{\left|h(0;\tau)\right|^2\right\}=E\left\{\left|h(\tau)\right|^2\right\} \tag{4}$$

substituting into (3) gives:

$$R_H(\Delta f)=\int_{-\infty}^{\infty} E\left\{\left|h(\tau)\right|^2\right\}e^{-j2\pi\Delta f\tau}d\tau \tag{5}$$

where $E\left\{\left|h(\tau\right|^2\right\}$ is the average PDP of the channel. So, under the above conditions, RH is the Fourier transform of the average PDP. Calculations of frequency selective fading in this work are based in these assumptions.

2.1 Coherence Bandwidth

The multipath effect of the channel, the arrival of different signals in different time delays causes that the statistical properties of two signals of different frequencies to become independent if the frequency separation is large enough. The maximum frequency separation for which the signals are still strongly correlated is called coherence bandwidth (B_c). Besides to contribute to the understanding of the channel, the coherence bandwidth is useful in evaluating the performance and limitations of different modulations and diversity models.

The coherence bandwidth of a fading channel is probed by sending two sinusoids, separated in frequency by $\Delta f = f_1 - f_2$ Hz, through the channel. The coherence bandwidth is defined as Δf, over which the cross correlation coefficient between $r1$ and $r2$ is greater than a preset threshold, say, $\eta_0 = 0.9$. Namely:

$$C_{r1,r2} = \frac{\operatorname{cov}(r1, r2)}{\sqrt{\operatorname{var}(r1)\operatorname{var}(r2)}} = \eta_0 \tag{6}$$

Then, using (2)

$$R(s,\tau) = \langle r1 r2 \rangle = \int_0^\infty \int_0^\infty r1 r2 \, p(r1, r2) dr_1 dr_2 \tag{7}$$

Where $p(r1,r2)$ is

$$\begin{aligned} p(r1, r2) &= \int_0^{2\pi} \int_0^{2\pi} p(r1, r2, \theta_1, \theta_2) d\theta_1 d\theta_2 \\ &= \frac{r1 r2}{\mu^2 (1-\lambda^2)} \exp\left[-\frac{r_1^2 + r_2^2}{2\mu(1-\lambda^2)} \right] I_0\left(\frac{r1 r2}{\mu} \frac{\lambda}{1-\lambda^2} \right) \end{aligned} \tag{8}$$

Where $I_0(x)$ is the modified Bessel function of zero order. Then, substituting (8) in (7) and integrating

$$R(s,\tau) = \frac{\pi}{2} b_0 F\left(-\frac{1}{2}, -\frac{1}{2}; 1; \lambda^2\right) \tag{9}$$

which may also be expressed as

$$R(s,\tau) = \frac{\pi}{2} b_0 \left(1 + \frac{\lambda^2}{4} \right) \tag{10}$$

$$\rho(s,\tau) = \frac{R(s,\tau) - \langle r1 \rangle \langle r2 \rangle}{\sqrt{\left[\langle r_1^2 \rangle - \langle r1 \rangle^2 \right] \left[\langle r_2^2 \rangle - \langle r2 \rangle^2 \right]}}$$

$$R(s,\tau) = b_0 (1+\lambda) E\left(\frac{2\sqrt{\lambda}}{1+\lambda} \right) \tag{11}$$

Where $E(x)$ is the complete elliptic integral of the second kind. The expansion of the hyper geometric function gives a good approximation to (9). After several reductions and considerations, the correlation coefficient becomes

$$\rho(s,\tau) = \frac{(1+\lambda) E\left(\frac{2\sqrt{\lambda}}{1+\lambda} \right) - \frac{\pi}{2}}{2 - \frac{\pi}{2}} = \lambda^2 = \frac{J_0^2(\omega_m \tau)}{1 + s^2 \sigma^2} \tag{12}$$

It is possible to see in this expression that the correlation decreases with frequency separation. This formula has been substituted by several practical expressions some of them are the following [4], [8], [9], [10].

$$B_{C=0.9} = \frac{1}{50\sigma_{rms}} \tag{13}$$

$$B_{C=0.5} = \frac{1}{5\sigma_{rms}} \tag{14}$$

$$B_{C=0.9} = \frac{1}{8\sigma_{mean}} \tag{15}$$

$$B_C = \frac{1}{2\pi\sigma_{rms}} \tag{16}$$

In general;

$$B_C = \frac{k}{\sigma_{rms}} \tag{17}$$

It will be shown, comparing with practical measurements that none of these expressions are accurate and it is difficult to obtain a comprehensive expression for all environments.

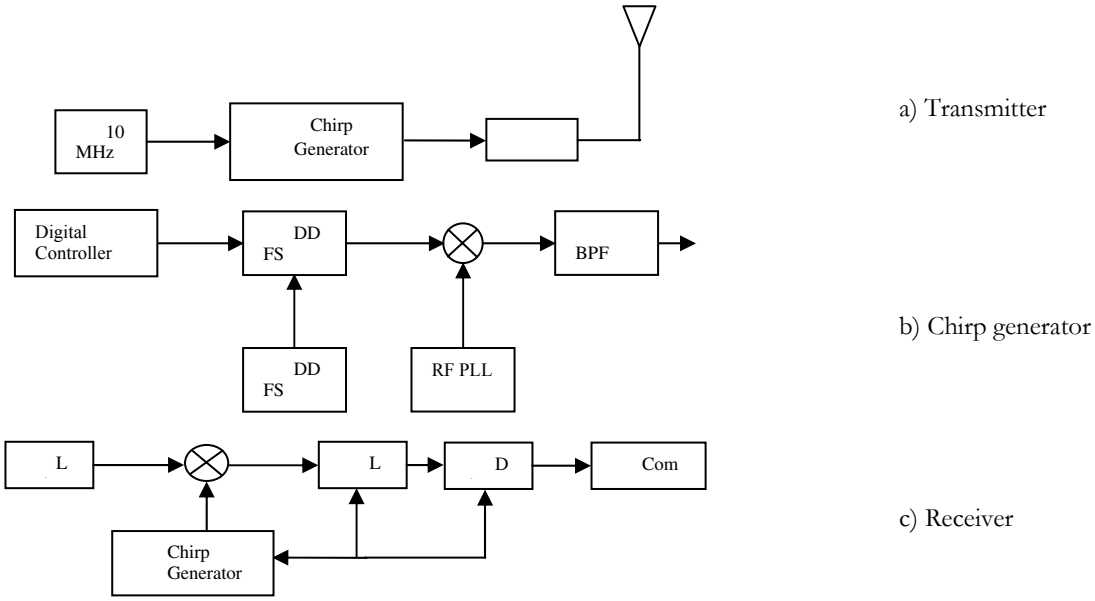


Fig. 1. *System block diagram*

3 Sounder Systems Characteristics and Environment Description

This sounder uses the FMCW or chirp technique. The generated chirp consists of a linearly frequency modulated signal with a bandwidth of 300 MHz and a carrier frequency of 2.35 GHz. The chirp repetition frequency is 100 Hertz, which allows having 50-Hertz Doppler range. The receiver has the same architecture as the transmitter. Nevertheless, in the receiver, the generated chirp is not transmitted but mixed with the incoming signal from the antenna, which are the multi-path components of the transmitted chirp. This mixing allows having the multi-path components at low frequencies, these low frequencies can be sampled, digitized and stored in a computer to perform the required analysis.

3.1 System Description

Figures 1 show the block diagram of the sounder system, the figure shows; a) the transmitter, c) the receiver and the common subsystem to both, b) the chirp generator. Figure 1a, shows the basic architecture of the transmitter, its three main parts are; a frequency reference of 10 MHz, a chirp generator and the RF part. A high stability Rubidium clock forms the frequency reference; a similar clock is used in the receiver. The chirp generator is shown in figure 1b, its main parts are a digital controller board, two PLL synthesisers, a mixer, a band pass filter and the DDFS. The DDFS requires as an input a 30 bits binary value, which finally will become the frequency at the output of the synthesiser, the digital controller provides these 30 bits binary values at the required rate for the DDFS. The two PLL synthesisers; one is used as the clock of the DDFS, this clock is set to 1.6 GHz. The other PLL is use as the RF carrier frequency, in this case to 2.35 GHz. The RF carrier and the output of the DDFS are mixed and single-sideband filtered. So, the output of the chirp generator will be the output of the DDFS sweep up converted to the RF carrier frequency. The RF part of the transmitter is formed from two amplifiers, one low level amplifier and a power amplifier, both amplifiers gets the output of the chirp generator to a suitable level for transmission. Figure 1c, shows the receiver block diagram, here it is possible to identify the following parts; a LNA, a mixer, a low pass filter and the data acquisition subsystem. Since the signals received through the channel arrive to the receiver with low level a low noise amplifier is required to the input of the receiver. After amplification the signal is mixed with a replica of the transmitted chirp, the output of the mixer are the down converted delayed replicas of the transmitted signal due to the multipath effect of the channel. These low frequency replicas called "*beats*" are filtered, digitised and stored in a personal computer for subsequent analysis.

3.2 Performance tests

Several performance tests had been conducted to make certain that the system achieves the expected specifications. The back-to-back test, two-tone test and the ambiguity function had been carried out on the system and the result is presented in figures 2. Figure 2a show the RF spectrum of the transmitter, here it can be highlighted that the flatness of the spectrum over the 300 MHz bandwidth is less than 0.5 dB. Figure 2b show the ambiguity function. In figure 2d is it possible to see that the Doppler resolution of the system is ± 1 Hertz and the Time Delay resolution is ± 10 nanoseconds up to 30 dBs w.r.t. the peak. Figure 2c shows a typical channel transfer function. Figures 23 and 2f show examples of RMS delay spread measurements and its CDF.

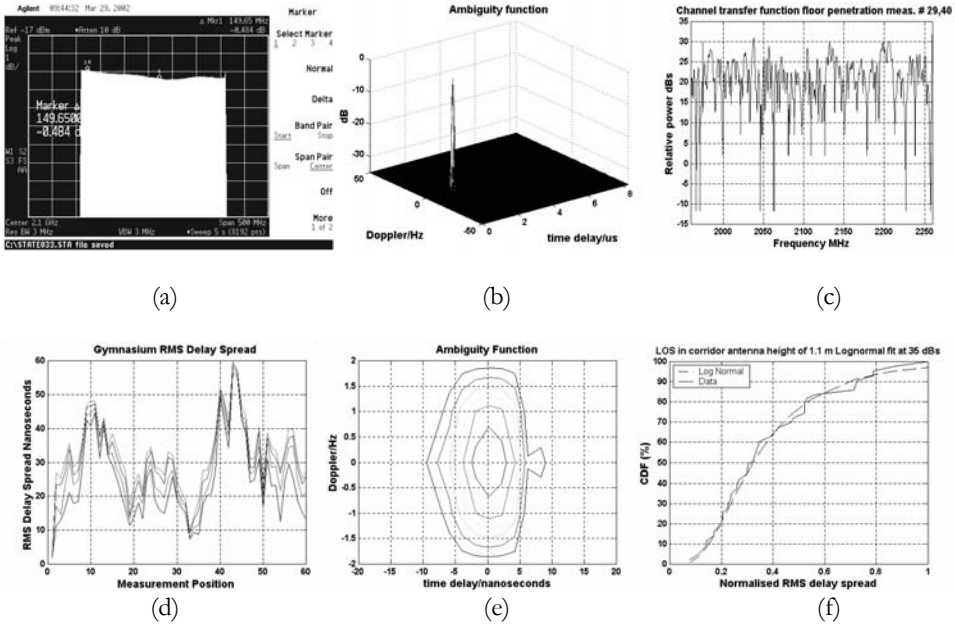


Fig. 2. *Performance tests.*

3.3 Environment description

To perform the evaluation of frequency selective fading, three environments where the measurements took place were the following: 1) A large laboratory, full of big tools and high power machinery. This environment has 612 square meters area and a 25 m height glassed ceiling, the area is divided into three stories at different levels. The transmitter was static in the first floor and the receiver was moved around the three floors at several locations around the environment. 2) Around a floor inside a building. This floor form a rectangle with four long corridors; two 66 m long and two 86 m long, the width of the corridors is 3 m, the total covered area was 912 square meters and the ceiling height is 5 meters. The transmitter was static and the receiver was moved around the corridors measurements were taken at specific distances in each corridor. 3) From building to building. These two buildings are eight stories high, they have the about the same high and they are separated by about 200 meters. The transmitter was located on top of one of the buildings and the receiver was moved around specific locations inside the second building in all eight different floors. 4) An outdoors environment a city center of a modern city, with large buildings. Each location in each environment was sampled during one second and 100 impulse responses were stored for that specific location. When the measurements were taken, every location was sampled with 100 impulse responses, that is to say, an impulse response (IR) was taken on that specific location every 10 milliseconds. The velocity of the receiver was about 2 m/s (fast walking speed.)

4 Frequency Selective Fading

To carry out the evaluation of frequency selective fading, each processed IR was split in sections of 60 kHz, which was the minimum sampled frequency, each sweep of 300 MHz bandwidth was sampled 5000 times every 10 milliseconds, a sample was taken every 2 microseconds. Each IR was averaged over one second; every frequency was averaged 100 times for each IR. The level of the frequency response of each 60 kHz point was calculated and recorded.

Calculations similar to the one mentioned in the previous paragraph, were done on the three environment measured, the fading characteristics in each of the environments were calculated. Each environment was sounded in about 50 different locations and 100 IRs were taken in each location. On each of the locations the fading characteristics were calculated; average delay spread, RMS delay spread, coherence bandwidth, channel transfer function and frequency correlation. Using the channel transfer function for each IR, specific fading characteristics for sub channels of 5 and 20 MHz were calculated. Figure 3 shows the fading characteristics for a 20 MHz sub channel, in this figure there are 15 different lines, each line corresponds to a 20 MHz sub channel in a 300 MHz bandwidth. To get this figure, the following tasks were performed; first, the channel response information of the complete 300 MHz bandwidth was divided in 15 sub-channels of 20 MHz each. Then, each sample that represents 60 kHz, was measured and the result was compared to the next sample, then the next sample was compared and so on up to the complete 20 MHz bandwidth was compared. After that, in a separate 20 MHz sub channel the same procedure was applied and so on up to the end of the 300 MHz bandwidth. To form figure 4, the same procedure was followed, but in this case, the sub channel bandwidth was only 5 MHz, then this figure has 60 different lines, which come from the 300 MHz bandwidth divided between 5 MHz segments.

Figure 3 shows that the maximum fading deep within the 20 MHz sub channel is lower than 14 dB in all the sub channels. On the other hand, in the 5 MHz bandwidth the maximum fading deep was of 18 dB. It is possible to see in figure 5, that most of the lines follows a pattern, which means that the fading in all sub channels is about the same. In figure 3, most of the lines stay below 5 dB and only a few lines go higher than 6 dB; this means that deep fades are rare.

Figure 5 shows the calculations of fading characteristics for the second environment with 15 different 20 MHz sub channels, this figure shows that the lines are more dispersed, which means that there are deeper fading in the responses of the IR. Since this environment is the measurement of the propagation of the signal in different floors in a building, higher delay spread, i.e. time dispersion, than the former environment was expected and therefore more fading. Figure 6 shows the fading characteristics for the same environment, in this case, the sub channel bandwidth was 5 MHZ, it can be seen in figure 5 that most of the fading stay below 10 dB, which is higher than that for the previous environment and the maximum values, are also higher. This value indicates that the variation of the signal level, it is expected to have deep and frequent fades.

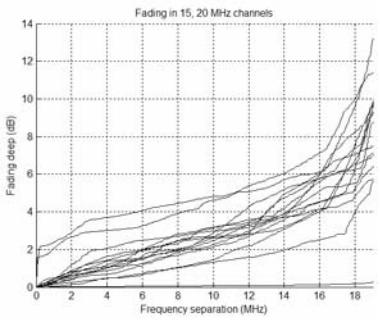


Fig. 3. *Laboratory environment fading for a 20 MHz channel.*

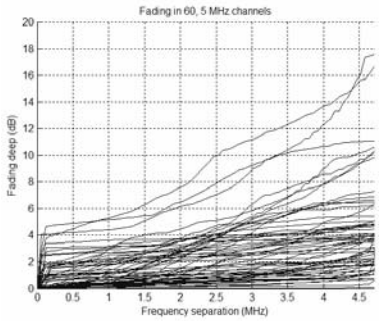


Fig. 4. *Laboratory environment fading for a 5 MHz sub channel.*

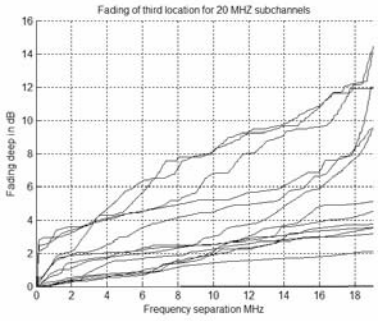


Fig. 5. *Floor environment fading for a 20 MHz sub channel.*

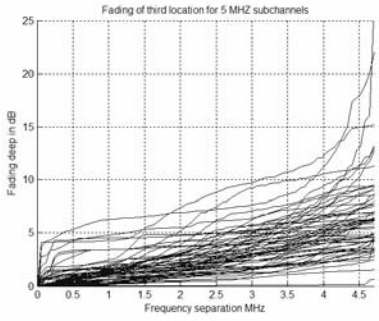


Fig. 6. *Floor environment fading for a 5 MHz sub channel.*

Other parameters that could be extracting from the fading behaviour of the environments measured are the statistical performance over the range of locations measured. The mean and standard deviation of the fading characteristics were calculated. To calculate the mean of all locations in this environment 100 IR for each location in a 20 MHz bandwidth were taken into account. Figure 5 shows the mean for each location of about 50 different locations in the building-to-building environment. Figure 5 shows that the average of fades in all the locations in the environment is not higher than 7 dB. Figure 6 shows the standard deviation for all locations for the same environment. In figure 6, it is possible to see that the variation of the fades will be within 8 dB most of the time.

Another way to look at the statistics of the fading is to calculate the CDF of this parameter. To make the calculations of these CDF's figures, the mean of all locations in the environment is used. Figure 7 shows the CDF of the building-to-building environment for both, the 20 and 5 MHz bandwidths, figure 7 shows that the fading deep for a 20 MHz sub channel is below 7 dB for 90% of the times. On the other hand, for the 5 MHz sub channel, the fades are below 5 dB for 90% of the times. Figure 8 shows the CDF's for the large laboratory environment; here both sub channels bandwidths; 5 and 20 MHz, are below 5 dB most of the times. Even though, the 5 MHz sub channel is below 3 dB for 90% of the times.

Multipath fading channels are usually classified into flat fading and frequency selective fading according to their coherence bandwidth relative to that of the transmitted signal. Coherence bandwidth is defined as the range of frequencies over which two frequency components remain in a strong amplitude correlation. Physically, it defines the range of frequencies over which the channel can be considered "flat". The analytic issue of coherence bandwidth was first studied by Jakes [1] where by assuming homogeneous scattering, his work revealed that the coherence bandwidth of a wireless channel is inversely proportional to its root-mean-square (rms) delay spread. The same issue was subsequently studied by various authors [4], [8], [9], [10]. Since many practical channel environments can significantly deviate from the homogeneous assumption, various measurements were conducted to determine multipath delay profiles and coherence bandwidths [19], [20], [21], [22], aiming to obtain a more general formula for coherence bandwidth. In this work the variations of this formula are reviewed and compared with actual results and a comparison is provided.

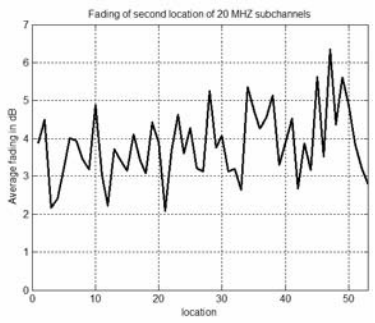


Fig. 7. *Mean fading deep for 20 MHz Sub Channel.*

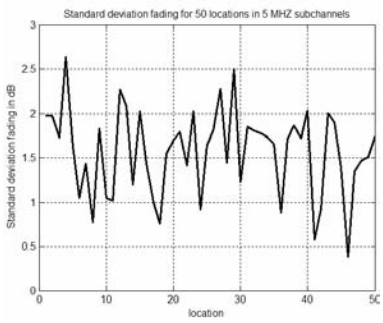


Fig. 8. *Standard deviation for 5 MHz Sub channel.*

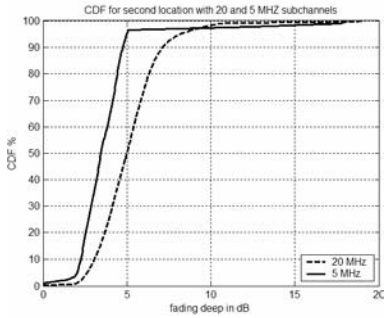


Fig. 9. *Fading CDF for building-to-building.*

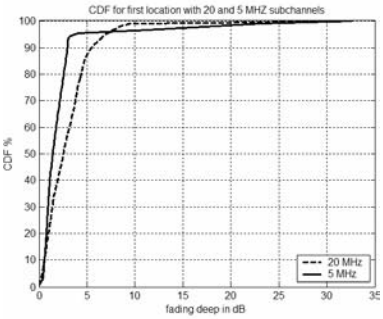


Fig. 10. *Fading CDF for large laboratory.*

5 Coherence Bandwidth Evaluation

Figure 11 shows the frequency correlation of all locations in the indoors environment. To make this figure the following was done; first the Power Delay Profile (PDP) of all locations was calculated. Then a Fourier transform was performed on the PDP, which gave us the frequency correlation for all locations. Then the frequency correlation for each location was plotted in figure 11. On this figure, the thick and dashed line is the line for the maximum coherence bandwidth, when the transmitter and receiver are connected directly. In figure 11, one can see that at 0.9 correlation coefficient, the coherence bandwidth (B_c) is lower than 10 MHz most of the locations. This is corroborated in figure 12, this figure shows the average B_c for all locations in the indoors environment. Figure 13, shows the RMS delay spread for all locations for the same environment. Quick calculations comparing figure 11 results and expression (13) show that, few calculated values of the versions of expression (13) match with the measured values of figure 13.

Figure 14 shows the frequency correlation for the outdoor to indoor environment, this figure shows that in this environment the B_c at frequency correlation of 0.9 is higher than the indoor environment, although the delay spread is not different is both environments. Figure 15, shows the average B_c for the outdoors to indoors environment, we can see in this figure, that the coherence bandwidth is higher than the indoor environment, which was expected, but the difference is higher than expected. In indoors the coherence bandwidth is not bigger than 20 MHz in average. In the other hand, in the outdoor to indoor, the average is about 100 MHz, here is relation of 5 to 1. The difference in RMS delay spread is 100 nS versus 200 nS, there is a relation of 2 to 1.

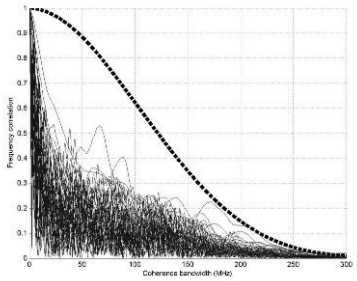


Fig. 11. *Coherence bandwidth for indoors.*

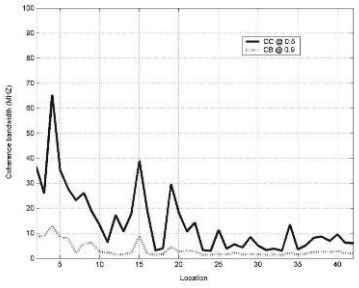


Fig. 12. *Average of coherence bandwidth for indoors.*

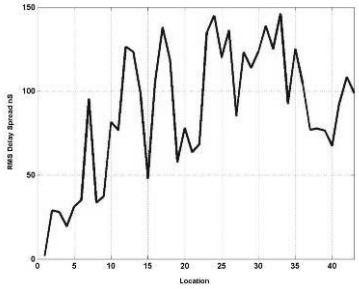


Fig. 13. *RMS delay spread for indoors.*

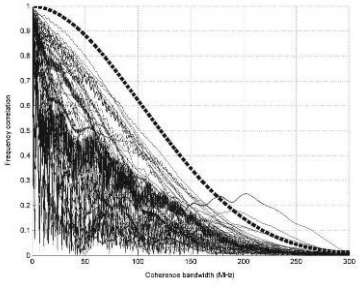


Fig. 14. *Coherence bandwidth for outdoor to indoor.*

Figure 17 shows the frequency correlation for the outdoors environment. Figure 18, shows the B_c at frequency correlation of 0.9. In this case the B_c can not be compared to the B_c for the other two environments, since in this environment a lower bandwidth is evaluated, 120 MHz instead of 300 MHz. Despite this difference and observing figures 17 and 18, B_c is not significantly lower even when we have higher distances and higher delay spread. In outdoors the B_c is not bigger than 2 MHz in average. In the other hand, the RMS delay spread is 1.5 μS in average.

Table 1, shows the comparisons of B_c for the three environments with the different versions of expressions 13 -16 and measured results. This table shows that the values of the expressions are always lower than the measured results, which induce to conclude that the expressions were underestimated, at least in these environments. Moreover, it is possible to conclude that these expressions were deduced with not enough measured results. Also, table1 shows that the relationship between delay spread and coherence bandwidth, not necessarily is a single constant.

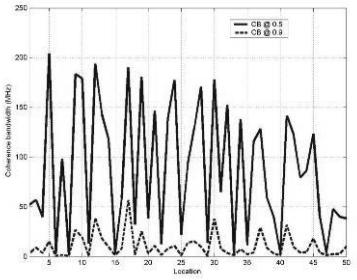


Fig. 15. *Average of coherence bandwidth for outdoors to indoors.*

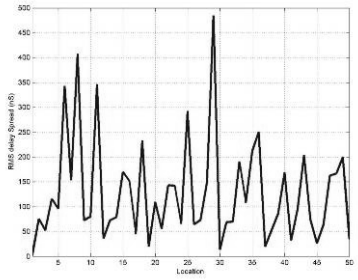


Fig. 16. *RMS delay spread for indoor to outdoor.*

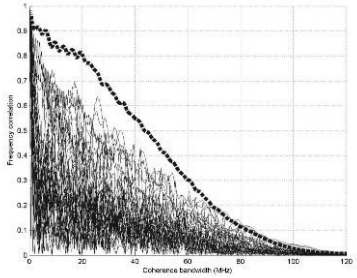


Fig. 17. *Coherence bandwidth for outdoors.*

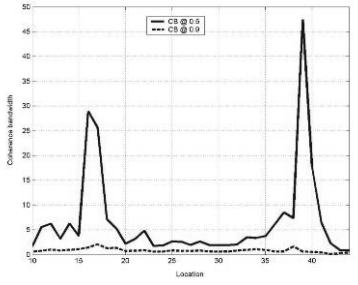


Fig. 18. *Average of coherence bandwidth for outdoors.*

6 Conclusions

A system for Indoor environment channel characterisation was presented, its functionality was discussed and the appropriate block diagrams were explained, several performance test were presented and these test show that the intended specifications of the systems were achieved; the instantaneous dynamic range is 35 dBs or better, the Doppler resolution is about 1 Hertz, there are no inter-modulation products and the spectrum flatness is better that 0.5 dBs. The actual measurements show the time delay resolution required (3.3 nanoseconds). Summarising the system works at the expected specifications or better. The measurements and analysis showed that the system works as expected. The results of the analysis show good consistency with previous works in similar environments. Also, in this work the results of analysis of frequency selective fading on two indoor and one outdoor environment have been presented. The three environments analyzed demonstrate that the fading is within specific limits, these results could help to the designers of adaptive receivers to estimate the channel more accurately. The division of the channel impulse bandwidth in segments of 20 and 5 MHz bandwidths, allow the calculation of fading in the bandwidth of interest for OFDM and WCDMA transmission. Plots of the frequency selective fading will help for this assessment. The analysis of coherence bandwidth show that the expressions accepted in the literature for its calculation are not accurate and the accepted direct relationship between delay spread and coherence bandwidth is not simple. Also, additional work is require on try to determine how much the combined effect of Doppler spread, time variability and frequency offsets affects the transmission on multi-carrier signals as the ones on OFDM y CDMA.

Table 1. *Coherence bandwidth calculations*

Value from	Indoors	Outdoors to indoors	Outdoors
(13)	400 kHz	200 kHz	30 kHz
(14)	4 MHz	2 MHz	300 kHz
(15)	3.3 MHz	2.5 MHz	250 kHz
(16)	3.2 MHz	1.6 MHz	212 kHz
$\text{Measured}_{0.9}$	5.3 MHz	12 MHz	300 kHz
$\text{Measured}_{0.5}$	19 MHz	72 MHz	5.6 MHz

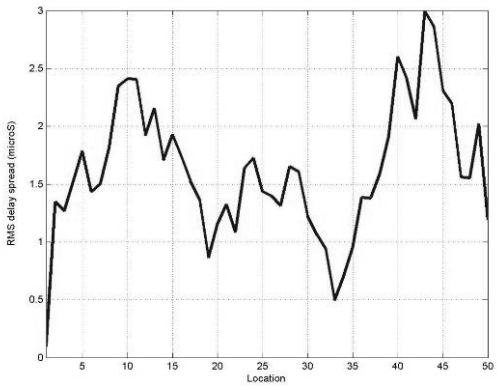


Fig. 19. *RMS delay spread for outdoors*

References

1. Jakes W. C., Microwave mobile communications, (Wiley, 1974).
2. Aurelian B, Gessler F, Queseth O, Stridh R, Unbehaun M, Wu J, Zander J, Flament M. "*4th-Generation Wireless Infrastructures: Scenarios and Research Challenges*", IEEE Personal Communications Magazine, 8(6), 25-31, Dec 2001.
3. Salous S, Hinostroza V," *Bi-dynamic indoor measurements with high resolution sounder"*, 5th. International Symposium on wireless multi-media Communications, Honolulu Hawaii USA, October 2002.
4. Golkap H., " Characterization of UMTS FDD channels ", *PhD Thesis*, Department of Electrical Engineering and Electronics, UMIST, UK 2002
5. Lee W. C. Y., *Mobile Communication Engineering*, (McGraw-Hill, 1998).
6. Bello P.A., "Characterization of randomly time-variant linear channels", *IEEE Transactions on Communications Systems*, December 1963, pp. 360-393.
7. Hehn T., Schober R.m and Gerstacker W., "Optimized Delay Diversity for Frequency Selective Fading Channels", *IEEE Transaction on Wireless communications*, September 2005, Vol. 4, No. 5, pp. 2289-2298.
8. Hashemi H., "The indoor radio propagation channel", *IEEE Proceedings*, Vol. 81, No. 81, July 1993, pp. 943-967.
9. Lee W. C. Y., *Mobile Communication Engineering*, (McGraw-Hill, 1999)
10. Rappaport T. S., Wireless communications, (Prentice-Hall, 2002, 2nd ed.)
11. Parsons J. D., *The mobile radio propagation channel*, (Wiley, 2000).
12. Morelli M., Sanguinetti L. and Mengali U., "Channel Estimation for Adaptive Frequency Domain Equalization ", *IEEE Transaction on Wireless communication*, September 2005, Vol. 4, No. 5, pp. 2508-2518.
13. Salous S., and Hinostroza V., " Bi-dynamic UHF channel sounder for Indoor environments ", *IEE ICAP 2001*, pp. 583-587
14. Biglieri E., Proakis J. and Shamai S., "Fading Channels: Information Theoretic and Communications Aspects", IEEE Transactions on Information Theory, October 1998, Vol. 44 , No. 6, pp. 2619-2692.
15. Al-Dhahir N., "Single Carrier Frequency Domain Equalization for Space-Time Blok-Coded Transmission over Frequency Selective Fading Channels", IEEE Communications Letters, July 2001, Vol. 5, No. 7, pp. 304-306.
16. Namgoong N, and Lehnert J., " Performance of DS/SSMA Systems in Frequency Selective Fading", *IEEE Transaction on Wireless communication*, April 2002, Vol. 1, No. 2, pp. 236-244.
17. TA0 X., et. al., " Channel Modeling of Layered
18. Space-Time Code Under Frequency Selective fading Channel", Proceedings of ICCT2003, May 2003, Berlin Germany.
19. Shayevitz O. and Feder M.," Universal Decoding for Frequency Selective Fading", IEEE Transactions on Information Theory, August 2005, Vol. 51 , N0. 8, pp. 2770-2790.

20. Sánchez M and García M, RMS Delay and Coherence Bandwidth Measurements in Indoor Radio Channels in the UHF Band, *IEEE Transactions on Vehicular Technology*, vol. 50, no. 2, march 2001
21. Jia-Chin Lin, Frequency Offset Acquisition Based on Subcarrier Differential Detection for OFDM
22. Communications on Doubly-Selective Fading Channels,
23. Yoo D. and Stark W. E., Characterization of WSSUS Channels: Normalized Mean Square Covariance, *IEEE Transactions on Wireless Communications*, vol. 4, no. 4, july 2005.
24. 22. Riback M., Asplund H., Medbo J., Berg J., Statistical Analysis of Measured Radio
25. Channels for Future Generation Mobile Comunication Systems,

Chapter 4

Quadruple Play: On the Wireless Communications Convergence in México

Ángel G. Andrade

Electrical Department, University of Baja California, Blvd. Benito Juarez s/n,
Mexicali, Baja California, México, Phone and Fax: (686) 5664270, e-mail: angel_andrade@uabc.mx

Abstract

At the start of the 21st century, wireless communications has witnessed an unprecedented growth fueled by information explosion and technology revolution. The WI-FI devices and the new mobile terminals based UMTS, change the way in how the users access to the internet. Actually the Mobility is consider the main aspect to the technological and services convergence. People talk about voice, data, video, and mobility services like a strategy of the quadruple play. The IP telephony is the new service that enable to users enhance the wireless communication in anywhere in the world.

Keywords: Wireless communications, convergence, IP telephony, Wi-Max, mobility.

1 Introduction

With the rapid growth of the wireless mobile applications, wireless voice has begun to challenge wireline voice, whereas the desire to access email, surf the web or download music wirelessly is increasing for wireless data.

In 2004, Nokia announced the 3200 GSM cell phone, which can make electronic payment (similar to a smart card) and place calls based on the RFID tags it encounters. For example, you could place your phone near an RFID tag attached to a taxi-stand sign, and your phone would call the taxi company's coordinator to request taxi at that location [1].

In [2], the authors present an Alcatel's helping service thinking in lonely elders. The system allows, just pushing a button, find to the nearest contact from an emergency list. The system use GPS technology, and it is capable to decide which is the best way to send the message, e.g. by mail or making a phone call.

The scenarios reported previously are the result for pursuing the convergence of the wireline and wireless telephony, also for the necessity of optimize the operator's infraestructure.

The ability to provide more spectrally efficient voice capacity and spectrally efficient high-speed wireless data has been the focus of third generation (3G) technologies. Actually, the mobile networks transport a lot of information due email service, IP video, music download, etc. The network's operators need tools for management bandwidth of the users and for developing killer applications.

Some research on mobile information services provision explores voice as the primary interaction modality. Because people frequently carry mobile phones wherever they go, distinct opportunities exist for harvesting profile and activity data and for delivering timely information services. Finally, many mobile phones currently feature wireless local networking capabilities that let them interact with other nearby devices.

In this work, I shall present a vision on the wireless communications convergence. Also, I present a comprehensive introduction of the mainstream wireless mobile technologies and their evolution paths to the future in terms of the expected voice and data spectral efficiency. In doing so, I will examine how all of this wireless technologies converge for inter-operating between them and offering to users ubiquitous service.

2 Technological Convergence

In the past, telecommunications, information technology (IT) and broadcasting all operated independently in terms of the technology used, the information transmitted and the networks employed. Television, radio, telephones and computers were used for discrete purposes and the services provided were regulated via separate laws, usually by different regulators, and with no obvious need for coherence between these separate laws and regulators. Technological convergence enables traditionally distinct voice and data transmissions to be transported over the same network and to use integrated consumer devices for purposes such as telephony, television or personal computing. The European Union (EU) defines convergence

as "the ability of different network platforms to carry essentially similar kinds of services, including the coming together of consumer devices such as the telephone, television and personal computing".

Traditional convergence is noted in the combination of the personal computer and the internet technology. This combination provides a convergence of data processing, images and audio services. Recent examples of new, convergent services include: Internet services delivered to TV sets via systems like Web TV; E-mail and World Wide Web access via digital TV decoders and mobile phones; Web casting of radio and TV programming on the Internet; Using the Internet for voice telephony.

The availability of carrier technology with high bandwidth means that, transmission is not limited to voice only, now data, picture and other multimedia and interactive media can be transported in one single carrier technology like the fibre optic cable and satellite technology.

The convergence of telecommunications and information technology has also led to a geographical convergence. This has led to theoretical conception of the world as a global village, where interactions and communication are no longer hindered by distance. The satellite helps facilitate communication irrespective of geographical location. Information sent from Iceland could be received instantaneously in Burundi via telecommunications link up. As the lines between data transmission, audio cast and voice transmission are eroded, regulators are faced with the task of how best to classify the converging segments of the telecommunication sector. Example of this is Voice Over Internet (VOI). The debate is whether to consider this as part of Internet services and if so, it resides in the domain of Internet Service Providers or as a voice transmission which is the market domain of the local telephone operator.

We live in a world where all networks are folding into one, where you get television on the Internet and Internet on your TV. Wireless usage nears saturation throughout the civilized world, and mobile phones can do almost anything. Communications companies, using the Internet as a foundation, are pursuing the quadruple play of voice, video, data and mobility. As part of this new connected landscape, intelligence in the network has moved from the core to the devices at the edge, where users have control over that power. Corporate dependence on networks is now absolute. Formerly wired services, such as phone calls, have switched to wireless, and now back to wired with voice over IP. Such "switch backs" occur the other way, too.

The Connected World names eight connectedness trends, they are: the all-IP enterprise, industry crossovers, bandwidth at the edge, networks in new places, new things being connected, liquid time and place, pervasive presence, and the next frontier of mobility. The goal in this connected world is a core network that brings all data, voice, video, mobility and applications over one network, wired or wireless. That core will be an IP network.

3 Next Step to Connectivity: Seamless Mobility

In the past decade, the Internet has spawned many innovations and services that stem from its interactive character. There are numerous indications that the ongoing process of adding mobility to interactivity will transform the role of the Internet and pave the way for yet another set of innovations and services. The convergence of computing and communication is a process that is about to turn phones and mobile terminals into powerful multimedia units.

Thanks to the convergence of telecommunications and data communication, future computer applications will rely on seamless wireless networking, and will thus be inherently mobile. This latest trend is now observable and a clear example is the convergence between two technologies that had developed separately during most of the nineties: wireless communication devices (pagers, mobile phones) and handheld devices (personal digital assistants, PDAs). Recently, a number of mobile phones and other wireless devices with PDA capabilities have been introduced; conversely, more and more handheld devices now come equipped with wireless capabilities.

All these new forms of interactive multimedia and communication offer new possibilities as to the way we learn, think, and communicate. The combination of handheld computing and wireless communication suggests enormous potential for education especially given how familiar most young students already are with these technologies.

4 Wireless Technologies: Myths and Realities

The most significant wireless technology to date is Wi-Fi, which is dotting the planet with hotspots and giving us access to the wireless network from the places we eat, work and play. But while Wi-Fi hotspots are limited to ranges of 300 feet, these hotspots are now being federated into a much larger area of coverage.

WiMAX, with several times the connection speed and range of Wi-Fi, is the next logical step in wireless technology. WiMAX can take broadband speeds to parts of the world where landlines had a tough time reaching. With its bandwidth, WiMAX could handle voice, data, video and mobility. But important questions regarding WiMAX, including the wireless spectrums WiMAX uses, still have to be figured out before it can reach broad adoption.

While Wi-Fi and its siblings are rooted in place, a good portion of wireless Internet traffic today comes from cellular phones, which can move anywhere. Millions of mobile users tap into broadband networks. These cellular networks are expanding in transmission speeds and function, making them a legitimate alternative to conventional wireless bandwidth.

4.1 Third Generation: 3G

Third-generation (3G) networks have brought some order and sufficient bandwidth to the cellular world, making heavy Internet use and video downloads a real option. But future upgrades to the cellular networks will bring packet switching and dramatic increases in speed.

The label "3G" means third-generation cellular network (it is also called UMTS in Europe), and it covers a bunch of technologies. The technologies tend to fall into two groups: One group includes those technologies that work on the CDMA networks, the other group includes those technologies that work on the GSM networks. In the CDMA world, the technologies are, from slowest to fastest, CDMA 1XRTT and CDMA 1XEVDO. In the GSM world, they are, again from slowest to fastest, GPRS, EDGE (often called WCDMA in Europe), HSDPA, and the planned HUDPA.□□ The CDMA2000 1XEVDO technology delivers speeds of 300Kbps to 1.2Mbps, depending on geographic location, network load, distance from the cell tower, and other factors. Most EVDO cards supplied for notebooks and handheld devices automatically switch to 1XRTT when EVDO service is unavailable.□□ The EDGE technology, which provides real-world access at 100Kbps to 200Kbps, though the faster HSDPA technology is now available in a few cities and offers speeds of 300Kbps to 1.2Mbps. EDGE and HSDPA also have a slower backup network in place for less populated areas: GPRS, which provides 50Kbps to 100Kbps connections. The HSDPA cards switch to EDGE when HSDPA is unavailable, and to GPRS when both EDGE and GPRS are unavailable. Most EDGE cards also switch to the GPRS networks when EDGE service is unavailable.□ It can note that you cannot typically roam between carrier networks. Also, most carriers limit your broadband usage each month, even for their "unlimited" plans, and forbid the use of streaming technologies like voice over IP (VoIP) and video, since their networks can't yet handle that amount of traffic. Of course, there are few issues with 3G. First is that 3G services cost approximately $80 per month for laptop-only service, or $50-60 per month in addition to your cell phone service. And while advertised as "unlimited" service, the contracts have lots of fine print about not using certain kinds of bandwidth-hungry applications such as voice-over-IP (VoIP) and streaming media (such as audio and video files). So they probably aren't a realistic choice as your only broadband service — though as a supplement to your home or business wired service, they could be worth the extra cost compared to Wi-Fi hot spot service due to the greater convenience. □□ Another issue is speed. 3G devices typically run at the bottom range of DSL speeds or even slower — 100Kbps to 400 Kbps — despite what the ads imply (they typically cite the maximum possible rates, which you'll get only occasionally). That speed is of course perfectly fine for e-mail, Web browsing, and working with files over the network, but it won't feel as responsive as wired broadband does (typically 1Mbps to 10Mbps), or even a Wi-Fi hot spot's speed (typically 500Kbps to 1Mbps).□□ Over time, 3G networks will get more capable and cheaper, so the price and speed issues will likely diminish.

4.2 Wi-MAX

WiMAX is a term coined to describe standard, interoperable implementations of IEEE 802.16 wireless networks, in a rather similar way to Wi-Fi being interoperable implementations of the IEEE 802.11 Wireless LAN standard. Defined as Worldwide Interoperability for Microwave Access, and is described by the WiMAX Forum as "a standards-based technology enabling the delivery of last mile wireless broadband access as an alternative to cable and DSL."

These days, WiMAX seems to be at the top of everyone's watch list and has become one of the most anticipated developments in the telecommunications world. The IEEE 802.16 Working Group has developed point-to-multipoint broadband wireless access standard for systems in the frequency ranges 10-66 GHz and sub 11 GHz. The standard covers both the Media Access Control (MAC) and the physical (PHY) layers.

At higher frequencies, line of sight is a must. This requirement eases the effect of multipath, allowing for wide channels, typically greater than 10 MHz in bandwidth. This gives IEEE 802.16 the ability to provide very high capacity links on both the uplink and the downlink. For sub 11 GHz non line of sight capability is a requirement. WiMAX has the potential to replace a number of existing telecommunications infrastructures. In a fixed wireless configuration it can replace the telephone company's copper wire networks, the cable TV's coaxial cable infrastructure while offering Internet Service Provider (ISP) services. In its mobile variant, WiMAX has the potential to replace cellular networks.

Some goals for WiMAX include a radius of service coverage of 6 miles from a WiMAX base station for point-to-multipoint, non-line-of-sight service. This service should deliver approximately 40 megabits per second (Mbps) for fixed and portable access applications. That WiMAX cell site should offer enough bandwidth to support hundreds of businesses with T1 speeds and thousands of residential customers with the equivalent of DSL services from one base station.

Mobile WiMAX takes the fixed wireless application a step further and enables cell phone-like applications on a much larger scale. For example, mobile WiMAX enables streaming video to be broadcast from a speeding police or other emergency vehicle at over 70 MPH. It potentially replaces cell phones and mobile data offerings from cell phone operators such as EvDo, EvDv and HSDPA. In addition to being the final leg in a quadruple play, it offers superior building penetration and improved security measures over fixed WiMAX. Mobile WiMAX will be very valuable for emerging services such as mobile TV and gaming.

However, WiMax is available in very limited areas, since it will require significant investment — billions of dollars — to deploy all the needed radio towers and/or to convert existing 3G towers to support it for true national coverage. Still, because it is wireless, WiMax does support usage anywhere in its coverage area — unlike DSL and cable service — so it's a much more flexible service than the wired options even if offered within a single city. But it's not as flexible as 3G in terms of how far you can roam — at least not today.

While Wifi hotspots have signals that reach at most a few hundred feet, WiMax signals reach as far as several miles depending on conditions. But 3G networks, thanks to aggressive rollouts by cellular companies, cover tens or even hundreds of miles around metropolitan areas throughout many areas of Europe, Asia and the Americas. Also, the 3G technologies support true mobility, so you can use your laptop in a moving train or bus (assuming the signal reaches you there), while the current version of WiMax (technically called 802.16-2004) is designed to work for a stationary device, such as a laptop sitting on a table in a café or in your living room. You could use WiMax to work offsite, such as at a café (assuming the signal reaches you there), but you shouldn't be continually moving around while working to get a reliable connection. A future version of WiMax (technically called 802.16e) will support true mobile usage. WiMax runs both in licensed spectrum and unlicensed spectrum, so you can have both services available at the same location. But it's possible that WiMax delivered over unlicensed spectrum could cause signal interference with other people's WiMax and Wifi networks. So expected unlicensed WiMax to be deployed in areas like airports, industrial zones, and downtowns where a business — or group of businesses — works to manage the network, minimizing the chances of interference.

4.3 Fourth Generation: 4G

While most wireless operators are still struggling to understand how to properly monetize their third generation wireless networks, the race for fourth generation network technologies has already begun. This is not a contradiction as current 3G networks will be operated and enhanced for many years to come. Furthermore, specification, development, rollout and mass production of 4G devices all take their time. Thus, most 4G systems are at least five years or more away from the mass market.

The primary question when looking at future 4G systems is why there is or will be a need for them. Looking back only a couple of years, voice telephony was the first application that was mobilized. The short message service (SMS) was the first data application that was mobilized as a mass market application. By today standards comparably simple mobile phones were required. Also, bandwidth requirements were very small. In a way, the SMS service was a forerunner for other data services like mobile eMail, mobile web browsing, mobile blogging, push to talk, mobile instant messaging and many others. These were enabled by the introduction of packet based wireless networks that could carry IP data on the one hand and more and more powerful mobile terminals that could cope with the requirements of these applications on the other. Today, current 3G and 3.5G networks are able to cope quite well with these applications as they offer a sufficient bandwidth per user. Also, network capacity is still not an issue as only few people use these services today. There are a number of trends which are already visible today which will increase bandwidth requirements in the future: here are two main goals of 4G wireless systems. First of all, more bandwidth will be required, secondly, 4G networks will no longer have a circuit switched subsystem as current 2G and 3G networks. Instead, the network is based purely on the Internet Protocol (IP). The main challenge of this design is how to support the stringent requirements of voice calls for constant bandwidth and delay. Having sufficient bandwidth is a good first step.

Currently, 3G networks are transforming into 3.5G networks as carriers add technologies such as High Speed Data Packet Access (HSDPA) and High Speed Uplink Packet Access (HSUPA) to UMTS. Similar activities can be observed in the EVDO world. Staying with the UMTS example, such 3.5G systems are realistically capable of delivering about 6-7 MBit/s in a 5 MHz band. However, these speeds can only be reached under ideal conditions (very close to the antenna, no interference, etc) which are rarely found in the real world.Increasing channel size and using MiMo will increase throughput by about 8-10 times. Thus speeds of 40 MBit/s per sector of a cell are thus possible. Sophisticated base stations use three or even four individual sectors which results in a total throughput of a single base station of up to 120 to 160 MBit/s. At some point current 2G and 3G network operators will migrate to a 4G network technology. As 4G network technology is based on IP only and includes no backwards compatibility for circuit switched services, current operators do not necessarily have to select the evolution path of the standard they are currently using.

For current UMTS network operators the most likely evolution path will be to LTE. Devices will most likely be backwards compatible to their existing 3G and 3.5G networks. Also, connectivity of the new LTE radio network to their

existing core network infrastructure, billing systems and services will be seamless. Also, current 3.5G networks offer enough capacity for a number of years to come. Thus, UMTS operators are currently in no hurry with 4G technologies. Having a predecessor technology already in place is a great help in introducing a new technology especially if new devices are backwards compatible to existing networks.. Thus, handsets and other mobile devices will not only work in LTE networks but also in 3G UMTS networks and most likely also in 2G GSM/GPRS/EDGE networks. This is especially important in the first few years of network deployments when coverage is still limited to big cities. EVDO Rev C. is likely to follow a similar path.

WiMAX on the other hand is not backwards compatible to any previous wireless network standard. Thus, it remains to be seen if devices will also include a 3G UMTS or EVDO chip. This is not only a question of technology but also a question of strategy. If a company with a previously installed 2G/3G network deploys WiMAX then they will surely be keen on offering such handsets. New alternative operators without an already existing network on the other hand might be reluctant to offer such handsets as they would have to partner with an already existing network operator. They might not have much of a choice though if they want to reach a wider target audience.

5 Conclusions

In the end I am quite convinced that at least two technologies will gain global traction. If WiMAX is one of them, and I am quite convinced that it will be, there will be even more competition in the wireless domain than today. The disadvantage of WiMAX of not having a network legacy could in the end be a major advantage. It will allow new companies to enter the market more easily and thus increase competition, network coverage, services and hopefully decrease prices. Research focus has thus shifted to what's currently known as fourth-generation (4G) or beyond-3G (B3G) mobile networks, seen as an extension of the current situation. Is a fact that Cellular and Wi-Fi networks should eventually merge. That will open up new uses for the network. Access is potentially capable of being all mobile and all IP, providing the ultimate in convergence and a foundation for new innovations at the network's edge.

References

1. Want Roy.: An Introduction to RFID Technology. IEEE Pervasive Computing Journal. 1 (2006) 25-33
2. Roussos, G., Marsh, A., Bruce, Maglavera, S.: Enabling Pervasive Computing with Smart Phones", IEEE Pervasive Computing Journal. 2 (2005) 20-27
3. Qi, Bi; Seymour, J.: The Future of Wireless Mobile Communications. Wireless Communications and Mobile Computing Journal. 3 (2003) 705-716
4. Zahariadis, T.: Migration toward 4G Wireless Communications. IEEE Wireless Communications. 3 (2004) 6-7.
5. Shadbolt, N.: Ambient Intelligence. IEEE Intelligent Systems. 4 (2003) 2-3.
6. Gustafsson, E., Jonsson, A.: Always best Connected. 1 (2003) 49-55.
7. Want, R.: Enabling Ubiquitous Sensing with RFID. Computer. 4 (2004) 84-86.

Chapter 5

Copying the Human Eye Strategies to Design Antenna Arrays

Diego Betancourt and Carlos del Rio

Public University of Navarra, Campus Arrosadía s/n 31006 Pamplona, Spain. Email: carlos@unavarra.es

Abstract

In this chapter, the behavior of the human eye is analyzed obtaining different solutions for the main common trade-offs of antenna array systems, in particular regarding the angular resolution and the signal/noise ratio.

Normally, to improve the angular resolution, more directive beams close to each other are needed. This is difficult because of the needed overlapping of the effective radiating areas of the different independent beams.

On the other hand, to solve to problem of signal/noise ratio, a heterodyne detection is preferred instead a direct detection, since the power balance of the system ensures more margin at the receptor position.

As we will see, the human eye uses very different techniques to solve these problems. In fact, both problems are solved with the same strategy, using many photo-receptors, cones, to create each of the beams, and re-using many of these cones for other neighboring beams. Thanks to this recycling technique of the cones, the effective radiating areas are practically overlapped and because are many cones detecting the same information coherently, the signal/noise ratio is also improved thanks to the spatial diversity used. The final result is well know for everyone and is that the "antenna system" of the human eye works really very well using very simple detection techniques.

1 Introduction

The Antenna Group of the Public University of Navarra has been studying the behavior of the human eye trying to identify and compare its behavior with an antenna array. We pay special attention to the impressive performance of the human eye regarding the visual acuity, that translated to antenna language will be the angular resolution of the system.

Usually, in the conventional antenna array, every detector is responsible of extracting the information from the area covered with the detector itself, the effective radiating area. The possibilities to increase the resolution of the obtained image need to improve the individual directivity of the detectors, increasing the radiating areas. Additionally, this improvement in the directivity also helps to improve the signal/noise ratio.

However, if the size of the detectors increases the number of pixels at the focal plane would decrease, limiting the resolution of the obtained image. A trade-off is clearly established.

In the case of the human eye, the evolutional solution to this problem, results to be a quite optimum solution regarding this compromise.

The angular resolution of the human eye, visual acuity, is commonly referred as the minimal angle of resolution (MAR) and it is established to be 1 minute of arc obtained experimentally. However, the real behavior of the human eye should be more complicated than this, since also experimentally has been reported a minimum visible threshold of approximately 1 second of arc. Additionally, it is also clear that our images are not discrete as the discrete disposition of the photoreceptors on the retina could suggest.

The hypothesis to explain the real working of the human eye of the authors of this chapter is based on the fact that we should need a bigger effective receiving areas being all of them highly overlapped to each other, to be able to justify the high directivities so close in angle in the far field. Since the size of the cones is fixed, the only possibility to define such bigger receiving effective areas is using a set of cones for each arrival direction, and the only possibility of overlapping these areas is that every cone would be effectively working to different arrival directions simultaneously. This hypothesis also could explain the impressive signal/noise ratio achievable by the human eye thanks to the use of the spatial diversity used to receive each particular direction.

It is quite easy to imagine the possibilities of applying this detecting philosophy to imaging systems, but also to the antenna systems where many different beams should be supported simultaneously.

2 State of The Art of Antenna Arrays

One of the most important requirements of all the antenna systems is the angular resolution achievable, to be able to distinguish two punctual sources placed quite close to each other in far field.

Normally, the antenna configuration is an array antenna system with many radiating elements which conforms a unique beam steerable in angle electronically modifying the complex coefficients of each of the radiating elements of the systems. In this kind of systems, ASAR antenna of ENVISAT [1] and ALMA experiment [2] could be excellence examples of this (Fig. 1 and 2); the angular resolution is given by the scanning capabilities of the system. In this case, all the efforts should be introduced not in the antenna elements but in the electronics behind the antennas to scan the beam.

Fig. 1. *ASAR FM antenna of the ENVISAT satellite during the radiation testing at ASTRIUM Portsmouth. (Image courtesy of ASTRIUM Ltd.)*

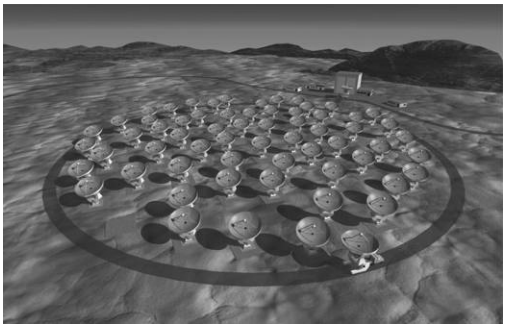


Fig. 2. *Artist view of the ALMA experiment*

Other possibility could be to try introducing more independent beam in the same antenna system, being possible to detect each point in the far field with different beams.

There are different possibilities: to use some reflector system with multiple feeds or introducing the concept of a Direct Radiating Array (DRA).

The first possibility is the easier one, since increasing the number of beams of the systems could only mean to introduce another feed at the focal zone of the reflector. In these systems the limitations arise since the focal zone is limited to a small area from which the distortion introduced to the beam is negligible. Additionally, the beams are generated by horn antennas, so the size of the feeders does not allow us to handle as many beams as we could need [3, 4, 5, 6].

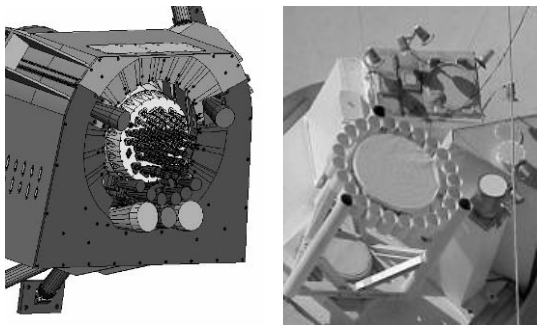


Fig. 3. *Multiple feed disposition at the focal plane of the PLANCK (left) [4] and VLA (right) [5] systems.*

Other possibility could be using a Direct Radiating Array antenna, where the different beams are created simultaneously by all the radiating elements of the antenna system. In this case the technological restrictions are not in the focal zone, but in the required control electronic systems.

With the actual technology, we should replicate the scanning electronics to be able to control each one of the beams independently. Under some specific conditions, common modulation, frequency and multiplexation by code; this replication could be performed by software.

In any case, if the number of beams to control increases, the complexity of the whole system could be so important that the system was simply not possible. The only things that we do not need to replicate are the antenna elements, but we will need additional power dividers and combiners to be able to drive all the independent signals to the radiators.

An example of this technique could be observed in the MIMO systems, where the antennas are essentially DRA's managing a small number of beams simultaneously.

At this point, it should be interesting to have a look to the human eye, since we have the capabilities that we ask the antenna systems to have. We are able to distinguish between two really close punctual sources at long distances. The strategy is not based on scanning any beam; the idea is to have many beams pointing to as much directions as possible, with the maximum directivity achievable in each one of the beams.

This chapter wants to be an invitation to study the working principle of the human eye as detector, to consider the viability of application of these principles to the antenna system design at lower frequencies.

3 The Human Eye

The human eyes have an impressive performance. They are able to focus near and far objects automatically; they have possibilities to see under bad circumstances (low light); they are able to control de entrance of light to prevent possible saturation of photoreceptors; and finally, the most impressive feature, the really high resolution of the obtained images.

But, it is even more impressive if we try to understand the human eye under antenna parameters. To be able to distinguish two points quite close to each other at certain distance, we should need some kind of really high directivity. Some experiments performed shown that the needed directivities should be close to 90dB, which means, that the main beam is subtended under angles of less that 1 minute of arc. But this is even more amazing if we think that many other beams should be simultaneously placed every 1 minute of arc, to be able to distinguish the changes of light intensity and color.

In Fig. 4, a simple experiment of what we could see if the directivity of our beam widths being equal to the angular separation was 10 times bigger than the normal vision.

Some people like to assign to the brain the capabilities to obtain such high resolution images, but it is clear that it could be difficult to "generate" small details that have never been received by the eye. It could be understandable some kind of interpolating post-processing technique to try to solve points placed between photoreceptors to obtain a continuity sensation of the images, but this never would generate additional details in the image increasing the resolution.

Fig. 4. *What we could see with a beam width and angular separation 10 times bigger (left) than normal vision (right).*

Thus, it is clear that some more investigation about the human eye could be very interesting to try to clarify these aspects that seems be not so clear.

First of all, we will collect some data from the human eye in order to illustrate the working principle of the human vision, and to evaluate the causes of the high resolution.

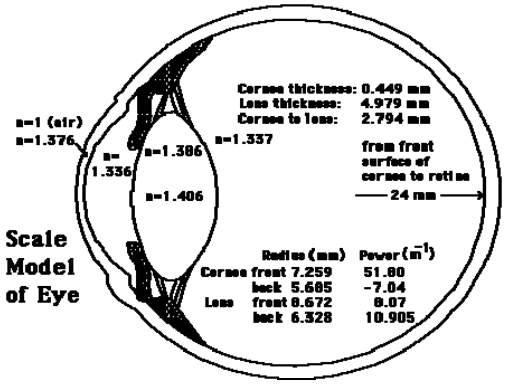


Fig. 5. *Scale model of the human eye, including the different lenses and their refractive indices.*

In principle, in the human eye, the light passes through the cornea and the crystalline having a focus situated just over the retina. The total refractive power of the human eye is measured in dioptres and it is determined by the inverse of the focal length, being 62 the considered normal value for a current human eye [7]. This means that the focal length is then 16 mm, and it corresponds with the separation between the crystalline and the retina (Fig. 5). Then, the images in front of the eyes (just at the focal plane outside the eye) will be inversely projected over the retina as in a photographic camera (Fig. 6).

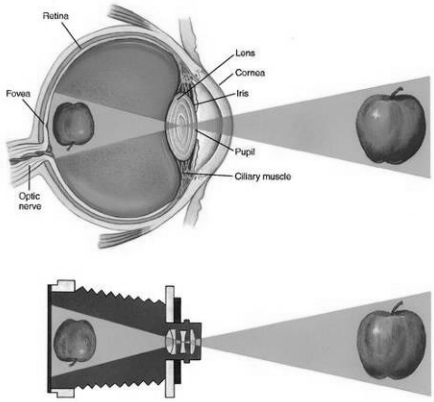


Fig. 6. *Optical focusing system of the human eye and a photographic camera, showing the similarities in the inverted projected image over the retina or the film respectively.*

Over the retina there are special cells, photoreceptors, specially prepared to receive the information modulated at optical frequencies introducing this information in the neural system to the brain to process. There are different kinds of photoreceptors: cones and rods. The cones responsible of the colored vision and the rods more related with the vision under poor lighting conditions.

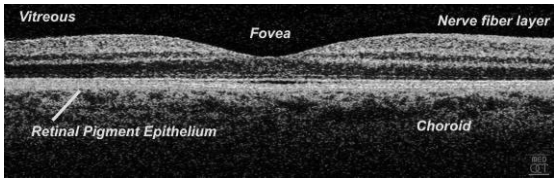


Fig. 7. *Optical Coherence Tomography at 800nm of the Macula in a human retina.*

It is also well known that we really have properly focused a small cone subtended in an angle of less than one degree, and the responsible of this is a small area on the retina just centered at the vision axis called Macula recognized as a depression of the retina surface being the deeper point knows as Fovea (Fig. 7).

In the Macula, the density of cones is 160.000 per square millimeter being the main responsible of the focused central vision. The Macula is in fact acting like a divergence lens, dispersing the parallel rays arriving close to the optical axis over a wider area of receptors. This divergence effect is not included properly in the optical system of the eye since the Fovea is only 100 m deep, and the dispersion effect is really small and only relevant for the increase of area of receptors affected.

It is also known that the cones diameter is approximately 1.5 m and the separation between two cones is about 0.5 m.

With these dimensions, if we apply the photography principle and we expect to obtain focused over the retina the inverted image, we could calculate the resolution of the retina image simply translating the separation between two cones outside the eye. To solve two different points outside, the retinal image should have an "unexcited" cone between two other excited ones by the two points respectively, given a distance of 4 m as the minimum distance necessary to solve two points on the retina. Translating this distance out of the eye up to a distance of 350mm (reading distance), these resolution step is 84 m.

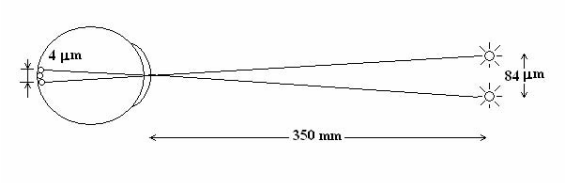


Fig. 8. *First approximation of human eye resolution.*

This could mean that we shouldn't be able to distinguish anything smaller than this resolution step, but really we are able to see below that limit, so something more should be consider.

4 Explanation of the Higher Angular Resolution of the Human Eye

In this point of the explanation it is important to think a little bit about the detection mechanism. It could be agreed that the cones are not detecting the carrier signal (optical frequencies); they are really detecting the modulated signal. Certainly, each type of cone has its own frequency response to optical frequencies, being more sensible to different colors: red, green and blue.

In a simplified model, we could think that the cones could be acting as integrators, just counting the number of photons received, and generating an output signal level proportional to that number.

The relative position over the retina surface jointly with the lens system will determine the arrival direction, so the beam forming network is performed by the optical system. Additionally, the retina surface follows a phase front of the optical system of the human eye, so the phase information is, under this perspective, totally negligible.

Some authors have tried to explain the higher resolution by means of the rapid vibrations of the eyes to avoid the saturation of the photoreceptors loosing the possibilities to see the objects in front of us. Under this assumption, it could be argued that it could really difficult to think that the brain could know exactly the position of the eye when each photon arrives to each cone to reconstruct the higher resolution image.

Thus, assuming the integration behavior of the cones just mentioned above, this vibrating phenomenon could really diffuse the image over the retina.

Furthermore, under the optical theory, the light travels in straight lines, but really some diffraction should be consider since the light is also a wave. This means that really the light will not focus in a single point but in an area around this point of impact. This phenomenon again generates some diffusion of the retina image.

But even more, if we study the chemistry reactions of the reception mechanism of a photon by a cone, some horizontal coupling between neighboring cones have been reported [7], in some kind of amplifying effect ensuring that all photons are properly received by the retina, improving the signal-noise ratio of the receiving system.

All these three phenomena generate some kind of distortion in the retina image since the information of one photon is effectively spread over some area of the retina, either by the diffraction of light, by the movement of the eye or by the chemistry mechanisms.

So, in summary, where we were expecting to obtain an image perfectly focused over the retina, since is placed just at the focal plane of the eye lenses (cornea and crystalline), we are really obtaining a totally de-focused image.

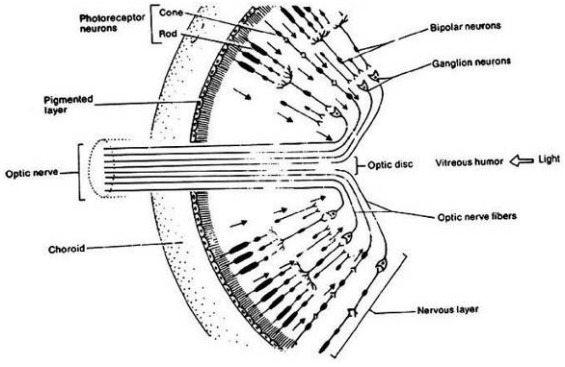


Fig. 9. *Detailed disposition of the neurons and the photoreceptors inside the retina of a human eye.*

Originally, we start this study to explain the higher resolution of the human eye. Could all this really explain the higher resolution? Or, on the contrary we should continue investigating other causes to justify the higher resolution.

Let's follow a little bit more inside the composition of the retina. Looking carefully at the total composition of the retina, we found two additional layers of neurons over the light receptors highly interconnected (Fig. 9). These two layers could perform some kind of image processing trying to build a higher resolution image to be sent to the brain through the optical nerve.

This could be possible thanks to the inherent coherence of the eye as a detector of photons. The effect of a photon over the retina finally excites a set of photoreceptor cells, as it was justified above. With the neural networks we could identify without any problem the impact point calculating the position of the maximum.

To do this in only one layer, we should have as much neurons as detecting cells over the retina, being every neuron connected to several retina cells (this number should corresponds with the number of excited cells by a single photon) and applying a threshold function.

Under the inherent coherence conditions of the detection mechanism, the linearity perfectly applies with any distortion phenomena, so by using a single layer of neurons we really could clarify the diffused retinal image, or even more, we could be able to increase the resolution of the received image having more neurons than cones.

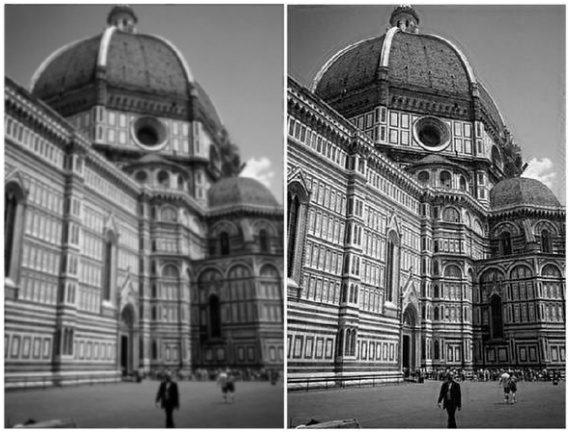


Fig. 10. *Blurred image hypothetically captured at the retina surface and deblurred by the neuron layers.*

In Fig. 10, and hypothetical diffused image captured at the retina surface (simulated blurring the image with a Gaussian mask) is clarified just applying the spatial convolution of the inverse mask with the blurred image. This could perfectly be performed by these one or two neural layers [8].

There are some advantages of having a diffuse image over the retina to be clarified afterwards, but the most important is that the robustness of the vision system is substantially improved since the information over the retina is redundant. By this method, it could be understandable that different types of cones and rods could be working together to define the small details of the image in front the eyes.

Coming back to the antenna design theory, it is clear that the retina could be consider as an antenna array, each of the cones acts as a detector, where to obtain a high directive beam, some area over the retina should be effectively used, having highly overlapped radiating areas that could justify very close high directive beams, and therefore, high angular resolution.

This is, every cone is used to define many different spots of the obtained image, or in other words we could also say that the information corresponding with every spot of the image have been effectively received by many cones, allowing a direct detection strategy with a reasonable signal to noise ratio, obviously obtaining a very simple detecting system.

5 Conclusions

In this chapter, the behavior of the human eye as an antenna system have been analyzed noting the different solutions for the main common trade-offs of antenna array systems, in particular regarding the angular resolution and the signal to noise ratio.

Normally, to improve the angular resolution, more directive beams close to each other are needed. With the conventional techniques, the minimum distance between two neighboring beams is determined by the size of the radiating elements used to create them.

On the other hand, to solve to problem of signal to noise ratio, historically the heterodyne detection is preferred instead a direct detection, since the power balance of the system ensures more signal to noise ratio at the receptor position.

As we have seen, the human eye uses very different techniques to solve these problems. In fact, both problems are solved with the same strategy, using many photoreceptors, cones, to create each of the beams, and re-using many of these cones for other neighboring beams. Thanks to this recycling technique of the cones, the effective radiating areas are practically overlapped and since there are many cones detecting the same information coherently, the signal/noise ratio also improves.

The final result is well know for everyone and is that the "antenna system" of the human eye works really very well using very simple detection techniques. So the final question is: Could we use these techniques in microwave and millimeter imagining systems?; and, what about using these techniques in Smart antennas, MIMO, or any kind of array antenna configuration? With a high probability the answer could be that we could have very much simpler systems with very similar or improved performances just copying the detecting strategies of the human eye.

Acknowledgements

This work have supported by the Spanish Government by the project TIC2003-09317-C03-01.

References

1. http://envisat.esa.int/
2. http://www.alma.nrao.edu/
3. http://www.naic.edu/
4. http://www.esa.int/science/planck
5. http://www.nrao.edu/
6. S.G. Hay, S.J. Barker, C. Granet, A.R. Forsyth, T.S. Bird, "Multibeam Earth Station Antenna for a European Teleport Application,", Int. Conf. on Antenna and Porpagation, pp. 300-303, July 2001.
7. Kaufman, P.L. and Alm, A., "Adler's Physiology of the eye", Edited by P. Kaufman, Mosby, tenth edition, ISBN 0-323-01136-5.
8. Gomez and C. del Río, "Obtaining images from CORPS systems", 1st European Conf. on Antennas and Propagation, EuCAP 2006, November 2006.

PART II

CONTROL

Chapter 6

Robust Stability of LTI Systems by Means of Roots Bounding

César Elizondo-González

Universidad Autónoma de Nuevo León, Facultad de Ingeniería Mecánica y Eléctrica.
Apartado Postal 139 F C.U. C.P. 66450, San Nicolás de los Garza, N.L. México
celizond@yahoo.com

Abstract

In this work, analysis is conducted for root bounding in characteristic polynomial with positive real coefficients corresponding to LTI system. Such bounds are applied to obtain a theorem which determinate conditions to be achieved in the polynomial as to its roots have the real part bounded in the left side of the complexes plane. It is presented an example employing the results and applying a recent stability theorem for LTI system.

Keywords: stability, relative stability, LTI.

1 Introduction

It is very well known that stability of a LTI system is obtained from the characteristic polynomial $p(s) = c_0 + c_1 s + c_2 s^2 + \cdots + c_n s^n$. The system is stable if and only if the roots of the characteristic polynomial have real negative part. Determination of a polynomial roots by analytic methods is impossible when grade is greater than 4, so historically it was decided to develop stability criterions, that without getting to know the roots, allowed us to know if they were found on the left side plane of the complexes, or how many are found on the right side. It may be said [1] that history of stability of nominal polynomials, c_i = constant, initiates with three algebraic criterion: *Hermite* in 1856 [1] 1854 [2], *Routh* in 1875 [1] 1877 [8] and *Hurwitz* in 1895 [8]. Recently, in 2001, C.Elizondo published a new theorem of stability with certain advantages on Routh´s and Hurwitz´theorems. The theorem is applied in this analysis and will be explained in the background section.

To apply Hermite-Biehler [2] criterion, the polynomial is separated in its even and odd parts $p(s) = p_p(s) + sp_i(s)$. It is necessary to obtain the roots of $p_p(j\omega)$ and $p_i(j\omega)$, so it is impractical. Routh through Cauchy indexes, Sturm generalized chains, and Sturm Theorem [8], [9], proves [8] his very well known theorem. There are several interesting results related to Routh´s findings, but in reality do not show works savings in numeric calculus with respect to Routh table, when determining a polynomial stability.

Hurwitz in 1895,based un Hermite work, and without knowing Routh works [8] achieves a result equivalent to Routh´s. He proposes the *Hurwitz Matrix* H created with the polynomial coefficients, existing a close relation between the main minors from *Hurwitz Matrix* H and elements $a_{i,1}$ from first column from Routh table.

The organization of this work is as follows. In section I, a brief description on the development of stability theory. In section II, appears a recent criterion on stability applicable to LTI systems, with fix parameters and parametric uncertainty. This criterion is used in the analysis. Additionally, describes recent results related with relative stability, as to say the stability of polynomials moved in the real axis in the complexes plane, obtaining criterion of "D" stability. In section III, we obtain algorithms to calculate coefficients of polynomials moved in the real axis in the complexes plane with respect to the original polynomials. In section IV, we obtain results to calculate coefficients of a polynomial moved with respect to other known polynomial moved, in addition a theorem is obtained that determines the restriction on the real part of the roots. In section V an example is presented to show application of the results of this work. In section VI, conclusions are presented.

2 Background

The new criterion for stability [6] has its fundamental in Cauchy indexes, Sturm Chains, and in [6] some sequences for functions are proposed, similar to Sturm Chains but with different properties. Using these mathematic bases, finally the following theorem is obtained.

Theorem 1: [6] Given a polynomial $p(s) = c_0 + c_1 s + c_2 s^2 + \cdots + c_{n-1} s^{n-1} + c_n s^n$ with real coefficients, the number of roots in the right side of the complexes plane, is equal to the number of variations of sign in the column in the following arrange.

σ_1	c_n	c_{n-2}	c_{n-4}	...
σ_2	c_{n-1}	c_{n-3}	c_{n-5}	...
σ_3	$e_{3,1}$	$e_{3,2}$	...	...
σ_4	$e_{4,1}$	$e_{4,2}$	...	
$\vdots$	$\vdots$	$\vdots$		
$\sigma_{(n-1)}$	$e_{(n-1),1}$	$e_{(n-1),2}$		
σ_n	$e_{n,1}$			
$\sigma_{(n+1)}$	$e_{(n+1),1}$			

$$e_{i,j} = (e_{i-1,1} e_{i-2,j+1} - e_{i-2,1} e_{i-1,j+1}), \forall\, 3 \le i \le n+1$$

$$e_{i,j} = c_{n+1-i-2(j-1)}, \forall\, i \le 2$$

$$\sigma_i = Sign(e_{i,1}) \forall i \le 2 \;\square\square\; \sigma_i = Sign(e_{i,1}) \prod_{j=1}^{(i+1-m)/2} Sign(e_{m+2(j-1),1}) \forall i \ge 3$$

$$m = 3 \text{ for } i \text{ even}, \; m = 2 \text{ for } i \text{ odd}$$

To calculate elements from column σ, is easy from the textual interpretation of the above expressions, as follows. The sign for the first two rows, Is the sign for coefficients c_n and c_{n-1} respectively, σ_i, is obtained multiplying the sign of element $e_{i,1}$, times the sign of the immediate superior element $e_{i-1,1}$ and times the sign of each superior element from column number one, jumping two by two. Example σ_9 is the result of multiplying the signs from $\sigma_{9,1}$ $\sigma_{8,1}$ $\sigma_{6,1}$ $\sigma_{4,1}$ $\sigma_{n-1,1}$.

In the case of an element $e_{i,1}$ be cero, then cero is substituted by $\varepsilon > 0$ and the calculation of the table is continued, In case that all elements $e_{i,j}$ from a row have a value of cero, then the complete row is substituted by the derivative of superior row

It can be noticed that elements $e_{i,j}$ from new table are obtained just from multiplication and subtraction from past elements, but there is not a division. When polynomial is fix, in other words that its coefficients c_i are constant numeric values, then the new theorem is easier than Routh's.

Example 1: Case of a fix polynomial. Given the characteristic polynomial $p(s) = s^5 + s^4 + s^3 + 3s^2 + 2s + 1$ corresponding to LTI system. Determine using theorem 1 the number of roots in polynomial $p(s)$ that are located on the right side of the complex plane.

Using theorem 1 elements $e_{i,j}$ from new table, were first calculated, and finally the signs of σ column were calculated, as showed ahead.

elements $e_{i,j}$

1	1	2
1	3	1
(1)(1) - (1)(3) = -2	(1)(2) - (1)(1) = 1	
(-2)(3) - (1)(1) = -7	(-2)(1) - (1)(0) = -2	
(-7)(1) - (-2)(-2) = -11		
(-11)(2) - (-7)(0) = 22		

signs σ_i C. Elizondo Table

$\sigma_1 = Sign(e_{1,1})$	$\sigma_1 = +$	1	1	2
$\sigma_2 = Sign(e_{2,1})$	$\sigma_2 = +$	1	3	1
$\sigma_3 = Sign(e_{3,1})Sign(e_{2,1})$	$\sigma_3 = -$	-2	1	
$\sigma_4 = Sign(e_{4,1})Sign(e_{3,1})Sign(e_{1,1})$	$\sigma_4 = +$	-7	-2	
$\sigma_5 = Sign(e_{5,1})Sign(e_{4,1})Sign(e_{2,1})$	$\sigma_4 = +$	-11		
$\sigma_6 = Sign(e_{6,1})Sign(e_{5,1})Sign(e_{3,1})Sign(e_{1,1})$	$\sigma_5 = +$	22		

It can be seen, that in σ column from new table, there are two changes un sign, from σ_2 to σ_3 and from σ_3 to σ_4, so then polynomiąl $p(s)$ has two roots on right side of complex plane, so the system is unstable. The difference of operations between the new theorem and Hurwitz is because in the new theorem operations done on a row, are used in the row below, while that in Hurwitz theorem the calculation of a minor Δ_i is not used in minor Δ_{i+1} The difference of operations between theorems, grows when the degree of the polynomial increases, as illustrated in following table.

Operations in n grade polynomials

grade	Hurwitz Theorem		C. Elizondo Theorem	
n	×	+ o -	×	+ o -
3	4	1	2	1
4	9	2	5	2
5	66	18	9	4
6	193	45	14	6
7	780	145	20	9

In industrial application, of control theory, normally it is necessary that the system besides of being stable, have a certain performance, for example that the roots from its characteristic polynomial, keep certain position in the complexes plane. With this in mind, in [7] we obtain the following two lemmas, that will be use in next sections.

Lemma 1*:* [7] Let $p(s) = c_0 + c_1 s + c_2 s^2 + \cdots + s^n$ be the characteristic polynomial with real positive coefficients corresponding to a LTI system, let a be a positive real number. Then roots of $p(s)$ are located to the left of $-a$ if and only if the polynomial $p(s-a)$ is asymptotically stable.

Proof.*:* General considerations: Given, $p(s) = c_0 + c_1 s + c_2 s^2 + \cdots + s^n$ then the polynomial can be expressed as $p(s) = \Pi_{i=1}^{n}(s + z_i)$ where z_i are the roots. Taking in consideration the above mention, then the polynomial $p(s-a)$ can be expressed as $p(s-a) = \Pi_{i=1}^{n}(s - a + z_i) = \Pi_{i=1}^{n}(s + (z_i - a))$ where the roots of $p(s-a)$ are $-(z_i - a) = -z_i + a$.

Necessity proof: If roots of polynomial $p(s)$ are located to left of $-a$ than all root $-z_i + a$ has a real negative part and then $p(s-a)$ is asymptotically stable.

Sufficiency Prof.; If $p(s-a)$ is asymptotically stable, then all its roots $-z_i + a$ have a real negative part, implying that roots of polynomial $p(s)$ are located to the left of $-a$.

Lemma 2 [7] Let $p(s)=c_0+c_1s+c_2s^2+\cdots+s^n$ the characteristic polynomial with real positive coefficients, corresponding to a system LTI, let b a real positive number. Then roots from $p(s)$ are located to the right of $-b$ if and only if the polynomial $p(-s-b)$ is asymptotically stable.

Proof.*:* General considerations. Given $p(s)=c_0+c_1s+c_2s^2+\cdots+s^n$, then the polynomial might be expressed as $p(s)=\Pi_{i=1}^{n}(s+z_i)$ where $-z_i$ are the roots for the polynomial. Considering the before mentioned, then the polynomial $p(-s-b)$ might be expressed like $p(-s-b)=\Pi_{i=1}^{n}(-s-b+z_i)=\Pi_{i=1}^{n}(-1)(s-(z_i-b))=(-1)^n\Pi_{i=1}^{n}(s-(z_i-b))$ where the roots of $p(-s-b)$ are (z_i-b).

Necessity proof. If roots of polynomial $p(s)$, $-z_i$, are located to the right of $-b$ implies that $\mathrm{Re}(-z_i)>-b$ and at the same time $0>-\mathrm{Re}(-z_i)-b$ that is equivalent to $0>\mathrm{Re}(z_i)-b$ deducting that $0>\mathrm{Re}(z_i-b)$, getting to all roots z_i-b has negative real part so $p(-s-b)$ is asymptotically stable.

Sufficiency Proof.; If $p(-s-b)$ is asymptotically stable, then all its roots z_i-b have negative real part, $\mathrm{Re}(z_i-b)<0$, where $\mathrm{Re}(z_i)+\mathrm{Re}(-b)<0$ and so $-b<-\mathrm{Re}(z_i)$ that is equivalent to $-b<\mathrm{Re}(-z_i)$, implying so that roots of the polynomial $p(s)$ are located to the right of $-\tilde{b}$.

In the case of a polynomial to analyze be of parametric dependence, than when applying lemma 1 or 2 trough theorem 1, elements from first column $e_{(i,1)}$ from new table will be multivariable polynomic functions, being necessary to prove its robust positivity. To cover this, a new mathematic tool was developed called Sign Decomposition [3], [4], [5]. Using this tool, and needed and sufficient conditions, the property of robustness from lemma 1 o 2 can be proved.

3 Preliminary results

Looking for easy and practical application of results shown in above sections, it is necessary to know the following results.

Lemma 3*:* Given the polynomial $p(s)=c_0+c_1s+c_2s^2+\cdots+s^n$ then $p(s-a)=\alpha_0(a)+\alpha_1(a)s+\alpha_2(a)s^2+\cdots+\alpha_n(a)s^n$ where $\alpha_0(a)=\sum_{j=0}^{j=n}c_j(-a)^j$ and $\alpha_i(a)=\frac{1}{i!}\sum_{j=0}^{j=n}((j!)/(j-i)!)c_j(-a)^{j-i}$

Proof.*:* (Coefficient $\alpha_0(a)$) Given the polynomial $p(s)=c_0+c_1s+c_2s^2+\cdots+\alpha_n(a)s^n$ that can be expressed as $p(s)=\sum_{j=0}^{j=n}c_js^j$, then $p(s-a)=c_0+c_1(s-a)+c_2(s-a)^2+\cdots+(s-a)^n$ that can be expressed as $p(s-a)=\sum_{j=0}^{j=n}c_j(s-a)^j$ and also as $p(s-a)=\alpha_0(a)+\alpha_1(a)s+\alpha_2(a)s^2+\cdots+\alpha_n(a)s^n$, then $\alpha_0=p(s-a)\big|_{s=0}$ and so $\alpha_0(a)=\sum_{j=0}^{j=n}c_j(s-a)^j\Big|_{s=0}$, getting finally $\alpha_0(a)=\sum_{j=0}^{j=n}c_j(-a)^j$.

(Coefficient $\alpha_i(a)$) Given the polynomial $p(s)=c_0+c_1s+c_2s^2+\cdots+s^n$ that can be expressed as $p(s)=\sum_{j=0}^{j=n}c_js^j$, then $p(s-a)=c_0+c_1(s-a)+c_2(s-a)^2+\cdots+(s-a)^n$ that can be expressed as $p(s-a)=\sum_{j=0}^{j=n}c_j(s-a)^j$. On the other hand, $p(s-a)$ can also be expressed as $p(s-a)=\alpha_0(a)+\alpha_1(a)s+\alpha_2(a)s^2+\cdots+\alpha_n(a)s^n$. From last expressions we get $\alpha_i(a)$ trough derivatives: $\alpha_i(a)i!=\frac{d^ip(s-a)}{ds^i}\Big|_{s=0}$ from where $\alpha_i(a)=\frac{1}{i!}\frac{d^ip(s-a)}{ds^i}\Big|_{s=0}$. Having $p(s-a)=\sum_{j=0}^{j=n}c_j(s-a)^j$ then $\alpha_i(a)=\frac{1}{i!}\sum_{j=0}^{j=n}\frac{j!}{(j-i)!}\ c_j(s-a)^{j-i}\Big|_{s=0}$, it must be noted that terms $c_j(s-a)^{j-i}$ will only exist to $j-i\geq 0$ since when $j-i<0$ means that we got derivative to constants, then the final expression is $\alpha_i(a)=\frac{1}{i!}\sum_{j=0}^{j=n}\frac{j!}{(j-i)!}c_j(-a)^{j-i}$, $j\geq i$.

Lemma 4*:* Given the polynomial $p(s)=c_0+c_1s+c_2s^2+\cdots+s^n$ then the polynomial $p(-s-b)$ will be $p(-s-b)=\beta_0(b)+\beta_1(b)s+\beta_2(b)s^2+\cdots+\beta_n(b)s^n$ where $\beta_0(b)=\sum_{j=0}^{j=n}(-1)^j c_j(b)^j$ and $\beta_i(b)=\frac{1}{i!}\sum_{j=0}^{j=n}\frac{j!}{(j-i)!}c_j(-1)^j(b)^{j-i}$.

Proof.*:* (Coefficient $\beta_0(b)$) Given the polynomial $p(s)=c_0+c_1s+c_2s^2+\cdots+s^n$ that can be expressed as $p(s)=\sum_{j=0}^{j=n}c_j s^j$, then $p(-s-b)=c_0+c_1(-s-b)+c_2(-s-b)^2+\cdots+(-s-b)^n$ that can be expressed like $p(-s-b)=\sum_{j=0}^{j=n}c_j(-s-b)^j$ and also $p(-s-b)=\beta_0(b)+\beta_1(b)s+\beta_2(b)s^2+\cdots+\beta_n(b)s^n$, then $\beta_0(b)=p(-s-b)\big|_{s=0}$, $\beta_0(b)=\sum_{j=0}^{j=n}$, $c_j(-s-b)^j\big|_{s=0}$, obtaining finally $\beta_0(b)=\sum_{j=0}^{j=n}c_j(-1)^j(b)^j$.

(Coefficient $\beta_i(b)$) Given the polynomial $p(s)=c_0+c_1s+c_2s^2+\cdots+s^n$ that can be expressed as $p(s)=\sum_{j=0}^{j=n}c_j s^j$, then $p(-s-b)=c_0+c_1(-s-b)+c_2(-s-b)^2+\cdots+(-s-b)^n$ that can be expressed like $p(-s-b)=\sum_{j=0}^{j=n}c_j(-s-b)^j=\sum_{j=0}^{j=n}c_j(-1)^j(s+b)^j$. On the other hand, $p(-s-b)$ can be expressed also as $p(-s-b)=\beta_0(b)+\beta_1(b)s+\beta_2(b)s^2+\cdots+\beta_n(b)s^n$. From last expressions, we get $\beta_i(b)$ trough derivatives $\beta_i(b)i!=\frac{d^i p(-s-b)}{ds^i}\Big|_{s=0}$ where $\beta_i(b)=\frac{1}{i!}\frac{d^i p(-s-b)}{ds^i}\Big|_{s=0}$. Taking that $p(-s-b)=\sum_{j=0}^{j=n}c_j(-1)^j(s+b)^j$ then $\beta_i(b)=\frac{1}{i!}\sum_{j=0}^{j=n}(j!/(j-i)!)(-1)^j c_j(s+b)^j\Big|_{s=0}$, it must be noted that terms $c_j(s+b)^{j-i}$ will only exist for $j-i\geq 0$ since $j-i<0$ means that a derivative at constant values was derived, then final expression is $\beta_i(b)=\frac{1}{i!}\sum_{j=0}^{j=n}(j!/(j-i)!)(-1)^j c_j(b)^{j-i}$, $j\geq i$.

4 Main results

The results shown below are described to easy and make possible the determination of the root bounding of characteristic polynomial to systems LTI.

Lemma 5*:* Given the polynomial $p(s)=c_0+c_1s+c_2s^2+\cdots+s^n$ then the coefficient of the polynomial $p(-s-b)=\beta_0(b)+\beta_1(b)s+\beta_2(b)s^2+\cdots+\beta_n(b)s^n$ are obtained from the coefficients of $p(s-a)=\alpha_0(a)+\alpha_1(a)s+\alpha_2(a)s^2+\cdots+\alpha_n(a)s^n$, trough $\beta_0(b)=\alpha_0(b)$ and $\beta_i(b)=(-1)^i\alpha_i(b)$.

Proof.*:* From lemma 3 $\alpha_0(a)=\sum_{j=0}^{j=n}c_j(-a)^j$ and from lemma 4 $\beta_0(b)=\sum_{j=0}^{j=n}(-1)^j c_j(b)^j$ getting $\beta_0(b)=\alpha_0(b)$. On the other hand, from lemma 3 $\alpha_i(a)=\frac{1}{i!}\sum_{j=0}^{j=n}(j!/((j-i)!))c_j(-a)^{j-i}$ $\forall j\geq i$ and from lemma 4 $\beta_i(b)=\frac{1}{i!}\sum_{j=0}^{j=n}(j!/((j-i)!))c_j(-1)^j(b)^{j-i}$ $\forall j\geq i,\quad \alpha_i(a)$ is equivalent $\alpha_i(a)=\frac{1}{i!}\sum_{j=0}^{j=n}(j!/((j-i)!))c_j(-1)^{j-i}(a)^{j-i}$ $\forall j\geq i$, multiplying both sides times $(-1)^i$ it is obtained $(-1)^i\alpha_i(a)=\frac{1}{i!}\sum_{j=0}^{j=n}(j!/((j-i)!))c_j(-1)^j(a)^{j-i}$ then $(-1)^i\alpha_i(b)=\frac{1}{i!}\sum_{j=0}^{j=n}\frac{j!}{(j-i)!}c_j(-1)^j(b)^{j-i}$ finally $\beta_i(b)=(-1)^i\alpha_i(b)$.

Theorem 2*:* Let $p(s)=c_0+c_1s+c_2s^2+\cdots+s^n$ be the characteristic polynomial with real positive coefficients corresponding a system LTI, let a and b be two positive real numbers. Then roots of $p(s)$ are located to the left of $-a$ and to the right of $-b$ if and only if the polynomials $p(s-a)=\alpha_0(a)+\alpha_1(a)s+\ \alpha_2(a)s^2+\cdots+\alpha_n(a)s^n$ and $p(-s-b)=\beta_0(b)+\beta_1(b)s+\beta_2(b)s^2+\cdots+\beta_n(b)s^n$ are asymptotically stables. Where $\alpha_0(a)=\sum_{j=0}^{j=n}c_j(-a)^j$, $\alpha_i(a)=\ \frac{1}{i!}\sum_{j=0}^{j=n}(j!/((j-i)!))c_j(-a)^{j-i}$, $\beta_0(b)=\ \alpha_0(b)$ and $\beta_i(b)=(-1)^i\alpha_i(b)$.

Proof.*:* From lemmas 1, 2 it is proved that roots of $p(s)$ are located to the left of $-a$ and to the right of $-b$ if and only if the polynomial $p(s-a)$ and $p(-s-b)$ are asymptotically stables. In lemma 3 and 5 are obtained the coefficients $\alpha_0(a)$, $\alpha_i(a)$, $\beta_0(b)=\alpha_0(b)$ and $\beta_i(b)=(-1)^i\alpha_i(b)$.

5 Example

Given that in the finality of this article is not contemplated the application de sign decomposition, developed in [3], [4], [5] to prove the positivity of multivariable polynomic functions, we provide an example depending of a single parameter.

(Case of variable gain). Given the characteristic polynomial $p(s)=s^4+10s^3+37s^2+(k+24)s+k$ corresponding to a system of fix coefficients and a variable gain. By theorems 1 and 2 determinate values for k as to the polynomial has roots in the left side of the complex plane between values of -1 and -4.

To apply theorem 2, first of all it is necessary to obtain coefficients α of polynomial $p(s-a)$. $\alpha_0(a)=c_0-c_1a+c_2a^2-c_3a^3+c_4a^4$, $\alpha_1(a)=c_1-\ \ 2c_2a+3c_3a^2-4c_4a^3$, $\alpha_2(a)=c_2-3c_3a+6c_4a^2$, $\alpha_3(a)=c_3-4c_4a$, $\alpha_4(a)=c_4$. When applying the coefficients of the polynomial $p(s)$ we obtain: $\alpha_0(a)=k-(k+24)a+37a^2-10a^3+a^4$, $\alpha_1(a)=k+24-74a+30a^2\ -4a^3$, $\alpha_2(a)=37-30a+6a^2$, $\alpha_3(a)=10-4a$, $\alpha_4(a)=1$, now applying the value of $a=1$ it is obtained: $\alpha_0(a)=k-(k+24)+37-10a^3+1=4$, $\alpha_1(a)=k+24-74+30-4\ =\ k-24$, $\alpha_2(a)=37-30+6=13$, $\alpha_3(a)=10-4=6$, $\alpha_4(a)=1$. From prior values it is obtained the polynomial $p(s-a)=\alpha_0(a)+\alpha_1(a)s+\alpha_2(a)s^2+\alpha_3(a)s^3+\alpha_4(a)s^4$ resulting in $p(s-a)=4+(k-24)s+13s^2+6s^3+s^4$. Applying now theorem 1 the following new table is obtained as shown below.

C. Elizondo table

1	13	4
6	$(k-24)$	
$102-k$	24	
$-2592+126k-k^2$		
$e_{5,1}$		

To asymptotic stability of polynomial $p(s-a)$ it is required that all elements from first column be positive, from third row it is obtained that $k<102$ and from fourth row it is obtained $25.892<k<100.11$. So carried on, it is obtain that the polynomial $p(s-a)$ is asymptotic stable if and if $25.892<k<100.11$. It must be noted that element $e_{5,1}$ there is no need to calculate it, if positivity of elements where it is obtained that is $e_{4,1}=-2592+126k-k^2$ and $e_{4,2}=24$. In similar way it is obtained for the polynomial $p(-s-b)$: $25.892<k<37.333$. Using the conditions obtained for $p(s-a)$ and $p(-s-b)$ is obtained that polynomial $p(s)=s^4+10s^3+37s^2+(k+24)s+k$ has its roots in the left plane of the complexes between the values 1 and 4 if and only if $25.892<k<37.333$.

6 Conclusions

In this work, it presented the bound for roots for characteristics polynomials with real positive coefficients corresponding to a system LTI, a theorem was obtain to determine the restriction on the real part of the roots for this type of polynomials, including an example where this theorem in used to determine the gain limits.

Acknowledgments

The author wants to acknowledge to José Eduardo Viera Hernández for language support to improve this work.

References

1. J. Ackermann, Robust Control Systems with Uncertain Physical Parameters, Springer Verlag, 1993.
2. S. P. Bhattacharyya and H. Chapellat and L. H. Keel, Robust Control the Parametric Approach, Prentice Hall, 1995.
3. César Elizondo-González, Estabilidad y Controlabilidad Robusta de Sistemas Lineales con Incertidumbre Multilineal, Programa Doctoral de la Facultad de Ingeniería Mecánica y Eléctrica de la Universidad Autónoma de Nuevo León, 1999.
4. C. Elizondo-González, Necessary and Sufficient Conditions for Robust Positivity of Polynomic Functions Via Sign Decomposition, 3 rd IFAC Symposium on Robust Control Design ROCOND 2000, Prague Czech Republic, 2000.
5. César Elizondo-González, Robust Positivity of the Determinant Via Sign Decomposition, The 5th World Multi-Conference on Systemics Cybernetics and Informatics SCI 2001, Orlando Florida USA, 2001.
6. César Elizondo-González, A New Stability Criterion On Space Coefficients, Conferences on Decision and Control IEEE, Orlando Florida USA, 2001.
7. César Elizondo-González, Efraín Alcorta-García, Análisis de Cotas de Raíces de Polinomios Característicos y Nuevo Criterio de Estabilidad, Congreso Anual de la Asociación de México de Control Automático 2005, Cuernavaca, 19 al 21 de Octubre de 2005.
8. F. R. Gantmacher, The Theory of Matrices, Chelsea Publishing Company, 1990.
9. M. Marden, The Geometry of the Zeros of a Polynomial in a Complex Variable, American Mathematical Society, 1949.

Chapter 7

Supervision of Electrical Transformers

Efraín Alcorta García,

FIME, Universidad Autónoma de Nuevo León, Apdo. Postal 140-F, Cd. Universitaria, San Nicolas de los Garza, 66450 Nuevo León, México.
ealcorta@mail.uanl.mx

Abstract

Transformers are used in electric networks to reduce loses due to the i^2r factor in the transmission of power. Faults in transformer could cause severe damage and lost of energy supply in a big region. Consequently, the operation supervision of transformers is a very important issue. A common way to achieve fault supervision is due to current differential protection, which is based on signal processing and the static model of the transformer. Inrush or load currents could change the operation point of the transformer to the nonlinear zone, in which case a protection will operate, even if no fault is present. This work discus different model-based approaches and stress one of the methods in particular the so called observer-based approach. This approach makes use of a dynamic model of the transformer to be supervised in order to reduce the rate of false alarms. The main advantage of observer-based methods is the required stability is achieved.

Keywords: Transformer, fault, analytical redundancy, observers, protection.

1 Introduction

A very important element in order to reduce loses in power transmission is the electric transformer. By transforming to high voltage (and low currents) loses due to i^2r can be reduced. However a fault in the operation of the transformers can cause severe damage to the electric network and to the service.

There are many approaches in the technical literature for fault diagnostic of transformers, but in general these methods can be grouped in two: the ones operating off-line as well as the ones operating in real time. The so called off-line approaches test the transformer when it is not in use and in some cases is requires to unplugged it in order to carried on the test. Examples of these approaches are the dissolved gas in oil analysis (DGA) [6], the frequency response analysis [3], [13], [19], the impulse test [25] as well as some of its variants [16].

The more expanded approach to the fault detection in transformers in real time (on-line) is based on the so called current differential protection [9]. Note that this approach is based basically on signal processing as well as in a static model of the transformer, i.e. the dynamic behavior of the transformer is depreciated. This fact makes the differential protection approach sensible to external faults and to nonlinear effects o the transformer.

An alternative on-line approach for fault detection in transformers is based on the mathematical model of the transformer. In this kind of approaches the nonlinear effect of the transformer is considered in advance and no false alarms due to these facts are occurring. Because of model based approaches only test the modeled part of the system considered, external faults are not producing effect on the detection. However, no modeled dynamics or modeling errors can produce false alarms.

In this work, some of the more relevant model-based results for the supervision of transformers are reviewed. Basically are reviewed approaches based on data, like neural network based, see for example [22] and [26]. Methods based on algebro-differential models are also reviewed. In this sense an interesting approach has been proposed in [11], [18], [21], [23], [24], where the voltage equations re combined in order to obtain only linear differential equations and not nonlinear algebra-differential equations. The nonlinear effects are considered in a implicit form. However, as reported in [14], the system equations to be supervised require being asymptotically stable (or stabilized if a feedback is used in the system). As will be shown later, in the approaches in [11], [18], [21], [23] and [24] the linear system model to be supervised is critically stable (and not asymptotically stable). Consequently, the selection of a threshold is difficult and small perturbations or modeling errors can cause false alarms. In a recent work [12], [1], an approach based in models and nonlinear observers is proposed in such a way that the requirements of [14] are satisfied. This approach uses results about the model given in [1].

2 Preliminaries

2.1 Model Based Fault Diagnosis

Model based fault diagnosis can be realized in different ways, the one considering here is the based on observers, i. e. the required redundancy for fault detection and isolation is obtained by an observer. Fault detection can be due in two steps [7]:

- Residual generation. Signals depending on the faults are generated. Ideally a residual is cero (or close to zero) if no fault is present and different from zero if a fault is present.
- Residual evaluation. The information about faults is extracted from the residual.

In order to isolate a fault, is necessary to design a residual sensible to a fault but insensible to other fault. This kind of residual is called structured.

2.1.1 Structured Residual Design

The sensitivity to a certain fault can be obtained by the use of the so called "Unknown input observers" (UIO), see for example [4] and [10]. Here is considered the approach reported in [10].

Consider a linear time invariant system

$$\dot{x} = Ax + Bu + E_a f$$
$$y = Cx + Du$$

Where $x \in \Re^n$, is the state vector, $u \in \Re^p$, is the input vector, $y \in \Re^m$, is the output vector, $f \in \Re^s$ is the fault vector and A, B, C, D, E_a are constant matrices of appropriated dimensions.

The UIO is designed in two steps: first the nominal system is transformed in such a way that the fault effect on the state is divided. As a result two subsystems are obtained, one in which the faults of no interest are directly connected with the states (there are so many states as faults of no interest f_a) and a second subsystem in which there are not unknown inputs, except the states of the first subsystem and the faults of interest (the faults of interest are those for which the residual should be sensible, i.e. f_b). The states of the first subsystem should b obtained from the output of the system. So the second subsystem will depend only on known inputs except because of the faults of interest (f_b). A residual generator can be realized by an observer for the second subsystem. Remembering that the design of observers for fault detection is based on the nominal subsystem, i.e. a subsystem with fault set to zero. See figure 1. In the case in which the first state can not be obtained from the system output, the decoupling procedure should be applied recursively, see [10] for details.

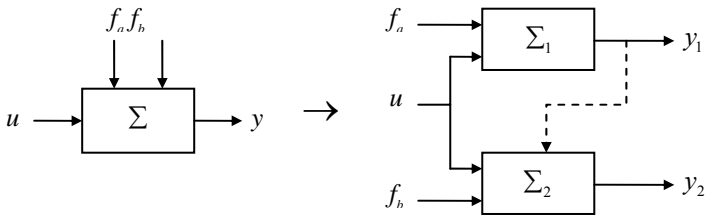


Fig. 1. Decoupling of the system for the design of unknown input observers.

The idea of using UIO's for fault detection has been used intensively; see for example the book [4]. Approaches for the nonlinear case can be found in [2].

2.1.2 Residual Evaluation

For the second step for fault detection, the residual evaluation, an evaluation function is required [15] as well as a threshold. A way to define an evaluation function is by using signal norms. The result of the evaluation function is compared with a threshold to decide when a fault has occurred. As an example of evaluation function consider the weighted sum of the absolute value of residual in a time window:

$$\Omega(r_k) = T\sum_{k=1}^{v} \omega_{k-i} r_{k-i}$$

For the threshold selection see [4] and [5].

2.2 Problem Formulation

Consider an electric single phase transformer and its model Σ with all no linear effects. Given the measurements of voltages and currents the task is to design a residual generator and evaluation step in order to have following result:

$$\Omega(r) < Th \qquad if\ f = 0$$
$$\Omega(r) \geq Th \qquad if\ f \neq 0$$

With *Th* the threshold function.

3 Transformer Model

In the literature on power systems there are different approaches to model the transformer, see for example [15]. Even if the many ways to obtain a model of a transformer presented in [15], here is considered the model as presented in [12]. That model is based on the equivalent electric circuit of the transformer and uses an approximation for the nonlinear effect of saturation and magnetic hysteresis. Different to early approaches presented in [15], the model considered in [12] is defined only by differential equations and not requires algebra-differential equations. This is paid with an error in the approximation. A schematic representation of the equivalent electric circuit for a single phase transformer is presented in figure 2.

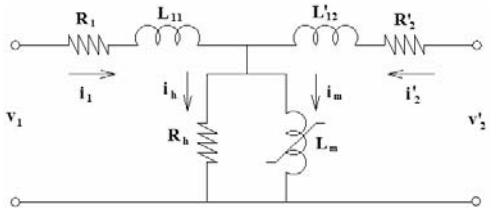


Fig. 2. Schematic representation of equivalent electric circuit of a single phase transformer.

In this work, as inputs are considered the voltages v_1 and v'_2 and as outputs the currents i_1 and i'_2. With this assignation the resulting model is multiple input - multiple output (MIMO). The corresponding equations of the system are given by:

$$\dot{x}(t) = Ax(t) + Bu(t) + \Phi(x(t))$$
$$y(t) = Cx(t)$$

with

$$x(t) = \begin{bmatrix} x_1(t) \\ x_2(t) \\ x_3(t) \end{bmatrix} = \begin{bmatrix} i_1(t) \\ i'_2(t) \\ i_m(t) \end{bmatrix};\ u(t) = \begin{bmatrix} u_1(t) \\ u_2(t) \end{bmatrix};$$

and the matrices:

$$A = \begin{pmatrix} \frac{R_1 + R_h}{L_{l1}} & -\frac{R_h}{L_{l1}} & \frac{R_h}{L_{l1}} \\ -\frac{R_h}{L'_{l2}} & \frac{R'_2 + R_h}{L'_{l2}} & \frac{R_h}{L'_{l2}} \\ \frac{R_h}{\alpha\beta} & \frac{R_h}{\alpha\beta} & -\frac{R_h}{\alpha\beta} \end{pmatrix};\ B = \begin{pmatrix} -\frac{1}{L_{l1}} & 0 \\ 0 & \frac{1}{L'_{l2}} \\ 0 & 0 \end{pmatrix};\ C = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$$

$$\Phi(x(t)) = \begin{pmatrix} 0 \\ 0 \\ \dfrac{\beta R_h}{\alpha}\left(x_1(t)x_3^2(t) + x_2(t)x_3^2(t) - x_3^3(t)\right) \end{pmatrix}$$

where the state x_1 correspond to the current in the primary, the state x_2 correspond to the current in secondary, the state x_3 correspond to the magnetizing current, $R_1>0$, $R'_2>0$ and $R_h>0$ are the resistance in the primary, in the secondary and due to the magnetization losses respectively, $L_{l1}>0$ and $L'_{l2}>0$ are the dispersion inductance in the primary and in the secondary respectively. The parameters α and β should be selected in such a way that the saturation curve of the core is approximated by $\lambda_m = \alpha \cdot \arctan(\beta \cdot i_m)$ [17].

3.1 Model Properties

The stability of the nonlinear model of the transformer will be studied in the next. Note that in the literature of fault diagnosis there is not equivalent result for nonlinear systems of the one given in [14] for the linear case, however it can be expected that nonlinear systems requires also such a condition.

In order to study the stability of the nonlinear model of the transformer, consider the transformer with input $u=0$, which equations are given by:

$$\dot{x}(t) = Ax(t) + \Phi\left(x(t)\right)$$
$$y(t) = Cx(t)$$

An equilibrium point of this equation is given by $x_e=0$. The main result respect to the model properties has been taken from [1]:

Theorem 1. *Consider the system*

$$\dot{x}(t) = Ax(t) + \Phi\left(x(t)\right)$$
$$y(t) = Cx(t)$$

Suppose that x(t)=0 is a stable equilibrium point of the system. Then the system is small-signal finite-gain Lp stable for each $p \in [1, \infty]$.

3.2 Fault Modeling

There are different kind of faults that could affect the electrical transformers: short circuit between turns, short circuits turn to earth, incipient faults (incipient short circuits because of isolation degradation). The model associated to the faults will not present here, because it is not required for fault detection. However, the model is required in order to have a simulation results given in the next sections.

The transformer fault model considered in this work has been taken from [27]. In the case of short circuit fault turn to earth, the resultant faulty model is similar to a transformer with three windings. In the case of short circuit between turns will reduce the resistance and inductance and in general. The short circuit will produce a big current in the short circuit and some damage could occur. For incipient faults the short circuit is partial and the loop is closed using a Resistance-capacitor (RC) circuit.

4 Analytical Redundancy Approaches to Fault Detection in Transformers

4.1 Neural Networks

In this case the transformer model is obtained from measurements of the transformer variables (voltages and currents). With the data a neural network is trained in order to reproduce the nominal (fault free) behavior of the transformer. Te residuals are formed as the difference between the measurements and the estimated values of the outputs. Fault detection is realized by classification of the residual behavior using wavelets [26]. A second approach use neural networks for the

design of an observer based on the neuro-fuzzy theory and dynamic neural networks [22]. Most of the results presented using neural networks are based on experimental verification, however no formal proof of convergence are given.

4.2 Flux Restrained Current Differential Relay

The idea is to computes and uses flux-current relationship of the transformer to obtain the restraint function. Computation of flux is based on the availability of voltage signals to the transformer relay. In a computer based hierarchical protection system, the voltages will be available as shared data from other protective devices in the substation. It is shown that this protection technique requires fewer computations as compared to the harmonic current computation, and hence can be implemented on a microcomputer of modest capability [18].

The principle has been tested in the laboratory on a model power transformer [18], however the conditions reported in [14] about the stability requirements are not complete satisfied. The initial condition is actually not a problem; however the resulting equations of the approach are critically stable. Here small model uncertainties will produce false alarms. This fact is not explicitly reported, however from the results in [18] can be appreciated.

4.3 Inverse Inductance Approach

This approach considers the representation of a transformer as a universal equivalent circuit composed of inverse inductances. Analysis of the inductances and inverse inductances matrices of actual transformers show that the values for the equivalent circuit in the case of internal faults are remarkably different from those of magnetizing inrush currents. By sampling voltage and current in each transformer terminal the elements can be calculated using an algorithm [11].

The parameters on-line calculation of the equivalent model from the samples uses a model which is not stable. Actually is critically stable. This means that the result from [14] about the requirements of the residual is again not satisfy.

4.4 Microprocessor based

The basic idea is almost similar to the above approaches: the model of the transformer is used. The start point of each of the above approaches is the same model; however, the different approaches manipulate the model in a different way to obtain the result. In the above approaches the manipulation of the equations is done in such a way that the resulting variables have a physical meaning (inductance, inverse inductance, flux). The manipulation used in this approach is more abstract: the equations are combined in order to obtain a new one which is free of the nonlinear effect of the transformers (this effect is implicitly considered) [21].

Unfortunately, the resulted reduced equations are no stable (actually are critically stable) and the detection is sensible to model uncertainties, as pointed out by [14].

4.5 Observer-based approach

The basic idea is to use the model of the transformer in order to estimate the nominal (fault free) output of the system in order to compare it with the actual output. If no fault is present in the transformer the resulting signal should be close to zero and very different from zero if a fault is present. This idea has been widely used; see for example the book [4]. In the case of nonlinear systems, like the transformer case, a review of methods could be found in [2].

The required transformer output estimation is realized via a non linear observer [2], which uses the mathematical model of the transformer. Consider the transformer model and the following assumptions:

- $\Phi(x(t))$ is Lipschitz, i.e.
  $\|\Phi(x_1(t)) - \Phi(x_2(t))\| \le \delta \|x_1 - x_2\|$. Where δ is the Lipschitz constant.
- The pair *(A,C)* is observable

The output estimator for the transformer could be defined as follows:

$$\dot{\hat{x}}(t) = A\hat{x}(t) + Bu(t) + \Phi(\hat{x}(t)) + L\left(y(t) - C\hat{x}(t)\right)$$
$$\hat{y}(t) = C\hat{x}(t)$$

In order to analyze the convergence let us define the estimation error is $e(t) = x(t) - \hat{x}(t)$. The dynamic of the estimation error could be obtained as:

$$\dot{e}(t) = (A - LC)e(t) + \Phi(x(t)) - \Phi(\hat{x}(t))$$
$$r(t) = Ce(t)$$

A sufficient condition for the convergence of the estimation error to zero is given by the following result:

Theorem 2. [20] If a gain matrix L can be chosen such that:

$$(A - LC)^T P + P(A - LC) + \gamma^2 PP + I < 0$$

For some positive definite matrix P, then this choice of L leads to asymptotically stable estimates by the observer for the system.

Remark 1. As show by [20], the above inequality can be re-written as a linear matrix inequality.

The observer-based approach has some grade of robustness; it means that the estimated output will be insensible to small transformer faults. Note that even if the Lipschitz constant has some value, it still possible to set other value (bigger) in order to have more robustness. It will reduce the sensitivity to small faults, but will make possible to support some perturbations.

5 Conclusions

In this work, different model-based approaches to fault detection in electrical transformers are discussed and the advantages of each method are reviewed. Because of the importance of the model, a novel approach to model electric transformers is discussed. The properties of the considered model are also presented. One of the principal advantages of the considered model is that only differential equations are required and not algebro-differential ones as in other models.

The last method discussed corresponds also to the more recent proposed in the literature. This approach is an observer-based one and corresponds to the analytical approaches to fault diagnosis. The main advantage is that nonlinear effect of the transformer is not affecting the detection, i.e. no false alarms are due to the nonlinearities of the transformer. A more reliable discrimination between internal transformer faults and external faults affecting the behavior of the transformer can be done by using observer-based methods. More that a competition with the "popular" differential current approach for transformer protection, the model-base method represents a complementary approach can be used in order to improve the fault detection task.

Acknowledgments

The author wants to acknowledge the grant support from CONACYT and PAICYT-UANL (the last under grant CA1286-06).

References

1. E. Alcorta García, C. Elizondo González, C. Pérez Rojas, A. Ávalos González. Diagnóstico de fallas en transformadores eléctricos. Conferencia Nacional de la Asociación de México de Control Automático, Cd. de México Octubre 19-21. (2006)
2. E. Alcorta García, P. M. Frank. Deterministic nonlinear observer based approaches to fault diagnosis: A survey. Control Eng. Practice, 5(5):663–670, May (1997).
3. J. Bak-Jensen, B. Bak-Jensen, S. D. Mikkelsen. Detection of faults and ageing phenomena in transformers by transfer functions. IEEE Transactions on Power Delivery, 10(1):308–314, Jan. (1995)
4. J. Chen, R. Patton. Robust model-based fault diagnosis for dynamic systems. Kluwer Academic publishers, Boston, (1999).
5. X. Ding, P. M. Frank. Frequency domain approach and threshold selector for robust model-based fault detection and isolation. In IFAC/IMACS Symposium SAFEPROCESS, pages 307–312, (1991).
6. M. Duval. A review of faults detectable by gas-in-oil analysis in transformers. IEEE Electrical Insulation Magazine, 18(3):8–17, May/Jun (2002).
7. P. M. Frank. Fault diagnosis in dynamic systems using analytical and knowledge-based redundancy - a survey. Automatica, 26:459–474, (1990).
8. J. Gertler. Analytical redundancy methods in fault detection and diagnosis. In IFAC / IMACS Symp. SAFEPROCESS, Baden-Baden, Germany, pages 9–21, (1991).
9. A. Guzman, S. Zocholl, G. Benmouyal, H. Altuve. A current based solution for transformer differential protection–part II: Relay description and evaluation. IEEE Trans. on Power Delivery, 17(4):886–893, October (2002).
10. 10.M. Hou, P. C. Müller. Fault detection and isolation observers. International Journal of Control, 60:827–846, (1994).

11. K. Inagaki, M. Higaki, Y. Matsui, K. Kurita, M. Susuki, K. Yoshida, T. Maeda. Digital protection method for power transformers based on an equivalent circuit composed of inverse inductances. IEEE Transactions on Power Delivery, 3(4):1501–1510, Oct. (1988).
12. F. M. Jorge Zavala, E. Alcorta García. Detection of internal faults in transformers using observers. IEEE-CCA03 Conference on Control Applications, pag. 195–199, July (2003).
13. J.-W. Kim, B. K. Park, S. C. Jeong, S. W. Kim, P. G. Park. Fault diagnosis of a power transformer using an improved frequency response analysis. IEEE Transactions on Power Delivery, 20(1):169–178, Jan. (2005).
14. M Kinnaert, R. Hanus, Ph. Arte. Fault detection and isolation for unstable linear systems. IEEE Transaction on Automatic Control, 40(4):740–742, April (1995).
15. J. A. Martínez, B. A. Mork. Transformer modeling for low- and mid- frequency transients-A review. IEEE Transactions on Power Delivery, 20(2):1625–1632, April (2005).
16. S. K. Pandey, L. Santish. Multiresolution signal decomposition: a new tool for fault detection in power transformers during impulse tests. IEEE Transactions on Power Delivery, 13(4):1194–1200, Oct. (1998).
17. C. Pérez Rojas. Fitting saturation and hysteresis via arctangent functions. IEEE Power Engineering Review, 20(11):55–57, November (2000).
18. A. G. Phadke, J. S. Thorp. A new computer-based flux restrained current differential relay for power transformer protection. IEEE Transactions on Power Apparatus and Systems, 102(11):3624–3629, Nov. (1983).
19. E. Rahimpour, J. Christian, K. Feser, H. Mohseni. Transfer function method to diagnose axial displacement and radial deformation of transformer windings. IEEE Transactions on Power Delivery, 18(2):493–505, April (2003).
20. R. Rajamani, Y. Cho. Observer design for nonlinear systems: stability and convergence. In 34rd IEEE Conference on Decision and Control, volume New Orleans, LA USA, pages 93–94, (1995).
21. M. S. Sachdev, T. S. Sidhu, H. C. Wood. A digital relaying algorithm for detection of transformer winding faults. IEEE Trans. on Power Delivery, 4(3):1638–1648, July (1989).
22. R. Shoureshi, T. Norick, D. Linder, J. Work. Electric power transformer diagnosis using neural-based observer. American Control Conf., pag. 2276–2281, June 4-6 (2003).
23. T. S. Sidhu, M. S. Sachdev. On-line identification of magnetizing inrush and internal faults in three-phase transformers. IEEE Trans. on Power Delivery, 7(4):1885–1891, Oct (1992).
24. T. S. Sidhu, M. S. Sachdev, H. C. Wood. Microprocessor based relay for protecting power transformers. IEE Proceedings, Part C Generation, Transmission and Distribution, 137(6):436–444, Nov. (1990).
25. L. R. Stuffle, R. E. Stuffle. A computerized diagnostic technique applicable to hv impulse test of transformers. IEEE Trans. on Power Delivery, 5(2):1007–1012, April (1990).
26. H. Wang and K. L. Butler. Detection of transformers winding faults using wavelet analysis and neural network. In IEEE International Conference on Intelligent System applications to Power Systems, April (1999).
27. H. Wang and K. L. Butler. Modeling transformers with internal incipient faults. *IEEE Transactions on Power Delivery*, 17(2):500–509, February 2002.

Chapter 8

Linear and Nonlinear Control Strategies to Stabilize a Vtol Aircraft: Comparative Analysis

Guillaume Sanahuja, Pedro Castillo, Octavio Garcia, and Rogelio Lozano

Université de Technologie de Compiègne, CNRS, Heudiasyc. B.P. 20529, Compiègne Cedex 60205, France. {gsanahuj, castillo, ogarcias, rlozano}@hds.utc.fr

Abstract

In this chapter a comparative analysis of a linear and nonlinear control laws to stabilize a VTOL aircraft is presented. A linear control law using LQR method is obtained and compared with respect to three nonlinear control strategies obtained using the well-known backstepping technique and the saturation functions. The performance of the controllers is compared by simulation but also in real-time experiences. The robustness of the control algorithms with respect to aggressive perturbations is illustrated with real-time experiments.

Keywords: UAVs, nonlinear control, linear control**,** real-time application.

1 Introduction

This Unmanned Aerial Vehicles (UAVs) have been referred to in many ways: RPV (remotely piloted vehicle), drone, robot plane, and pilotless aircraft are a few such names. Most often called UAVs, they are defined by the USA Department of Defense (DOD) as powered, aerial vehicles that do not carry a human operator, use aerodynamic forces to provide vehicle lift, can fly autonomously or be piloted remotely, can be expendable or recoverable, and can carry a payload. There are a number of reasons why UAVs have only recently been given a higher priority. Technology is now available that wasn't available just a few short years ago. Some say that the services' so-called ``silk scarf syndrome" of preferring manned aviation over unmanned, has diminished as UAVs entered the mainstream. UAVs might have gained momentum earlier if a crisis had occurred, such as an extreme shortage of surveillance and reconnaissance aircraft during a conflict. The lack of such a crisis, along with the paradigm shift that needed to occur before unmanned vehicles were accepted, meant that UAVs have evolved as technology has become available.

More recently, a growing interest for unmanned aerial vehicles has been shown among the research community [2], [10], [7], [9], [13]. We focus our study on the Planar Vertical Take Off and Landing (PVTOL) aircraft. The system of the PVTOL (Planar Vertical Take-Off and Landing) aircraft is based on a mathematical model of simplified plane having a minimal number of states and inputs, which involves in a vertical plane. It also represents the longitudinal mode of the helicopter. The PVTOL aircraft is a topic of interest of the automatic community for these applications and its nonlinear behavior. Several methodologies for controlling the forces and moments have been proposed in the literature to control the PVTOL.

An algorithm to control the PVTOL based on an approximate I-O linearization procedure was proposed in [9]. Their algorithm achieves bounded tracking and asymptotic stability. A non-linear small gain theorem was proposed in [5] which can be used to stabilize a PVTOL. He has proved the stability of a controller based on nested saturations. An extension of the algorithm proposed by [9] was presented in [20]. They were able to find a flat output of the system that was used for tracking control of the PVTOL in presence of unmodeled dynamics. The forwarding technique developed in [21] was used in [12] to propose a control algorithm for the PVTOL. This approach leads to a Lyapunov function which ensures asymptotic stability. Other techniques based on linearization were also proposed in [12].

Marconi [13] proposed a control algorithm of the PVTOL for landing on a ship whose deck oscillates. They designed an internal-model-based error feedback dynamic regulator that is robust with respect to uncertainties. Olfati-Saber [16] presented an algorithm to stabilize a VTOL aircraft with a strong input coupling using a smooth static state feedback. In recent works, see [22] and [23], the PVTOL control is based in the knowledge of the parameter ε (coupling between the rolling moment and the lateral acceleration of the PVTOL). It is well known that, this parameter is very small $(\varepsilon << 1)$ and not always well-known. In addition, from our real-time experience this parameter can be tuned experimentally to be very

small but not known. Recently, an extension of the control algorithm proposed by [5] was presented in [2]. They showed in real-time experiences the performance of their controller.

In this chapter we present a comparative analysis of a linear and the most common nonlinear control strategies to stabilize a VTOL aircraft. The controllers have been tested in numerical simulations, but also in a real-time application. We applied these control algorithms to control the roll angle and the horizontal displacement of a radio-controlled electrical quad-rotor helicopter.

The chapter is organized as follows: in Section 2 the system dynamics of a VTOL aircraft are presented. Section 3 is devoted to establish the control strategies. In section 4, we compared the performance of the controllers. Experimental results are shown in Section 5, and conclusion is given in 6.

2 Problem Statement

The mathematical model to describe the dynamics of a VTOL aircraft assuming small angles around a given position, can be obtained by representing the aircraft as a solid body evolving in a three dimensional space and subject to the thrusts and torques, and using a classical method: The Euler-Lagrange approach.

$$L(q,\dot{q}) = T_{trans} + T_{rot} - U \tag{1}$$

where $T_{trans} = \frac{m}{2}\dot{\xi}^T\dot{\xi}$ is the translational kinetic energy, $T_{rot} = \frac{1}{2}\Omega^T I\Omega$ is the rotational kinetic energy, $U = mgz$ is the potential energy of the rigid object, z is the object altitude, m denotes the mass of the rigid object, and g is the acceleration due to gravity, $q = \begin{pmatrix}\xi & \eta\end{pmatrix}^T$ is the vector of the generalized coordinates of the aircraft.

The model of the full aircraft dynamics is obtained from Euler-Lagrange equations with external generalized forces

$$\frac{d}{dt}\frac{\partial L}{\partial \dot{q}} - \frac{\partial L}{\partial q} = \begin{bmatrix} F_\xi \\ \tau \end{bmatrix} \tag{2}$$

where $F_\xi = R\widehat{F} \in \Re^3$ is the translational forces applied to the rotorcraft, $\tau \in \Re^3$ represents the yaw, pitch, and roll moments, and R denotes the rotational matrix $R(\psi,\theta,\phi) \in SO(3)$ representing the orientation of the aircraft relative to a fixed inertial frame. Since the Lagrangian contains no cross terms in the kinetic energy combining $\dot{\xi}$ with $\dot{\eta}$ the Euler-Lagrange equation can be partitioned into dynamics for ξ coordinates and the η coordinates. The Euler-Lagrange equation for the translation motion is

$$\frac{d}{dt}\left[\frac{\partial L_{trans}}{\partial \dot{\xi}}\right] - \frac{\partial L_{trans}}{\partial \xi} = F_\xi, \tag{3}$$

As for the η coordinates we can rewrite:

$$\frac{d}{dt}\left[\frac{\partial L_{rot}}{\partial \dot{\eta}}\right] - \frac{\partial L_{rot}}{\partial \eta} = \tau, \tag{4}$$

where $L_{trans} = T_{trans} - U$, and, $L_{rot} = T_{rot}$. One example of a VTOL aircraft is the well known PVTOL aircraft, see Figure 1. The PVTOL is a mathematical model of a flying object that evolves in a vertical plane, see [2]. This aircraft has three degrees of freedom, (x,z,ϕ) corresponding to its position in the plane and roll angle. The PVTOL has two independent thrusters that produce a force and a moment and thus it is underactuated since it has three degrees of freedom and only two inputs.

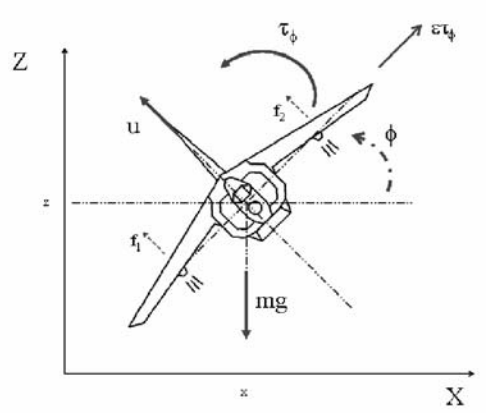


Fig. 1. *The PVTOL aircraft.*

From figure 1 and using the Lagrangian and the general forces applied to the aircraft, we can obtain the basic equations of motion for the PVTOL aircraft (see [9]):

$$\begin{aligned} m\ddot{x} &= -u\sin\phi + \varepsilon\tau_\phi\cos\phi \\ m\ddot{z} &= u\cos\phi + \varepsilon\tau_\phi\sin\phi - mg \\ m\ddot{\phi} &= \tau_\phi \end{aligned} \tag{5}$$

where x, z, denote the horizontal and the vertical position of the aircraft center of mass, m is the mass of the aircraft and ϕ is the roll angle that the aircraft makes with the horizon. The control inputs $u = f_1 + f_2$ and τ_ϕ are the thrust (directed out the bottom of the aircraft) and rolling moment. g is the gravitational acceleration. The parameter ε is a small coefficient which characterizes the coupling between the rolling moment and the lateral acceleration of the aircraft. Neglecting ε, the simplified model is, see [16] :

$$\begin{aligned} m\ddot{x} &= -u\sin\phi \\ m\ddot{z} &= u\cos\phi - mg \\ m\ddot{\phi} &= \tau_\phi \end{aligned} \tag{6}$$

3 Control Strategies

In this section four control strategies to stabilize a VTOL aircraft are proposed. The first control law is a linear control strategy obtained using the LQR method. We next propose a nonlinear control strategy using the backstepping approach and finally two nonlinear control strategies using saturation functions and the Lyapunov analysis are presented. Note that, the PVTOL aircraft is an underactuated system with three DOF and only two inputs. To stabilize this aircraft first, we will stabilize the altitude forcing it to satisfy the dynamics of a linear system, then we propose:

$$u = \frac{-a_1\dot{z} - a_2(z - z_d) + mg}{\cos\phi} \tag{7}$$

where a_1, a_2 are positive constant and z_d is the desired altitude. Introducing (7) into (6) we obtain :

$$m\ddot{y} = -a_1\dot{z} - a_2(z - z_d) \tag{8.1}$$

$$m\ddot{x} = -(-a_1\dot{z} - a_2(z - z_d) + mg)\tan\phi \tag{8.2}$$

$$\ddot{\phi} = \tau_\phi \tag{8.3}$$

Note from (8.1) that for a time T large enough, y and $\dot{y}$ are arbitrarily small and (8.2) reduces to :

$$\ddot{x} = -g \tan \phi \tag{9}$$

The comparative analysis of the control laws will been done for the subsystem (x, ϕ).

3.1 Linear control law

Linearizing the subsystem (x, ϕ), i.e., (9)-(8.3), we obtain:

$$\begin{aligned} \ddot{x} &= -g\phi \\ \ddot{\phi} &= \tau_\phi \end{aligned} \tag{10}$$

Defining $X = (x_1, x_2, \varphi_1, \varphi_2)^T$, we obtain:

$$\dot{X} = AX + B\tau_\varphi \tag{11}$$

with:

$$A = \frac{1}{2}\begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & -g & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \qquad B = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} \tag{12}$$

Propose:

$$\tau_\phi = -KX \tag{13}$$

such that,

$$\dot{X} = (A - BK)X = \bar{A}X \tag{14}$$

where $\bar{A} = (A - BK)$. Note that, $K \in \Re_{1,4}$ is obtained by the Matlab command $place(A, B, P)$, where $P \in \Im_{4,1}$ is the matrix containing the desired poles. Choosing $P = (-1+i, -1-i, -1+i/2, -1-i/2)^T$, we obtain $K = (-0.2548, -0.6626, 7.25, 4)$. Thus, the proposed control law yields :

$$\tau_{L_\varphi} = 0.2548\mathrm{x}_1 + 0.6626\mathrm{x}_2 - 7.25\varphi_1 - 4\varphi_2 \tag{15}$$

3.2 Nonlinear Control Law

Backstepping technique

Rewriting (9)-(8.3), we have

$$\begin{aligned} \dot{x}_1 &= x_2 \\ \dot{x}_2 &= -g \tan \phi_1 \\ \dot{\phi}_1 &= \phi_2 \\ \dot{\phi}_2 &= \tau_\phi \end{aligned} \tag{16}$$

Define the error e_1 as

$$e_1 = x_1 - x_1^d, \tag{17}$$

and propose the following positive function

$$V_1 = \frac{k_1}{2} e_1^2 \tag{18}$$

where $k_1 > 0$ is a constant, we have

$$\dot{V}_1 = k_1 e_1 (\dot{x}_1 - \dot{x}_1^d) = k_1 e_1 (x_2 - x_2^d) \tag{19}$$

Let define x_2^v as the virtual control input, such that,

$$x_2^v = x_2^d - e_1 \tag{20}$$

then, $\dot{V}_1$ yields

$$\dot{V}_1 = k_1 e_1 (x_2 - x_2^v - e_1) = -k_1 e_1^2 + k_1 e_1 (x_2 - x_2^v) \tag{21}$$

Now, define the error e_2 as

$$e_2 = x_2 - x_2^v \tag{22}$$

such that, we get

$$\dot{V}_1 = -k_1 e_1^2 + k_1 e_1 e_2 \tag{23}$$

Propose the following positive function

$$V_2 = \frac{k_2}{2} e_2^2 \tag{24}$$

Differentiating the above, we obtain

$$\dot{V}_2 = k_2 e_2 (\dot{x}_2 - \dot{x}_2^V) = k_2 e_2 (-g \ \tan\phi_1 - \dot{x}_2^V) \tag{25}$$

Defining the virtual control input $(g \ \tan\phi_1)^v$, we get

$$(g \ \tan\phi_1)^v = \delta_1^v = -\dot{x}_2^v + \frac{k_1}{k_2} e_1 + e_2 \tag{26}$$

and $\dot{V}_2$ yields

$$\dot{V}_2 = -k_2 e_2^2 - k_1 e_1 e_2 + k_2 e_2 (\delta_1^v - g \ \tan\phi_1) \tag{27}$$

Define the error e_3

$$e_3 = \delta_1^v - g \ \tan\phi_1 \tag{28}$$

then, V_2 yields :

$$\dot{V}_2 = -k_2 e_2 - k_1 e_1 e_2 + k_2 e_2 e_3 \tag{29}$$

Now, we propose the following positive function

$$V_3 = \frac{k_3}{2} e_3^2 \tag{30}$$

Differentiating V_3, we have

$$\dot{V}_3 = k_3 e_3 \dot{e}_3 = k_3 e_3 (\delta_1^v - g(1 + \tan^2 \phi_1)\phi_2) \tag{31}$$

Let us define $(g(1 + \tan^2 \phi_1)\phi_2)^v$ as the virtual control input, such that

$$(g(1+\tan^2\phi_1)\phi_2)^v = \delta_2^v = \dot{\delta}_1^v + \frac{k_2}{k_3}e_2 + e_3 \tag{32}$$

then, $\dot{V}_3$ yields

$$\dot{V}_3 = -k_3e_3^2 - k_2e_3e_2 + k_3e_3(\delta_2^V - g(1+\tan^2\phi_1)\phi_2) \tag{33}$$

Defining the error e_4 as

$$e_4 = \delta_2^v - g(1+\tan^2\phi_1)\phi_2 \tag{34}$$

such that, V_3 yields

$$\dot{V}_3 = -k_3e_3^2 - k_2e_3e_2 + k_3e_3e_4 \tag{35}$$

Proposing the following positive function

$$V_4 = \frac{k_4}{2}e_4^2 \tag{36}$$

thus,

$$\dot{V}_4 = k_4e_4(\dot{\delta}_2^v - g(1+\tan^2\phi_1)(\tau_\varphi + 2\phi_2^2\tan\phi_1)) \tag{37}$$

From the above, we can propose a control input τ_t, such that $\dot{V}_4 = -k_4e_4^2 - k_3e_4e_3$. So, this control law is given by

$$\tau_{B\phi} = \frac{1}{g(1+\tan^2\phi_1)}(\dot{\delta}_2^v + \frac{k_3}{k_4}e_3 + e_4) - 2\phi_2^2\tan\ \phi_1 \tag{38}$$

Now, we propose a Lyapunov function as

$$V = V_1 + V_2 + V_3 + V \tag{39}$$

and $\dot{V}$ yields

$$\dot{V} = -k_1e_1^2 - k_2e_2^2 - k_3e_3^2 - k_4e_4^2 \le 0 \tag{40}$$

Rewriting (38), we have

$$\begin{aligned}\tau_{B_\phi} = &\frac{1}{g(1+\tan^2\phi_1)}(\dddot{x}_2^d - 4\ddot{x}_2^d - \bar{k}_1\dot{x}_2^d + \bar{k}_2(x_2 - x_2^d) \\ &+\bar{k}_3(x_1 - x_1^d) - \bar{k}_4 g\ \tan\phi_1 - 4g(1+\tan^2\phi_1)\phi_2) - 2\phi_2^2\tan\phi_1\end{aligned} \tag{41}$$

where

$$\begin{aligned}&\bar{k}_1 = \bar{k}_4 = k_\phi + 6\ , \quad \bar{k}_2 = 2k_\phi + 4 \\ &\bar{k}_3 = k_\phi + \frac{k_3k_1}{k_4k_2} + 1\ ,\ k_\phi = \frac{k_1}{k_2} + \frac{k_2}{k_3} + \frac{k_3}{k_4}\end{aligned} \tag{42}$$

Nested Saturation Functions Technique

In this subsection we present the nonlinear control law proposed in [2], to stabilize the sub-system (9)-(8.3). Note that, the control strategy is based on nested saturation and the analysis of convergence is done using the Lyapunov analysis, more details see [2]. In addition, the control strategy is such that, ensure a small bound on $|\phi|$ in such a way that $\tan\phi \approx \phi$.

$$\tau_{NS\phi} = -\sigma_{\phi 1}(\dot{\phi} + \sigma_{\phi 2}(\phi + \dot{\phi} + \sigma_{\phi 3}(2\phi + \dot{\phi} - \frac{\dot{x}}{g} + \sigma_{\phi 4}(\dot{\phi} + 3\phi - 3\frac{\dot{x}}{g} - \frac{x}{g})))) \tag{43}$$

Saturation Functions Technique

Recently, in [19] a nonlinear control strategy to stabilize a PVTOL aircraft using saturation functions has been proposed. The control algorithm is also obtained using the Lyapunov analysis but the final control law does not have nested saturation. The performance of the controller is better than the other one using nested saturation functions. Then, from [19] we have

$$\tau_{S\phi} = -\sigma_{b4}(k_4 z_4) - \sigma_{b3}(k_3 z_3) - \sigma_{b2}(k_2 z_2) - \sigma_{b1}(k_1 z_1) \tag{44}$$

with

$$\begin{aligned} &z_4 = \dot{\phi}, && z_3 = k_4\phi + \dot{\phi} \\ &z_2 = z_3 + k_3\dot{\phi} - \frac{k_3 k_4}{g}\dot{x}, && z_1 = z_2 + k_2\phi - \frac{k_2 k_3 k_4}{g}x - \frac{k_2(k_3 + k_4)}{g}\dot{x} \end{aligned} \tag{45}$$

Table 1. *Gain values used in the control laws (15), (41), (43) and (44)*

Controller	Control parameters	Value
Linear ($\tau_{L\phi}$)	k_1	0.2548
	k_2	0.6626
	k_3	7.25
	k_4	4
Backstepping ($\tau_{B\phi}$)	$\bar{k}_1 = \bar{k}_4$	9
	$\bar{k}_2$	10
	$\bar{k}_3$	5
Nested saturations ($\tau_{NS\phi}$)	b_1	0.523
	b_2	0.261
	b_3	0.13
	b_4	0.06
Separated saturations ($\tau_{S\phi}$)	b_1	0.06
	b_2	0.1
	b_3	0.17
	b_4	0.193
	k_1	1
	k_2	0.1
	k_3	1
	k_4	1

4 Simulation Results

In this section, we present the simulation results when applying the control strategies (15), (41), (43) and (44) to stabilize the pitch angle and the horizontal displacement of the PVTOL aircraft. The parameters used for the different controllers are shown in Table 1. The values for the saturated controllers have been chosen such that the input τ_ϕ is bounded by the same value for the two controllers (see Table 1), i.e.

$$\begin{aligned} \left|\tau_{NS\phi}\right| &\le b_{1NS\phi} = 0.523 \\ \left|\tau_{S\phi}\right| &\le b_{S\phi} = b_1 + b_2 + b_3 + b_4 = 0.523 \end{aligned} \tag{46}$$

Figures 2-3 show the performance when applying the controllers, (15), (41), (43) and (44), to stabilize the (x,ϕ) subsystem of the PVTOL aircraft. The initial conditions were: $x(0)=3,\ \dot{x}(0)=0.5,\ \phi(0)=\pi/4,\ \dot{\phi}(0)=-1$. Note that, the convergence using a classical linear controller or a backstepping controller is faster than the others controllers. This characteristic is normal because we choose the bound of the saturation so small to guarantee a convergence when the initial positions are so far to the desired positions (see Figure 4). To increase the convergence speed of these controllers, we need to increase the bound of the saturation. From Figure 3, note that, the input $\tau_{\gamma\phi}\ \forall\gamma = L,B$ is not bounded, and it takes high value which is not realistic for experimentation. Therefore, using the controllers $\tau_{\gamma\phi}\ \forall\gamma = NS,S$ the control input is bounded. Note from Figure 2 that, the convergence is faster with the separated saturation functions than the nested ones. Moreover, Figure 4 shows that when the system is far from the equilibrium point then the linear controller diverges.

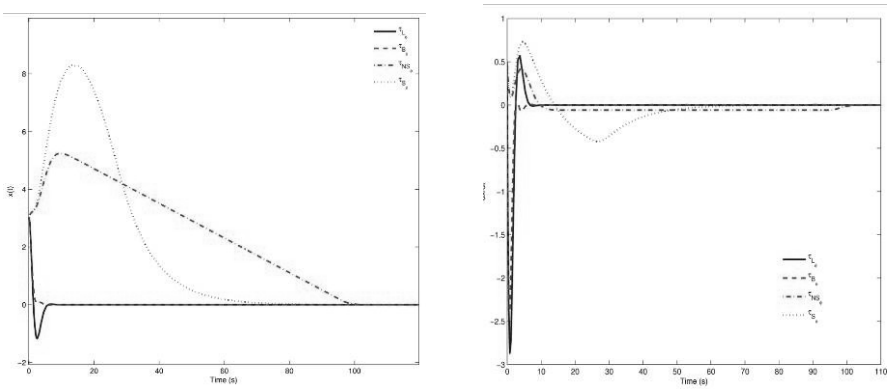


Fig. 2. *States* $x(t)$ *and* $\dot{x}(t)$ *of the PVTOL aircraft. The initial conditions were:* $x(0)=3, \dot{x}(0)=0.5$

In Figures 2-5 the solid line represent the linear control law, $\tau_{L\phi}$, the dash line represent the backstepping control strategy, $\tau_{B\phi}$, the dashdot line represent the nonlinear control law using nested saturation functions, $\tau_{S\phi}$ and the dot line represent the nonlinear control algorithm using separated saturation functions, $\tau_{S\phi}$. The initial conditions for figures 4 and 5, were: $x(0)=100, \dot{x}(0)=0.5, \phi(0)=\pi/4, \dot{\phi}(0)=-1$. Provided $|\phi|<\pi/2$ (see Figure 4), the backstepping controller is always able to reach the x desired position, therefore, it is not enough to reach the angular desired position, see Figure 5.

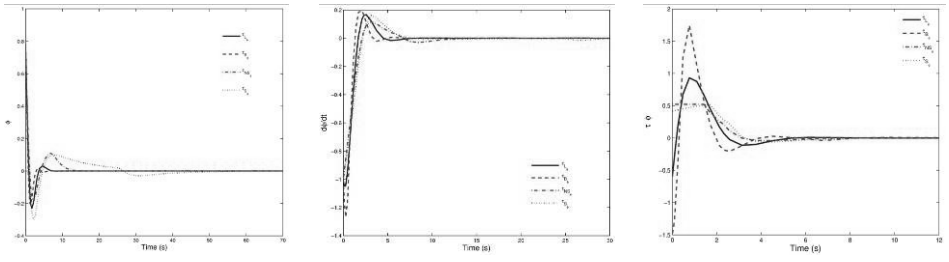


Fig. 3. *States* ϕ, $\dot{\phi}$ *and control input* τ_ϕ *of the PVTOL aircraft. The initial conditions were:* $\phi(0)=\pi/4,\ \dot{\phi}(0)=-1$.

Note also from this figure, that the angular velocity is not bounded using a Linear controller and a backstepping controller. Figure 5 shows that using the backstepping controller the control input is bigger than these one using the bounded controllers. This fact is interesting in real applications because it shows that using a backstepping controller the system could

diverge before to reach the desired position. Notice that, the time scale is not the same in the Figures. This fact is due to show more details of the figure.

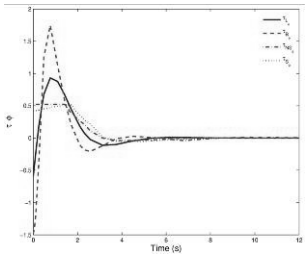


Fig. 5. *Control input* τ_ϕ *of the PVTOL aircraft.*

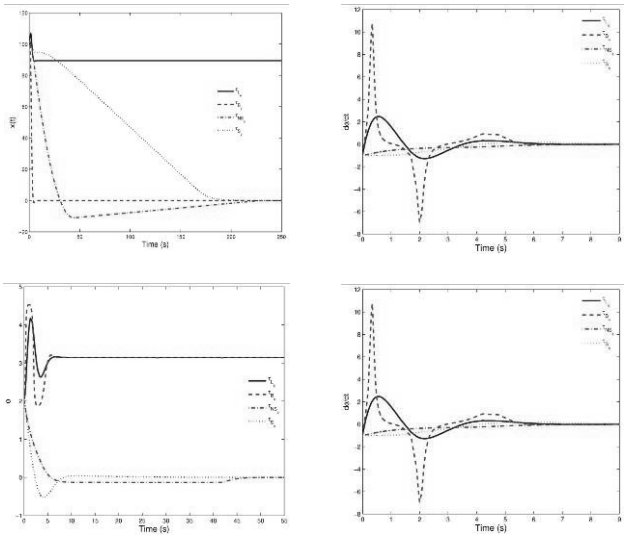


Fig. 4. *States* $x(t)$, $\dot{x}(t)$, $\phi(t)$ *and* $\dot{\phi}(t)$ *of the PVTOL aircraft.*

5 Experimental Results

In this section we present the real-time experimental results obtained when applying the proposed controllers, (15), (41), (43) and (44), to a quad-rotor. The experimental platform is composed of a Draganflyer helicopter, a Futaba radio, two PCs and a three-dimensional tracker system (Polhemus) for measuring the position and orientation of the quad-rotor. The software employed is the real-time XPc-Target from Matlab. In our experiment the pitch and yaw angles are controlled using a controller proposed in [2]. The throttle input is controlled manually while the roll and lateral position are controlled using the strategies presented in this chapter. When applying the linear control law using the nominal parameters, see Table 1, the helicopter cannot make hover. We have tuned these parameters, see Table 2, by trial and error until obtain a desired performance, see Figure 6. The desired values are $x_d = 0$ cm and $\phi_d = 0°$. Note that k_2 and k_4 are so small, this fact is because the helicopter has three onboard gyros that help stabilize the mini-rotorcraft. Thus, these three angular velocity feedbacks provided by the on-board gyros are considered as an in-built inner control loop. Notice that this gyro stabilization is not enough for performing hover autonomously. In Figure 6 the dotted lines represent the desired position. Note in the middle-figure that we add two manuals perturbations and as the helicopter is close to the equilibrium point the system remains stable. Figure 6 shows also the altitude of the quad-rotor helicopter.

Table 2. *Gain values used in the experimentations (15) and (43)*

Controller	Control parameters	Value
Linear $(\tau_{L\phi})$	k_1	0.005
	k_2	0.0001
	k_3	0.1
	k_4	0.0001
Separated saturations $(\tau_{S\phi})$	b_1	0.06
	b_2	0.1
	b_3	0.17
	b_4	0.193
	k_1	1.3
	k_2	0.1
	k_3	0.3
	k_4	0.1

Figure 7 shows the response of the system when applying the nonlinear controller (44). The control parameters used for this experience are shown in table 2. Note that error of the lateral displacement, x, is smaller than these obtained when using the linear controller. Note also from this Figure that, the angular displacement is closer to the desired position. Thus, the aircraft remains closer to the equilibrium point. Figure 7 shows also the altitude of the aircraft for this experience.

The control algorithm using nested saturation, $\tau_{NS\phi}$, is well known [5], [2], some experiences with this controller have been made including aggressive perturbations in the roll or pitch angle. Figure 8 shows the performance of the helicopter when applying this controller. In this experience we use the same parameters used in simulations. Note from Figure 8 that, the lateral and angular position remain closer to the desired position. The backstepping control technique is very popular in the control community [15]. Some researchers use this technique to stabilize an UAV. In our knowledge, there are not a lot of applications of this control law in UAVs. This fact is due to the nature of the parameters in the control law. These parameters are not easy to tune. We tried to apply the controller (41) in a real-time experience using the nominal parameters and we didn't arrive to do it. After, we try to tune the gains in the control algorithm but we didn't arrive to have a good performance of the system. Figure 9 shows the system response of the helicopter when applying this controller. For the moment we did not manage to find gains to have a good stability with this law. Note from this Figure that, the big error in the angular position. Note also that, the control input is bigger than the others control inputs, $\tau_{L\phi}$ and $\tau_{S\phi}$.

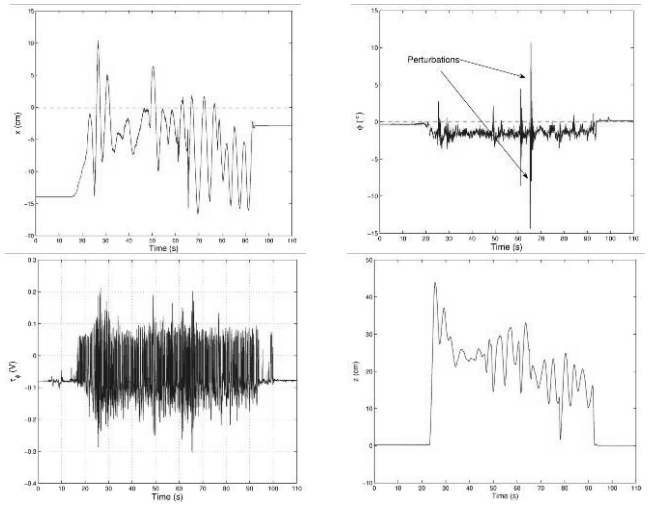


Fig. 6. *Stabilization of the VTOL aircraft when applying the LQR controller (15). The LQR gains are manually adjusted to improve the performance of the mini-rotorcraft until obtain a desired performance. These parameters are show in Table 2. The desired position is* $x_d = 0\ cm,\ \phi_d = 0°$.

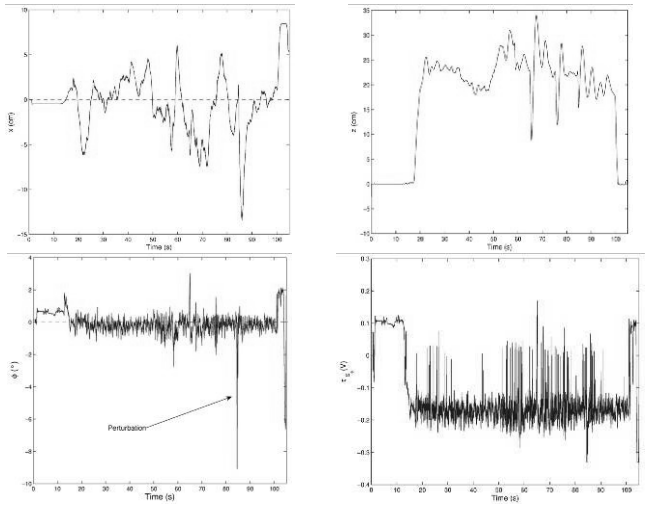


Fig 7. *Stabilization of the VTOL aircraft using the controller (44). The dotted lines represent the desired position.*

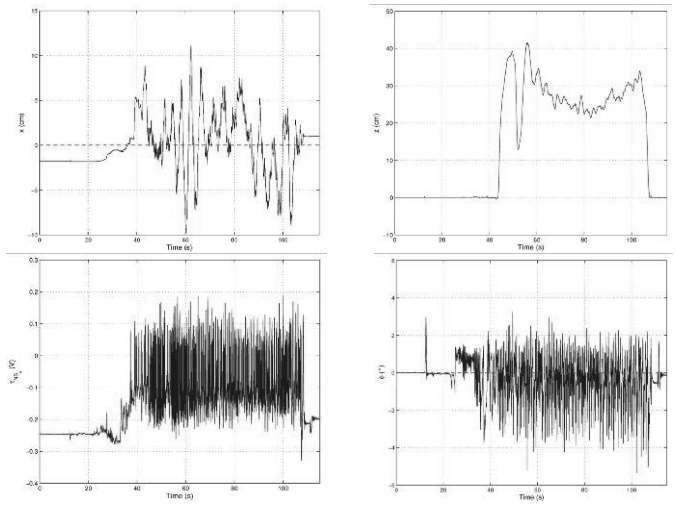


Fig 8. *Stabilization of the VTOL aircraft using the controller (43). The dotted lines represent the desired position.*

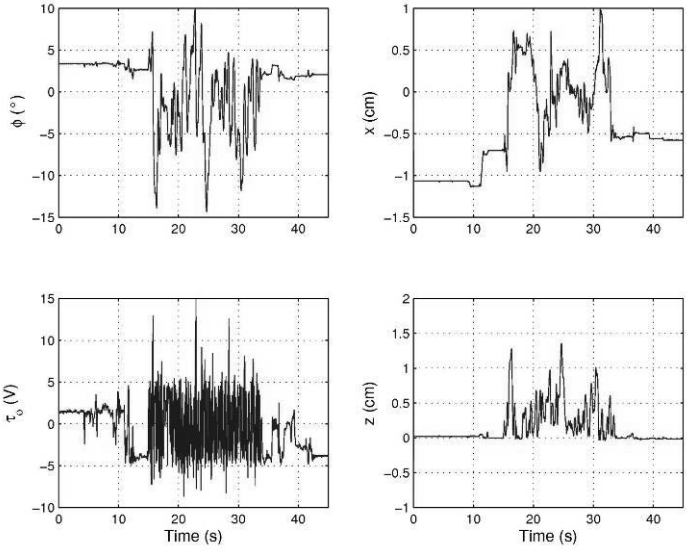


Fig. 9. *Unstable response of the quad-rotor for the backstepping control law applied to the* (x, ϕ) *subsystem. The oscillations in the roll angle* ϕ *prevent the helicopter from taking off. The backstepping gains are manually adjusted to improve the performance of the mini-rotorcraft, although the performance is not adequate for hovering.*

6 Conclusions

In this chapter a comparative analysis of linear and nonlinear control strategies to stabilize a VTOL aircraft has been presented. In addition, a backstepping control law has been developed step by step. The linear control strategy is a classical LQR and the nonlinear control laws are the backstepping controller and two controllers using saturation functions.

The controllers have been tested in simulation and in real-time experiences in a quad-rotor helicopter. The experimental results show the well performance of the linear controller and the nonlinear controls law using saturation functions even in presence of manual perturbations. The backstepping controller performs well in simulation but to apply it in real applications is so hard.

References

1. Kaliora G. and Astolfi A., "Nonlinear control of feedforward systems with bounded signals". IEEE Trans. Automatic Control, 49(11), pp. 1975–1990, 2004.
2. P. Castillo, R. Lozano and A. Dzul, "Modelling and Control of Mini-Flying Machines", Springer-Verlag in *Advances in Industrial Control*, July 2005. ISBN: 1-85233-957-8
3. B.P. Ickes, "A new method for performing control system attitude computation using quaternions", AIAA J., Vol. 8, pp. 13-17,1970.
4. T.R. Kane, "Solution of kinematical differential equations for a rigid body", J. Applied Mechanics, pp. 109-113, 1973.
5. A.R. Teel, "Global stabilization and restricted tracking for multiple integrators with bounded controls", Syst. & Contr. Lett., vol. 18, pp. 165-171, 1992.
6. T.S. Alderete, "Simulator aero model implementation" [Online], NASA Ames Research Center, Moffett Field, California.
7. B. Etkin and L. Duff Reid, *Dynamics of Flight*, John Wiley and Sons, Inc., New York, 1959. ISBN 0-471-03418-5
8. H. Goldstein, *Classical Mechanics*, Addison Wesley Series in Physics, Adison-Wesley, U.S.A., second edition, 1980.
9. J. Hauser, S. Sastry and G. Meyer, "Nonlinear control design for slightly nonminimum phase systems: Application to V/STOL aircraft", *Automatica*, vol. 28, no. 4, pp. 665-679, 1992.
10. B.W. McCormick, *Aerodynamics Aeronautics and Flight Mechanics*, John Wiley & Sons, New York, 1995.
11. R. Lozano, B. Brogliato, O. Egeland, B. Maschke, *Passivity-based control system analysis and design*. Springer-Verlag, Communications and Control Engineering Series, 2000.
12. Fantoni, R. Lozano, *Control of Nonlinear Mechanical Underactuated Systems*. Springer-Verlag, Communications and Control Engineering Series, 2001.
13. L. Marconi, A. Isidori, A. Serrani, "Autonomous vertical landing on an oscillating platform: an internal-model based approach", *Automatica*, vol. 38, pp. 21-32, 2002.
14. Quanser, 3 DOF Hover, [Online]. Available at: http://www.quanser.com/english/downloads/ products/3DOF Hover.pdf
15. Tarek H., Mahony R., Lozano R. and Ostrowski J., "Dynamic modelling and configuration stabilization for an X4-flyer", In Proc. of IFAC World Congress, Barcelona, Spain, 2002.
16. Olfati-Saber R., *Global configuration stabilization for the VTOL aircraft with strong input coupling*. In proceedings of the 39th IEEE Conf. on Decision and Control, Sidney, Australia, Dec. 1999.
17. J. T-Y. Wen and K. Kreutz-Delgado, "The attitude control problem", IEEE Transactions on Automatic Control, Vol. 36, No. 10, pp. 1148-1162, 1991.
18. B. Wie, H. Weiss and A. Arapostathis, "Quaternion feedback regulator for spacecraft eigenaxis rotations", AIAA J. Guidance Control, Vol. 12, No. 3, pp. 375-380, 1989.
19. Sanchez, P. Castillo, and R. Lozano, "Simple real-time control strategy to stabilize the PVTOL aircraft using bounded inputs", European Control Conference 2007, Kos, Greece 2-5 July 2007.
20. Martin P.,Devasia S. & Paden B., A different look at output tracking: control of a VTOL aircraft, Automatica, 32(1):101-107, 1996.
21. F. Mazenc and L. Praly, Adding integrations, satured controls, and stabilization for feedforward systems, IEEE Transactions on Automatic Control, vol. 41, issue 11, pp. 1559-1578, 1996.
22. G. Notarstefano, J. Hauser and R. Frezza, Trajectory Manifold Exploration for the PVTOL aircraft. In proceedings of the 44th IEEE Conference on Decision and Control and the European Control Conference, Seville, Spain, Dec. 2005.
23. H. Rodriguez, A. Astolfi and R. Ortega, On the construction of static stabilizers and static output trackers for dynamically linearizable systems, related results and applications. In proceedings of the 43th IEEE Conference on Decision, Atlantis, Paradise Island, Bahamas, Dec. 2004.

Chapter 9

A Mechatronics Methodology

Efren Gorrostieta[1], Juan Manuel Ramos[2], and J. Carlos Pedraza[3]

1 Universidad Autónoma de Querétaro,
2 Universidad Tecnológica de San Juan del Río,
3 Centro de Ingeniería y Desarrollo Industrial
4 Springer-Verlag, Computer Science Editorial, Tiergartenstr. 17, 69121 Heidelberg, Germany {Alfred.Hofmann, Ingrid.Beyer, Christine.Guenther, Frank.Holzwarth, efrengorrostieta@gmail.com, jramos@mecamex.net, jpedraza@cidesi.mx

Abstract

In the present work a proposal of methodology for the development of projects in mechatronics field appears which has been applied in several projects where it has valued the development and its behavior. The iteration of the different disciplines appears on projects development and obtaining in this way one a better integration, also this methodology has been used in the development of new industrial machinery. A mechatronics design method is proposed as a part of the research and engineering interaction activities, but also the manufacture aspects and complex mechanical adjustments are considered. The methodology has been applied to the neural networks and fuzzy logic control of a pneumatic valve used in a flexible manipulator robot.

Keywords: Mechatronics methodology, fuzzy logic and neural networks, control system.

1 Introduction

The objective to count with a methodology to develop mechatronics projects helps by one hand, to facilitate the interaction between each one of the disciplines that take part in the project [1], [2], and on the other hand facilitates the development and shortens the time of development [3]. In this case of study, a mechatronics method is used for robotics design, and also for academic research projects.

In principle, a flexible manipulator robot with pneumatic actuator is light, cheap, and has the advantage to handle a higher power-weight, with respect to robots that have electric actuators [4]. Along this research line, a Thermo-mechanical model has been developed [6].

The modeling of the flexible manipulators has been made for almost 35 years [7][8], where almost all the works use electrical or hydraulic actuators, and those of pneumatic type are less used due to their non linear nature.

In this proposal, the pneumatic actuators are used to give the robot a dielectric characteristic and to avoid the induction of high tensions in the system. In addition, these robots present advantages of economy, cleaning and weight. However, they have a highly non linear behavior, due to the air compressibility and internal friction [9], while the controller development is more complex.

The pneumatic control story starts in 1968, with the work of Burrows [6], and currently the development of this type of controllers and techniques have taken to the use of adaptive control methods [11] to be used in the electro pneumatic actuator position control [12], even it is used in biological applications like muscle actuators [11]. Based on this type of works, a flexible manipulator with pneumatic actuator was developed, where a mechanical system and pneumatic behavior are involved, to give a flexible arm movement [14], which leads to the Thermo-Mechanical model. From the point of view of the control engineering, this model allows to predict the behavior of different variables that take part in the physical process to control.

The pneumatic control and the work with flexible manipulators have been used until this moment as separated lines, but the necessities of the project requires that both lines work in a united way, therefore the experience in the interaction between both lines is the contribution in the present work. It is important to mention that in this proposal we want to control the angle of the arm of the manipulator without considering the flexibility of the used material. A Thermo-Mechanical model is used, because it allows us to predict the behavior of the different variables involved in the physical process that we are controlling, considering the air compressibility effects, the internal friction forces, the effect of the damping at the ends of the cylinder, the flow mass and energy conservation. It also allows us to know the pressures instantaneously, which depend on the piston rod position, and considers the geometric characteristics of the manipulator structure. A control for

this model is proposed, since in [6] is not made any proposal of control. Due to the highly non linear system, a fuzzy logic algorithm for the design of the controller is used, and a good performance of the movement of the manipulator is obtained.

2 Mechatronics Methodology Design

The proposed methodology consists on three levels. The first one is the conceptual development of the project, where the mathematical modeling based on physical principles is the most important part, as can be observed on figure 1. On the first level, the computer simulation, kinematics and dynamics of the system as well as control model are involved. The second level consist on control implementation, where the mechanical design, control design, integration and test are considered. The third level is when we have the project already developed and the communication with the environment must be done, In the third level, parameter monitoring is involved considering the sensors, actuators and so on.

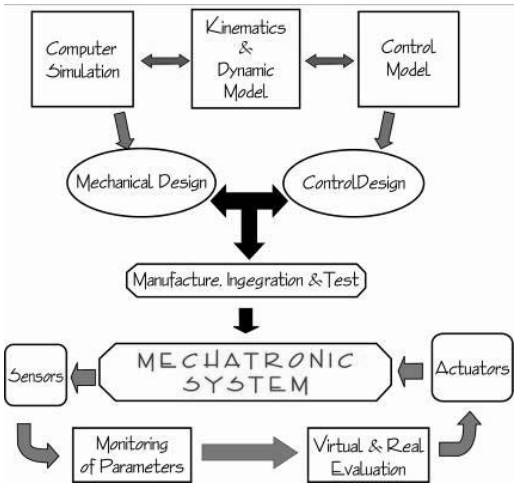


Fig.1. *Mechatronics Methodology*

3 Mathematical Model

On figure 2(a) a diagram of the pneumatic actuator can be observed, where $X, \dot{X}, \ddot{X}$ are the position of the piston actuator, the speed and the acceleration of the piston rod with respect to the cylinder, respectively; additionally we have the internal pressures P_{a1}, P_{c1}, P_{c2} and P_{a2}, that appear at the left side pad of the cylinder, in the chamber of the piston at the left side of the rod, the chamber at the right side rod, and at the pad of right side of the cylinder, respectively; the actuator force F_a; A_1, A_2 and A_3 are the free air flow area of the valves at the left side of the cylinder, at the right side of the cylinder and the air return; a 5/2 valve is used. Although the model needs the three valves, we use only two, considering the valves A_1 and A_2 with the same value, due that when the rod comes down, the valve enabled is A_2 and when it goes up, A_1 is enabled, so the configuration finally used is like figure 2(b).

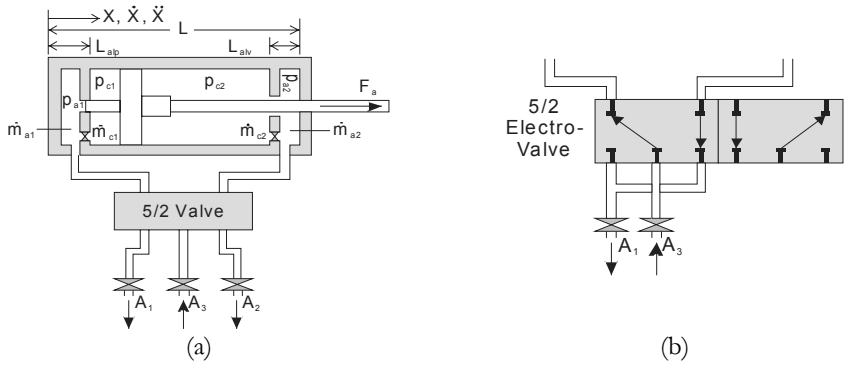


Fig. 2. *Diagram of a pneumatic piston with damping on both sides. (a) The three valves needed for the mathematical model. (b) Using only two valves for simulation and practice process, where $A_1 = A_2$.*

Figure 3 shows a scheme of the pneumatic actuator assembly, to give the movement to the flexible arm. The mechanical system output, that we call *plant*, is the θ angle, and its value depends on the displacement of the piston rod of the pneumatic actuator, just called *actuator*.

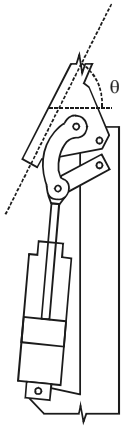


Fig. 3. *Assembly of the pneumatic actuator with the flexible arm. The output of the plant is the angle θ.*

As a starting point, we considered the work developed by Kiyama [6], where the integral Thermo-mechanical model is obtained, however, the work by Kiyama does not include any control proposals of the system, therefore in this work a control with fuzzy logic is proposed.

The set of equations (1) to (10) shows the Thermo-Mechanical model, that describes the dynamics of the pneumatic actuator. The model calculates the changes of the internal pressures of the cylinder; P_{a1}, P_{a2}, P_{c1} and P_{c2}, that depends of the cylinder rod position. Due to the effects generated by the damping pads at the ends of the cylinder, it is necessary to divide the Thermo-Mechanical model in three intervals: pad of the piston side ($0 \le X < L_{alp}$), middle ($L_{alp} \le X < L - L_{alv}$), and the pad of the piston rod side ($L - L_{alv} \le X \le L$), as observed in the equation set (1) to (10).

For the interval $0 \le X \le L$:

$$\dot{X} = \frac{d}{dt} X \tag{1}$$

$$D\dot{X} = \frac{d^2}{dt^2} X \tag{2}$$

For the interval $0 \le X \le L_{alp}$:

$$\dot{P}_{a1} = \frac{kRT_0}{A_{ap}\left(X + \frac{\Delta A_p}{A_{ap}}\right)}\left[\dot{m}_{1a} - \dot{m}_{1c} - \frac{A_{ap}}{RT_0}P_{a1}DX\right] \tag{3}$$

$$\dot{P}_{c1} = \frac{kRT_0}{\left(A_p - A_{ap}\right)X}\left[\dot{m}_{1c} - \frac{\left(A_p - A_{ap}\right)}{RT_0}P_{c1}DX\right] \tag{4}$$

For the interval $L_{alp} < X \leq L$:

$$\dot{P}_{a1} = \frac{kRT_0}{A_p\left(X + \Delta\right)}\left[\dot{m}_{1a} - \frac{A_p}{RT_0}P_{a1}DX\right] \tag{5}$$

$$\dot{P}_{c1} = \frac{kRT_0}{A_p\left(X + \Delta\right)}\left[\dot{m}_{1c} - \frac{A_p}{RT_0}P_{c1}DX\right] \tag{6}$$

For the interval $0 \leq X < (L - L_{alv})$:

$$\dot{P}_{c2} = \frac{kRT_0}{\left(A_p - A_v\right)\left(L - X + \Delta\right)}\left[\dot{m}_{2c} + \frac{\left(A_p - A_v\right)}{RT_0}P_{c2}DX\right] \tag{7}$$

$$\dot{P}_{a2} = \frac{kRT_0}{\left(A_p - A_v\right)\left(L - X + \Delta\right)}\left[\dot{m}_{2a} + \frac{\left(A_p - A_v\right)}{RT_0}P_{a2}DX\right] \tag{8}$$

For the interval $(L - L_{alv}) \leq X \leq L$:

$$\dot{P}_{c2} = \frac{kRT_0}{\left(A_p - A_v\right)\left(L - X\right)}\left[\dot{m}_{2c} + \frac{\left(A_p - A_v\right)}{RT_0}P_{c2}DX\right] \tag{9}$$

$$\dot{P}_{a2} = \frac{kRT_0}{L - X + \frac{\Delta A_p}{A_{av} - A_v}}\left[\frac{\dot{m}_{2a} - \dot{m}_{2c}}{A_{av} - A_v} + \frac{P_{a2}DX}{RT_0}\right] \tag{10}$$

Where A_p, A_v, A_{av} and A_{ap}, are the free air flow area from the outside inwards of the left side of the cylinder, from the outside inwards of the right side of the cylinder, at the left pad, and at the right pad of the cylinder, respectively.

From the set of equations, the dependency of the areas of free air flow can be observed, through the sections and the valves.

4 Computer Simulations

In recent days, to predict the behavior of the systems it is necessary to perform a computer simulation that involves all the mathematical modeling as well as the physical phenomena description associated to the system. If we know all the parameters, then the system response can be obtained by numerical approach (numerical simulation) and then a system control can be proposed.

Our idea is to have a mechanical model of a flexible manipulator drawn on any CAD software, and once the design is parameterized the model is exported into a VRML format. Later, the VRML file is converted to a virtual model using the OpenGL libraries and C++ language. The virtual model can be observed on figure 4. Then the OpenGL model is used together with the mathematical description of all the manipulator variables and the simulation is performed. It is important to say that the mathematical description includes the direct and inverse kinematics of the manipulator. A methodology for modeling and simulation of the flexible manipulator was developed [15].

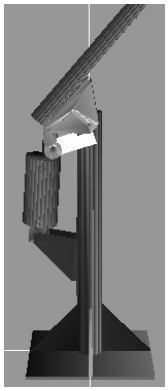


Fig. 4. *Robot Simulation*

5 Control System

The used control system for this mechatronic project, is divided in two parts, the fuzzy logic and the neural networks control. These parts are proponed due to the system behavior and the system to control doesn't present linearities as we can see on ecs. (1) to (10).

5.1 Fuzzy Logic Control

Figure 5 shows the block diagram of the control proposal, where a PID control is involved, using Fuzzy Logic feedback, and an adjustment is carried out on the control parameters.

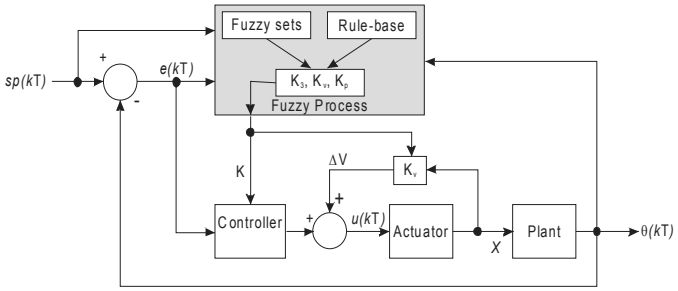


Fig. 5. *Diagram to blocks of the fuzzy control for the Thermo-Mechanical model.*

In figure 4, θ_p is the set point; θ is the current position; u is the vector of control variables, formed by the free air flow on the valves, as shown in the equation (10); also e is the position error in the output angle of the mechanism at the time T_k, and it includes the proportional error e_p, integral error e_i, and derivative error e_d, as shown in the equations (12) to (14); ΔV is the speed difference at two intervals of time, defined in the equation (15); K is the set of control values for the PID controller (16).

$$u = \left[A_1, A_2, A_3\right] \tag{11}$$

where A_1, A_2 and A_3, are the areas of the external valves, used in the controller, as shown in the figure 2.1(a). The valve A_3 is the pressure input, and the valves A_1 and A_2 are the air return to the atmosphere.

$$e_p(T_k) = \theta_p - \theta \tag{12}$$

$$e_i(T_k) = e(T_k) + e(T_{k-1}) + e(T_{k-2}) \tag{13}$$

$$e_d = e(T_k) - e(T_{k-1}) \tag{14}$$

$$\Delta V = V(T_k) - V(T_{k-1}) \tag{15}$$

$$K = [K_p, K_d, K_i, \mathrm{K_v}] \tag{16}$$

The K vector represents the fuzzy adaptation process of the proportionality constant, derivative constant, integral constant and speed constant, obtained by a fuzzy logic algorithm to improve the behavior of the system. In this control scheme, the feedback of the speed changes is important, due to the fact that a better behavior at the plant output was obtained; avoiding abrupt changes in the displacement speed generated by the pneumatic actuator, as well as the angular speed behavior of the plant, and therefore a vibration problem in the system is avoided. The control equation that defines the air free flow area is defined in the equation (17).

$$A_j(T_k) = A_{j0} + K_{pj}e_p(T_k) + K_{ij}e_i(T_k) + K_{dj}e_d(T_k) + K_v\Delta V(T_k) \tag{17}$$

Where T_k represents the sample time, considered of 50 ms; *j* represents the valve which is controlled; K_p, K_d, and K_i, are the constants of the law of the PID control, and K_v is the speed change constant.

The *K* vector values are obtained with a fuzzy logic algorithm.

To find the fuzzy sets needed to control the system, an initial values for K_v, K_3, K_i and K_d are close to zero, and only apply the fuzzy process to K_p. When a good response is obtained, a fuzzy process is applied to K_i, later to K_d, K_v, and finally to K_3. Next, is necessary to do a few changes in order to obtain a better system behavior.

The fuzzy process for K_p needs three inputs and one output. The fuzzy sets for all inputs and outputs use a triangular membership functions for each class. The classes for the input theta are extreme, low and positive. The classes for the input error are; negative big, negative half, negative small, zero, positive small, positive half and positive big. The classes for the input set point are; set point 1, and set point 2. The classes for the output are; small, half, big and very big. This parameter defines the system speed. The fuzzy rules are shown in table 1.

Table 3.1. *Rules used for K_p*

Rule 1	If **e** is **NSmall** then K_p is **Big**
Rule 2	If $\boldsymbol{\theta_p}$ is **SP1** then $\mathbf{K_p}$ is **Big**
Rule 3	If **θ** is not **Extreme** and **e** is **Zero** then $\mathbf{K_p}$ is **Small**
Rule 4	If **θ** is not **Positive** and **e** is **PSmall** then $\mathbf{K_p}$ is **Small**
Rule 5	If **θ** is not **Positive** and **e** is **PHalf t**hen $\mathbf{K_p}$ is **VBig**
Rule 6	If **θ** is not **Positive** and **e** is **PBig** then $\mathbf{K_p}$ is **Big**
Rule 7	If **θ** is **Positive** and **e** is **PSmall** then $\mathbf{K_p}$ is **Big**

The fuzzy process for K_v has the same inputs than parameter K_p, as show in figure 6, using triangular membership functions for each class. The classes for the input theta are; low and high. The classes for the input error are; negative and positive. The classes for the input set point are; negative 5, negative 4, negative 3, negative 2, negative 1 and positive. The classes for the output are; zero, very small, small, half, regular and high. The K_v parameter helps us to avoid an abrupt change of the speed in the system movement.

The fuzzy process for K_3 needs only one input and one output, using triangular membership functions for each class. The classes for the input set point are; very down, negative down, regular down, down and up. The classes for the output are; very few, few, half, high, very high and all. This parameter helps to control the flow air input of the system, and it is important to minimize the overshoot at the plant output, θ.

The optimal value of those parameters happens to be zero. To obtain it, the control parameters K_d and K_i are not used in control equation (17).

The equation (18) shows the final control equation. It is necessary to control the air return of the system, to get a better behavior; this valve is important to stop the piston displacement when the arm is close to the set point in the up and down directions.

$$A_1(T_k) = A_{10} + K_p e_p(T_k) + K_v \Delta V(T_k)$$
$$A_2(T_k) = A_1(T_k) \tag{18}$$
$$A_3(T_k) = K_3 A_1(T_k)$$

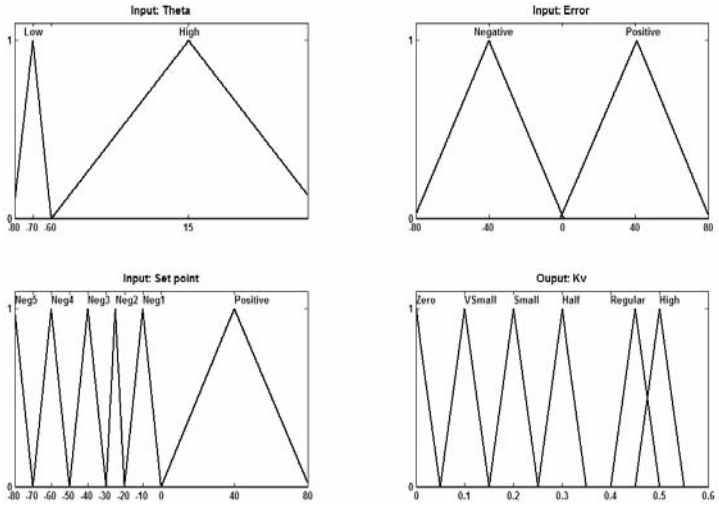


Fig. 6. Fuzzy sets for K_v control constant

5.2 Neural Networks Control

The structure of the proposed neural control system is presented on figure 7, as can be observed the system presents a non linear behavior showed on section 2. This non-linearity is due to the air compression of the pneumatic actuator. A neural control is proposed, as can be seen on figure 7.

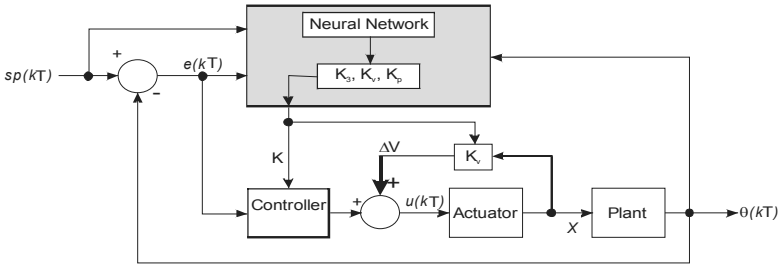


Fig. 7. *Diagram to blocks of the neural network control*

The neural nets are of the perceptron type, with a hidden layer which has only a neuron with a sigmoid activation function. The output layer is linear and has also a single neuron. The outputs of the nets are respectively ΔK_p ΔK_i, ΔK_d and ΔK_v, that is the increments of proportional integral and derivative gains.

The neural network can be calculated easily in real time and serve to adapt the values of proportional coefficients Kp. A similar derivation can be used to find the adaptation equations for derivative coefficients K_d , K_i, K_v.

6 Mechatronic System

Figure 8 shows the flexible manipulator photograph with pneumatic actuator, where the arm is made of PVC material. Figure 1 is a scheme that shows the air flow in both directions, and the air return is controlled by two independent proportional valves. These valves are used to avoid that some of the cameras remains without pressure, generating a braking effect to the advance of the piston.

Fig. 8. *Flexible manipulator robot with pneumatic actuator*

The figure 9, shows the rod movement, also is possible to observe the angle θ, it is greater than 0, the system has a different behavior when θ is smaller than 0. If θ is greater than 0, the system needs 5 seconds to arrive at 98% of set point, when is going down and 2 seconds when it goes up. Also if θ is smaller than 0, the systems needs 6 seconds to get a 98% of set point going in down direction, and 3 seconds in up direction.

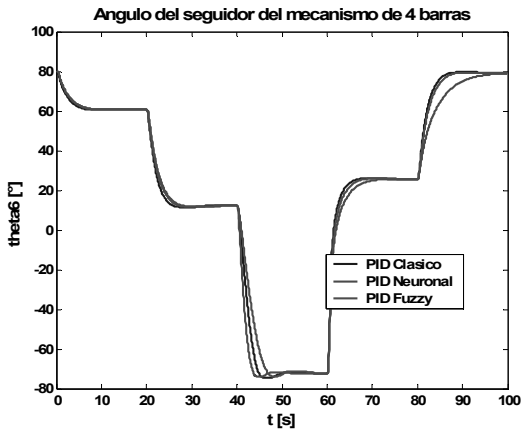


Fig. 9. *Control results of the flexible manipulator robot and comparison between fuzzy, Neuronal control and PID control*

The reference values used to test the control are: 60º, 12º, -72º, 25º and 79º. The results are shown in figure 9, including a comparison of results between classic control and fuzzy control.

The figure 5.2 is a comparison between the PID control, PID neuronal and fuzzy control response, shows a better control behavior when the neuronal algorithm is used than the PID control algorithm, especially when the set point is close to -80º. At this point, the rod position is close to 0.1 m, and the rod is close to the damping area, and it must support the inertial forces generated by the mechanism weight

7 Conclusions

The present methodology has been carried out in through the development of mechatronic projects, obtaining good result in the integration process as well as shortens the development time. Therefore, the main mechatronic principles can be observed and also give the importance to the mathematical simulation as well as the theoretical principles of each part involved in every project..

References

1. Bradley, D.A., The What, why and how of mechatronics: Engineering Science and Education Journal April (1997) 81-88.
2. Comerfor Richard: Mecha What? IEEE spectrum August (1994) 46-49.
3. Geoff Robert: Intelligence Mechatronics: Engineering Science and Education Journal April (1999) 81-88.
4. Heimann Bodo, Gherth Wilfried and Popp Karl: Mechatrinik. Fachbuchverlang Leipzig (2001) 1-16.
5. C. Mavroidis: Development of Advanced Actuators Using Shape Memory Allows and Electrorheological Fluids; Research in Nondestructive Evaluation; Springer New York, Volume 14, Number 1,(2002). 1-32.
6. F. F. Kiyama and E. Vargas; Dynamic Model Analysis of a Pneumatically Operated Flexible Arm; WSEAS Transactions on Systems, Vol. 4, No 1, (2005) 49-54.
7. Mirro, John; Automatic Feedback Control of a Vibrating Flexible Beam; MS Thesis, Department of Mechanical Engineering, Massachussets Institute of Technology, (1972).
8. Whitney, D. E., Book, W. J. And Lynch, P. M.; Design and Control Considerations for Industrial and Space Manipulators; Proceedings of the Joint Automatic Control Conference, ,(1974) June.
9. Moore P. y J. Pu; Progression of servo pneumatics toward advanced applications; Fluid Power Circuit, Component and System Design; K. Edge and C. Burrows, Eds. Boldock, U. K.: Research Studies Press; (1993) 347-365
10. Burrows C.R., Webb C.R.; Simulation of an On – Off Pneumatic Servomechanism; Automatic Control Group, (1968)
11. Jaydeep Roy, and Louis L. Whitcomb; Adaptive Force Control of Position/Velocity Controlled Robots: Theory and Experiment; IEEE Transactions on Robotics and Automation, vol. 18, no. 2, april (2002)
12. Paul D. Henri, John M. Hollerbach, Fellow, IEEE, and Ali Nahvi; An Analytical and Experimental Investigation of a Jet Pipe Controlled Electropneumatic Actuator; IEEE Transactions on Robotics and Automation, vol. 14, no. 4, august (1998)
13. Caldwell, D. G. Medrano-Cerda, G. A. Goodwin, M.; Characteristics and Adaptive Control of Pneumaic Muscle Actuators for a robotic Elbow; IEEE International Conference on Robotics and Automation, 1994 Proceedings; ISBN 0-8186-5330-2; Volume 4, (1994) 3558-3563.
14. Feliu V. y A. García; Gauge-Based tip Position Control of a New Three Degree of Freedom Flexible Robot;The International Journal of Robotics Research; vol. 20, no. 8, (2001) 660-675
15. M. Gamiño, J.C. Pedraza, J.M. Ramos and E. Gorrostieta; Matlab - C++ Interface for a flexible arm manipulator simulation using multi-language techniques;5th Mexican International Conference on Artificial Intelligence; Proceedings, (2006) 369-378

Chapter 10

New Results on Robust Stability of Interval Plants with Time Delay

Gerardo Romero, Irma Pérez, Luís García, Diego Castillo,
Iván Díaz, David Lara, and José Rivera

Unidad Académica Multidisciplinaria Reynosa Rodhe
Universidad Autónoma de Tamaulipas
Carr. Reynosa-San Fernando cruce con Canal Rodhe, Colonia Arco Iris, Cd. Reynosa Tamaulipas.
Apdo. Postal 1460, C.P. 88779, México. Tel: +52 (89) 21-33-00 Ext. 8313.
E-mail: gromero@uat.edu.mx

Abstract

This chapter presents sufficient condition to verify the robust stability property of a class of time delay systems which are described as an interval plant with uncertain time delay. The main result is obtained on the basis of two polynomials that can be easily computed using Kharitonov's polynomials.

Keywords: Stability, robust stability, time-delay systems.

1 Introduction

Time delay systems are interesting to many scientific researchers from different areas; this is due to the fact that this type of systems have multiple applications in industrial processes such as: thermal processes, electric power systems, biologic processes and many more, see [6], [9], [11], [18]. One of the most important qualitative properties to consider in the analysis of time delay systems is the stability property that is obtained from the mathematical representation of a physical process. It is well known that the mathematical representation of a physical process does not accurately portraits its dynamic behavior, this in turn implies that the stability property can not be precisely obtained; this problem has been addressed by including dynamic uncertainty [10] or parametric uncertainty [1], [2] in the mathematical model; the stability property that considers uncertainty in the mathematical model is defined as robust stability, this property will be discussed in the present chapter.

The stability property has been analyzed from different points of view, some authors [12], [17], [26] present papers where the robust stability property is based on the Lyapunov theory, linear matrix inequalities (LMI) and μ synthesis theory for time delay systems described in time domain. Time delay systems present very complex mathematical model that are difficult to analyze therefore some authors prefer to make some transformations to simplify the process of obtaining the robust stability property, see [19], [23], [24]; however, the attained conditions for these transformed systems, in some cases, result to be very conservative which could cause them to be not very useful; a good paper to find out more about this issue is presented in [13].

Another way to analyze the time delay systems is based on the frequency representation where the robust stability property is verified in terms of a kind of function called quasipolynomials. There exist some background related to this chapter, in [8] a generalization of the edge theorem [5] for a kind of time delay systems is presented. The same type of time delay systems are considered in [4] where necessary and sufficient conditions are presented in terms of a function defined as H_∞ that is computed using the value set construction, see [2]. A more recent result is provided in [14] where an extension of the convex directions concept is presented, see [20]. Some other results related to this type of systems are found in [7], [16], [25].

This chapter will present sufficient conditions to verify the robust stability property of interval plants with uncertain time delay; this result is based on the verification of some properties of a pair of adequately selected polynomials. This chapter is organized as follows: in section II the problem statement will be presented; section III includes a mathematical background; the main result will be presented in section IV and finally, the conclusions.

2 Problem Statement

The analysis presented in this chapter is performed for interval plants with time delay that are defined as follows:

Definition 1 ([2], [3]): An interval plant is a transfer function with parametric uncertainty that has the following structure:

$$g(s,q,r)=\frac{n(s,q)}{d(s,r)}=\frac{\sum_{i=1}^{m}[q_i^-,q_i^+]s^i}{s^n+\sum_{i=1}^{n-1}[r_i^-,r_i^+]s^i} \tag{1}$$

where $m<n$, i.e. $g(s,q,r)$ is a set of strictly proper rational functions, Q and R are sets that represent the parametric uncertainty and are defined as follows:

$$\begin{aligned} R&\equiv\left\{r=\begin{bmatrix}r_1 & \cdots & r_n\end{bmatrix}^T : r_i^- \le r_i \le r_i^+\right\}\\ Q&\equiv\left\{q=\begin{bmatrix}q_1 & \cdots & q_n\end{bmatrix}^T : q_i^- \le q_i \le q_i^+\right\}\end{aligned} \tag{2}$$

These type of sets are known as boxes by the way they are defined; the name of interval plants is used because the coefficients of the transfer function are uncertain values that belong to a closed interval. It is clear that interval plants represent an infinite number of transfer functions; however, the result of this chapter will be expressed in terms of eight important elements of such interval plant, these elements are known as Kharitonov's polynomials of the numerator and denominator of the interval plant which are defined as follows, see [15]:

$$\begin{aligned} n_1(s)&=q_0^-+q_1^-s+q_2^+s^2+q_3^+s^3+\ldots\\ n_2(s)&=q_0^++q_1^-s+q_2^-s^2+q_3^+s^3+\ldots\\ n_3(s)&=q_0^++q_1^+s+q_2^-s^2+q_3^-s^3+\ldots\\ n_4(s)&=q_0^-+q_1^+s+q_2^+s^2+q_3^-s^3+\ldots\\ d_1(s)&=r_0^-+r_1^-s+r_2^+s^2+r_3^+s^3+\ldots\\ d_2(s)&=r_0^++r_1^-s+r_2^-s^2+r_3^+s^3+\ldots\\ d_3(s)&=r_0^++r_1^+s+r_2^-s^2+r_3^-s^3+\ldots\\ d_4(s)&=r_0^-+r_1^+s+r_2^+s^2+r_3^-s^3+\ldots\end{aligned} \tag{3}$$

by adding a time delay to the interval plant it is obtained the type of structure of the type of systems that will be analyzed in this chapter, these are defined next:

Definition 2: An interval plant with time delay is a time delay system that has the following structure:

$$g(s,q,r)=\frac{n(s,q)}{d(s,r)}e^{-\tau s};\tau\in[0,\tau_{\max}] \tag{4}$$

We are interested in analyze the robust stability of control systems that are represented by the next block diagram:

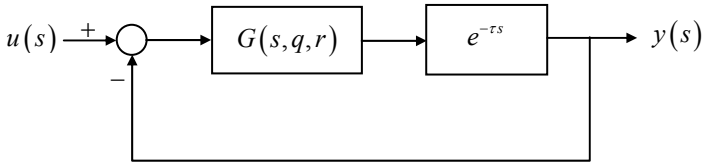


Fig. 1. *Interval plant with time delay.*

The property of robust stability is determined in terms of the following characteristic equation:

$$p(s,q,r,e^{-\tau s})=d(s,r)+n(s,q)e^{-\tau s} \tag{5}$$

These kind of functions are defined as quasipolynomials. It is clear that the former characteristic equation (5) represents an infinite amount of characteristics equations that have to be considered to verify the robust stability property; this family will be defined as follows:

$$P_\tau \equiv \left\{ p(s,q,r,e^{-\tau s}) : q \in Q, r \in R, \tau \in [0, \tau_{max}] \right\} \tag{6}$$

The robust stability property is guaranteed if and only if the following equation is satisfied:

$$p(s,q,r,e^{-\tau s}) \neq 0; \forall s \in C_+ \tag{7}$$

where C_+ is the set of complex numbers with real part greater than or equal to zero. From the equation (5) it can be seen that the robust stability property of the dynamic system is a property very difficult to verify due to the fact that this equation represents an infinite number of equations. The objective of this work is to present simpler results to verify the robust stability property of time delay systems as the ones shown in figure 1.

3 Preliminaries

The result presented in this chapter is based on the value set characterization of the family of characteristics equations P_τ that is defined as follows:

Definition 3: The value set of P_τ, noted by $V_\tau(\omega)$, is the graph in the complex plane of $p(s,q,r,e^{-\tau s})$ when $s = j\omega$ is substituted; this is:

$$V_\tau \equiv \left\{ \begin{matrix} p(j\omega,q,r,e^{-j\tau\omega}) : q \in Q, r \in R, \\ \tau \in [0, \tau_{\max}]; \omega \in \Re \end{matrix} \right\} \tag{8}$$

It is clear that the value set of P_τ is a set of complex numbers plotted on the complex plane when values to q_i, r_i, ω, τ are assigned inside the defined boundaries. Being able to characterize the value set $V_\tau(\omega)$ is of great relevance because by applying additional results it is possible to determine the robust stability property. This characterization was presented in [21] through the following lemma:

Lema 4 ([21], [22]): The value set $V_\tau(\omega)$ is composed, for each value of ω, by octagons that change their shape with respect of the time delay τ. The vertices of each octagon have the following coordinates in the complex plane:

$$\begin{aligned} v_{i+1} &= d_{i+1}(j\omega) + n_k(j\omega)e^{-j\omega\tau} \\ v_{i+5} &= d_{i+5}(j\omega) + n_h(j\omega)e^{-j\omega\tau} \end{aligned} \tag{9}$$

where:

$$\begin{aligned} &i = 01,2,3 \\ &k = (\gamma + i) \bmod_4 + 1 \\ &h = (\gamma + i + 1) \bmod_4 + 1 \\ &\gamma \equiv \left\{ \begin{matrix} 0 & 2n\pi \le \omega\tau < \frac{\pi}{2} + 2n\pi \\ 1 & \frac{\pi}{2} + 2n\pi \le \omega\tau < \pi + 2n\pi \\ 2 & \pi + 2n\pi \le \omega\tau < \frac{3\pi}{2} + 2n\pi \\ 3 & \frac{3\pi}{2} + 2n\pi \le \omega\tau < 2\pi + 2n\pi \end{matrix} \right\} \\ &n = 0,1,2,\ldots \end{aligned}$$

the function $(x)\,\mathrm{mod}_4$ represents the entire module base four function as shown in the following examples: $(2)\,\mathrm{mod}_4 = 2$. For any fixed τ and ω, the value set has the following form:

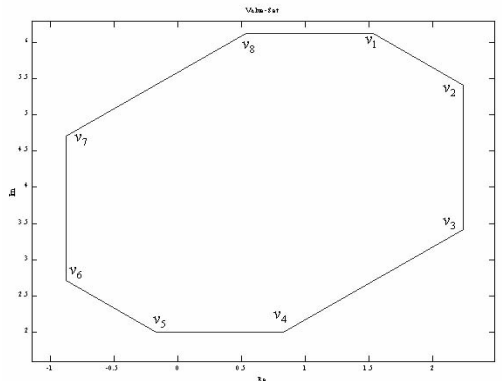


Fig. 2. *Value set for* ω *and* τ *fixed.*

for different values of ω and τ the value set is presented in the following figure:

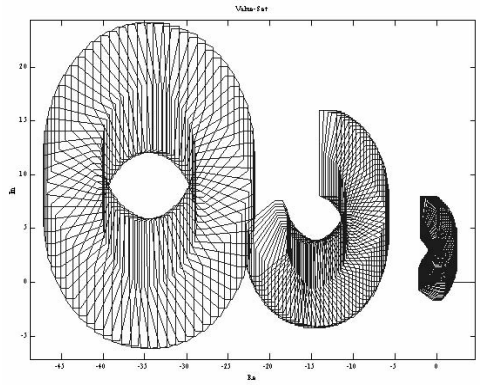


Figure 3. Values set for different values of ω and τ.

It is wort to mention that although the value set may appear as describing a circle, it is not. From the value set definition it can be clearly observed that the figure above represents all the values that the family P_τ can take when $s = j\omega$, this is, if the complex plane origin is contained in the value set $V_\tau(\omega)$ this means that P_τ has roots on the imaginary axis $j\omega$ for some values of $\omega \in \Re$, this in turn provokes instability in the time delay system. From the above mentioned it can be seen that the value set can be a tool to aid in verifying the robust stability property. An important question is how from a sweep over $j\omega$ it can be determined that P_τ does not have roots on the right half plane which was defined as C_+. The answer to this question is found in the result known as the zero exclusion principle, see [2].

4 Main Result

In this section the main result will be presented which consists in finding a set $\hat{V}_\tau(\omega)$ that is more simple to analyze and that completely contains the value set $\hat{V}_\tau(\omega)$ in such a way that sufficient conditions of robust stability for the quasipolynomials family previously defined as P_τ can be determined based on the value set $\hat{V}_\tau(\omega)$. The results of this chapter are supported by the definitions of the following polynomials:

$$c_d(s) = \frac{1}{4}\sum_{i=1}^{4} d_i(s) \tag{10}$$

$$c_n(s) = \frac{1}{4}\sum_{i=1}^{4} n_i(s) \tag{11}$$

$$c_0(s) = c_d(s) + c_n(s)e^{-\tau s} \tag{12}$$

where $d_i(s)$ and $n_i(s)$ are Kharitonov's polynomials as defined in (3). The next value set $\hat{V}_\tau(\omega)$ properties are important to be mentioned in order to provide the main result.

Lemma 5: Consider a fixed value of $\omega \in \Re$, then $c_0(j\omega$ is the center of the octagons that compose the value set $\hat{V}_\tau(\omega)$.

The previous lemma can be clearly seen in the preceding figure which shows the value set $\hat{V}_\tau(\omega)$ for a fixed frequency $\omega \in \Re$.

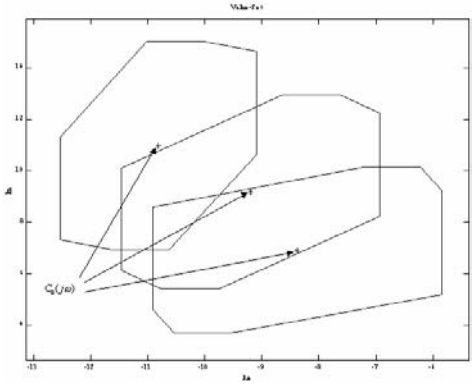


Fig. 4. *Center of* $\hat{V}_\tau(\omega)$ *for some fixed* ω.

From the definition of $c_0(s)$ it is possible to visualize that the center of the octagons that compose the value set $\hat{V}_\tau(\omega)$ takes the shape of archs of circumferences centered in the polynomial $c_0(j\omega)$, these archs have angles and radio equal to $\omega\tau$ and $|c_n(j\omega)|$, respectively. This property is important in the construction of $\hat{V}_\tau(\omega)$ and will be defined next:

Definition 6: The value set $\hat{V}_\tau(\omega)$ is the set of complex numbers $c_e = x_e + jy_e$ that satisfy the following condition:

$$(x_e - h)^2 + (y_e - k)^2 \le r_e^2(j\omega)$$

$$y_e + \tan(\phi(j\omega)) = 0 \tag{13}$$

$$\forall \phi(j\omega) \in [\phi_{\min}(j\omega), \phi_{\max}(j\omega)]; \omega \in \Re$$

where:

$$r_e(j\omega) = |c_n(j\omega| + r_0(j\omega)$$

$$r_0(j\omega) = \frac{|d_4(j\omega) - d_2(j\omega)| + |n_4(j\omega) - n_2(j\omega)|}{2}$$

$$h = \mathrm{Re}\left[c_d(j\omega)\right]; k = \mathrm{Im}\left[c_d(j\omega)\right]$$

$$\phi_{\min}(j\omega) = \arg\left(c_n(j\omega)\right) - \omega\tau_{\max} - \operatorname{arcsen}\left(\frac{r_0(j\omega)}{\left|c_n(j\omega)\right|}\right)$$

$$\phi_{\max}(j\omega) = \arg\left(c_n(j\omega)\right) + \operatorname{arcsen}\left(\frac{r_0(j\omega)}{\left|c_n(j\omega)\right|}\right)$$

And $\phi_{\min}(j\omega)$, $\phi_{\max}(j\omega)$ are computed using Kharitonov's polynomials (3). The previous definition describes a set of solid disc segments on the complex plane.

Lemma 7: The value set $\hat{V}_\tau(\omega)$ is completely contained in $\hat{V}_\tau(\omega)$; this is:

$$V_\tau(\omega) \subset \hat{V}(\omega) \tag{14}$$

This is a relevant result in order to achieve a more simple test to verify the robust stability property of the time delay system shown in figure 1. Before presenting the next result it is essential to introduce the following definitions.

Definition 8: The set W_c is defined as follows:

$$W_c \equiv \left\{\omega \in (0,\infty) : \left|c_d(j\omega)\right| = r_e(j\omega)\right\} \tag{15}$$

Definition 9: The value of maximum delay is defined as:

$$\bar{\tau} = \min\left\{\tau_1, \tau_2, \ldots, \tau_n\right\} \tag{16}$$

where n represents the number of elements that the set W_c has and the values of τ_i are attained by using the next formula:

$$\tau_i = \frac{\pi + \phi_{\max}(j\omega_i) - \arg\left(c_d(j\omega_i)\right) - 2\operatorname{arcsen}\left(\dfrac{r_0(j\omega_i)}{\left|c_n(j\omega_i)\right|}\right)}{\omega_i} \tag{17}$$

$$(i = 1, 2, \ldots, n) \forall\, \omega_i \in W_c$$

If the set W_c is the empty set, then τ takes an infinity value τ_i.

Theorem 10: P_τ is robustly stable if at least one member of P_τ is stable and $\tau_{max} < \bar{\tau}$.

It is worth to mention that the prior result offers only sufficient conditions due to the strict enclosure of $V_\tau(\omega)$ by $\hat{V}_\tau(\omega)$, however, this result is less conservative than many others that utilize some other techniques for their analysis.

5 Conclusions

This chapter presented sufficient conditions of robust stability for a class of time delay system which is described by interval plants with time delay. These results may be considered as conditions of robust stability dependent of delay where the property is verified based in two polynomials adequately defined on the basis of Kharitonov's polynomials of the interval plant. A recommendation for future research is to find sufficient and necessary conditions based on simple tests as the result presented in this chapter.

References

1. Ackermann J., Robust Control, *Springer-Verlag*, 1993.
2. Barmish B.R., New Tools for Robustness of Linear Systems, *Macmillan*, 1994.
3. Barmish B.R., Hollot C.V., Kraus F.J., Tempo R., "Extreme Point Result for Robust Stabilization of Interval Plants with First Order Compensators", *IEEE T-AC*, Vol. 37, No. 6, pp. 707-714, 1992.
4. Barmish B.R., Shi Z., "Robust Stability of Perturbed Systems with Time Delay", *IFAC Automatica*, Vol. 25, No. 3, pp. 371-381, 1989.
5. Bartlett A.C., Hollot C.V., Huang L., "Roots Locations of an Entire Polytope of Polynomials: It Suffices to Check the Edges", *Mathematics of Control Signals and Systems*, pp. 61-71, 1988.
6. Bellman R., Cooke K.L., Differential-Difference Equations, *Academic Press*, 1963.
7. Boese F.G., "Stability in a Special Class of Retarded Difference-Differential Equations with Interval-Valued Parameters", *Z. Angew. Math. Mech.*, 72, pp. 84-87, 1992.
8. Fu M., Olbrot A.W., Polis M.P., "Robust Stability for Time Delay Systems: The Edge Theorem and Graphical Test", *IEEE T-AC*, Vol. 34, No. 8, pp. 813-820, 1989.
9. Gorecki H., Fuksa S., Grabowski P., Korytowski, Analysis and Synthesis of Time Delay Systems, *John Wiley & Sons*, 1989.
10. Green M., Limebeer D.J.N., Linear Robust Control, *Prentice Hall*, 1995.
11. Hale J.K., Verduyn-Lunel S.M., Introduction to Functional Differential Equations, *Springer-Verlag*, 1993.
12. Huang Y.P., Zhou K., "Robust Control of Uncertain Time Delay Systems", *Proceedings of the 38th Conference on Decision and Control*, pp. 1130-1135, December 1999.
13. Kharitonov V.L., Melchor D.A., "Some Remarks on Transformations used for Stability and Robust Stability Analysis of Time-Delay Systems", *Proceedings of the 38th Conference on Decision and Control*, pp. 1142-1147, December 1999.
14. Kharitonov V.L., Zhabko A.P., "Robust Stability of Time Delay Systems", *IEEE T-AC*, Vol. 39, No. 12, pp. 2388-2397, 1994.
15. Kharitonov V.L., "Asymptotic Stability of an Equilibrium Point Position of a Family of Systems of Linear Differential Equations", *Plenum Publishing Corporation*, pp. 1483-1485, 1979.
16. Kogan J., Leizarowitz A., "Frequency Domain Criterion for Robust Stability of Interval Time-Delay Systems", *IFAC Automatica*, Vol. 31, No. 3, pp. 463-469, 1995.
17. Kolmanovskii V.B., Niculescu S.I., Gu K., "Delay Effects on Stability: A Survey", *Proceedings of the 38th Conference on Decision and Control*, pp. 1993-1998, December 1999.
18. Malek-Zavarei M., Jamshidi M., Time-Delay Systems, *North-Holland*, 1987.
19. Niculescu S.I., Chen J., "Frequency Sweeping Tests for Asymptotic Stability: A Model Transformation for Multiple Delays", *Proceedings of the 38th Conference on Decision and Control*, pp. 4678-4683, December 1999.
20. Rantzer A., "Stability Conditions for Polytopes of Polynomials", *IEEE T-AC*, Vol. 37, No. 2, pp. 79-89, 1992.
21. Romero G., Collado J., "Robust Stability of Interval Plants with Perturbed Time Delay", *Proceedings of the 1995 American Control Conference*, pp. 326-327, June 1995.
22. Romero G., Analysis of Robust Stability for Time Delay Systems, *Ph.D. Dissertation in Spanish*, Univesidad Autonoma de Nuevo Leon, June 1997.
23. Thowsen A. "The Routh-Hurwitz Method for Stability Test for a Class of Time Delay Systems", *International Journal Control*, Vol. 33, No. 5, pp. 991-995, 1981.
24. Thowsen A., "An Analytic Stability Test for a Class of Time-Delay Systems", *IEEE T-AC*, Vol. 26, No. 3, pp. 735-736, 1981.
25. Xin X., Feng C., "Robust Stability of Control Systems with Parametric Uncertainties", *Proceedings of the 31st Conference on Decision and Control*, pp. 1559-1564, December 1992.
26. Zhang J., Knospe C.R., Tsiotras P., "A Unified Approach to Time-Delay System Stability via Scaled Small Gain", *Proceedings of the 1999 American Control Conference*, pp. 307-308, June 1999.

Chapter 11

Adaptive Exact Tracking Error Dynamics Passive Output Feedback for the Sensorless Control of a DC Motor

Hebertt Sira-Ramírez and Enrique Barrios-Cruz

CINVESTAV-IPN
Mechatronics Section
Electrical Engineering Department
México City, México
E-mail: hsira,ebarrios@cinvestav.mx

Abstract

A high gain reduced order observer with integral estimation error injections is used for the fast determination of the unknown constant mechanical load in a DC motor. The motor is controlled by an exact tracking error dynamics passive output feedback control scheme which demands knowledge of the load parameter in the feed-forward terms alone and the on-line availability of the armature current. The determination of the feed-forward control is based on a certainty equivalence differential parametrization of the motor armature current and voltage input variables in terms of the angular velocity desired trajectory. The scheme results in a sensor-less, certainty equivalence, adaptive trajectory tracking feedback control scheme jointly exploiting the energy dissipation structure of the system and its flatness. The results are illustrated by means of digital computer simulations.

1 Introduction

Robust control of DC motors subject to unforseen loads, even if of constant nature, constitutes an area of active research and one which has received numerous contributions in the past (See Bowes et al. [1]). The fundamental problem is the lack of control input channel matching in the manner in which the load affects the motor dynamics. The need for a robust feedback control scheme becomes more evident when a sensor-less feedback regulation scheme is desired (See Ko et al. [3] and [4]). This means that the feedback control law should be based only on the measured armature current while angular velocity measurements are deemed inconvenient or expensive (See the work of Park et al. [5] and that of Pai et al. [6]).

Passivity based control constitutes an interesting sensor-less alternative for the control of DC motors, given that the armature current qualifies as the passive output of the system thus providing to the feedback scheme useful energy management properties. We show in general terms, for the case of linear systems, that a dissipation matching condition not only represents a natural restriction for damping injection in passive systems, but it also determines the systems energy dissipation respecting properties that need to be complemented by the feedback control actions in any regulation setup. This dissipation matching condition, which seems to have been largely overlooked in the literature, is central in energy based control alternatives yielding linear, time varying, feedback control laws in switched power electronics devices modeled as bilinear systems (see Sira-Ramírez and Silva-Ortigoza [8]).

The exact state tracking error dynamics inherits the energy managing structure of the system and, in the case of linear systems, preserves the nature of the passive output definition (this is not the case in bilinear systems and in nonlinear systems). This property is responsible, in the DC motor case, for a rather simple linear passive output error feedback control scheme solving trajectory tracking tasks. The presence of constant perturbation input loads cannot be naturally handled by the static passive output error feedback control. For this reason, we provide a certainty equivalence differential parametrization of the input and passive outputs, in terms of the flat output angular velocity, that takes into account the load influence in the feed-forward terms. Adaptation of the on-line estimated load parameter, as provided by an observer with integral output estimation error injections, in the feed-forward terms bestows the required robustness to the proposed feedback controller.

This article is organized as follows: Section II is devoted to establish the Exact Tracking Error Dynamics Passive Output Feedback (ETEDPOF) controller design technique for a class of perturbed linear systems. We remark that such static passivity based technique has been extensively used also in the linear, time-varying, feedback control of bi-linear systems, of the type encountered in many switched Power Electronics devices (See Sira-Ramírez and Silva-Ortigoza [8]). Section III considers, in detail, the application of the proposed approach to the angular velocity trajectory tracking regulation for a

DC motor. Illustrative simulation results are presented in that section. The conclusions and outlook for further work are presented in the last section.

2 Problem Formulation

We deal with the following perturbed dynamics for a DC motor

$$\begin{aligned} L\frac{di_a}{dt} &= u - R_a i_a - k\omega \\ J\frac{d\omega}{dt} &= -B\omega + ki_a - \tau_L \\ y &= i_a \end{aligned} \tag{1}$$

where i_a is the measured armature circuit curren, ω is the angular velocity of the rotor. The parameters L, J, R_a, k and B are, respectively, the armature circuit inductance, the motor inertia, the armature circuit resistance, the motor gain, and the viscous damping coefficient, all assumed to be strictly positive and perfectly known. The constant parameter τ_L represents the unknown load torque. We assume that only the armature circuit current is measured.

Given the motor dynamics (1) and a desired angular velocity reference trajectory $\omega^*(t)$*, find, for any constant, unknown, load parameter* τ_L*, a linear output feedback control law* $u = \phi\left(i_a, \omega^*(t), \hat{\tau}_L\right)$*, based on an on-line estimate* $\hat{\tau}_L$ *of the load torque* τ_L*, such that the origin of the trajectory tracking error space, defined by* $e_\omega = \omega - \omega^*(t)$*, is a globally asymptotically exponentially stable equilibrium point for the closed loop system, with internal stability of the armature current.*

3 Exact Tracking Error Dynamics Passive Output Feedback (ETEDPOF)

3.1 The exact tracking error dynamics passive output feedback (ETEDPOF)

Consider the following state space model for a SISO, n-dimensional, time invariant, linear system

$$A\dot{x} = Jx - Rx + bu - \varepsilon(t) \tag{2}$$

where A is a positive definite symmetric matrix, J is a skew symmetric matrix, R is a symmetric positive semi-definite matrix, The vector $\varepsilon(t)$ represents external perturbation inputs. The vector b is a constant vector. The vector $x \in R^n$ is the state of the system and u is the control input.

System (2) exhibits a natural energy managing structure in which Ax represents the total stored energy rate of the system (or total power), Jx represents the conservative field, or conservative forces, $-Rx$ is the dissipative field and b is the control acquisition term. The passive output associated with (2) is formally considered to be $y = b^T x$.

Consider a nominal trajectory $x^*(t)$ produced by a nominal open loop controller $u^*(t)$ satisfied by the certainty equivalence perturbed relation:

$$\begin{aligned} A\dot{x}^* &= Jx^* - Rx^* + bu^* - \varepsilon(t) \\ y^*(t) &= b^T x^*(t) \end{aligned} \tag{3}$$

We are temporarily assuming that $\varepsilon(t)$ is a known quantity. We proceed to derive a perturbation dependent controller and in closing the feedback loop we will replace it by its on-line estimated value.

Define $e = x - x^*(t)$ as the state trajectory tracking error, $e_u = u - u^*(t)$ as the control input error and $e_y = y - y^*(t)$ as the passive output tracking error. In view of (2) and (3), the exact tracking error dynamics is given by

$$\begin{aligned} A\dot{e} &= Je - R\mathrm{e} + be_u \\ e_y(t) &= b^T e \end{aligned} \tag{4}$$

3.2 Some geometric facts

We assume that the following *dissipation matching* condition is valid:
For any strictly positive real parameter γ the symmetric matrix $R+\gamma bb^T$ satisfies the following positive definite condition:

$$R + \gamma bb^T > 0 \tag{5}$$

Let $\ker(M)$ denote the null space of a linear operator represented by the matrix $M : {}^n \to {}^n$. We denote by $\mathrm{Im}(M)$ the range space of M. If M is a symmetric matrix then $\ker^{\perp}(M) = \mathrm{Im}(M)$ and, hence, $\ker(M) \oplus \mathrm{Im}(M) = {}^n$ We have the following proposition:

Proposition 1. Let R be a symmetric positive semi-definite matrix. Then, the dissipation matching condition (5) is valid if, and only if,

$$\mathrm{Im}(R) \oplus \mathrm{Im}(b) = {}^n \tag{6}$$

equivalently, the dissipation matching condition is valid if and only if

$$\ker(R) \cap \ker(b) = \{0\} \tag{7}$$

Proof. The dissipation matching condition (5) states that for all $e \in {}^n$ we have: $e \in e^T \left[R + \gamma bb^T \right] e > 0$. Let $e \in \ker(R)$ then: $e \in e^T \left[R + \gamma bb^T \right] e = \gamma e^T bb^T e > 0$. Hence, $e \notin \ker(b^T)$ i.e., $e \in \ker^{\perp}(b^T)$ and, therefore: $e \in \mathrm{Im}(b^T)$. It follows that, $\ker(R) \subset \mathrm{Im}(b)$. Also, let $e \in \ker(b^T)$, then $e^T \left[R + \gamma bb^T \right] e > 0$ implies that $e^T Re > 0$ and $e \in \mathrm{Im}(R)$. It follows that $\ker(b^T) \subset \mathrm{Im}(R)$ which is an alternative form of the previously established relation. Take orthogonal complements in $\ker(b^T)$ to obtain that $\mathrm{Im}(R) \oplus \mathrm{Im}(b) = {}^n$. An orthogonal complement of this relation yields the relation (7).
The result follows.

Remark 2: The dissipation matching condition simply means that there is a complementarity between the system dissipation structure, represented by R and that provided by the possibilities of feedback, which enter through the subspace Imb. In other words, whatever is not naturally dissipated by the system, then it must be in the image of the control input channel, so that feedback control actions can take care of it and the system enjoys full dissipation properties. In its boolean complement form, the dissipation matching condition says that whatever does not go through to impress the passive output of the system (i.e., it is blocked by the passive output map), it better be dissipated by the natural system dissipation structure if the system is to enjoy full dissipation.

3.3 The ETEDPOF controller

Consider the static passive output error linear feedback control

$$e_u = -\gamma e_y = -\gamma b^T e, \quad \gamma > 0 \tag{8}$$

The closed loop state tracking error system results in

$$A\dot{e} = Je - \left[R + \gamma bb^T \right] e \tag{9}$$

Take as a Lyapunov function candidate the positive definite scalar function $V(e)=\frac{1}{2}e^T Ae$ representing the total stored energy. The time derivative of $V(e)$ along the solutions of the closed loop tracking error dynamics yields:

$$\dot{V}(e)=e^T\left[R+\gamma bb^T\right]e<0 \tag{10}$$

in agreement with our fundamental assumptions. We have then the following result:

Proposition 3: Given a desired nominal trajectory, $x^*(t)$, for system (2) which satisfies (3). Then if $E(t)$ is a known quantity, the static linear passive output error feedback control law:

$$u=u^*(t)-\gamma\left(y-y^*(t)\right)=u^*(t)-\gamma b^T\left(x-x^*(t)\right) \tag{11}$$

globally asymptotically exponentially stabilizes the origin of the state trajectory tracking error space, $e(t)=0$, provided the dissipation matching condition

$$R+\gamma bb^T>0 \tag{12}$$

is satisfied for any $\gamma>0$.

4 Application to the Control of a DC Motor with Unknown Load

We consider the following model of a DC motor

$$\begin{aligned} L\frac{di_a}{dt}&=u-R_a i_a-k\omega\\ J\frac{d\omega}{dt}&=-B\omega+ki_a-\tau_L\\ y&=i_a \end{aligned} \tag{13}$$

where i_a is the measured armature circuit current, which is the passive output, ω is the angular velocity of the rotor. The parameters L, J, R_a, k and B and B are all assumed to be perfectly known. The constant parameter τ_L represents the unknown load torque.

4.1 The ETEDPOF controller

Certainty equivalence passivity based controller

Let $i_a^*(t)$ and $\omega^*(t)$ be two desired nominal reference signals for the motor obtained from the following certainty equivalence nominal dynamics

$$\begin{aligned} L\frac{d}{dt}i^*{}_a(t)&=u^*(t)-R_a i_a^*(t)-k\omega^*(t)\\ J\frac{d}{dt}\omega^*(t)&=-B\omega^*(t)+ki_a^*(t)-\tau_L\\ y^*&=i_a^*(t) \end{aligned} \tag{14}$$

Defining the tracking errors,

$$e_1=i_a(t)-i_a^*(t),\qquad \mathrm{e}_2=\omega-\omega^*(t) \tag{15}$$

we readily obtain the following exact tracking error dynamics in matrix form:

$$\begin{pmatrix} L & 0 \\ 0 & J \end{pmatrix} \frac{d}{dt} \begin{pmatrix} e_1 \\ e_2 \end{pmatrix} = \begin{pmatrix} 0 & -k \\ k & 0 \end{pmatrix} \begin{pmatrix} e_1 \\ e_2 \end{pmatrix} - \begin{pmatrix} R_a & 0 \\ 0 & B \end{pmatrix} \begin{pmatrix} e_1 \\ e_2 \end{pmatrix} + \begin{pmatrix} 1 \\ 0 \end{pmatrix} e_u + \begin{pmatrix} 0 \\ -1 \end{pmatrix} \tau_L$$

$$e_y = e_1 \tag{16}$$

where $e_u = u - u^*(t)$ and $e_y = y - y^*(t)$.

The passive output of the system is given by the armature circuit current $y = i_a$ since this is a *relative degree* 1 output with the following asymptotically stable perturbed zero dynamics

$$J \frac{d}{dt} \zeta(t) = -B\zeta(t) - \tau_L \tag{17}$$

with $\zeta = \omega$ for $i_a = 0$.

The dissipation matching condition adopts, in this case, the form:

$$\begin{pmatrix} R_a + & 0 \\ 0 & B \end{pmatrix} > 0 \tag{18}$$

The choice of the control input error, e_{u}, as a feedback of the passive output of the trajectory tracking error dynamics:

$$e_u = -\gamma e_y = -\gamma \left(i_a - i_a^*(t) \right) \tag{19}$$

yields the feedback control law

$$u = u^*(t) - \gamma \left(i_a - i_a^*(t) \right) \tag{20}$$

4.2 Certainty equivalence trajectory planning

The linear ETEDPOF controller demands, for its synthesis, the knowledge of the nominal state and input trajectories. We generate the nominal trajectories $i_a^*(t)$ and $u^*(t)$ on the basis of the specification of a desired nominal angular velocity $\omega^*(t)$, which is the flat output of the system. Letting $\omega^*(t) = F^*(t)$ we have the following certainty equivalence differential parametrization of the motor variables:

$$\begin{aligned} i_a^*(t) &= \frac{J}{k} \left[\frac{d}{dt} F^*(t) + \frac{B}{J} F^*(t) + \frac{\tau_L}{J} \right] \\ u^*(t) &= \left(\frac{JL}{k} \right) \frac{d^2 F^*}{dt^2} + \left(\frac{BL}{k} + \frac{R_a J}{k} + k \right) \frac{dF^*}{dt} \\ &+ \left(\frac{R_a B}{k} \right) R^*(t) + \left(\frac{L}{k} \right) \frac{d}{dt} \tau_L + \frac{1}{k} \tau_L \end{aligned} \tag{21}$$

We take $d\tau_L / dt$ to be identically zero due to the assumed constant value of τ_L. However, when sudden load torques appear at an unpredicted moment of time, the time derivative of this torque represents an impulsive function which may be ignored thanks to the corrective action of the on-line portion of the proposed feedback controller.

4.3 Certainty Equivalence Trajectory Planning

Consider the following *fake*, or auxiliary expression, for the indirect measurement of the angular velocity in terms of the time derivative of the armature circuit current:

$$\omega = \frac{1}{k} \left(u - R_a i_a - L \frac{di_a}{dt} \right) \tag{22}$$

A reduced order Luenberger observer including integral action for the angular velocity, ω, is proposed as:

$$J\frac{d\hat{\omega}}{dt} = -B\hat{\omega} + ki_a + \lambda_1\left(\omega - \hat{\omega}\right) + \lambda_2\int_0^t \left(\omega(\sigma) - \hat{\omega}(\sigma)\right)d\sigma \tag{23}$$

where, naturally, ω is to be substituted by the artificial measurement ω given in (22). The analysis of the dynamics of the estimation error: $e_\omega = \omega - \hat{\omega}$, may be, nevertheless, carried out directly in terms of (23) given by

$$\dot{e}_\omega = -\left(B + \lambda_1\right)e_\omega - \lambda_2\int_0^t e_\omega(\sigma)d\sigma - \tau_L \tag{24}$$

The integro-differential equation (24) is equivalent to the following second order linear system with unknown initial conditions

$$\begin{aligned}\dot{e}_\omega &= -\left(B + \lambda_1\right)e_\omega - \lambda_2\rho_\omega \\ \dot{\rho}_\omega &= e_\omega\end{aligned} \tag{25}$$

where

$$\rho_\omega(t) = \int_0^t e_\omega(\sigma)d\sigma + \frac{\tau_L}{\lambda_2} \tag{26}$$

and $\rho_\omega(0) = \tau_L/\lambda_2$. The characteristic polynomial of (25) is clearly given by

$$p(s) = s^2 + \left(B + \lambda_1\right)s + \lambda_2 \tag{27}$$

and the appropriate choice of $\lambda_1, \lambda_2 > 0$ guarantees the asymptotic exponential stability of the equilibrium point, $e_\omega = 0$, for the angular velocity estimation error. Moreover, since: $\lim_{t\to\infty}\rho_\omega(t) = 0$, then

$$\lim_{t\to\infty}\left[-\lambda_2\int_0^t e_\omega(\sigma)d\sigma\right] = \tau_L \tag{28}$$

It follows that the unknown load torque τ_L may be asymptotically estimated.

Substituting the value of ω given in (22) into the injection term of the reduced order observer (23) one obtains:

$$J\frac{d\hat{\omega}}{dt} = -B\hat{\omega} + ki_a + \frac{\lambda_1}{k}\left(u - R_a i_a - L\frac{di_a}{dt} - k\hat{\omega}\right) + \frac{\lambda_2}{k}\int_0^t\left(u - R_a i_a - L\frac{di_a}{dt} - k\hat{\omega}\right)d\sigma \tag{29}$$

Define:

$$\xi = \left[J\hat{\omega} + \left(\frac{\lambda_1 L}{k}\right)i_a\right] = \left[J\hat{\omega} + \left(\frac{\lambda_1 L}{k}\right)z\right] \tag{30}$$

Algebraic manipulations of the observer expression leads, in terms of the defined variable ξ to the following modified observer dynamics

$$\begin{aligned}\dot{\xi} &= -\left(\frac{B + \lambda_1}{J}\right)\xi + \frac{\lambda_1}{k}u - \left(\frac{\lambda_1 R_a J - \lambda_1^2 L - BL\lambda_1 - Jk^2}{Jk}\right)z \\ &\frac{\lambda_2}{k}\int_0^t\left[u - \left(R_a - \frac{\lambda_1 L}{J}\right)z - \frac{k}{J}\xi\right]d\sigma + \frac{\lambda_2 L}{k}\left(z - i_a(0)\right)\end{aligned} \tag{31}$$

The estimated angular velocity $\hat{\omega}$ is thus obtained from the solution of (31) for ξ, from arbitrary initial conditions and

$$\hat{\omega} = \frac{1}{J}\left(\xi - \frac{\lambda_1 L}{k} z \right) \tag{32}$$

The unknown load torque, τ_L, is then asymptotically estimated from the expression:

$$\begin{aligned} \hat{\tau}_L &= -\frac{\lambda_2}{k}\int_0^t \left[u(\sigma) - R_a i_a(\sigma) - L\frac{di_a(\sigma)}{d\sigma} - k\hat{\omega}(\sigma) \right] d\sigma \\ \hat{\tau}_L &= -\frac{\lambda_2}{k}\int_0^t \left[u(\sigma) - \left(R_a - \frac{\lambda_1 L}{J} \right) z(\sigma) - \frac{k}{J}\xi(\sigma) \right] d\sigma + \frac{L\lambda_2}{k}\left[z - i_a(0) \right] \end{aligned} \tag{33}$$

The asymptotic load parameter estimate (33) is to be substituted on the feed-forward terms of the passivity based controller given by (20) as demanded by the nominal armature current and nominal control input expressed in (21).

5 Simulation Results

In this section we present the simulations obtained for the responses of the DC motor system to our proposed observer-feedforward adaptive passivity based controller.

We used the following motor parameters values:

Armature resistance: $R_a = 6.14$ [Ohm], armature inductance: $L = 8.9$ [mH]. Viscous friction coefficient: $B = 40.923$ [μ(Nm-s)/rad]. Motor constant $K = 0.04913$ [N-m/A]. Motor moment of inertia: $J = 7.95$ [μKg-m²]. Input voltage bound $E = 24$ [V]. The unknown load torque was chosen to be: $\tau_L = 0.06$ [N-m].

In the simulations, we set as a reference trajectory an angular velocity maneuver taking the response from 0 [rad/s] to 300 [rad/s] in 0.3 seconds. Figures 1 and 2 depict the armature current response and the applied control input. The figures clearly depict the effect of a sudden appearance of a constant load torque at time $t = 0.7$ [s]. The angular velocity response rapidly recovers from the load perturbation input as it is depicted in Figure 3, where the desired reference angular velocity trajectory is shown to differ little from the controlled trajectory.

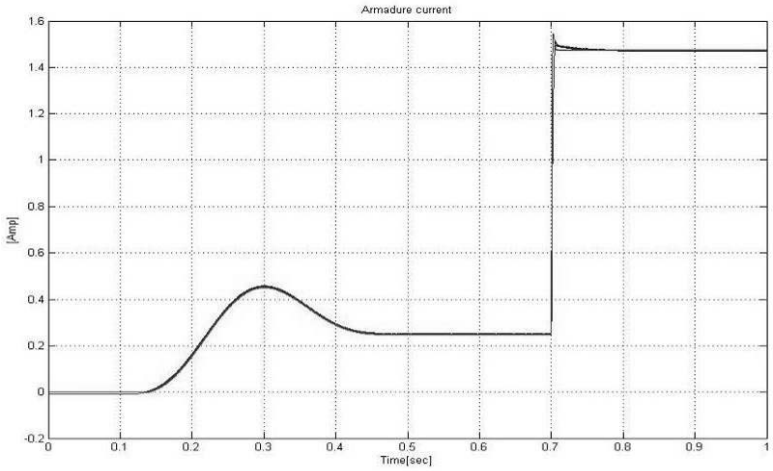


Figure 1. Armature current.

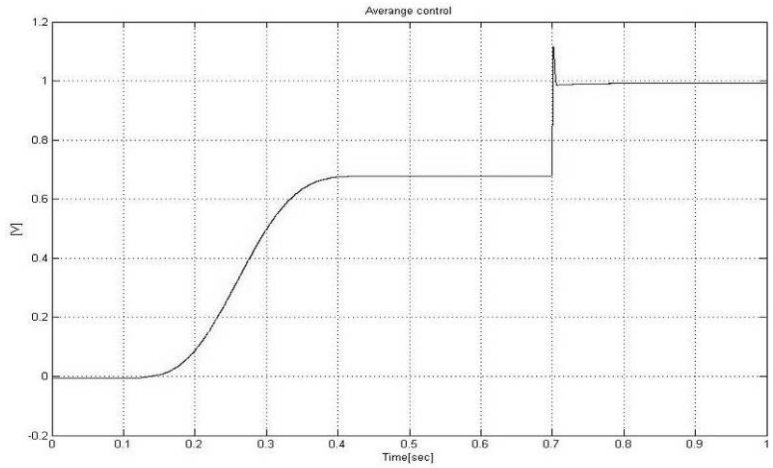


Figure 2. *Control Input.*

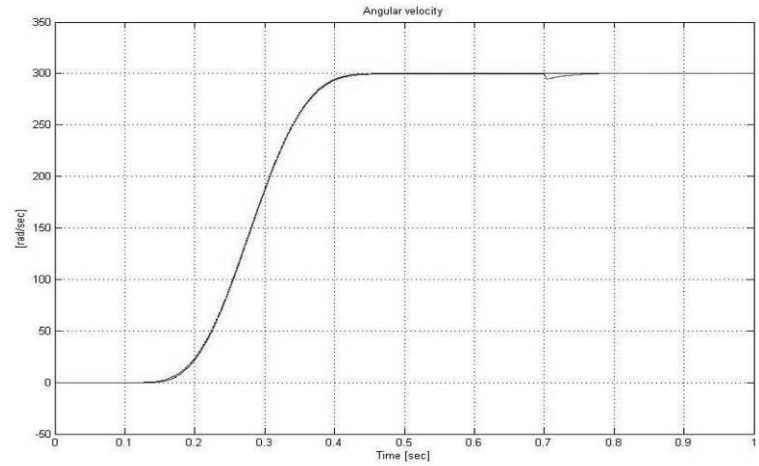


Figure 3. *Angular velocity response.*

6 Conclusions and Future Works

In this article we have proposed an adaptive feedback control scheme for a DC motor with unknown constant loads, based on the Exact Tracking Errror Dynamics Passive Output Feedback (ETEDPOF) control methodology. The unknown motor load is asymptotically estimated by means of a reduced order angular velocity observer including integral output estimation error injections. The flatness property of the DC motor system is used to readily obtain, in terms of the angular velocity reference trajectory, a certainty equivalence differential parametrization defining the nominal measured passive output and control input variables reference trajectories. These quantities are explicitly demanded by the ETEDPOF linear controller feedforward terms. The estimated load torque is on-line fed into the nominal feedforward terms in the controller. As a result we obtain a robust sensorless feedback control scheme, based only on armature circuit current tracking error feedback and feedforward load torque adaptation, for the angular velocity trajectory tracking task in a DC motor.

The reduced order observer is allowed to include a reasonably high injection gain for achieving fast estimation and, consequently, fast feedforward adaptation.

7 Future works

Encouraged by the presented simulation results, future publications will include experimental tests on the proposed linear passivity based controller with feedforward adaptation. We will establish comparisons with some other recent fast parameter estimation techniques. In particular, we will be using, for these comparison purposes, results obtained via algebraic

perturbation estimation techniques (See Fliess and Sira-ramírez [2] for details). These last are based on derivative estimations of noisy measured outputs (See Reger et al. [7]).

References

1. S. R. Bowes, A. Sevinç, and D. Holliday "New Natural Observer Applied to Speed-Sensorless DC Servo and Induction Motors" *IEEE Transactions on Industrial Electronics*, Vol. 51, NO. 5, October 2004.
2. M. Fliess and H. Sira-Ramírez "An algebraic framework for linear identification" *ESAIM, Control, Optimization and Calculus of Variations*, Vol 9, pp. 151-168, January. 2003.
3. J. S. Ko, "Asymptotically stable adaptive load torque observer for precision position control of BLDC motor," *Proc. IEE-Electr. Power Applications*, vol. 145, no. 4, pp. 383-386, 1998.
4. J. S. Ko, J.H. Lee, S.K. Chung and M.J. Youn " A robust position control of brushless DC motor with dead beat load torque observer", *IEEE Trans. Ind. Electron*., Vol. 40, No. 5, pp. 512-520, 1993.
5. K-H. Park, T-S. Kim, S-Ch. Ahn; D-S. Hyun; "Speed control of high-performance brushless DC motor drives by load torque estimation" *IEEE 34th Annual Power Electronics Specialist Conference (PESC '03)*. Volume 4, June 15-19, 2003. pp. 1677-1681.
6. D. A. Pai, M. Purnaprajna and R. Rao, "Nonlinear Observer Based Sensorless Direct Torque Control of Induction Motor." *In Proceedings The Third International Power Electronics and Motion Control Conference 2000 (PIEMC 2000)*, Beijing, China. 2000. Vol.1, pp. 440-445.
7. J. Reger, H. Sira-Ramírez and M. Fliess, " On non asymptotic estimation of nonlinear systems" *44th. IEEE Conference on Decision and Control*, Sevilla, Spain, December 2005.
8. H. Sira-Ramírez and R. Silva-Ortigoza Control Design Techniques in Power Electronics Devices, Springer, Power Systems Series. London, 2006.

PART III

EVOLUTIONARY COMPUTATION

Chapter 12

Evolutionary Computation Techniques for Two Computational Biology Problems

Carlos A. Brizuela-Rodríguez, Milton Rodríguez-Zambrano and Jorge E. Luna-Taylor

Centro de Investigación Científica y de Educación Superior de Ensenada,
Km. 107 Carr. Tijuana-Ensenada, Ensenada, B.C.
Email: {cbrizuel, milton, lunat}@cicese.mx

Abstract

This chapter deals with the design of evolutionary algorithms for two well known computational biology problems: whole-genome shotgun assembly and a simplified model of the protein folding. The folding problem is one of most challenging open problems in biology, and any clue on how to solve it or its approximations will be very valuable. Evolutionary algorithms have shown their suitability for complex optimization problems. The idea to solve the problems is to use a genetic algorithm tailored to specific problem knowledge. The proposed method has proven to be an effective alternative to solve both problems.

Keywords: Shotgun Assembly, Protein Folding, Genetic Algorithms, Monte Carlo, Local Search.

1 Introduction

In this chapter we present two very important computational biology problems: DNA assembly and an approximation to the protein folding. The latter constitutes one of the most challenging open problems in molecular biology.

DNA sequencing is the process to decipher the precise order of bases along an unknown DNA chain. With today's technology it is possible to sequence 300 to 1000 bp[1] [1], in a single experiment. However, whole genomes are many times longer, for instance the human genome, without being the largest is in the order of 10^9 bp. To overcome this situation the whole-genome shotgun sequencing has been developed. This technique developed by Sanger and Coulson [2] was applied to sequence the *lambda* bacteriophage.

The shotgun method consists on randomly fragmenting the original DNA sequence. Each of the resulting fragments (reads) are then sequenced, and finally an assembly stage takes place, long fragments or contigs of the original sequence are reconstructed.

Many strategies have been proposed to deal with this problem, among these, greedy methods have been the most widely used. Here we propose a genetic algorithm to deal with the assembly problem, motivated mainly by the success on the application of this method to a wide class of combinatorial optimization problems, category where the assembly problem belongs.

There are many available programs to deal with this problem, among them we can cite: PHRAP[2], TGI[3], CAP3 [3], ARACHNE [4] and EULER[4] [5], the first four are based on greedy algorithms while the last one abandon what is known as the Overlap-Layout-Consensus (OLC) approach and is based on the D'brujin graph representation where the goal is to find a super Eulerian path.

Previous attempts to solve the problem by evolutionary computation have been the proposal of Parsons et al. [6] and most recently the one by Luque et al. [7]. Luque et al. proposed a parallel genetic algorithm in a ring topology where k subpopulations evolve independently. In this model a number of individuals is interchanged among these subpopulations with a pre-specified frequency.

On the other hand, one of the most important open problems in computational biology is the protein folding problem. The main goal in this problem is to predict the three-dimensional (3D) structure of proteins in their native states based

[1] base pairs

[2] http://www.phrap.org, March-2005

[3] http://www.tigr.org/tdb/tgi/software, March-2005

[4] http://www-cse.ucsd.edu/groups/bioinformatics, April-2005

solely in the linear sequence of amino-acids [8]. The research effort in this area has started in the 50's when Linus Pauling proposes the existence of a thermodynamically stable state with a helix structure for a particular class of proteins. This fact was confirmed experimentally in 1958 [1] for the protein known as myoglobin.

In the last decade several efforts to solve a simplified variant to this problem have been carried out. In this variant only hydrophobic interactions are taken into account, and the twenty amino-acids are classified into two groups: hydrophobic (H) and polar (P). In spite of these efforts, the problem has been solved only for short amino-acids sequences. There are many algorithms for the 2D model as well as for the 3D model. Some of these algorithms are exact [8, 10]. The algorithms are capable of finding optimal solutions in 3D cubic lattices for sequence of up to 88 amino-acids, with computation times ranging from minutes to hours. There exist also approximation algorithms [11-13]. These algorithms guarantee an approximation factor to the optimal solution. To date the best approximation ratio are as follows $1/3$ for 2D HP and $3/8$ for 3D HP. There also exist heuristics based algorithms. These algorithms can be divided into two categories: Evolutionary Algorithms [14-19], and Monte Carlo algorithms [20-24]. Although these algorithms do not guarantee the optimum they obtain good solutions efficiently.

Both problems, under their simplest models belong to the NP-hard class of problems [25], [32], [33], which basically implies that there is no known method to solve the problem in polynomial time in the number of fragments for the assembly problem, and in the number of residues for the folding problem.

The remainder of this chapter is organized as follows. Section 2 introduces the assembly problem and explains the genetic algorithm proposed to deal with this problem. Section 3 introduces the protein folding problem. Section 4 briefly describes the experiments and results achieved with the proposed algorithms. Finally, Section 5 states the conclusions and some ideas for future research.

2 The DNA Assembly Problem

Pevzner [26] proposes the following analogy to the problem of DNA fragment assembly: imagine we have many copies of a book; each of these copies is chopped in many, said ten million, pieces. Assume that 10% of these pieces are removed and the rest are splashed with ink. The target is to build a complete copy of the book. The pieces corresponds to the DNA fragments, the ink on some pieces corresponds to the base errors in the fragments, the missing pieces corresponds to the missing fragments. The goal of building a copy of the book corresponds to assembling the original DNA sequence.

A definition for the DNA assembly problem is as follows:

Definition 1. The input is a set of DNA fragments over the four letter alphabet $\sum = \{A,T,C,G\}$, the fragments can have different lengths from hundreds to thousands of bases. The output or solution is the shortest common superstring of this set $\sum$.

This definition does not consider that each of the input fragments is a result of a sequencing process which does not produce error free fragments. The base errors can be classified in substitutions and indels (insertions/deletions). Base substitution errors are in the range of 0.5 to 2% [4]. Furthermore, there is still another factor that makes the problem more complicated, namely the fragments orientation. With shotgun the fragments can come from any of the strands in the double-helix.

Taking into account all these factors we can redefine the problem as follows:

Definition 2. The input is a set P of reads over the alphabet $\sum = \{A,T,C,G\}$, each fragment p_i is of length d_i. The orientation of each p_i is unknown and each fragment can contain base errors. The output is a permutation of the p_i's of length $|P|$, such that the alignment score F(Π) among them is maximized.

$$F(\Pi) = \frac{\sum_{i=0}^{n-2} w(\Pi[i], \Pi[i+1])}{c(\Pi)} \tag{1}$$

where,

- d_i is the length of fragment i.
- Π is a permutation of the fragments.
- n is the total number of fragments.
- $\Pi[i]$ outputs the index of the fragment at position i within the permutation Π.

- $w(\Pi[i],\Pi[i+1])$ is a conditioned local alignment (alc) between fragment $\Pi[i]$ and the $\Pi[i+1]$.
- $c[\Pi]$ is the number of contigs given by permutation Π.

Equation 1 forces us to search for a small number of well-overlapped contigs.

We consider here a graph based model of the problem. We start by transforming the input of the problem into a complete directed graph G=(V,E) as follows:

- **Vertexes:** correspond to each of the k initial fragments that are denominated $p_1, p_2, ..., p_k$. Since the orientation for each of these fragments is unknown they cannot be assembled correctly unless all of them come from the same strand, which is very unlikely to happen. Due to this, it is proposed to add k segments that are the reverse complement of each fragment in the input to which we denominate $p_{1'}, p_{2'}, ..., p_{k'}$. Each of these $2k$ vertexes in V is assigned a weight of one.
- **Arcs:** Each arc $(p_i, p_j) \in$ E of the graph G has an assigned gain $g_{i,j}$ that represents the alc between the fragments p_i and p_j.
- **Objective:** To find a path r in the graph, such that the sum of visited arcs' gains is maximized, and the sum of the vertexes weights do not exceed k. This constraint is to guarantee that we are using no more than the k fragments originally in the input. For any path r a fragment p_i is considered to be in the path whenever the fragment $p_{i'}$ was not previously included in r, the same applies to $p_{i'}$.

Figure 1 shows a graph constructed from an input of three reads, the vertexes p_1, p_2 and p_3 represents to these fragments, the graph is completed with the vertexes: $p_{1'}$, $p_{2'}$ and $p_{3'}$ which represent the reverse complement of the three original fragments.

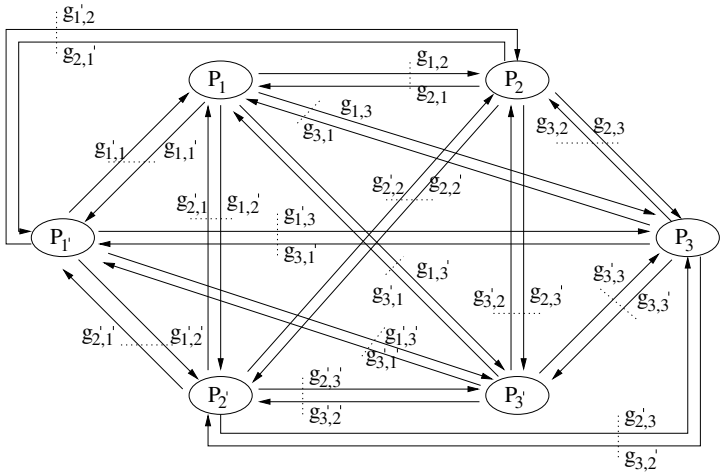


Fig. 2. *A graph constructed from an input of three DNA fragments (p_1, p_2, p_3) the reverse complement of these fragments are also considered ($p_{1'}, p_{2'}, p_{3'}$)*

2.1 The proposed Genetic Algorithm

The methodology proposed to solve the problem is based on a genetic algorithm, which constitutes one of the paradigms in evolutionary computation [27]. Next we show the pseudocode for the proposed algorithm (Algorithm 1).

Algorithm 1. GAssembler

Input: A set of DNA fragments
Output: A set of Contig(s) maximizing Equation 1

- **Preprocessing** { Generation of reverse complements; Overlaps Detection; Elimination of redundant fragments }

"GA starts"

- Generation of initial population
- Fitness Function Evaluation
- **Cycle starts**
- Selection
- Crossover
- Mutation
- Fitness Function Evaluation
- **After l iterations** { Fragments elimination Phase and individuals′ adjustment }
- **Cycle ends**
- Consensus

We present a brief explanation for the components of Algorithm 1.

Preprocessing. This preprocessing consist of three parts: the reverse complement computation, overlaps detection, and elimination of fragments. The first part is in charge of computing the reverse-complement sequence for each fragment in the input. The second part needs to compute the overlaps between fragments taking into account the substitution errors. Finally, the third part consist on eliminating all fragments that are subsegments, prefix, or suffix of any other fragment in the input.

Representation. An important issue to apply a GA in the resolution of a problem is to decide what representation to use. It is easy to see that a solution for the problem is a permutation of the fragments indexes, a permutation of $2k$ numbers. The codification we use is known as adjacency based representation [28].

Selection. We select the parents with the stochastic universal sampling algorithm ([27], page 62).

Crossover. A greedy crossover operator [29] is used. This operator produced high quality results for the sequencing by hybridization problem.

Mutation. The operator known as swap mutation for permutations is used ([27], page 45).

Fragment Elimination and adjustments of individuals. After l iterations we search for the longest contigs and analyze the fragments used on each of them. Once a fragment is used in a contig, if its reverse-complement appear anywhere downwards, then the latter is eliminated. After this operation is performed the individuals need to be adjusted by erasing their genes corresponding to reverse-complements that were already in contigs upwards.

The details for each component of the proposed algorithm can be found in [30].
In the next section we introduce the second problem to solve.

3 The Hydrophobic Polar (HP) model

An import nt assumption under this HP model is that the most significant force acting on a protein, and that contributes a great deal in the changes of the free energy, is the hydrophobic interaction [31]. This force is related to the property of certain molecules for repelling or attracting molecules of water.

When the amino-acids merge to form the peptide bond they free a water molecule. The resulting monomers are denominated *residues*, i.e. the residues are the amino-acids after having lost some atoms as a consequence of the binding. Depending on their side chain structure, these residues can be hydrophobic (repels water) or polar (attract water). Since many proteins, in their native states, are in an aqueous solution, the hydrophobic residues tend to concentrate in the interior part of the structure, defining in this way a particular folding.

Under this consideration the first simplification is undertaken. The 20 amino-acids are represented by only two monomers H or P, depending upon their affinity with water. The second simplification in this model is that each residue can occupy only the intersecting points in a 2D square lattice or a 3D cubic lattice. All self-avoiding walks on this lattice are allowed, i.e. the monomers are not allowed to collide with each other.

Under this model an H-H *contact* is defined as any pair of residues H's that are neighbors in the lattice but not in the sequence. To such contacts an energy of -1 is assigned. Since the natural state (native conformation) of a protein is given by is minimal energy free conformation, the optimal folding in the HP model is the one with the maximum number of H-H contacts (Figure 2).

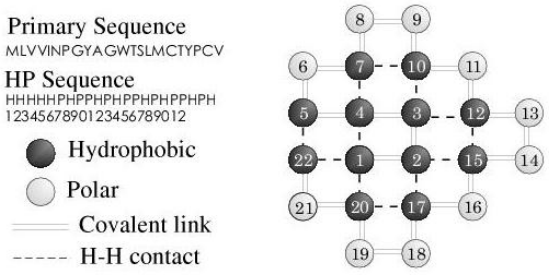


Fig. 3. *Optimal folding for the sequence MLVVINPGYAGWTSLMCTYPCV (HHHHHPHPPHPHPPHPHPPHPH), with a total of 12 H-H contacts, under the 2D HP model.*

3.1 Problem Definition

Under the HP model an amino-acid sequence of length n can be represented as $S = s_1 s_2 \ldots s_n$ where $s_i \in \{H, P\}, 1 \le i \le n$. The 3D structure of S is defined by a sequence of directions for the folding, starting in the first residue of the sequence. The codification to represent 3D conformations in the cubic lattice uses the symbols U, D, L, R, F, and B to denote the folding directions Up, Down, Left, Right, Front, and Back, respectively (see Figure 3).

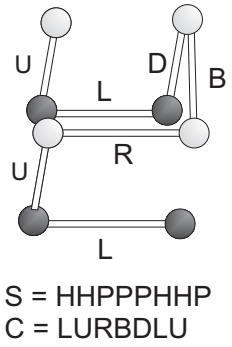


Fig. 4. *Optimal folding for the sequence S and its corresponding codification C, under the 3D HP model. The sequence starts with a hydrophobic residue located at the bottom right.*

A valid folding is the one where each location in the lattice has at most one residue, and each residue is connected to its neighbor(s) in the sequence, which is at the same time, a neighbor location(s) in the lattice.

Definition 3. Under the 3D HP model we can define the folding problem from an input/output point of view as follows:

Input: A sequence $S = s_1 s_2 s_3 \ldots s_n$ where $s_i \in \{H, P\}, 1 \le i \le n$, which represents a sequence of residues.

Output: A conformation or valid folding C on a 2D or 3D lattice that has the maximum number of H-H contacts.

In spite of all simplifications, it was proven that the problem in 2D HP is HP-hard [32], the same holds for the 3D HP model [33].

Yue and Dill have [8] proposed an algorithm to solve the problem. In their method they first find what is known as the optimal H-core. This core contains all hydrophobic residues in the sequence. Once they determine the volume and shape of this core, a combinatorial algorithm based on exhaustive search is applied to find the layout of the H residues inside the core. The latter problem of finding the layout of H residues given the optimal core is the second problem we are dealing with in this chapter.

Definition 4. Given the optimal H-core find the layout of hydrophobic residues inside the core, such that all amino-acids in the sequence remain connected.

3.2 The Proposed Algorithm

The algorithm has three main components: a global GA which evolves the layouts inside the core, a Monte Carlo method which helps to avoid exhaustive search of all possible combinations inside the core, and finally, the internal GA that helps to connect polar segments (outside the core) between pairs of hydrophobic residues (inside the core).

Algorithm 2. P-Folder

1. define the frame and the H-core
2. repeat n times
3. initialize the population
4. Start global GA: Initialize the Population
5. Fitness Evaluation
6. search for paths connecting H-H residues
7. run internal GA to compute the total number of connections
8. select the parents
9. apply MC based crossover (MC to handle constraints)
10. Fitness Evaluation
11. search for paths connecting H-H residues
12. run internal GA to compute the total number of connections
13. apply MC based mutation (MC to handle constraints)
14. end global GA
15. apply local optimization
16. end repeat
17. start refinement phase
18. select the best $\alpha\%$ individuals from the previous phase
19. compute the longest connected segment
20. repeat global GA with fixed segments
21. end refinement phase
22. output the best individual

The details for the proposed algorithm to deal with the 3D HP protein folding problem can be found in [34].
In Line 1 we need to define the frame and the H-core. This is achieved by the method proposed by Yue and Dill [8].

Generation of Initial Population. Once the H-core is defined a random layout of the H residues inside the core is performed for each individual. Note that it is not mandatory that all residues are connected at this stage.

Fitness Computation. The fitness is given by the total number of connected residues. Since the number of possible paths between pairs of residues grows exponentially with the longest length of a P segment, an internal GA is used to compute the connections. That is, the objective function is computing by running an internal genetic algorithm.

4 Computational Experiments and Results

4.1 Results for the DNA assembly problem

Engle and Burks [35] proposed a set of tools denominated *genfrag* to generate instances to test new assemblers. Here we use an implementation of this generator. We deal with sequences of up to 20000 base pairs.

The results were evaluated considering the number of sequences generated by the algorithm and the similarity of these sequences with the original one. On instances where the error was at the level of 5% the genetic algortihm clearly outperforms the TIGR [30].

4.2 Results for the protein folding problem

The results for the protein folding problems are also encouraging, out of 13 sequences 12 were laid out in their optimal positions. Among these; 10 sequences are of 48 residues, one of 64, one of 67, and one of 88 residues. The instance with 88 residues were the only one on which the algorithm could not find the optimal layout [34].

5 Conclusions

New strategies based on genetic algorithms and heuristics to tackle the DNA assembly and a simplified model of the protein folding problems have been proposed. The use of problem specific knowledge inside the algorithm has shown to be very effective for solving complex problems. The protein folding problem, under the 3D cubic lattice model, remains a computationally challenging problem

Future work will be focused on applying these methods to solve bigger instances for both problems. Another important open Bioinformatics problem to explore with these ideas has to do with: given a 3D structure of a protein, predict its function.

Acknowledgments

This research was partially supported by CONACyT under grant C01-45811 and by the Laboratotio Franco-Mexicano de Informática.

References

1. Mihai Pop, L. Salzberg, and M. Shumway. Genome sequence assembly: algorithms and issues. *IEEE computer*, 35(7):47--54, 2002.
2. F. Sanger and A. R. Coulson. The use of thin acrylamide gels for dna sequencing. *FEBBS*, 87:107--110, 1978.
3. Xiaoqiu Huang and Anup Madan. Cap3: a dna sequence assembly program. *Genome Research*, 9:868--877, 1999.
4. Ken Stanley Jonathan Butler Sante Gnerre Evan Mauceli Bonnie Berger Jill P. Mesirov Serafin Batzoglou, David B. Jaffe and Eric S. Lander. Arachne: A whole-genome shotgun assembler. *Genome Research*, (12):177--189, 2002.
5. Pavel A. Pevzner, Haixu Tang, and Michael S. Waterman. An eulerian path approach to dna fragment assembly. *PNAS*, 98(17):9748--9753, 2001.
6. Rebecca J. Parsons, Stephanie Forrest, and Christian Burks. Genetic algorithms, operators, and dna fragment assembly. *Machine learning*, 21:11--33, 1995.
7. Gabriel Luque, Enrique Alba, and Sami Khuri. *Parallel algorithms for solving the fragment assembly problem in DNA strands*, chapter Parallel computing for bioinformatics and computational biology. John Wiley & Sons. New York, pages 287--304. En: A. Y. Zomaya (ed.), 2005.
8. K. Yue and K. Dill. Sequence-structure relationships in proteins and copolymers. *Phys. Rev. E*, 48(3):2268--2278, Septiembre 1993.
9. [1] J. Kendrew, G. Bodo, H. Dintzis, R. Parrish, H. Wyckoff, and D. Phillps. A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. *Nature*, 181(4610):662--666, Mar 1958.
10. K. Yue and K. Dill. Forces of tertiary structural organization in globular proteins. *Proc Natl Acad Sci*, 92(1):146--150, Enero 1995.
11. W. Hart and S. Istrail. Fast protein folding in the hydrophobic-hydrophilic model within thee-eights of optimal. In *Proceedings of the twenty-seventh annual ACM symposium on theory of computing*, pages 157--168, 29 de mayo al 1 de junio, Las Vegas, Nevada, Junio 1995.
12. A. Newman. A new algorithm for protein folding in the hp model. In *Proceedings of the thirteenth annual ACM-SIAM symposium on discrete algorithms*, pages 876--884, 6 al 8 de enero, San Francisco, CA, Enero 2002.
13. A. Newman and M. Ruhl. Combinatorial problems on strings with applications to protein folding. In *Proceedings of the sixth Latin American symposium on theoretical informatics*, pages 369--378, 5 al 8 de abril, Buenos Aires, Argentina, Abril 2004.
14. T. Jiang, Q. Cui, G. Shi, and S. Ma. Protein folding simulations of the hydrophobic-hydrophilic model by combining tabu search with genetic algorithms. *Journal of Chemical Physics*, 119(8):4592--95, 2003.
15. R. Konig and T. Dandekar. Improving genetic algorithms for protein folding simulations by systematic crossover. *BioSystems*, 50(1):17--25, 1999.
16. N. Krasnogor, W. Hart, J. Smith, and D. Pelta. Protein structure prediction with evolutionary algorithms. In *Proceedings of the genetic and evolutionary computation conference*, pages 1596--1601, 13 al 17 de julio, Orlando, Florida, Julio 1999.
17. A. Patton, W. Punch, and E. Goodman. A standard ga approach to native protein conformation prediction. In *Proceedings of the sixth international conference on genetic algorithms*, pages 574--581, julio, San Francisco, CA, Julio 1995.

18. A. Shmygelska and H. Hoos. An improved ant colony hoptimisation algorithm for the 2d hp protein folding problem. In *Proceedings of te sixteenth conference of the Canadian society for computational studies of intelligence*, pages 400--417, junio, Halifax, Canada, Junio 2003.
19. R. Unger and J. Moult. A genetic algorithm for 3d protein folding simulations. In *Proceedings of the fifth international conference on genetic algorithms*, pages 581--588, 17 al 22 de julio, Illinois, Julio 1993.
20. T. Beutler and K. Dill. A fast conformational search strategy for finding low energy structures of model proteins. *Protein Sci.*, 5:2037--2043, 1996.
21. U. Bastolla, H. Frauenkron, E. Gerstner, P. Grassberger, and W. Nadler. Testing a new monte carlo algorithm for protein folding. *Proteins*, 32(1):52--66, 1998.
22. H. Hsu, V. Mehra, W. Nadler, and P. Grassberger. Growth algorithms for lattice heteropolymers at low temperatures. *Journal of Chemical Physics*, 118(1):444--451, Enero 2003.
23. F. Liang and W. Wong. Evolutionary monte carlo for protein folding simulations. *Journal of Chemical Physics*, 115(7):3374--3380, 2001.
24. R. Ramakrishnan, B. Ramachandran, and J. Penky. A dynamic monte carlo algorithm for exploration of dense conformational spaces in heteropolymers. *Journal of Chemical Physics*, 106(6):2418--2424, 1997.
25. Dan Gusfield. *Algorithms on strings, trees, and sequences: computer science and computational biology*. Cambridge University Press, Nueva York, NY, 1997.
26. P. A. Pevzner. *Computational molecular biology: An algorithmic approach*. The MIT Press, London, 2000.
27. A.E. Eiben and J.E. Smith. *Introduction to evolutionary computing*. Springer, Alemania, 2003.
28. John J. Grefenstette, Rajeev Gopal, Brian J. Rosmaita, and Dirk Van Gucht. Genetic algorithms for the traveling salesman problem. In John J. Grefenstette, editor, *Proceedings of the 1st International Conference on Genetic Algorithms*, pages 1:160--168. Lawrence Erlbaum Associates, 1985.
29. C. Brizuela, L. González, and H. Romero. An improved genetic algorithm for the sequencing by hybridization problem. In *Proceedings of the 3rd European Workshop on Evolutionary Computation in Bioinformatics, EvoBio, LNCS*, volume 3005, pages 11--20. Springer-Verlag, 2004.
30. Milton Rodriguez-Zambrano. *Un algoritmo genetico para el ensamble de secuencias de ADN*. CICESE, Ensenada, 2005.
31. K. Dill. Dominant forces in protein folding. *Biochemistry*, 29(31):7133--7155, Agosto 1990.
32. Pierluigi Crescenzi, Deborah Goldman, Christos Papadimitriou, Antonio Piccolboni, and Mihalis Yannakakis. On the complexity of protein folding. *Journal of Computational Biology*, 5(3):423--466, 1998.
33. B. Berger and T. Leighton. Protein folding in the hydrophobic-hydrophilic (hp) model is np-complete. In *Proceedings of the second annual international conference on computational molecular biology*, pages 30--39, 22 al 25 de marzo, New York, Marzo 1998.
34. Jorge E. Luna-Taylor. *Un algoritmo evolutivo hibrido para el problema de plegamiento de proteinas bajo el modelo hidrofobico polar en tres dimensiones*. CICESE, Ensenada, 2006.
35. Michael L. Engle and Christian Burks. Artificially generated data sets for testing dna sequence assembly algorithms. *Genomics*, 16:286--288, 1993.

Chapter 13

Design of Non-uniform Phased Linear Arrays using a Multi-objective Genetic Algorithm

Marco A Panduro

Reynosa-Rhode Multidisciplinary Academic Center, University of Tamaulipas (UAT)
Carretera Reynosa-San Fernando, Reynosa, Tamaulipas, 88779 México
Phone: (52) 899.921.3300, Fax: (52) 899.921.3301, E-mail: mamendoza@uat.edu.mx

Abstract

This chapter deals with the design of non-uniform phased linear arrays for smart antenna systems. The design problem is modeled as a multi-objective optimization problem with nonlinear constraints. A multi-objective genetic algorithm denominated NSGA-II is employed as the methodology to solve the resulting optimization problem. The main goal and contribution of this chapter is to extend a previous result on linear antenna array design to a related and more complicated problem. The addressed problem considers a driving-point impedance restriction and the design of non-uniform arrays to have a steerable radiation pattern while the previous result does not. Experimental results show the effectiveness of the NSGA-II for the design of non-uniform phased linear arrays.

Keywords: Non-uniform arrays, multi-objective, genetic algorithms, side lobe level, main beam width, radiation pattern.

1 Introduction

In mobile and wireless communications systems, the antenna array performance over a certain steering range is of primary concern. Phased antenna arrays used at the base station are able to reduce interference, due to their beam directivity. In CDMA systems like UMTS, the Multiple Access Interference (MAI) reduction provided by phased antenna arrays is translated into either more users in the system, or higher bit-rates for the existing users [1]. The performance of these systems depends strongly on the antenna array design.

Generally speaking, the problem of designing antenna arrays is characterized by different and conflicting requirements (beam width, side lobe level, directivity, noise sensitivity, robustness) to be satisfied. It is an example of a multi-objective problem. In this chapter two design criteria are considered to evaluate the performance of antenna arrays: the criteria of minimum main beam width and minimum side lobe level. In this case, the antenna array design problem consists of finding weights and antenna element positions that make the radiation pattern satisfy the desired characteristics: a minimum main beam width and a minimum side lobe level, so the direction of the main beam can be steered at will.

The design problem aims at minimizing at least two (main beam width and side lobe level) conflicting objectives at the same time. This is, therefore, a natural multi-objective problem.

Recently, heuristic-iterative methods such as simulated annealing [2], [3] and genetic algorithms [4]-[10] have been applied to the design of electronically steerable antenna arrays. These works consider design of steerable antenna arrays to be a problem optimizing a single objective function. In most cases these works consider the minimization of the side lobe level at a fixed main beam width and the behavior of the trade-off between side lobe level and main beam width is not illustrated, i.e., in this process, the antenna engineer specifies the desired antenna parameters using a synthesis approach then the optimization algorithm attempts to find the best antenna design for the intended application.

In the context of this chapter instead of aiming to find a single solution, we will try to produce a set of good compromises or "trade-offs" from which the decision maker, i.e, the antenna designer, will select one. A set of trade-off solutions would provide, under a set of criteria given (main beam width and side lobe level), a complete view of the behavior of radiation characteristics for a given design case in order to select which design option is the most appropriate to achieve the design goal. Furthermore, the application of multi-objective methods in the computation of trade-off curves of antenna arrays has been rather scarce. In [11] a multi-objective genetic algorithm denominated NSGA-II is applied to deal with the antenna array design problem. In [11] the efficiency of the NSGA-II in the computation of non-dominated fronts between main beam width and side lobe level is illustrated. However, the work presented in [11] has only considered optimizing an

array to have a radiation pattern fixed in broadside (90°). In applications as smart antenna systems the antenna array design must be optimum to have a steerable not fixed radiation pattern.

This chapter applies the efficient NSGA-II algorithm [12] in the computation of design trade-off curves for non-uniform phased linear antenna arrays. The purpose and contribution of this chapter is to extend a previous result on linear antenna array design to a model of problem that includes design of a linear array to have a steerable radiation pattern. This model considers a driving-point impedance restriction placed on each element in the array. Considering driving-point impedance makes matching antenna element impedances easier and can prevent the occurrence of blind angles during steering [13]. From the optimization point of view this consideration makes the problem more restrictive and therefore more difficult to solve.

The remainder of the chapter is organized as follows. Section II states the antenna array design problem we are dealing with. Then a description of the used algorithm is presented in Section III. Following this description the experimental setup and results are presented in Section IV. Finally, the summary and conclusions of this work along with some future line of research are presented in Section V.

2 Problem Statement

If the elements in the linear array are taken to be isotropic sources, the radiation pattern of this array can be described by its array factor [14]. The array factor for the linear array shown in Figure 1 is given by

$$AF\left(\theta,\theta_0,\mathbf{I},\mathbf{dm}\right)=\sum_{n=1}^{N} I_n \exp\left(jkd_n\left(\cos\theta-\cos\theta_0\right)\right) \tag{1}$$

where $\mathbf{I} = [I_1, I_2, ..., I_N]$, I_i represents the amplitude excitation of the ith element of the array, $\mathbf{dm} = [dm_1, dm_2,..., dm_{N-1}]$, dm_i represents the distance from element i to element i+1, i.e., $d_1 = 0$; $d_2 = dm_1$; $d_3 = d_2 + dm_2$; $d_4 = d_3 + dm_3$; ...; $d_N = d_{N-1} + dm_{N-1}$, $k = 2\pi/\lambda$ is the phase constant and θ is the angle of incidence of a plane wave, λ is the signal wavelength, and θ_0 is the steering direction.

The excitations I_i (i = 1, ..., N) at the input terminals are related to the terminal voltages of the antenna elements by the impedance matrix **Z**:

$$\mathbf{V} = \mathbf{I} \cdot \mathbf{Z} \tag{2}$$

where

$$\mathbf{V} = [V_1, V_2, ..., V_N], \mathbf{I} = [I_1, I_2, ..., I_N]$$

and

$$Z=\begin{bmatrix} Z_{11} & Z_{12} & \cdots & Z_{1N} \\ Z_{21} & Z_{22} & \cdots & Z_{2N} \\ \vdots & & & \vdots \\ Z_{N1} & Z_{N2} & \cdots & Z_{NN} \end{bmatrix}.$$

Through (2), the terminal voltage of any one element can be expressed in terms of the currents flowing in the others:

$$V_n=\sum_{i=1}^{N} Z_{in} I_i \qquad n=1,\ldots,N, \tag{3}$$

where Z_{in} is the mutual impedance between elements i and n; Z_{ii} is the self impedance of element i.

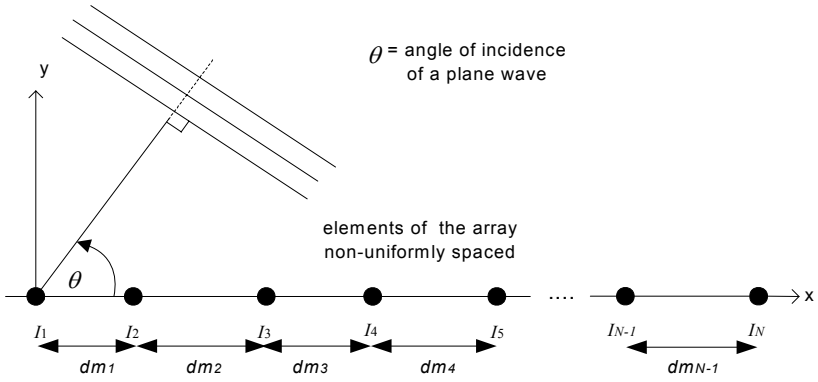


Fig. 1. *Geometry and notations used for non-uniform phased linear array.*

In general, numerical techniques such as method of moments can be used to obtain the mutual impedance matrix **Z**. For dipoles, however, **Z** can be determined using classical induced electromotive force (EMF) method. For the side-by-side configuration and dipole lengths $l = \lambda/2$, an element of the mutual impedance matrix Z_{in}, where $1 \le i, n \le N$, is given by [14]

$$z_{in}(\mathrm{dm}, l) = \begin{cases} 30\left[0.5772 + \ln(2kl) - C_i(2kl)\right] + j\left[30\left(S_i(2kl)\right)\right] & i = n \\ 30\left[2C_i(u_0) - C_i(u_1) - C_i(u_2)\right] - j\left[30\left(2S_i(u_0)\right)\right] & i \neq n \end{cases} \tag{4}$$

where $u_0 = kd_{in}$, $u_1 = k\left(\sqrt{d_{in}^2 + l^2} + l\right)$, $u_2 = k\left(\sqrt{d_{in}^2 + l^2} - l\right)$, and d_{in} is the distance between elements i and n, i.e., $d_{in} = d_n - d_i$ where $d_1 = 0$; $d_2 = dm_1$; $d_3 = d_2 + dm_2$; $d_4 = d_3 + dm_3$; ...; $d_N = d_{N-1} + dm_{N-1}$. $C_i(u)$ and $S_i(u)$ are the cosine and sine integral equations, respectively, and are defined as $C_i(u) = \int_{\infty}^{u} \left(\cos(x)/x\right) dx$ and $S_i(u) = \int_{0}^{\infty} \left(\sin(x)/x\right) dx$.

The driving point impedance of the nth antenna element is given by

$$Z_n^{in} = \frac{V_n}{I_n} \tag{5}$$

where

V_n the terminal voltage

I_n the current flowing in element n.

To design a linear array having a radiation pattern with minimal values of the main beam width and side lobe level during scanning, one must optimize the array when it is steered to the maximum desired scan angle [6], [7], [15] from broadside (90°), see Figure 2.

Because of its geometry and symmetry, a linear array is used to give coverage in a 120° sector. Therefore, to create a linear array to have a steerable radiation pattern with minimal values of the main beam width and side lobe level during scanning in a 120° sector, i.e., in the range [30°, 150°], the antenna array will be optimized when it is steered to 30°.

An interesting open problem is to solve the design problem for other geometries where the latter condition, does not hold.

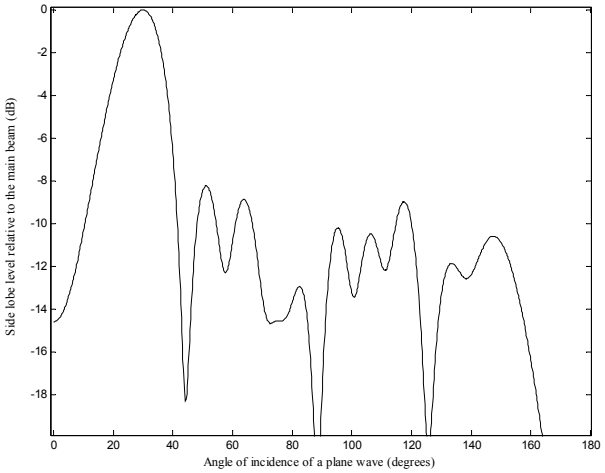


Fig. 2. *Normalized radiation pattern of a linear steered to 30° (N=8, dm_1=0.8326λ, dm_2=1.0544λ, dm_3=0.4937λ, dm_4=1.5031λ, dm_5=0.4561λ, dm_6=0.4467λ, dm_7=0.4432λ; I_1=0.6571, I_2=0.7137, I_3=0.7346, I_4=0.7381, I_5=0.5135, I_6=0.5376, I_7=0.5369, I_8=0.6368).*

We now need to formulate the objective functions we want to minimize. Let us introduce first a couple of definitions [11] for the steerable radiation pattern. Let $m = \{\theta \in \Xi \mid |\mathrm{AF}(\theta, \mathbf{I}, \mathbf{dm})| \geq |\mathrm{AF}(\omega, \mathbf{I}, \mathbf{dm})| \ \forall\ \omega \in \Xi\}$ be the angle where global maximum is attained in Ξ=[0, π], i.e., the maximum desired scan angle to provide a steerable radiation pattern in a 120° sector. Let $q = \{\theta \in \Xi\text{-m} \mid (\partial |\mathrm{AF}| / \partial \theta) = 0\}$ be the set of angles where local maxima excluding the global maximum are attained, and $p = \{\theta \in \Xi \mid |\mathrm{AF}(\theta, \mathbf{I}, \mathbf{dm})| = 0\}$ the set of angles where the radiation pattern crosses zero. Based on these definitions the objective functions can be written as:

$$f_1 = \max_q (|\mathrm{AF}(q, \mathbf{I}, \mathbf{dm})| \ / \ |\mathrm{AF}(m, \mathbf{I}, \mathbf{dm})|), \text{ and } f_2 = \min_p \{|m - p(\mathbf{I}, \mathbf{dm})|\}.$$

Thus the optimization task is then in the first component: minimization of the maximum side lobe (f_1), in the second component it is the minimization of the null-to-null beam width (f_2). The problem can be defined as:

$$\textit{Minimize } (f_1, f_2)$$

$$\textit{subject to } \mathbf{dm} \in \mathbf{D}, \mathbf{I} \in \mathbf{\Lambda}, \mathrm{Re}\{\mathbf{Z}^{\mathrm{in}}\} \in \mathbf{\Gamma},$$

where $\mathbf{D} = (0, 2\lambda]^{N-1}$ is imposed to take the physical size of the antenna array into account; $\mathbf{\Lambda} = [W_{\min}, W_{\max}]^N$ is the range of the weight coefficients imposed for practical implementation of the attenuators; $\mathbf{Z}^{\mathrm{in}} = [Z^{\mathrm{in}}_1, Z^{\mathrm{in}}_2, ..., Z^{\mathrm{in}}_N]$, Z^{in}_n represents the driving-point impedance of the nth element of the array and $\mathbf{\Gamma} = [20\Omega, 250\Omega]^N$ will result in reasonable component values for matching networks that do not tend toward the extremes of very large or very small [6].

The next section presents the method we use to obtain the trade-off curve between the main beam and side lobe level.

3 The Proposed Algorithm

We are interested in the trade-off curve computation for non-uniform phased linear antenna arrays. For this purpose we propose to use the Non-dominated Sorting Genetic Algorithm-II proposed by Deb et al. [12]. The genetic algorithms are specially well suited for multi-objective problems since they are designed to handle a multi-set of solutions in a single iteration [16], [17]. We chose this algorithm for its easiness of implementation and its efficient computation of non-dominated ranks. The procedure for NSGA-II (Fig. 3) is described as follows.

The function **Generate Initial Population** randomly and uniformly generates a set of individuals.

The main idea in **Classify Individuals** is to rank the individuals according to their dominance relation, i.e., the set of non-dominated individuals are said to be in front 0. After removing them the remaining non-dominated solutions are in front 1. The procedure continues until all individuals are assigned to a front.

In order for this explanation to be self-contained we just need to explain what a nondominated individual is. Let a^1, $a^2 \in \Re^n$ be two n-dimensional vectors. We say that a^1 dominates a^2 if and only if $a_i^1 \leq a_i^2\ \forall\ i$ and $a_i^1 < a_i^2$ for at least one i. Given

a set of vectors A, if there is not a single vector in A that dominates the vector $a^i \in A$ then we say that a^i is a nondominated vector of A (for details see [18], page 147-148). In our case $a^i = [f_1(i)\ f_2(i)]^T$ is the vector of objective functions for individual i.

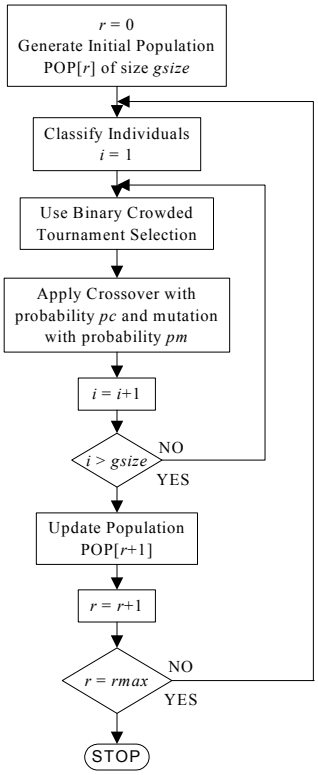


Fig. 3. *Flow chart for the evolutionary multi-objective optimization algorithm (NSGA-II)*

In the **Binary Crowded Tournament Selection** two individuals are randomly selected and the winner is the one with the highest rank. If both individuals have the same rank, then the individual with the highest local crowding distance [17, pp. 235] wins the tournament. This distance measure gives us an idea of how crowded (in the criterion space) is the volume around the given individual.

The function **Update Population** assigns ranks to individuals in the population generated by the union of parents and children. The procedure starts copying individuals into the new population considering first those belonging to the lowest index front as long as the number of individuals in the front does not overflow the population size (*gsize*). In the last front to be copied, individuals are sorted according to their crowding distance, eliminating those individuals with smaller crowing distance until the total number of individuals (*gsize*) is completed. Deb [17, pp. 233-241] explains the procedures involved at each step of this algorithm in detail. The individual representation as well as the crossover and mutation operators are explained in the following subsections.

3.1 Individual Representation and Decoding

Each individual is in general represented by two vectors of real numbers. One vector of real numbers restricted to be on the range $[W_{min}, W_{max}]$, i.e. $\mathbf{I} = [I_1, I_2, ..., I_N]$, where I_i is the amplitude excitation of element i, and another one restrained on the range $(0, 2\lambda]$, i.e. $\mathbf{dm} = [dm_1, dm_2, ..., dm_{N-1}]$, where dm_i is the distance from element i to element $i+1$.

3.2 Genetic Operators

The used genetic operators are standard, the well known two point crossover [19] along with a single mutation where a locus is randomly selected and the allele is replaced by a random number uniformly distributed in the feasible region.

The results of using this algorithm for design of non-uniform phased linear antenna arrays are described in the next section.

4 Experimental Setup and Results

The method described in the previous section was implemented to determine the Pareto front of the main beam width and the side lobe level for non-uniform phased linear arrays. Several experiments were carried out with different number of antenna elements (N = 6, 8, 10, 12). In these experiments the driving-point impedance of each element was calculated using the induced electromotive force (EMF) method outlined in [14] for arrays of side-by-side half-wave dipoles. In the experiments the algorithm parameters, after a trial and error procedure, were set as follows: maximum number of generations *rmax* = 500, population size *gsize* = 200, crossover probability *pc* = 1.0 and mutation probability *pm* = 0.1. The obtained results are explained below.

Figure 4 illustrates the behavior of the trade-off curve between main beam width and side lobe level for different numbers of antenna elements obtained by the NSGA-II algorithm. The values for the main beam width are average values considering the direction of the main beam (θ_0) swept in the range [30°, 150°]. The side lobe level is the maximum side lobe level over entire scan range of the array.

The results shown in Figure 4 indicate that the design trade-off improves as the number of antenna elements is increased. The behavior of the trade-off curve for non-uniform phased linear arrays illustrates that a steerable radiation pattern with small values of side lobe level can only be obtained with a considerable increase in the main beam width.

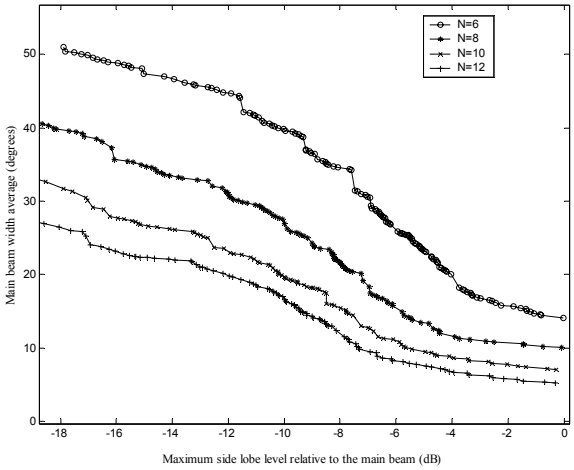


Fig. 4. *Computation of the trade-off curve between main beam width and side lobe level for non-uniform phased linear arrays obtained by the NSGA-II algorithm with different numbers of antenna elements.*

Table 1 shows examples of the element distribution and the resulting excitation distribution. In this case, the weightings for the array elements, I_1, I_2, …, I_N, are normalized using $\max(I_i) \leq 1$. Maximum and minimum values of the spatial aperture found by the NSGA-II are given. This table illustrates that the largest spatial aperture found by the NSGA-II is for small values of side lobe level considering different numbers of antenna elements.

Figure 5 shows a comparison between the trade-off curves for a non-uniform linear array to have a radiation pattern fixed at broadside [11] and the non-uniform phased linear array to have a steerable radiation pattern (in a 120° sector) obtained by NSGA-II, considering the same number of elements in both cases (N=8). As it can be observed, the trade-off curve for the phased linear array presents higher values of main beam width and side lobe level than the trade-off curve for a linear array whose radiation pattern is fixed at broadside. This is because the radiation characteristics (main beam width and side lobe level) of a linear array degrade as the array is optimized to operate in a range further from broadside [6], [20].

Table 1. *Examples of element distributions and the resulting excitation distribution for non-uniform phased linear antenna arrays obtained by the NSGA-II algorithm for different numbers of antenna elements.*

N	SLL_{max}(dB)	BW_{ave}(deg)	dm_1, dm_2, dm_3, dm_4 … dm_{N-1}; I_1, I_2, I_3, I_4 … I_N	Aperture
6	-17.99	51.15	0.4638λ, 0.3843λ, 0.4174λ, 0.4467λ, 0.3793λ; 0.1940, 0.4665, 0.6457, 0.6605, 0.5280, 0.4161	Min: 2.09λ
	-13.92	45.54	0.4055λ, 0.4219λ, 0.4376λ, 0.4467λ, 0.4668λ; 0.5683, 0.5874, 0.7335, 0.6205, 0.5939, 0.5866	2.17λ
	-8.79	38.13	0.4638λ, 0.4219λ, 0.9999λ, 0.4467λ, 1.4628λ; 0.8926, 0.8915, 0.7335, 0.6677, 0.7321, 0.7135	3.79λ
	-0.025	14.05	1.9514λ, 1.9726λ, 1.9924λ, 1.9881λ, 1.8928λ; 0.9977, 0.1653, 0.1666, 0.1854, 0.3711, 1.0000	Max: 9.79λ
8	-18.23	40.89	0.4549λ, 0.4095λ, 0.4937λ, 0.3740λ, 0.4561λ, 0.3974λ, 0.3818λ; 0.2490, 0.7137, 0.7346, 0.7381, 0.5135, 0.5376, 0.5369, 0.6368	Min: 2.96λ
	-12.74	33.69	0.4317λ, 0.5234λ, 0.4197λ, 0.5330λ, 0.4007λ, 0.4720λ, 0.4564λ; 0.8058, 0.5381, 0.7260, 0.6369, 0.6739, 0.5724, 0.7558, 0.3872	3.23λ
	-8.713	26.59	0.8326λ, 1.0544λ, 0.4937λ, 1.5031λ, 0.4561λ, 0.4467λ, 0.4432λ; 0.6571, 0.7137, 0.7346, 0.7381, 0.5135, 0.5376, 0.5369, 0.6368	5.22λ
	0.000	10.02	1.9836λ, 1.9805λ, 1.9498λ, 1.9408λ, 1.9997λ, 1.9732λ, 1.9612λ; 0.9891, 0.4240, 0.1568, 0.2016, 0.1702, 0.1962, 0.1798, 0.9930	Max: 13.7λ
10	-18.51	32.43	0.4487λ, 0.4676λ, 0.4710λ, 0.5040λ, 0.4576λ, 0.4272λ, 0.4652λ, 0.5193λ, 0.3260λ; 0.4607, 0.6504, 0.6793, 0.7819, 0.8359, 0.7551, 0.8192, 0.4211, 0.4129, 0.1901	Min: 4.08λ
	-14.86	27.54	0.4487λ, 0.4669λ, 0.4710λ, 0.5040λ, 0.4576λ, 0.4272λ, 0.4652λ, 0.5193λ, 0.3787λ; 0.8010, 0.6504, 0.6793, 0.7819, 0.9000, 0.7243, 0.7437, 0.8760, 0.5466, 0.6676	4.13λ
	-11.80	21.97	0.4663λ, 0.3653λ, 0.4710λ, 0.5040λ, 0.4576λ, 0.4272λ, 0.9258λ, 1.0282λ, 0.4531λ; 0.5686, 0.7491, 0.6793, 0.6558, 0.8359, 0.7243, 0.8192, 0.8760, 0.5466, 0.8805	5.09λ
	-0.005	7.270	1.9572λ, 1.9682λ, 1.9956λ, 1.9751λ, 1.9804λ, 1.9483λ, 1.9829λ, 1.9647λ, 1.9518λ; 0.9941, 0.1398, 0.0696, 0.0039, 0.0710, 0.0350, 0.0320, 0.0231, 0.0315, 0.9086	Max: 17.7λ

Table 1. *(Cont.)*

12	-9.935	17.54	0.4139λ, 1.4000λ, 1.3664λ, 0.6039λ, 0.9207λ, 0.3941λ, 0.4824λ, 0.3459λ, 0.9191λ, 1.4835λ, 0.9423λ; 0.5837, 0.6785, 0.5220, 0.7030, 0.4791, 0.5036, 0.7017, 0.5080, 0.7685, 0.6523, 0.8129, 0.7576	Min: 9.27λ
	-6.695	10.05	0.4139λ, 1.7561λ, 1.9924λ, 1.9744λ, 1.9274λ, 1.9900λ, 1.9772λ, 1.9503λ, 0.9191λ, 1.4835λ, 0.9423λ; 0.7766, 0.7954, 0.1703, 0.1812, 0.5274, 0.4525, 0.4239, 0.5254, 0.1787, 0.9968, 0.6346, 0.9601	17.32λ
	-4.403	7.478	0.9514λ, 1.9878λ, 1.9924λ, 1.9744λ, 1.9274λ, 1.9900λ, 1.9772λ, 1.9805λ, 1.9595λ, 1.9011λ, 1.5191λ; 0.9621, 0.6339, 0.1703, 0.1539, 0.2280, 0.1697, 0.1610, 0.3070, 0.1787, 0.1501, 0.8129, 0.9925	20.16λ
	-0.980	6.102	1.8827λ, 1.9878λ, 1.9924λ, 1.9744λ, 1.9274λ, 1.9900λ, 1.9772λ, 1.9503λ, 1.9595λ, 1.9768λ, 1.5191λ; 0.9621, 0.7250, 0.1703, 0.1539, 0.1610, 0.1697, 0.1610, 0.1766, 0.1675, 0.1501, 0.9780, 0.9925	Max: 21.1λ

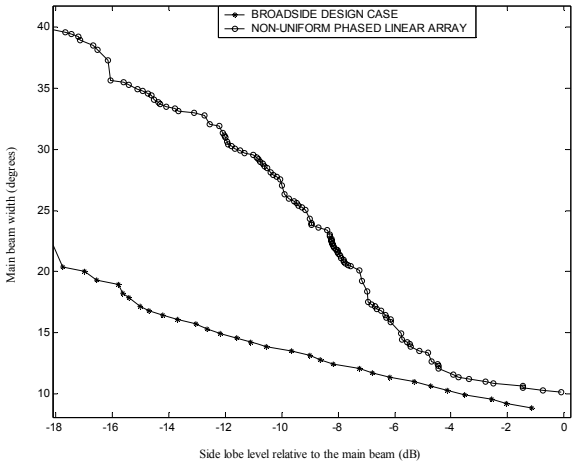


Fig. 5. *Comparison between the trade-off curves for a non-uniform linear array to have a radiation pattern fixed at broadside [16] and the non-uniform phased linear array to have a steerable radiation pattern (in a 120° sector) obtained by NSGA-II.*

The effectiveness of the NSGA-II for the design of non-uniform phased linear arrays is shown by means of experimental results with a set of design options that provide a steerable radiation pattern (in a 120° sector) with radiation characteristics (main beam width and side lobe level) physically attainable.

Finally, Table 2 presents the results for a diversity measure denominated *spacing* (*S*) (see [17], pages 313-314). This table illustrates good *spacing* values for experimental results shown in Figure 4. These results show that the diversity measures improve as the number of antenna element increases.

Table 2. *Computation of Spacing for the nondominated fronts by the NSGA-II method*

Number of antenna elements (N)	Spacing (S)
6	0.3581
8	0.2779
10	0.2454
12	0.2137

From the results shown previously, it is illustrated the application of NSGA-II to the design of non-uniform phased linear arrays. This genetic algorithm efficiently computes an approximation to the set of Pareto-optimal solutions for non-uniform phased linear arrays. The decision maker, i.e., the antenna array designer, will select a simple solution in accordance with the design goal. In this case, the criteria specified will help to select the most appropriate design option. This is in order to meet potentially a reduction of the antenna system cost.

Furthermore, the design trade-off for a non-uniform phased linear array presents higher values of main beam width and side lobe level than the trade-off for a non-uniform linear array to have a radiation pattern fixed in broadside. Antenna arrays designers should take this information into account in order to select design options that help to achieve networks that maximize capacity and improve quality and coverage.

In this chapter only linear arrays are dealt with. However, there are other antenna array structures that could provide a wider coverage than the linear geometry such as circular and planar geometries. These geometries could provide coverage in the whole x-y plane (360°). The NSGA-II method can be applied to the design of these antenna array structures or other geometries proposed in order to meet the design configuration most appropriate for mobile and wireless communications systems.

5 Conclusions

This chapter illustrates how to model the design of non-uniform phased linear arrays as a multi-objective optimization problem. This model is considered realistic due to the driving-point impedance restriction placed on each element in the array. This restriction makes matching antenna element impedances easier and prevents the occurrence of blind angles during steering. The well-known NSGA-II algorithm is proposed as the solution for this problem. The NSGA-II algorithm efficiently computes the design trade-off curves between main beam width and side lobe level for non-uniform phased linear arrays. These trade-off curves could allow antenna array designers to decide which design option is the most appropriate in accordance with the design goal of mobile and wireless communications system. This is in order to meet potentially a reduction of the antenna system cost and the control complexity.

Future research will be aimed at dealing with other geometries and constraints. Many different areas of antenna design and analysis require a feasible and versatile procedure, being able to perform array synthesis by tuning antenna characteristics and parameters. Because of the versatility of the NSGA-II it seems a good candidate to face this problem.

Acknowledgements

This work was supported by Mexican National Science and Technology Council, CONACyT, under grant J50839-Y.

References

1. J. C. Liberti and T. S. Rappaport, Smart Antennas for Wireless Communications: IS-95 and Third Generation CDMA Applications, Prentice Hall, New Jersey, 1999.
2. V. Murino, A. Trucco, and C. S. Regazzoni, "Synthesis of unequally spaced arrays by simulated annealing", IEEE Transactions on Signal Processing, Vol. 44, No.1, pp. 119–123, 1996.
3. Trucco and V. Murino, "Stochastic optimization of linear sparse arrays", IEEE J. Ocean. Eng., Vol. 24, pp. 291-299, 1999.
4. J. H. Bae, K. T. Kim, J. H. Lee, H. T. Kim and J. I. Choi, "Design of steerable non-uniform linear array geometry for sidelobe reduction", Microwave and Optical Technology Letters, Vol. 36, No. 5, pp. 363-367, 2003.
5. J. H. Bae, K. T. Kim, C. S. Pyo and J. S. Chae, "Design of scannable non-uniform planar array structure for maximum side-lobe reduction", ETRI Journal, Vol. 26, No. 1, pp. 53-56, 2004.
6. [6] M. G. Bray, D. H. Werner, D. W. Boeringer and D. W. Machuga, "Optimization of thinned aperiodic linear phased arrays using genetic algorithms to reduce grating lobes during scanning", IEEE Transactions on Antennas and Propagation, Vol. 50, pp. 1732–1742, 2002.

7. R. Haupt, "Thinned arrays using genetic algorithms", IEEE Transactions on Antennas and Propagation, Vol. 42, No. 7, pp. 993-999, 1994.
8. P. K. Varlamos and C. N. Capsalis, "Electronic Beam Steering Using Switched Parasitic Smart Antenna Arrays", Progress in Electromagnetics Research PIER, Vol. 36, pp. 101-119, 2002.
9. P. K. Varlamos and C. N. Capsalis, "Design of a Six-sector Switched Parasitic Planar Array Using the Method of Genetic Algorithms" Wireless Personal Communications Kluwer, Vol. 26, No. 1, pp. 77-88, 2003.
10. P. K. Varlamos and C. N. Capsalis, "Direction of Arrival Estimation (DoA) Using Switched Planar Arrays and the Method of Genetic Algorithms", Wireless Personal Communications Kluwer, Vol. 28, No. 1, pp. 59-75, 2004.
11. M. A. Panduro, D. H. Covarrubias, C. A. Brizuela and F. R. Marante, "A Multi-objective Approach in the Linear Antenna Array Design", AEU International Journal of Electronics and Communications, Vol. 59, No. 6, 2005.
12. K. Deb, S. Agrawal, A. Pratap and T. Meyarivan, "A fast elitist non-dominated sorting algorithm for multi-objective optimization: NSGA-II", Parallel Problem Solving from Nature – PPSN VI, Springer, Berlin, pp. 849-858, 2000.
13. R. C. Hansen, Phased Array Antennas, Wiley, New York Chichester Weinheim Brisbane Singapore Toronto, 1998.
14. Balanis, Antenna Theory-Analysis and Design, 2nd Ed., New York: Wiley 1997.
15. K. Chang, X. Ma and H. B. Sequeira, "Minimax-maxmini: a new approach to optimization of the thinned antenna arrays", Proc. IEEE Antennas and Propagation Society Int. Symp Seattle, WA, pp. 514-517, 1994.
16. Coello, D. A. van Veldhuizen, G. B. Lamont, Evolutionary algorithms for solving multi-objective problems, Kluwer Academic Publishers, Boston Dordecht London.
17. K. Deb, Multi-Objective Optimization using Evolutionary Algorithms, John Wiley & Sons, Chichester New York Weinheim Brisbane Singapore Toronto, 2001.
18. R. Steuer, Multiple criteria optimization: theory, computation and application, John Wiley & Sons, 1986.
19. E. Golberg, Genetic algorithms in search, optimization, and machine learning, Addison-Wesley, Massachusetts, 1989.
20. Y. T. Lo and S. W. Lee, Antenna Handbook: Theory, Applications, and Design, New York: Van Nostrand Reinhold, 1988.

PART IV

PHOTONIC TECHNOLOGY

Chapter 14

Specialty Optical Fibres in Laser and Sensing Applications

Romeo Selvas-Aguilar1, Eduardo Pérez-Tijerina1, Ismael Torres-Gomez2, and Julian Estudillo-Ayala3

1 FCFM- Universidad Autonoma de Nuevo León, San Nicolas de los Garza, N.L., 66450, México
2 Centro de Investigaciones en Óptica, Lomas del Bosque 115, Col. Lomas del Campestre, León, Gto., 37150, México,
3 FIMEE-Universidad de Guanajuato, Salamanca, Gto., 36150, México, rselvas@fcfm.uanl.mx

Abstract

This chapter gives an overview of the investigation carried out on specialty optical fibre during the last 40 years and what has been done in Mexico. Rare-earth-doped optical fibres and photonic crystal fibres are well described in this article as those kinds of fibres have important applications in laser and sensor systems.

Keyword: Lasers, optical fibres, optical fibre sensors, photonic crystal fibres.

1 Introduction

The low propagation loss of optical fibre is an outstanding feature that has made them an unrivalled transmission medium in modern telecommunication systems [1]. An even earlier development was the use of rare earth doped fibre in fibre lasers. Since then, fibre lasers and their relations, fibre amplifiers, of a wide variety of formats with a wide variety of properties have found applications in many different areas. In particular, the invention of the optical fibre amplifier revolutionised optical communications, thanks to its superb amplification characteristics.

The development of the fibre optics back to the 60´s when the researcher Charles Kao [2] who working with ITT company (currently cited as BT –British Telecomm) proposes the construction of a two concentric cylindrical waveguides made by silica. The first results obviously revealed a high optical insertion loss but he predicted that a fibre optic with low transmission losses is still possible and thus a practical device shall open a vast of applications in the telecom industry, parallel, another important device was also discovered at that time and it consisted in the the first laser which was proposed by the scientific Ted Maiman [3]. This device also help in the progress of the telecomm. So far, in the 80´s and part of the 90´s many companies were involved in this frenetic career. The first company to fabricate an optical fibre with industrial features was the American Company called Corning with a working team leaded by Robert Maureer. Figure 1 show the scheme of an optical fibre and there can be noted the distribution of the refractive index at both the core and in the cladding of an optical fibre.

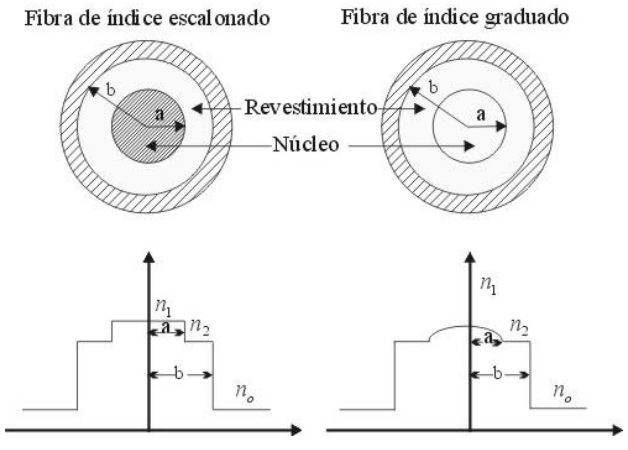


Fig. 1. *Optical Fibres (Refractive index distribution).*
(a: step index fibre , b: graded index fibre)

Simplicity is a major attraction of active fibre devices. By simply incorporating a dopant into the core of an optical fibre, a powerful gain-medium is realised. The fibre is practically always made from glass, often a high silica glass. The first rare earth doped fibre devices go also back to the 1960´s when C Koester and E. Snitzer developed a flash lamp pumped neodymium doped fibre amplifier and they published this experiment in the Journal of Applied Physics with the titled "proposed fibre cavities for optical masers" in 1961 [4]. The potential of rare earth doped fibres for practical devices with unique properties became clear in 1985, with the first demonstration of low loss single mode rare earth doped fibre, fabricated via the widely available modified chemical vapour deposition method. The characteristics of the fibre itself were first reported by S Poole, D Payne [5] and M Fermann, at the University of Southampton. In a subsequence paper S Poole et al. demonstrated a single mode fibre laser. Soon afterwards, Snitzer et al. demonstrate the first diode pumped fibre laser. Then other rare earth doped fibres were also proved in this fascinating career [7-10]. The Erbium doped fibre amplifiers was therefore demonstrated by the scientific David N Payne, Robert Mears and Laurence Reekie [6], and parallel the French Scientific Emmanuel Desurvire carried out some work in EDFAs. Desurvire finally written the widely well-know book titled "Erbium doped Fibre Lasers and Amplifiers" [10] which is one of the best ever made scientific book on this topic.

However a breakthrough idea, referred to as cladding-pumping (using double-clad fibres), patented by James Kafka at Spectra Physics, changed the perspective. A double-clad fibre, indeed, enables a good match with the output beam from broad-stripe diodes. These are multi-mode devices that can generate much higher powers than single-mode ones can. Thus, with cladding-pumping, a double-clad fibre laser can produce much higher output power, in a beam that nevertheless can be diffraction-limited. More advanced, higher power multi-mode diode pump sources such as multi-diode arrangements, diode bars and stacks provide even higher powers than single-emitter broad-stripe diodes, and can also be used for fibre pumping (if necessary together with some sort of beam-shaping). This opened up a vast opportunity to scale output power, reaching 110 W and beyond [11-16], in devices that combined, at the same time, high efficiency and excellent spatial beam quality.

Cladding-pumping has revolutionised fibre lasers over the last decade. Recently, a Japanese research consortium involving HOYA, Hamamatsu, and the Electrocommunications University (Ken-Ichi Ueda) demonstrated a 1 kW *cw* fibre embedded laser in a multi-mode design, pumped by multiple diode sources [17]. Three fibre lasers were cascaded to reach 1 kW. In addition, in the pulsed regime, cladding-pumped fibre lasers also offer excellent characteristics. For example, multi-kW peak-powers in milli-joule level pulses can be reached in fibre lasers with small footprints.

Cladding-pumped fibre lasers are currently considered for many applications, where high laser powers are required. Beside fibre lasers, there are also other options for high-power lasers, with so-called bulk (not waveguiding) solid-state lasers being the prime ones. They are typically made with a crystal host, and like optical fibre lasers, can be doped with impurities such as Nd^{3+}, Er^{3+}, Yb^{3+}, and others. In addition, when pumped by an appropriate light source they can generate even higher output powers than fibre lasers. While the diode pump sources needed for fibre lasers are more expensive than the lamp pump source traditionally used for high-power bulk lasers, diode pumped solid-state lasers (including fibre lasers) provide many performance advantages and are becoming increasingly attractive as the cost of pump diodes decreases.

Currently, there are two main companies dedicated to the sale of high power fibre laser, and these are IPG (German) [18] and Southampton Photonics (England). For example IPG claims to be the first in sale a fibre laser with a 6KW power [19] to the industry (automobile German company) although the beam quality of this laser was multi-mode. SPI, on the other hand, claims to be the first to demonstrate a pure single mode 1KW fibre laser and that fact was done in Feb 2003. Since then, both companies and others were demonstrating interesting approached for a KW fibre laser [20], and it is showed in Figure 2.

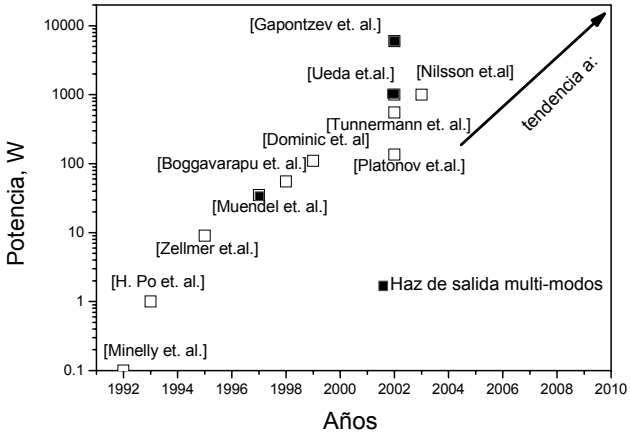


Fig. 2. *Reported power for a optical fiber laser (mainly single-mode output and some with multi-moded output beams).*

As it can been seen in the review, fibre lasers are inside of the multi-million industry of lasers. Trans-national company like the American, Spectral Physics (2004) added to its web pages new products consisting in high power fibre lasers.

2 Optical Fibre Lasers

Historically, diode-pumped fibre lasers were initially simple structures with a single core for guiding both the signal and the pump light. In the early days, a principal attraction of fibre lasers (and amplifiers) was the very low threshold and the high gain efficiency. This is made possible, in particular, by a single-mode core that provides tight beam confinement over a sufficient length to absorb the pump. A single-mode core implies the use of single mode pump sources. The limited power of single-mode diodes has then limited the output powers of core-pumped fibre lasers to ~1 W, or even less.

Consequently, the cladding-pumping scheme has been developed as a method to overcome this limitation. Cladding-pumped fibre lasers do not require single-mode pump sources, but can still produce a single-mode laser output. Cladding-pumped fibre lasers use double-clad fibres (or possibly some more advanced structure) that can guide light in the inner cladding. A double-clad fibre has a primary waveguide (the core) for guiding the signal. It is typically doped with a rare earth. The core is surrounded by a lower-index inner cladding. Both of these are normally made from glass. The inner cladding also forms the "core" for a large, highly multi-moded, secondary waveguide that can guide pump light. The core (primary waveguide) is located within the inner cladding and forms a part of the pump waveguide, so pump light propagating in the pump waveguide reaches the core and excites the laser-active rare-earth ions. The inner cladding is surrounded by an outer cladding of lower refractive index to facilitate wave-guiding. The large inner cladding allows the use of high-power diode bars and arrays providing up to many hundreds of watts of light, as pump sources. At such power levels, a small core is no longer an advantage. Instead, it is a serious obstacle to power-scaling and, besides the double-clad fibre structure a large core is often the most important design feature of a high-power fibre laser.

The researches carried out in Mexico in this topic go back to the 2003 and there are two main institutions dedicated to the research of high power fibre lasers. The National Institute for Astrophysics Optics and Electronics (INAOE) and the Optical Research Centre (CIO) which also were the first to report experiments to the community [21-24]. In 2005, the UANL-FCFM added efforts in the construction of new scheme for fibre lasers. CIO has generated some work in YDF, while for INAOE produced much work in EDF. A recent demonstration of 5W in cw regime as it is showed in Fig. 3 was the first important result in Mexico for a high power fibre laser. It was used a virtual square-inner-clad ytterbium-doped fibre laser as a gain medium and simple mirrors cavity as resonator.

Moreover, there are some reports in tuneable fibre laser, and recently some of these works showed photonic crystal fibres in the setup. In general, there are two important techniques demonstrated so far and both shows tuneable range up to 10nm and output power as high as 1 W. The UANL as it was mentioned before is mainly dedicated to the construction of new schemes for optical devices in order to increase the tuneability as well as the pumping management capabilities of laser setups.

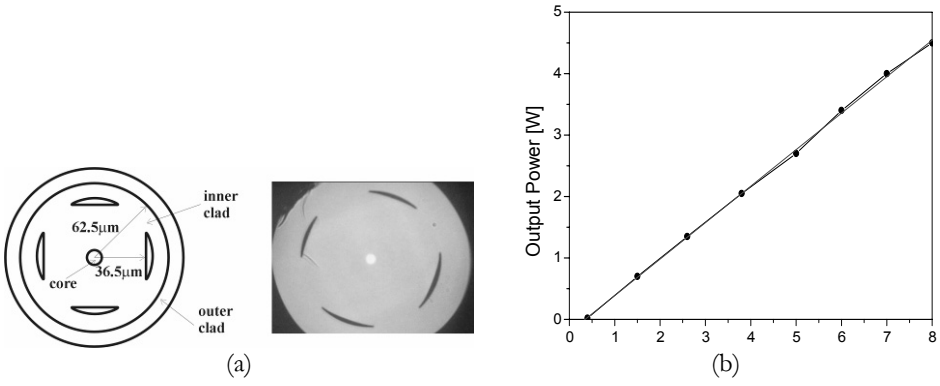


Fig. 3. *Reported power by the CIO research group, (a) design of the specialty optical fibre, (b) optical output power vs absorbed pump power.*

The interesting fact is that all together are working in building up fibre lasers with unique characteristics for example to build up cost effect devices with very simple setups and expanding other emission ranges with other rare earth doped fibres such as Nd (Neodymium), Tm(Thulium) and Er-Yb fibres.

3 Specialty Optical Fibres in Sensing Applications

Fibre optic sensors are often loosely grouped into two basic classes referred to as extrinsic or hybrid fibre optic sensors and intrinsic or all fibre sensor. Many of the intrinsic and extrinsic sensors may be multiplexed, offering the possibility of large numbers of sensors supported by a single fibre optic line. The most commonly employed techniques are time, frequency, wavelength, coherence, polarization and spatial multiplexing. The developed of fibre sensors go also backs to the 70´s as soon as the commercial optical fibres appear in the market. Consequently, the research carried out in optical fibre sensors has impacted the modern world.

One of the areas of greatest interest has been the development of high performance interferometric fibre optic sensors. Substantial efforts have been undertaken on Sagnac interferometers, ring resonator, Mach-Zehnder and Michelson interferometers, as well as dual-mode polarimetric, grating and etalon based interferometers[25]. The principle of operation is based in the recombination of two split beam paths and the interferometric fringes generated brings the information required to know.

There is also another group of fibre sensors and these are the one based on Fibre Bragg grating as a sensing element [26]. Fibre grating sensors can be configured to have gauge lengths from 1mm to approximately 1cm, with sensitivity comparable to conventional strain gauges. Substantial efforts are being also made by laboratories around the world to improve the manufacturing of fibre grating because they have the potential to be used to support optical communications as well as sensing technology.

Once the fibre grating has been fabricated, the next major issue is how to extract information. When used as a strain sensor, the fibre grating is typically attached to or embedded in, a structure, As the fibre grating is expanded or compressed, the grating period expands or contracts, changing then the grating spectral response. Obviously one can expand even more in this review but as we have only interests in others types of fibre sensors we are not included more information on it.

In 1990, a new class of optical fibre based on the properties of photonic crystal was invented as it was referred as photonic crystal fibre. These fibres have finding applications in fibre optics communications, nonlinear devices, fibre lasers, high power transmission, highly sensitive gas sensors, and other areas.

In general, such fibres have a cross section micro-structured from two or more materials, most commonly arranged periodically over mucho of the cross section, usually as a cladding surrounding a core where light is confined. These are constructed by the same general principles as other optical fibres: first one constructs a perform on the scale of centimetres in size, and then heats the perform and draws it down to a much smaller diameter. PCF may be divided into two categories, high index guiding fibres and low index guiding fibres. Similar to conventional fibres, high index guiding fibres are guiding light in a solid core by the modified total internal reflection principle. Low index guiding fibre guide light by the photonic bandgap effect. The light is confined to the lo index core as the PBG effect makes propagation in the micro-structure cladding region impossible. The strong wavelength dependency of the effective refractive index and the inherently large design flexibility of the PCFs allow for a whole new range of novel properties. Such properties include endlessly single mode fibres, extremely nonlinear fibres and fibres with anomalous dispersion in the visible wavelength region [27-30].

The material surrounding the core is typically formed in a hexagonal arrangement, [31] such as is showed in Fig. 4, where *d* is for the hole-diameter and Δ is for the period of the net. Another class of micro-structure fibres is the the jacketed-air-clad fibre which was also developed [32]. Surrounding the inner cladding, this fibre has an outer cladding, which is primarily of air. The outer cladding is in turn surrounded by a silica jacket, which protects the wave-guiding region of the fibre and gives the fibre strength. A micro-structured mesh in the outer cladding connects the inner cladding to the jacket. This allows for high NA pump waveguides with small cross-sectional areas.

In Mexico, there is a place where is possible to fabricated this kind of fibres, and it is in Leon Gto. (CIO), which are the only facilities in Mexico with a pulling optical fibre machine.

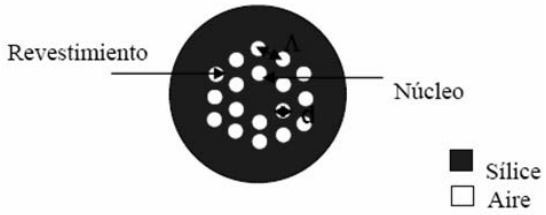


Fig. 4*. Transversal view of a photonic crystal fibre with high core index*

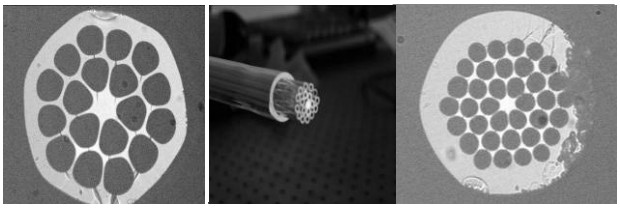


Fig. 5. *From left to right, (a) Photonic crystal fibre with large holes, (b) fabrication process of a PCF, (c) Photonic Crystal fibre with small core.*

As a sample of this, we put in Figure 5 some of the fibre fabricated in Mexico.

In 2002, the research group has been explored important applications of this type of fibre, and we reported a wide number of works to the community and support with this the new research lines of researching of this group.

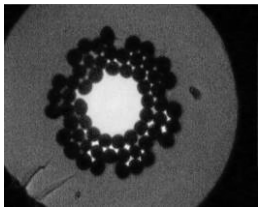


Fig. 6. *Rare-earth-doped Photonic Crystal Fibre and its potential uses as a sensor.*

Some of these research works have been published and consequently demonstrated applications in amplification and laser generation [32-33]. By 2005, we start with applications in sensing [34-35]. Recently the nano-science group of the UANL has joined to the research line of sensing applications with the fabrication, synthesis and management of nano magnetic particles in order to construct gas and micrometric sensor.

By using the characteristic that an ytterbium doped fibre and the mesh of a photonic crystal fibre (Fig. 6) is possible to deposit nano particles of magnetic spheres inside of the 10 microns holes surrounded the core of this specialty fibre and therefore manipulate the generation of the modes of the fibre itself when the superfluorescent of stimulated emission is created.

This work is still undergo and has the objective to group a team of researcher with multidisciplinary ideas with a unique goal. The possibility to build up a bio-photonic sensor is then viable. On the other hand, other kind of sensors developed by the authors are the voltage sensors presented in one of the important conference on fibre optics, in which a piezo-actuator and the optical characteristics of a fibre are joined in one simply voltage sensor [36].

4 Conclusions

Light has been played an important role in our everyday life style and the invention of laser systems has invaded our quality of life with many interesting applications.

In this chapter, we showed subsequent events in terms of scientific and technological on optical fibres. We began with the fabrications of the first optical fibre laser. Ytterbium doped fibres were the focusing of this work and it was review the experimental work of the last five years in Mexico, given more emphasises in the research work carried out by the authors. Notably, the advances in high power fibre lasers show the transcendental of this topic and it is prognostic that fibre lasers shall substitute the solid-state-based laser in the competitive industry within the multi-million industry market as the fibre lasers are more simple and low cost effective.

Finally, we explained the characteristics of the photonic crystal fibres and their uses as a lasers as well as sensing device. More than 10 years of continuo research on this fibre was also described and notably the potential uses as a fibre sensors and its new approaches that the authors have recently demonstrated. PCF are therefore potential sensors for gas, intense magnetic fields, biomaterials sensing, etc.

Acknowledgement

The authors thank PAICYT-UANL for the financial support and also appreciated for the useful discussion with Alejandro Martinez-Rios (CIO), and Rene Dominguez Cruz (UAT).

References

1. Kapany, N.S., Philips, B.G.: Fiber Optics in low light level systems, Low Light level imaging systems, Proc. of two-day seminar, (Mar 1970) 126-35.
2. Encyclopedia4U, Encyclopedia 4U.com, Charles K Kao bio, WWW http://www.encyclopedia4u.com/c/charles-k-kao.html, 2000 y también en: Por Jim Ericksom and Julanda Chung, Asian of the Century web page, WWW http://www.asiaweek.com/asiaweek/features/aoc/aoc.kao.html, (1999).
3. Friedman, Gregg: OE-reports, Technology and Trends for the International Optical Engineering Community, Inventing the light fantastic: Ted Maiman and the world´s first laser, [WWW] http://www.spie.org/web/oer/august/aug00/maiman.html, (2000).
4. Snitzer, E.: Proposed Fiber Cavities for Optical Masers, Journal of Applied Physics, Vol. 32(1), (1961) 36-39.
5. Poole, S.B., Payne, D.N., and Fermann, M.E.; Fabrication of low loss optical fibers containing rare earth ions, Electron. Lett. Vol. 21, (1985) 737-738.
6. Mears, R.J., Reekie, L., Jauncey, I.M., and Payne, D.N.: Low-noise erbium-doped fibre amplifier operating at 1.54 um, Electron. Lett. Vol. 23, (1987) 1026-1028.
7. Yamamoto, T., Miyajima, Y., and Komukai, T.: 1.9 um Tm-doped silica fiber laser pumped at 1.57 um, Electron. Lett., vol. 30., (1994).
8. Weber, T., Luthy, W., Weber, H.P.: Side pumped fiber laser, Appl. Phys. B., 63, (1996) 131-134.
9. Goldberg, L., Cole, B., Snitzer, E.: V-groove side-pumped 1.5 □ m fiber amplifier, Electron. Lett. 33, (1997) 2127-2129.
10. Desurvier, E.: Erbium-doped fiber amplifiers: principles and applications (NY, John Wiley & Sons, Inc.), (1994) Ch. 1&2.
11. Seller, H., Willamowski, U., Tunnermann, A., Welling, H., Unger, S., Reichel, V., Muller, H.R., Kirchhof, J., and Albers, P.: High power cw neodymium-doped fiber laser operating at 9.2 W with high beam quality, Opt. Lett. Vol. 20, (1995) 578-580.
12. Dulling III, I.N., Moeller, R.P.O., Burns, W.K., Villarruel, C.A., Goldberg, L., Snitzer, E., and Po H.: Output characteristics of diode pumped fiber ASE sources, IEEE J Quatum. Electron., Vol. 27, (1999) 995-1003.
13. Golderg, L., Koplow, J., Kliner, D.: High efficient 3 W side-pumped Yb-doped fiber amplifier and laser, in Proc. Conference on Lasers and Electro-Optics, Baltimore, USA, (1999) 11-12.
14. Broderick, N.G.R., Offerhaus, H.L., Richardson, D.J., Sammut, R.A., Caplen J., and Dong, L.: Large mode area fibers for high power applications, Opt. Fiber Technol. Vol. 5, (1999). 185-189.
15. Sahu, J.K., Jeong, Y., Richardson, D.J., Nilsson, J.: 103 W Erbium-ytteribum co-doped large core fiber laser, Optics Communications, Vol. 227, (2003) 159-163.
16. Dominic, V., MacCormack, S., Waarts, R., Sanders, S., Bicknese, S., Dohle, R., Wolak, E., Yeh, S., and Zucker, P.E.: 110 W fibre lasers, Electron. Lett. Vol. 35, (1999) 110.
17. Ueda, K-I., Sekiguchi, H., and Kan, H.: 1 KW cw output from fiber embedded lasers, in Proc. Conference on Lasers and Electro-Optics, Long Beach, USA, post-deadline paper CPDC4, (2002).
18. Platonov, N.S., Gapontsev, D.V., and Shumilin, V. P.: 135 W cw fiber laser with perfect single mode output, in Proc. Conference on Lasers and Electro-Optics, Long Beach, USA, post-deadline paper CPDC3, (2002).
19. Shiner, B. & Lopresti A.: Press Room IPG web page, IPG Photonics Announces shipment of first 6 kilowatt fiber laser, WWW http://www.ipgphotonics.com/html/180_IPG_Photonics_Announces_Shipment_of_First_6_kilowatt_Fiber_Laser.cfm, 2002.

20. Jeong, Y., Sahu, J.K., Payne, DN., and Nilsson, J.: Ytterbium-doped large-core fiber laser with 1 KW continuous wave output power, Electron. Lett., Vol. 40, (2004) 470-472.
21. Selvas, R., Torres, I., Martinez-Rios, A., Alvarez-Chavez, J.A., May-Arrioja, D.A., LiKamWa, P., Mehta, A., and Johnson, E.G.: Wavelength tuning of fiber lasers using multimode interference effects, Opt. Express Vol. 13, (2005) 9439-9445.
22. Ceballos-Herrera, D.B., Torres-Gomez, I., Martinez-Rios, A., Alvarez-Chavez, J.A., Selvas, R., Sanchez-Mondragon, J.: Ultra-widely tunable long-period-holey fiber grating by the use of mechanical pressure, Applied Optics 46 Vol. 3, (2007): 75-77.
23. Martinez-Rios, A., Selvas, R., Torres-Gomez, I., Mendoza-Santoyo, F., Po, H., Starodumov, A.N., and Wang, Y.: Double-clad Yb-doped fiber lasers with non-circular cladding geometry, Opt. Commun. 246, (2005) 385-392.
24. Torres-Gomez, I., Martinez-Rios, A., Anzueto-Sanchez, G., and Selvas, R.: Multi-wavelength switchable double-clad ytterbium-doped fiber laser based on reflectivity control of fiber Bragg gratings by induced bend loss, Optical Review Vol. 12, (2005) 65-68.
25. Grattan, L.S.: Optical Fiber Sensor Technology: Advanced applications – Bragg Gratings and distributed sensors, (Kluwer Academic Publishers), Ch. 1, (2000).
26. Raman, Kashyap: Fiber Bragg Grating (Academic Press), Ch. 1 & 6, (1999).
27. Bircks, T.A., Knight, J.C., and Russell, P.St. J.: Endlessly single-mode photonic crystal fiber, Opt. Lett., Vol. 22, No. 13, (1997) 961.
28. Knight, J. C. , Birks T. A. , Cregan, R. F., Russell P. St. J.and de Sandro, J. P.: Large mode area photonic crystal fiber, Electron. Lett., Vol. 34, No. 13, (1998) 1347.
29. Lee, J.H., Yusoff, Z., Belardi W., Ibsen, M., Monro, T.M., and Richardson, D.J.: Investigation of Briullouin effects in small-core holey optical fiber: lasing and scattering, Opt. Lett., Vol. 27, No. 11, (2002) 927.
30. Mortensen, N.A., Folkenberg, J.R., Skovgaard, P.M.W., and Broeng, J.: Numerical Aperture of Single-Mode Photonic Crystal Fibers, (2002).
31. Bjarklev, A., Broeng, J., and Bjarklev, A.S.: Photonic Crystal Fibres, Kluwer Academic Publishers, Ch. 2, 2003. y Zolla Ederic, Foundations of Photonic Crystal Fibres, GBR: Imperial College Press, Ch. 1, 82005).
32. Sahu, J.K., Renaud, C.C., Furusawa, K., Selvas, R., Alvarez-Chavez, J.A., Richardson, D.J., Nilsson, J.: Jacketed air-clad cladding pumped ytterbium-doped fibre laser with wide tuning range, Electron. Lett., 37 vol.18, (2001) 1116-1117.
33. Nilsson, J., Selvas, R., Belardi, W., Lee, J.H., Yusoof, Z., Monro, T.M., Richardson, D.J., Park, K.D., Kim, P.H., and Park, N.: Continuos-wave pumped holey fiber Raman laser, in proc. Optical Fiber Communications, WR6, (2002).
34. Torres-Gomez, I., Stolen, R., Kominsky, D., Martínez-Gamez, A., Martinez-Rios, A., Selvas-Aguilar, R.: Fibras de Cristal Fotónico, XLVII Congreso Nacional SMF/XVI Reunion Anual AMO, Hermosillo, Son., Octubre del (2004).
35. Ceballos-Herrera, D.E., Torres-Gomez, I., Martinez-Rios, A., Alvarez-Chavez, J.A., Selvas-Aguilar, R., Sanchez-Mondragon, J.: A Simple Widely Tunable Band-rejection Filter in Holey Fiber, Proc. OSA Optical Fiber Sensor Congress (2006), paper ThE3.
36. Castillo-Guzman, A., Selvas-Aguilar, R., Castañeda-Rodriguez, D., Calles-Arriaga, C., Martinez-Rios, A., Torres-Gomez, I., May-Arrioja, D.A.: Simple Optical Fiber Voltaje Sensor Based on an U-Groove Fiber Alignment System, Proc. OSA Optical Fiber Sensor Congress (2006), paper TuE59.

Chapter 15

Integrated InP Photonic Switches

Daniel A. May-Arrioja[1], and Patrick LiKamWa[2]

1 Photonics and Optical Physics Laboratory, Optics Department, INAOE
Apdo. Postal 51 y 216, Tonantzintla, Puebla 7200, México
2 CREOL and FPCE: The College of Optics and Photonics, University of Central Florida, Orlando, FL 32816-2700 USA
dmay@inaoep.mx, patrick@creol.ucf.edu

Abstract

An integrated 1x3 optical switch that operates using the principle of carrier-induced refractive index change in InGaAsP multiple quantum wells is demonstrated. The device is very simple, only requiring currents to be applied to two electrodes for complete operational control. An area-selective zinc in-diffusion process is characterized to channel the currents into the multiple quantum wells, thereby enhancing the efficiency of the carrier-induced effects. The zinc depth and profile can be easily controlled by careful control of the diffusion parameters and background doping concentration. The net result is that very low electrical power consumption is achieved, allowing the switch to be operated uncooled and under d.c. current conditions. The crosstalk between channels is better than –17 dB over a range of 50 nm centered at 1565 nm.

Keywords: Beam steering, Carrier induced, Switch, Photonic switch, Integrated, InP, Multiple quantum wells.

1 Introduction

The demand for information transmission capacity has increased dramatically over the past two decades. The primary impetus has been the phenomenal growth of telecom applications such as the Internet. Deployment of high capacity networks are required given the rapidly growing number of users, each one consuming more and more bandwidth due to data transfers that involve image and video, as well as the introduction of bandwidth consuming services such as video-on-demand and interactive media. A challenge for these networks is to offer very high-speed routing capabilities that can handle such massive amounts of information. This can be achieved by exploiting the recent technological explosion in advanced integrated photonic devices.

Optical crossconnects (OXCs) are becoming an interesting approach for future optical networks because they can provide efficient routing functionalities. Several OXCs have been demonstrated using different platforms. In fact, OXCs based on micromechanical devices (MEMS) are now commercially available [1]-[3]. However, a significant amount of research has been devoted to the development of photonic integrated circuits, in particular InP-based technologies, because this can lead to a drastic reduction in the volume, and the interconnection costs, of complex optical circuits. In the case of planar lightwave circuits, the use of a large number of photonic switches to fabricate OXCs is a widely used technique [4]-[6]. One promising approach is to cascade several 1xN or NxN photonic switches in order to achieve NxN operation, because a device with a smaller footprint can be obtained in this manner [7]-[8]. The realization of a versatile and compact 1XN switch becomes critical in this case, as it will ultimately determine the complexity and size of the OXCs.

In this work we demonstrate a versatile 1x3 photonic switch based on InGaAsP multiple quantum wells (MQW). Switching was accomplished by steering the launched optical beam [9] and redirecting it to any one of the three output waveguides. The switch was integrated using an area selective zinc in-diffusion process, which not only optimize the current consumption but allows for control of the free carrier absorption due to the zinc concentration. A switch crosstalk better than –17 dB was experimentally observed over a 50 nm range. Since the switch is patterned on an InP platform, it can be easily integrated with other photonic devices in order to implement a more sophisticated OXC.

2 Zinc In-Diffusion in InP

Diffusion has been the primary method of introducing impurities because is a very simple and economical process. The most common p-type diffusants for III-V compound semiconductors are zinc and cadmium. Zinc atoms occupy substitutional group III sites, behaving as acceptors and, thus, producing p-type electrical properties. Since zinc diffuses one to two orders of magnitude faster than cadmium, it is the preferred diffusant for the realization of deep p-n junctions.

2.1 Interstitial-Substitutional Diffusion Mechanism

The diffusion of Zn in InP has been extensively investigated, and it is generally accepted that the diffusion mechanism is governed by an interstitial-substitutional mechanism [10-12]. The Zn diffusion process in InP is believed to be dominated by the highly mobile Zn interstitials in chemical equilibrium with the substitutional Zn. The overall amount of the interstitial Zn is assumed negligible in comparison to the substitutional Zn. These fast diffusing Zn interstitials transport Zn atoms from high to low concentrations regions and then convert into substitutional or neutral Zn through chemical reactions with In vacancies, with In lattice atoms, or with phosphorus vacancies. This diffusion mechanism leads to a concentration dependent diffusion process. The one-dimensional Zn diffusion process is therefore modeled by the diffusion equation with a concentration dependent diffusion coefficient [13],

$$\frac{\partial C_s}{\partial t} = \frac{\partial}{\partial x}\left(D_{eff} \frac{\partial C_s}{\partial x} \right) \tag{1}$$

where C_s is the concentration of substitutional ions, and D_{eff} is the concentration dependent diffusion coefficient,

$$D_{eff} = D_0 C_s^n \tag{2}$$

with D_0 being a diffusion constant.

The equation can be solved for the case of constant surface concentration and a semi-infinite medium. Shown in Fig. 1 are the normalized diffusion profiles obtained from Eq. 1 for the cases of $n = 0, 1, 2, and\ 3$, with the depth axis normalized to $n = 1$. The case of n=0 corresponds to the typical complementary error function solution that is obtained for a constant diffusion coefficient. The main feature in Fig. 1 is that the concentration dependent diffusion profiles exhibit a sharper diffusion front when compared to the constant diffusion case. In fact, the diffusion front turns out to be steeper as the power dependence of the diffusion coefficient is increased. This is a highly desirable profile since a sharper junction can be easily obtained. Therefore, a square or cubic dependence on the diffusion coefficient would be beneficial in our devices in order to avoid free-carrier absorption.

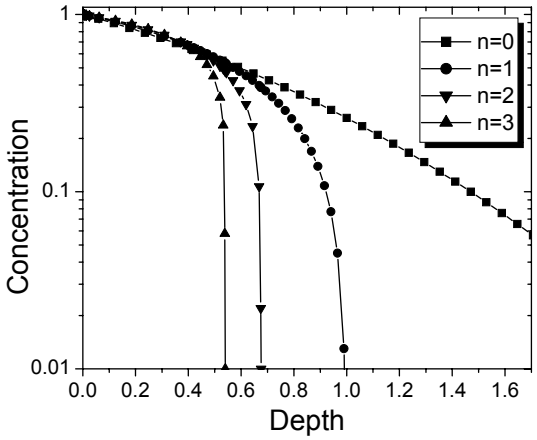


Fig. 5. *Zinc diffusion profiles for different powers of the concentration dependent diffusion coefficient.*

The diffusion of zinc in InP is a complex process, and the concentration dependent diffusivity has been shown to be derived from the charge state of the interstitial. It has also been demonstrated that the charge of the interstitial diffusion species is highly dependent on the initial background donor concentration of the wafer, and this effectively determines the strength of the concentration dependent diffusion process [14]. Highly n-type doped InP wafers (~$1x10^{18}$ cm^{-3}) typically exhibit a cubic dependence of the diffusion constant (n=3), whereas un-doped wafers (~$1x10^{15}$ cm^{-3}) exhibit a linear dependence (n=1) [15]. Intermediate doping concentration on the order of $1x10^{17}$ will reveal a quadratic dependence (n=2) [14-16]. This allows a simple way of controlling the diffusion front. Therefore, in our wafer, an n-type doping concentration of $2x10^{17}$ cm^{-3} was selected for the top InP cladding.

2.2 Semi-sealed Open Tube Zinc Diffusion

The open-tube technique is particularly attractive since it can be easily scaled to accommodate bigger samples making it suitable for processing large wafers. In our experiments, a semi-sealed open-tube diffusion technique was implemented because it is a relatively simple process, and the crystal quality after the diffusion is not degraded as compared to the sealed ampoule technique [17]. The experimental diffusion setup consisted of a conventional linear tube furnace and a graphite box that contains the sample and the Zn source. The graphite box is covered with a loose fitting graphite lid to achieve a higher Zn vapor pressure. The use of the graphite box, instead of a sealed ampoule, also provides a consistent Zn vapor pressure for every run which translates into a more reproducible Zn concentration and diffusion profile.

During the diffusion process the boat is initially placed in the unheated zone of the furnace as the latter is purged with high purity nitrogen for 20 min. The nitrogen flow is then reduced to a trickle flow of 25 cm3/s, and the furnace temperature control is set to the required temperature. After the temperature has been fully stabilized, the boat is then pushed into the heated zone for the required diffusion time. After the required diffusion time, the graphite boat is pulled back to the low temperature zone and cooled down to room temperature with the aid of external fans. Typically, a diffusion temperature higher than 420 °C (melting point) is required to achieve enough vapor pressure and a good diffusion rate. At this elevated temperature the vapor pressure above III-V materials is dominated by that of the more volatile group V element, phosphorous (P) in the case of InP. Therefore, elemental zinc can not be used as the diffusion source because thermal decomposition of the wafer surface occurs due to P depletion that results in indium droplet nucleation. A simple solution for this problem is the use of zinc phosphide (Zn3P2) as the zinc source. When the material is heated, the phosphorous component provides an over-pressure of phosphorus that prevents the surface decomposition of the InP based compounds, and allows for further processing of the wafer.

2.3 Experimental Zinc Diffusion Profile

The wafer structure used in this work was grown by metal organic chemical vapor deposition (MOCVD) on an n^+ InP substrate with a doping concentration of 3×10^{18} cm^{-3}. The whole epitaxial structure was doped n-type at 2×10^{17} cm^{-3} except for the MQW core region that was nominally undoped. The first layer was a 1 μm thick InP buffer layer, followed by the MQW region that was clad by a 1.6 μm thick InP top layer which was in turn capped by a 0.1 μm thick InGaAs layer. The undoped MQW guiding layer consisted of 14 pairs of 100 Å thick InGaAsP (E_g = 0.816 eV) quantum wells and 100 Å thick InGaAsP (E_g = 1.08 eV) barriers. The quantum well width and composition is selected such that an effective bandgap of 0.855eV (1.45 m) could be obtained as confirmed by the room temperature photoluminescence (PL) spectrum. This is designed to provide a 100 nm detuning from the operating wavelength at 1.55 μm, so as to minimize material absorption losses.

The diffusion process was carried out at a temperature of 500 °C using 100 mg of Zn_3P_2 as the Zn source. These values were kept constant for every run, leaving the diffusion time as the only variable. After the diffusion, the depth profile of the zinc concentration that was incorporated into the sample was measured by secondary ion mass spectrometry (SIMS). The Zn profile for diffusion time of 30 min. is shown in Fig. 2 (left). The phosphorous concentration is also shown in this figure as a reference, since the multiple quantum wells are clearly resolved. We can see that the zinc concentration is slightly higher than 10^{18} cm^{-3} at the surface, and decreasing only slightly as it goes deeper, reaching a point where a sharp drop to the background level of the SIMS is observed over a further 20 nm of depth. It was also observed that by changing the diffusion time we can easily control the depth of the Zn diffusion front. From different diffusion times an average diffusion rate of 0.18 μm/min$^{1/2}$ was calculated, which is consistent with previous experimental results in n-type InP. Also shown in Fig. 2 (Left) is a theoretical fitting of the experimental results. As expected, the best fit corresponds to a quadratic dependence of the diffusion coefficient which is typically observed for an acceptor concentration on the order of 10^{17} cm^{-3}.

To selectively define p-n junctions using this process, a suitable diffusion mask is required. This is achieved by depositing a 200 nm thick Si_3N_4 film using PECVD. The areas where p-n junctions will be created are defined using standard photolithography, and the Si_3N_4 is etched away using reactive ion etching (RIE). The diffusion process is then performed using the semi-sealed open tube diffusion technique. Finally, the Si_3N_4 is removed using RIE leaving the wafer ready for further processing. We found the process to be very reliable and reproducible as long as the temperature, Zn source, and time are carefully controlled. The absorption losses resulting form the Zn diffusion process were characterized by fabricating 1 mm long and 2 μm wide straight waveguides that pass through 500 μm length of zinc diffused material. The diffusion was carried out for several different durations, and the losses were measured using the scanning Fabry-Perot resonances technique. As shown in Fig. 2 (Right), the waveguiding loss increases almost exponentially as the Zn depth is increased. This is explained by a larger spatial overlap between the Zn dopant impurities and the optical mode profiles. Therefore, the extent of the Zn depth has to be carefully controlled to minimize total insertion loss.

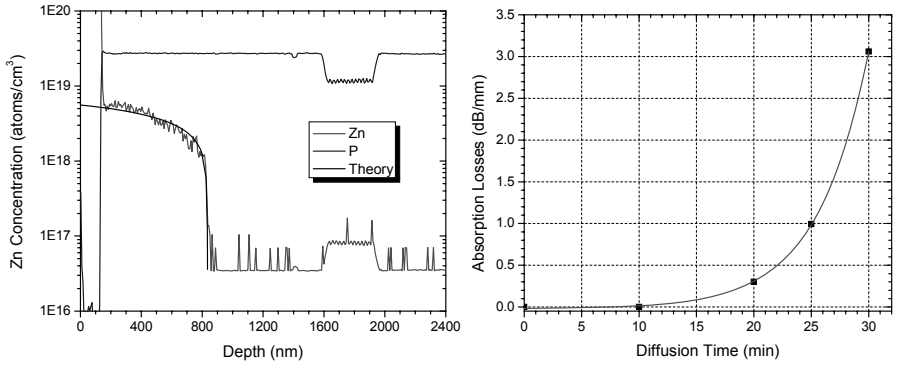


Fig. 6. Experimental zinc diffusion profile and theoretical fitting (n=*2) for 30 min. diffusion (Left), and Measured free carrier absorption losses as a function of diffusion* time (Right).

3 Integrated 1x3 Optical Switch

In this section the operation and fabrication of the proposed optical switch is fully described.

3.1 Principle of Operation

As illustrated in Fig. 3, the 1×3 optical switching device can be divided into two primary sections, the beam steering section and the output waveguides. The beam steering section consists of a 2-μm wide by 500-μm long single-mode input waveguide followed by an 800-μm long slab waveguide. A laser beam entering the 2 μm wide input waveguide is launched into the exact center of the steering region, which consists of a slab waveguide with two parallel Ti/Au/Zn/Au top layer contact stripes, separated by 21 μm as measured from the inner edge of each stripe. The contact stripes are both 800 μm long and 10 μm wide. The output waveguides comprise the second section of the device. Each one is 3 μm wide and has a length of 500 μm. Initially each output waveguide is separated by 3 μm, but as they spread out from the beam steering region, this value increases to 6 μm at the output facet.

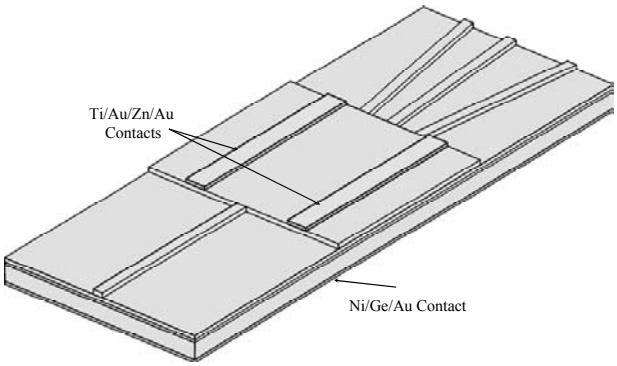


Fig. 7. *Schematic of the 1x3 photonic switch.*

The wafer structure is grown on an n+ InP substrate. It is composed of a 1 μm thick n-type InP buffer layer on which are grown 14 pairs of 100 Å thick, undoped, InGaAsP (E_g=0.816 eV) quantum wells interspaced with 100 Å InGaAsP (E_g=1.08 eV) barriers. The MQWs are topped by a 1.6 μm, n-type, InP cladding layer and a 100 nm InGaAs capping layer. The resulting slab waveguide is essentially symmetric, with all layers grown by metal organic chemical vapor deposition (MOCVD).

The device operates using the carrier induced refractive index change in semiconductors [9]. When no current passes through the contact stripes, a 1.55 μm wavelength laser beam, launched into the steering region by the input waveguide, will expand into a slab mode. Lateral control and confinement of this beam is achieved by the application of an electrical current to each stripe. Zinc is previously diffused underneath the contact stripes in order to control the carrier spreading within the active region of the device. These diffused regions act to channel the electrons into the MQW layer and consequently

enhance their efficiency in providing optical confinement and waveguiding. As current is applied, electrons are injected into the MQW layer where they then spread laterally through carrier diffusion. The areas within the active region that become saturated with electrons experience a corresponding decrease in refractive index. Carrier concentration is highest in the areas directly underneath the stripes, and decreases with lateral distance. This effectively results in the formation of a graded index channel waveguide between the two stripes in the steering region. Careful adjustment of the ratio between the currents applied to the two parallel contact stripes allows the waveguide to be shifted across the entire available range, thereby steering the signal beam. The input optical beam can then be directed to any of the output waveguides.

3.2 Device Fabrication

Device fabrication began with the deposition of a silicon nitride diffusion mask for creating the zinc diffused regions. Plasma enhanced chemical vapor deposition (PECVD) was used to deposit a 200 nm thick silicon nitride film on the substrate surface. For each device, two 10 μm x 800 μm diffusion windows were defined in the nitride film using conventional photolithography and CF_4 plasma-based reactive ion etching. The zinc in-diffusion process previously described was then performed for a time duration of 30 min. After the diffusion process was complete, the silicon nitride mask was removed and Ti/Au/Zn/Au p-type contacts were patterned on top of the zinc diffused areas by evaporation and lift-off. Photolithography was then used to pattern the input waveguide and output waveguide structures. Wet chemical etching with an H_3PO_4:H_2O_2:DI water (1:1:38) mixture was employed to selectively remove the InGaAs top layer. The remaining portions of the InGaAs layer were then used as a mask for the selective wet etching of the InP using an HCl:H_3PO_4:$CH_3CHOHCOOH$ (2:5:1) mixture. A 10 nm InGaAsP etch stop layer, located 190 nm above the MQW layer, allowed precision control of the etch depth and yielded InP ridges of constant height, in addition to a smoothly etched surface. After etching was finished, the substrate was lapped to a thickness of 150 μm and polished to a mirror finish. Finally, the n-type contact consisting of Ni/Ge/Au was deposited via thermal evaporation and annealed-in. The device sample was then cleaved and mounted on a copper header for testing. A top view of the fabricated device is shown in Fig. 4.

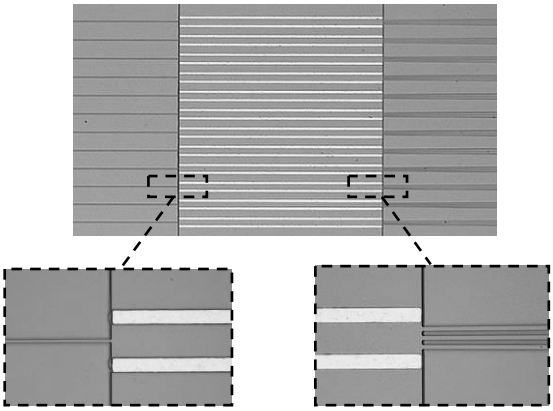


Fig. 8. *Top view of the fabricated 1x3 photonic switch.*

4 Experimental Results

In order that the performance of the switch could be optimized the beam steering unit was initially fabricated and tested as a separate unit before evaluating the complete 1×3 optical switch. This allowed us to optimize the optimum zinc diffusion depth, and thus optimum current consumption.

4.1 Beam Steering Section

The devices were tested using a laser beam from a fiber pigtailed tunable laser operating at 1550 nm and collimated through a fiber collimator that was end-fired coupled to the input waveguide via a 40x microscope objective. The light passed through the steering region and the near-field pattern of the output facet was imaged onto a CCD camera using a 20x microscope objective, with the image displayed on a TV monitor. Located after the 20x microscope objective was a beam splitter, which directed a portion of the output to a detector that feeds into a lock-in amplifier. This branch of the experimental setup was used for measuring the crosstalk between the output waveguides in the completed 1x4 optical switch. A mechanical chopper, operating in conjunction with a lock-in amplifier, was positioned before the 40x microscope objective.

When in use, the chopper modulated the input beam at a frequency of 980 Hz. Precise control of the applied currents was implemented by using a separate laser diode driver for each contact stripe

Thermal effects were precluded in the initial testing of the beam steering section by modulating the laser beam with an in-line integrated Mach-Zehnder modulator to produce optical pulses of 0.7 μs duration at a repetition time of 20 μs. Modulation of the electrical currents directed to the contact stripes was carried out at the same repetition time, but with the pulse duration at 2.5 μs. The optical and electrical pulses were temporally synchronized so that the free carrier distributions were fully stabilized when the optical pulses passed through the steering section.

Optimization of the beam steering section was done by testing the device performance for different zinc diffusion depths, such that the beam could be steered along the gap with the lowest possible current. A zinc depth of 0.8 μm proved successful at providing better localization of the injected current and in limiting the carrier spreading to within the device active region. The maximum applied current required to fully shift the beam within its entire 20 μm range was 12.5 mA. At these low current levels, thermal effects were negligible and uncooled operation of the beam steering section under d. c. current injection was realized. The final results from these optimization studies are shown in Fig. 5.

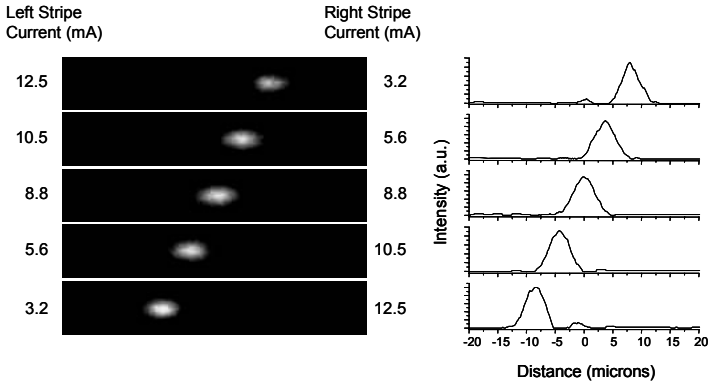


Fig. 9. *Left: Images of guided beam at the output facet showing the beam being steered as the ratio of the currents is changed. Right: The corresponding near-field intensity profiles of the output beam indicating a total steering range of 17μm.*

4.2 Reconfigurable 1x3 Photonic Switch

The 1x3 optical switch was tested using the setup described above. The only variant was the use of a 40x microscope objective, instead of the 20x, to facilitate measuring the crosstalk between the output channels using the detector in the beam splitter defined branch and the lock-in amplifier. A 100 μm diameter pinhole at the detector aperture guaranteed that only one channel could be measured at a time. The positioning of this detector could be adjusted along both the x- and y-axes, with the z-axis denoted as being parallel to the laser beam, so that the intensity in each of the four channels could be measured in turn. The approach employed in measuring the crosstalk required determining the current values needed to direct the laser beam to each of the four output waveguides. Then the current driver values were set to correspond to the first output waveguide and measurements of the light intensity in each of the four channels was measured through the lock-in amplifier. The current values were changed to send light to the next waveguide and the process repeated until all channels had been likewise evaluated. As with the beam steering section, pictures of the output facet of each waveguide and intensity profiles for each channel were acquired. The data is shown in Fig. 6.

As can be seen in the above picture, the applied current values required to direct the laser beam into the various output waveguides of the 1×3 optical switch were higher than those used for the beam steering subunit. These higher currents are attributed to the fact that the beam steering section was optimized to achieve maximum steering with the lowest possible currents, not to achieve any predetermined confinement specification. In contrast, the current values selected for directing the laser beam to the individual output waveguides of the 1×3 switch were optimized to reduce the crosstalk. Therefore, higher currents were necessary in order to achieve a more confined mode. Crosstalk from the switched channel to either one of the two remaining channels was measured and was found that crosstalk levels better than -17 dB could be obtained over a range of 50 nm, with a center wavelength of 1565 nm. This crosstalk level is easily obtained for light switched to either channel. If the wavelength range is decreased, it is possible to further reduce the crosstalk, and also the required current is reduced. The results demonstrate the versatiliy of the switch. Furthermore, since the switch is developed on an InP semiconductor platform, integration with other photonic and electronic devices is a very attractive option.

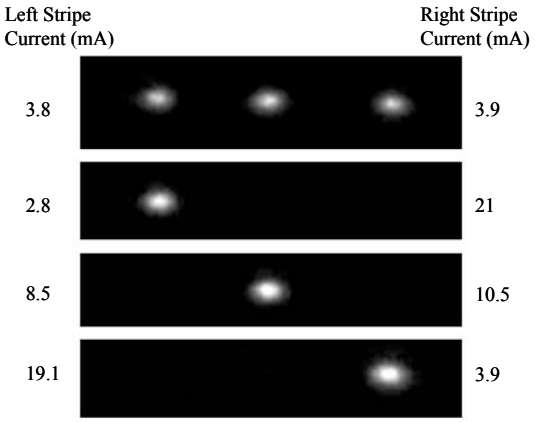


Fig. 10. *Switching characteristics of the 1x3 photonic switch.*

5 Conclusions

A novel 1×3 optical switching device that operates with low driving current values using the principle of carrier-induced refractive index change in semiconductors has been demonstrated. The low power consumption resulted from the use of an area selective zinc in-diffusion process that acted to channel the current into the MQWs, thereby enhancing the efficiency of the carrier induced effects. A simple beam steering device was employed to direct the laser beam to the appropriate output channel. The lowest crosstalk obtained in our device was better than –17 dB over a range of 50 nm.

References

1. A. Olkhovets, P. Phanaphat, C. Nuzman, D. J. Shin, C. Lichtenwalner, M. Kozhevnikov, and J. Kim, "Performance of an Optical Switch Based on 3-D MEMS Crossconnect", *IEEE Photon. Technol. Lett.*, Vol. 16, pp. 780-782, No. 3, March 2004.
2. Bishop D. J., Giles C.R., and Austin G. P., "The Lucent LambdaRouter: MEMS technology of the future here today", *IEEE Communications Magazine*, Vol. 40, pp. 75-79, No. 3, March 2002.
3. Patrick B. Chu, Shi-Sheng Lee, and Sangtae Park, "MEMS: The Path to Large Optical Crossconnects", *IEEE Communications Magazine*, Vol. 40, pp. 80-87, No. 3, March 2002.
4. Takashi Goh, Akira Himeno, Masayuki Okuno, Hiroshi Takahashi, and Kuninori Hattori, "High-Extinction Ratio and Low-Loss Silica-Based 8x8 Strictly Nonblocking Thermooptic Matrix Switch", *J. Lightwave Technol.*, Vol. 17, pp. 1192-1199, No. 7, July 1999.
5. Gregory A. Fish, Beck Mason, Larry A. Coldren, and Steven P. DenBaars, "Compact, 4 x 4 InGaAsP–InP Optical Crossconnect with a Scaleable Architecture", *IEEE Photon. Technol. Lett.*, Vol. 10, pp. 1256-1258, No. 9, September 1998.
6. Toshio Kirihara, Mari Ogawa, Hiroaki Inoue, Hiroshi Kodera, and Koji Ishida, "Lossless and Low-Crosstalk Characteristics in an InP-Based 4 x 4 Optical Switch with Integrated Single-Staged Optical Amplifiers", *IEEE Photon. Technol. Lett.*, Vol. 6, pp. 218-221, No. 2, Feb.1994.
7. M. P. Earnshaw, J. B. D. Soole, M. Cappuzzo, L. Gomez, E. Laskowski, and A. Paunescu, "8 x 8 Optical Switch Matrix Using Generalized Mach–Zehnder Interferometers", *IEEE Photon. Technol. Lett.*, Vol. 15, pp. 810-812, No. 6, June 2003.
8. M. P. Earnshaw, J. B. D. Soole, M. Cappuzzo, L. Gomez, E. Laskowski, and A. Paunescu, "Compact, Low-Loss 4x4 Optical Switch Matrix Using Multimode Interferometers", *Electron. Lett.*, Vol. 37, pp. 115-116, No. 2, January 2201.
9. D. A. May-Arrioja, N. Bickel, and P. LiKamWa, "Optical Beam Steering using InGaAsP Multiple Quantum Wells", *IEEE Photon. Technol. Lett*, vol. 17, no. 2, pp. 333-335, February 2005.
10. U. Schade and P. Enders, "Rapid Thermal-Processing of Zinc Diffusion in Indium-Phosphide," *Semiconductor Science and Technology* 7(6), 752-757 (1992).
11. B. Tuck, "Atomic diffusion in III-V semiconductors " *Bristol [Avon]*, (1988).
12. G. J. Vangurp, T. Vandongen, G. M. Fontijn, J. M. Jacobs, and D. L. A. Tjaden, "Interstitial and Substitutional Zn in Inp and Ingaasp," *Journal of Applied Physics* 65(2), 553-560 (1989).
13. S. N. G. Chu, R. A. Logan, M. Geva, and N. T. Ha, "Concentration-Dependent Zn Diffusion in Inp during Metalorganic Vapor-Phase Epitaxy," *Journal of Applied Physics* 78(5), 3001-3007 (1995).
14. G. J. Vangurp, P. R. Boudewijn, M. N. C. Kempeners, and D. L. A. Tjaden, "Zinc Diffusion in N-Type Indium-Phosphide," *Journal of Applied Physics* 61(5), 1846-1855 (1987).
15. H. B. Serreze and H. S. Marek, "Zn Diffusion in Inp - Effect of Substrate Dopant Concentration," *Applied Physics Letters* 49(4), 210-211 (1986).

16. I. Yun and K. S. Hyun, "Zinc diffusion process investigation of InP-based test structures for high-speed avalanche photodiode fabrication," *Microelectronics Journal* 31(8), 635-639 (2000).
17. T. H. Weng, "A comparative study of p-type diffusion in III-V compound semiconductors," *Proceedings Electron Devices Meeting*, 120-122 (1997).

Chapter 16

Non-Linear Optical Effects in Liquid Crystals

René Domínguez-Cruz[1], Abel Padilla-Mijares, and Adolfo Rodríguez-Rodríguez.

Autonomous University of Tamaulipas, UAT-UAMRR; Reynosa, Tamaulipas, México
Apdo. Postal 88779.
rfdominguez@uat.edu.mx.

Abstract

In this work we present the recent experimental results about the characterization of non-linear properties in a liquid crystals. these results are obtenided througth z-scan tecnhique using a simple low power he-ne laser in cw operation. we observe a great positive kerr nonlinearity in the sample obtained in the regime temperature of nematic phase. for upper temperature, we observe a notable decreasing of the optical nonlinearity due to sample change to isotropic phase. in adition, we observe the presence of a nonlinear absorption contribution.

Keywords: Electrical susceptibility, liquid crystal, nematic liquid crystal.

1 Introduction

Recently, there are a relevant interest to search for new optical materials with large nonlinear coefficients and fast response for photonics applications. It is due to the wide applications in several areas like high speed optical switching devices, real-time coherent optical signal processors and holographic storage. All of them, the operations are based on the refraction index dependent of illumination intensity[1]. These elements are currently begin performed but there are an expectation that they may eventually to be improve their functions with new optical materials. Recently, in several areas of opto-electronics have been a huge interest for organics materials because the possibility of optimization of these nonlinearities through manipulation of their composition and aggregation state[2]. In particular, the organic liquid crystals are known to exhibit large optical nonlinearities which have been the subject of considerable study in last years[2-4]. The liquid crystals have shown being competitive materials for photonics applications, especially for their potential use as inexpensive nonlinear optical element and for optical switches[5] and photorefractive holographic storage[6]. In consequence, is a priority to measure the nonlinear properties in materials to determine the optical limit device for the suitable application.

In adition, there are numerous techniques for the measurement of non linear refraction index in materials. Non-linear interferometry[7], degenerate four wave mixing[8], and beam distortion measurements, known as Z-scan[9], are the frequently techniques reported. The first two methods, interferometry and wave mixing, are potentially sensitive techniques but require a complex experimental set-up. The Z-scan technique, on the other hand, is based on the principles of spatial beam distortion due to self-focusing (or defocusing) processes derivated from the combination of a intensity-dependent refractive of the media. With this technique is possible to measure with a single beam the sign and magnitude of the refractive nonlinearities offering simplicity as well as high sensitivity. As examples, this technique has been used to determine the optical nonlinearity in a wide type of materials, as photorefractive media, liquid crystals, photopolymers and biological materials[11-13].

In this chapter we report the preliminary results on a Z-scan detection technique applied to measure the sing of a new organic liquid 5CB sample doped with methyl-yellow at 1%. Also, we measure the magnitude of the optical nonlinearity and we observe a great optical nonlinearity ten times more than other results previously reported in liquid crystals. This phenomena is obtained in the nematic regime.

2 Theoretical Description

When a high irradiance laser beam is propagated through any transparent medium, photo-induced refractive index variations may be lead to self-focusing of the beam. These photo-induced refractive index variations are commonly described by the simple relationship:

$$n = n_0 + \Delta n = n_0 + \gamma I \tag{1}$$

where n_0 is the linear refractive index, I is the intensity of the incident light, expressed in W/m^2 and γ is defined as the nonlinear refractive index., expressed in m^2/ W. Frequently, the term $\Delta n = \gamma I$ is called refractive index change. The expression (1) is valid for centrosymetric media and considering only the influence of the third order electrical susceptibility $\chi^{(3)}$ in the term of the polarization density ***P.*** This equation defines the optical Kerr effect.

In addition, is well known the Z-scan technique[9] is a simply method to obtain the sign and magnitude of the non-linear refractive index besides other optical properties, like the nonlinear absorption, in optical materials. This method is based on the spatial distortion of a laser beam passed through a nonlinear optical material, as is showed in the figure 1.

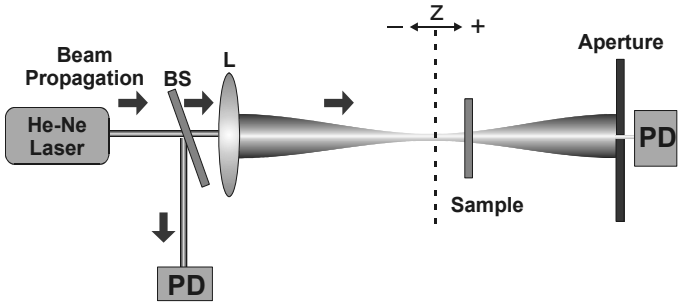


Fig 1. *Experimental set-up to perform Z-scan measurements.* ***PD:*** *Photodetector,* ***L:*** *lense,* ***BS:*** *beam splitter.*

The sample is situated near the beam waist and moved along z direction. Lately the transmitted intensity is measured through a finite aperture in the far field as a function of sample position z, measured respect focal plane. If the sample has a positive non linearity and the sample is moved to one side of the beam waist, the detected intensity increased to a peak. When the sample is moved to the other side of the waist, the detected intensity decreases to a valley. The difference of the intensity from the peak to the valley has been shown to be proportional to the non-linear index refraction γ [9]. In the other case, when the sample has a positive non-linearity, we obtain a inverse Z-scan curve, i.e. a valley followed by a peak. The Figure 2 describes those situations. Consequently the Z-scan permits the calculations of non linear index refraction for different materials by a comparatively simple method.

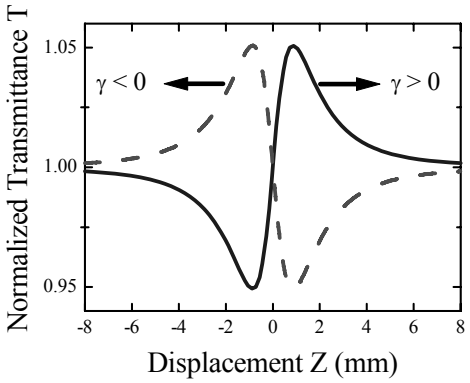


Fig. 2. *The Z-scan theoretical curves for the transmittance as function of the z distance. These curves are obtained by equation (2) considering only a Kerr nonlinearity. Continuous line corresponds to positive non- linearity and dashed line to negative non linearity behavior.*

The intensity distribution of the beam induces a position dependent $\Delta n(r,z) = \gamma\, I(r,z)$ change of refraction inside the sample, where γ is the coefficient of nonlinear refraction. This causes a divergence or a convergence of the laser beam, depending on the sample position and the sign of γ. In particular, when a sample is moving along the propagation direction z inside the focal region of a focused laser beam. The nonlinear refractive index of the material induces a variable phase shift across the profile of the laser beam exiting from the sample. The variation is proportional to the incident irradiance on the sample and depends on its position relative to the beam focus. Finally, the far field pattern of the beam can be obtained by

virtue of Huygens–Fresnel's principle[14,15], through a zeroth-order Hankel transform[14–16] of the electric field at the exit plane of the sample.

In other hand, one of the advantages of the Z-scan technique is the possibility of separation of the contributions of several nonlinearities when they are presented simultaneously. In general, when both nonlinear refraction and absorption are presented, the normalized transmittance $T_{\Delta\Phi}$ of sample placed in the closed-aperture scheme for only Kerr nonlinearity can be written following a similar analisys as[16]:

$$T_{\Delta\Phi} = 1 + \frac{4x}{(x^2+9)(x^2+1)}\Delta\Phi_0 \tag{2}$$

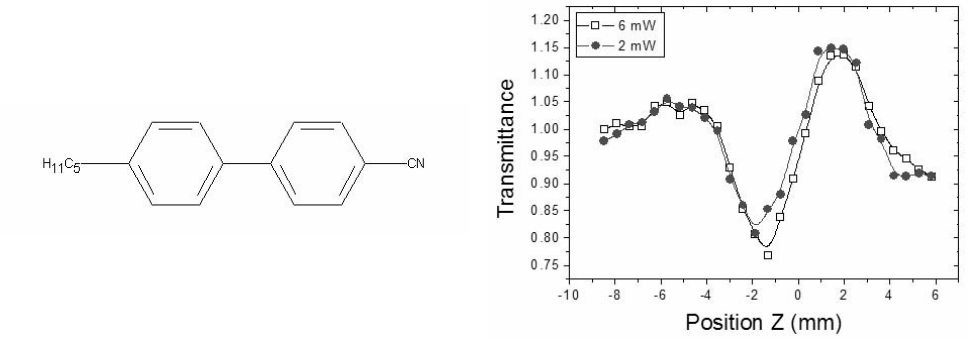
where $x = z/z_0$, $\Delta\Phi_0$ is the parameter which is related to phase shift near the focal point as a result of nonlinear refraction, $\Delta\Phi_0 = k\gamma I_0 L_{eff}$, I_0 is the laser-radiation intensity at the focal point, $k = 2\pi/\lambda$ is the wave number, γ is the nonlinear refraction index $L_{\text{eff}} = [1-\exp(-\alpha_0 L)]/\alpha_0$ is the effective length of the sample, α_0 is the linear absorption coefficient, and L is the sample length. Considering a radial phase shift $\Delta\Phi_0$ at the aperture less to π, the final formula, giving a bridge between the normalized transmittance difference from peak to valley, $\Delta T_{\text{p-v}}$ and the on-axis refractive index change Δn in the beam waist, takes the form[10]:

$$\Delta n_0 = \frac{\Delta T_{p-v}}{0.406(1-S)^{0.25} kL_{eff}} \tag{3}$$

where $S = 1-\exp[(-2r_a/w_a)^2]$, r_a is the beam radius at the aperture and w_a is its radius. The equation (3) is usually applied for calculations of the refractive index change of an optical nonlinear medium using the Z-scan experimental data.

3 Experimental Methodology and Results

The experimental array consist in the typical Z-scan configuration showed in Figure 1. The illumination source is a cylindrical He-Ne laser with linear polarization in a CW operation at λ=632.8 nm and FHWM= 0.45 mm. The laser beam is divided by a beam splitter in two arms, on of them is used as reference incident beam. The second one passes through a positive lens f=5 cm and illuminate the liquid crystal sample is situated near to the focus. The intensity transmitted is monitored by the silicon photodetector provided with an small aperture. The liquid 5CB sample doped with methyl-yellow at 1% was fixed in the laboratories of the Engineering and Science Opto-electronics Group, at INAOE. The chemical structure is showed in the figure 3.a. The sample was located into a termal box isolated, which the temperature is controlled electronically. We measure the transmitted intensity by the sample for negative and positive position z respect the focal length and finally we calculate the transmittance respectively. These measurements were performed for 2 different temperatures. The experimental results are shown in figure 3.



(a) (b)

Fig. 3. (a) The chemical structure of the 5CB *sample. (b) Experimental Z-scan curves for the 5CB sample doped with methyl-yellow at 1%. A positive Kerr nonlinearity is observed for two illumination powers. Both of them were obtained at room-*temperature.

The Figure 3 shows the normalized transmittances as functions of sample position in the closed-aperture Z-scan scheme. We observe a positive nonlinearity in the sample and a small increment in the peaks of the characteristic Z-scan curve for temperature P= 6 mW. Next, we measure the Z-scan transmittance as a function of the temperature.

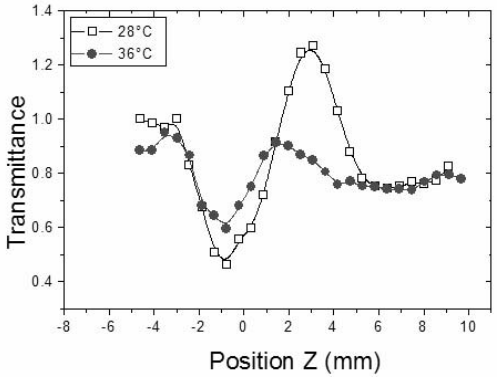


Fig. 4. *Experimental Z-scan curves for the 5CB sample at two different temperature. P= 2 mW , f=5 mW*

This behavior could be explained as following. The liquid crystals are a state of the matter between a solid and liquid[18]. They flow like a liquids, but possess some physical properties characteristic of crystals, such as birefringence. Liquid crystals are composed of anisotropic-shaped organic molecules and as a result all their physical properties, such as the dielectric and magnetic susceptibilities, are anisotropic. Most of the more popular liquid crystals are composed of molecules that are strongly elongated in one direction[3] so that they appear as a collection of road-like molecules, as is showed in the figure (4).

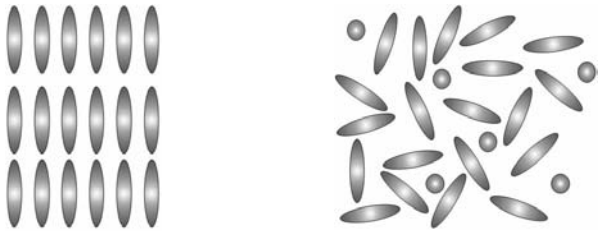


Fig. 5. *Molecule array in a liquid crystal in a) nematic phase ($T<T_c$) and b) isotropic phase ($T>T_c$). The circles indicate rods that are pointing out of or in the plane of the page.*

In the isotropic phase the axes of the molecules are randomly distributed, whereas in the nematic phase the configuration of the lowest energy is reached when all the molecules are on average aligned aling a single direction. In the nematic phase, the orientation is keeping still a critical temperature T_c[17]. For temperature over T_c, the orientations of the molecules is broken and it adopt a random orientation. The typical transition temperature T_c for the usual liquid crystals is reported in the range 35-80°C[15]. For our measurements showed in figure 1, we can observe s similar behavior.In contrast, for a temperature 22°C the characteristic Z-scan curve has a significant increasing, probably due to the experimental measurement was perform in the nematic phase where the anisotropy is more strong.

4 Conclusions

In this chapter we report the observation of a positive optical nonlinearity in a new sample liquid crystal using a simple low power He-Ne laser in CW operation at 632 nm. This phenomena is due the experiments was perform under transition temperature T_c. For upper temperatures to T_c , we observe a decreasing in the Kerr nonlinearity. Nevertheless, we consider that is necessary future experiments about a nonlinear absorption contribution and to examine both effects in samples with different concentrations of dopant and the polarization of the molecules.

Acknowledgments

We acknowledge to Mexican program PROMEP-SEP to support this project through Grant 103.5/03/2524. Also, gratefully acknowledge Dr. Ruben Ramos-García to provide the sample and for his the valuable support.

References

1. C. Yeh. *Applied photonics,* Academic Press. San Diego, 1994.
2. P.N. Prassad, D.J. Williams. *Introduction to nonlinear optical effects in molecules and polymers,* John Wiley and Sons. New York, 1991.
3. I.C. Khoo. *Liquid Crystals: Physical properties and non-linear optical phenomena,* John Wiley and Sons. New York, 1995.
4. F. Simoni, *Nonlinear optical properties of liquid cristals and Polymers-dispersed liquid cristals*, World Scientific Publishing. Singapore, 1997.
5. L. Lucchetti, M. Di Fabrizio, O. Francescangeli, and F. Simoni, *Optics Communications*. 233, 417-424. 2004.
6. I.C. Khoo et al, Dye-doped photorefractive liquid cristals for dinamic and storage holographics grating formation and spatial light, *Proceddings of IEEE*, 87, 11, November 1999. 1897-1911.
7. M.J. Moran, C.Y. She, R.L. Carman. *IEEE Journal of Quantumm Electronics*. 11, (1975) 259.
8. S.R. Friberg, P.W. Smith, IEEE *Journal of Quantumm Electronics*. 23, (1987) 2089.
9. M. Sheik-Bahae, A.A. Said, E.W.V. Stryland. *Optics Letters*. 14, (1989) 995.
10. M. Sheik-Bahae, A.A. Said, T.H. Wei D.J. Hagan E.W.Van Stryland. IEEE J. Quantum Electron. 26, (1990) 760.
11. S. Bian, *Optics Communications*. 141, (1997) 292.
12. M. Tremblay, T.V. Galstyan, M.M. Denariez-Roberge, R.A. Lessard. Proccedings SPIE 3294 (1997) 78.
13. P.A. Márquez-Aguilar, J.J. Sánchez-Mondragón, S. Stepanov G. Bloch. 118, (1995) 165-174.
14. A.E. Siegman, Lasers, University Science Books, Mill Valley, California, 1986.
15. R.E. Samad, N.D. Vieira Jr., J. Opt. Soc. Am. B 15 (1998) 2742.
16. A.E. Siegman, Opt. Lett. 1 (1977) 13.
17. S. McConville, D. Laurent, A. Guarino, S. Residon. Am. J. Phys. 73, 5, 425-432, (2005)
18. P.G. De Gennes and J. Prost, *The Physics of Liquid Crystals*, 2nd Ed. Oxford Science, New York, 1992.

Esta obra se terminó de imprimir en septiembre del 2007
en los talleres de Ultradigital Press, S.A. de C.V.
Centeno 162 - 3, Col. Granjas Esmeralda
CP 09810, México, D.F.