# *Telecommunications: Advances and Trends in Transmission, Networking and Applications*

**Edited by**

Charles Casimiro Cavalcante
*Federal University of Ceará, Brazil*

Ricardo Fialho Colares
*University of Fortaleza, Brazil*

Paulo Cezar Barbosa
*University of Fortaleza, Brazil*

UNIFOR
ENSINANDO E APRENDENDO

# Telecommunications: Advances and Trends in Transmission, Networking and Applications


**Edited by**

Charles Casimiro Cavalcante
*Federal University of Ceará, Brazil*

Ricardo Fialho Colares
*University of Fortaleza, Brazil*

Paulo Cezar Barbosa
*University of Fortaleza, Brazil*

**Authors**

André L. F. de Almeida
F. Rodrigo P. Cavalcanti
Bernadette Dorizzi
Rafael Ferrari
M. Koshiba
Josef A. Nossek
Vitaly F. Rodríguez-Esquerre
Ricardo Suyama

Romis R. F. Attux
Fabiano de S. Chaves
Leonardo T. Duarte
Hugo E. Hernández-Figueroa
João Cesar M. Mota
Raimundo A. de Oliveira Neto
João Marcos T. Romano

Charles C. Cavalcante
Mérouane Debbah
Gérard Favier
Cynthia Junqueira
Aline de O. Neves
Moisés V. Ribeiro
Ricardo B. dos Santos
Benno Zerlin



Fortaleza, CE - Brazil
2006

T267t Telecommunications : advances and trends in transmission , networking and applications / edited by Charles Casimiro Cavalcante, Ricardo Fialho Colares e Paulo Cezar Barbosa. - Fortaleza : Universidade de Fortaleza - UNIFOR, 2006.
187 p.

ISBN: 85-98876-18-6

1.Telecomunicações. I. Cavalcante, Charles Casimiro. II. Colares Ricardo Fialho. III. Barbosa, Paulo Cezar.

CDU 621.39

*To Casimiro, my loved father, in memoriam*
C. C. C.

*To my lovely wife and our expected son*
R. F. C.

*To my lovely wife, daughters and granddaughters*
P. C. B

# Contents

# Preface

In the last years, the area of telecommunications has been experiencing an effervescent period in its development. A wide number of new strategies, techniques and applications have arisen from the communications-related societies. The development has been noticed in several directions: wireless, wireline, signal processing, optics and electromagnetics. Theoretical advances with the possibilities to turning research results and proposals into industrial/commercial applications allowed a fertile field for improvements on many topics of telecommunications. Further, this offers a good opportunity to focus on key developments in the area during the last years and outline emerging trends which may drive research and industrial activities in the next years.

During the organization of the VI International Telecommunications Symposium (ITS2006), co-sponsored by the Brazilian Telecommunication Society (SBrT) and IEEE Communications Society (COMSOC), we have had the opportunity to reviewing papers and tutorials of several subareas in telecommunications and the privilege of knowing how much outstanding work is being done all over the world, providing a great state-of-art of edge-like subjects. This book is a collection of manuscripts selected from those ones accepted for publication into the Proceedings of the ITS2006. Since our main idea is to provide a general, didactic and deep material on hot topics, we have chosen some of the proposals for tutorials and requested extended manuscripts for being chapters of this piece. We are deeply indebt with the General Chairman of the ITS2006, Prof. João Cesar Moura Mota, by providing us the best conditions to go further with our project. We could not forget to thank the cooperation and help of all contributors of the book, such distinguished researchers, who have, in essence, made this work possible. This collection is the representation of their work and we made our best in order to allow their work to have the merit they deserve. We hope we could fit their expectations.

This book is aimed at the advanced level and is suitable for graduate students and researchers. The material can be used as advanced special topics to be used for independent study, for example, an introduction for graduate students who wish to conduct research in the areas covered by the chapters of the book. Engineers may appreciate the up-to-date tutorial exposition to timely topics. Some technologies reviewed here are being (or even already) incorporated into commercial systems.

The book begins with a chapter by M. Debbah on wireless channel modeling and the limits of performance by means of the use of information-theoretic tools. Such material is imprescriptible for those interested in capacity, limits of performance and evaluation of wireless systems.

The second chapter by Suyama *et al* discusses the concept of unsupervised signal processing in its different branches. Starting from equalization of single user systems they cover up to the multiuser MIMO case the blind recovering information strategies in a very general background allowing the application in different contexts of information systems.

Chapter three, by Almeida *et al* states the modeling of wireless communication systems by means of tensor approach. Such an elegant tool allows the generalization of many different problems in telecommunication into the same framework.

Oliveira Neto *et al* study the problem of power control in wireless systems in Chapter four. The aspects of optimization of such important resource in wireless networks is discussed through a number of strategies showing the great potential of those methods.

Chapter five by Zerlin and Nossek discusses the topic of cross-layer optimization. The rationale of such approach differs from the most used ones by the fact of consideration of the interrelations among the different layers of the system to provide a global gain of performance.

The very updated subject of power line communications is covered by M. V. Ribeiro in chapter six. The importance of such information transmission strategy is discussed by means of the different possibilities of handling with the main

issues which arise in such systems as well as the countermeasures for them, based on intelligent signal processing, such as fuzzy filtering.

B. Dorizzi covers in Chapter seven the most advanced trends in biometrics for user recognition. This subject has been receiving an increasing interest of researchers from several countries and different communities (e.g. biomedical, signal processing, machine learning, communications) due to the great impact such techniques and applications have into access control and monitoring. The tutorial presented here states the state-of-art of those methods.

Finally, Rodríguez-Esquerre *et al* in Chapter eight discuss about the devices and applications of photonic structures. The number of possible applications, e.g. couplers, wave guides and lasers, makes this topic a very important one in telecommunication, specially when we are observing the convergence of very different services in telecommunications requiring very high data rates.

Fortaleza Ceará, Brazil,
September, 2006

*Charles Casimiro Cavalcante*
*Ricardo Fialho Colares*
*Paulo Cezar Barbosa*

# The Editors

**Charles Casimiro Cavalcante** was born in 1976 and obtained the B.S. and the M.S. degrees in Electrical Engineering in 1999 and 2001, respectively, both from the Federal University of Ceará (UFC) in Fortaleza, Brazil. In 2004 he received the Ph.D. from the State University of Campinas (UNICAMP) in São Paulo, Brazil, with a thesis in the subject of blind source separation strategies for multiuser processing. Dr. Cavalcante has been working on signal processing strategies for communications where he has several papers published. He has been working as technical coordinator on research projects for future wireless systems with private and governmental partners. He is now an Associate Professor at Teleinformatics Engineering Departament of UFC and researcher of the Wireless Telecommunications Research Group (GTEL). He is member of the IEEE Signal Processing Society, Institution of Engineering and Technology (IET) and Brazilian Telecommunications Society. His main research interests are signal processing for communications, information theory and information geometry, blind source separation, wireless communications and statistical signal processing.

**Ricardo Fialho Colares** was born in Fortaleza, Brazil, in 1969, graduated in Electrical Engineering from Federal University of Ceará-UFC, Brazil, in 1992, obtained a Master degree in Electronics and Communications from Unicamp, Brazil, in 1997 and Doctorate degree in Telecommunications and Telematics (Unicamp, 2000). Adjunct Professor at University of São Francisco-USF, Itatiba-SP, in 2000. Full Professor at University of Fortaleza-Unifor and Coordinator of the Telecommunication Engineering Undergraduate Course, since 2001. Member of the Teaching, Research and Extension Council-CEPE at Unifor from 2003 to 2005. Member of the Editorial Committee of the Unifor scientific periodical Revista Tecnologia, since 2001. Author of scientific papers presented on congress meetings in Telecommunications and Digital Signal Processing promoted by IEEE and the Brazilian Telecommunications Society. Coordinator of various R&D project sponsor by Brazilian and multinational companies (SANMINA-SCI, SIEMENS, MARCONI, ARMTEC). Member of the Brazilian Telecommunications Society since 1994. Member of the organizing committee of various local and International scientific conference, workshop and symposium (SBrT 2001, WCSF2004, ICT2004, Tech Week at Unifor 2002-2006, ITS2006).

**Paulo Cezar Barbosa** was born in Fortaleza in 1949, graduated in Physics from the Federal University of Ceará-UFC in 1974, obtained Master Studies in Physics from the Physics Department of the Federal University of Ceará-UFC in 1980. Retired adjunct Professor from the Physics Department of the Federal University of Ceará. Coordinator of the Physics Course at Federal University of Ceará 92/95, Head Professor of the University of Fortaleza (UNIFOR). Coordinator of the Initiating Research Program Meetings held at the University of Fortaleza (UNIFOR), co-author of scientific papers presented at both National and International Condensed Matter Physics. Post-grad and Research advisor for the University of Fortaleza (UNIFOR). Member of the organizing committees of National and International meetings (Meeting of Physicists from the North-Northeastern Regions -1994, SBrT 2001, ICT2004, ITS2006).

# List of Contributors

**André L. F. de Almeida**
I3S Laboratory/CNRS
University of Nice, France
lima@i3s.unice.fr

**Romis R. F. Attux**
Lab. of Signal Processing for Communications (DSPCom)
FEEC - Unicamp, Brazil
romisri@decom.fee.unicamp.br

**Charles C. Cavalcante**
Wireless Telecommunications Research Group (GTEL)
Federal University of Ceará, Brazil
charles@gtel.ufc.br

**F. Rodrigo P. Cavalcanti**
Wireless Telecommunications Research Group (GTEL)
Federal University of Ceará, Brazil
rodrigo@gtel.ufc.br

**Fabiano de S. Chaves**
Lab. of Signal Processing for Communications (DSPCom)
FEEC - Unicamp, Brazil
fabiano@decom.fee.unicamp.br

**Mérouane Debbah**
Mobile Communications Group
Institut Eurecom, France
debbah@eurecom.fr

**Bernadette Dorizzi**
Équipe Intermédia
Institut National des Télécommunications (INT), France
Bernadette.Dorizzi@int-evry.fr

**Leonardo T. Duarte**
Lab. of Signal Processing for Communications (DSPCom)
FEEC - Unicamp, Brazil
ltduarte@decom.fee.unicamp.br

**Gérard Favier**
I3S Laboratory/CNRS
University of Nice, France
favier@i3s.unice.fr

**Rafael Ferrari**
Lab. of Signal Processing for Communications (DSPCom)
FEEC - Unicamp, Brazil
rferrari@decom.fee.unicamp.br

**Hugo E. Hernández-Figueroa**
Department of Microwave and Optics
University of Campinas, Brazil
hugo@dmo.fee.unicamp.br

**Cynthia Junqueira**
General-Command of Aerospace Technology (CTA/IAE), Brazil
cynthia@iae.cta.br

**M. Koshiba**
Information Communication Systems
Hokkaido University, Japan
koshiba@ist.hokudai.ac.jp

**João Cesar M. Mota**
Wireless Telecommunications Research Group (GTEL)
Federal University of Ceará, Brazil
mota@deti.ufc.br

**Aline de O. Neves**
Lab. of Signal Processing for Communications (DSPCom)
FEEC - Unicamp, Brazil
aline@decom.fee.unicamp.br

**Josef A. Nossek**
Institute for Circuit Theory and Signal Processing
Munich University of Technology (TUM), Germany
nossek@tum.de

**Raimundo A. de Oliveira Neto**
Wireless Telecommunications Research Group (GTEL)
Federal University of Ceará, Brazil
neto@gtel.ufc.br

**Moisés V. Ribeiro**
Electrical Circuit Department
Federal University of Juiz de Fora, Brazil
mribeiro@ieee.org

**Vitaly F. Rodríguez-Esquerre**
Departament of Electro-Electronics Technology (DTEE)
Federal Center of Technological
Education of Bahia (CEFET-BA), Brazil
vitaly@cefetba.br

**João Marcos T. Romano**
Lab. of Signal Processing for Communications
(DSPCom)
FEEC - Unicamp, Brazil
romano@decom.fee.unicamp.br

**Ricardo B. dos Santos**
Wireless Telecommunications Research Group (GTEL)
Federal University of Ceará, Brazil
brauner@gtel.ufc.br

**Ricardo Suyama**
Lab. of Signal Processing for Communications
(DSPCom)
FEEC - Unicamp, Brazil
rsuyama@decom.fee.unicamp.br

**Benno Zerlin**
Institute for Circuit Theory and Signal Processing
Munich University of Technology (TUM), Germany
zerlin@tum.de

# 1

# Information Theory and Wireless Channel Modeling

Mérouane Debbah[1]

Mobile Communications Group, Institut Eurecom, 2229 Route des Cretes, B.P. 193, 06904 Sophia Antipolis CEDEX, France
debbah@eurecom.fr

## 1.1 Introduction

The problem of modelling channels is crucial for the efficient design of wireless systems [1, 2, 3]. The wireless channel suffers from constructive/destructive interference signaling [4, 5]. This yields a randomized channel with certain statistics to be discovered. Recently ([6, 7]), the need to increase spectral efficiency has motivated the use of multiple antennas at both the transmitter and the receiver side. Hence, in the case of i.i.d Gaussian entries of the MIMO link and perfect channel knowledge at the receiver, it has been proved [8] that the ergodic capacity increase is min($n_r$,$n_t$) bits per second per hertz for every 3dB increase ($n_r$ is the number of receiving antennas and $n_t$ is the number of transmitting antennas) at high Signal to Noise Ratio (SNR)[1]. However, for realistic[2] channel models, results are still unknown and may seriously put into doubt the MIMO hype. As a matter of fact, the actual design of efficient codes is tributary of the channel model available: the transmitter has to know in what environment the transmission occurs in order to provide the codes with the adequate properties: as a typical example, in Rayleigh fading channels, when coding is performed, the Hamming distance (also known as the number of distinct components of the multi-dimensional constellation) plays a central role whereas maximizing the Euclidean distance is the commonly approved design criteria for Gaussian channels (see Giraud and Belfiore [9] or Boutros and Viterbo [10]).

As a consequence, channel modelling is the key in better understanding the limits of transmissions in wireless and noisy environments. In particular, questions of the form: "what is the highest transmission rate on a propagation environment where we only know the mean of each path, the variance of each path and the directions of arrival?" are crucially important. It will justify the use (or not) of MIMO technologies for a given state of knowledge.

Let us first introduce the modelling constraints. We assume that the transmission takes place between a mobile transmitter and receiver. The transmitter has $n_t$ antennas and the receiver has $n_r$ antennas. Moreover, we assume that the input transmitted signal goes through a time variant linear filter channel. Finally, we assume that the interfering noise is additive white Gaussian.

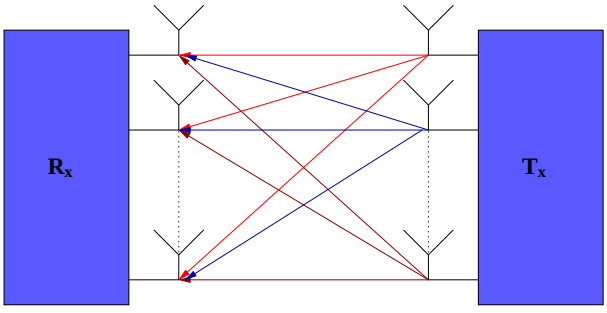



**Fig. 1.1.** MIMO channel representation.

The transmitted signal and received signal are related as:

[1] In the single antenna Additive White Gaussian Noise (AWGN) channel, 1 bit per second per hertz can be achieved with every 3dB increase at high SNR.

[2] By realistic, we mean models representing our state of knowledge of reality which might be different from reality.

$$\mathbf{y}(t) = \sqrt{\frac{\rho}{n_t}} \int \mathbf{H}_{n_r \times n_t}(\tau, t)\mathbf{x}(t-\tau)d\tau + \mathbf{n}(t) \tag{1.1}$$

with

$$\mathbf{H}_{n_r \times n_t}(\tau, t) = \int \mathbf{H}_{n_r \times n_t}(f, t)e^{j2\pi f\tau}df \tag{1.2}$$

$\rho$ is the received SNR (total transmit power per symbol versus total spectral density of the noise), $t$, $f$ and $\tau$ denote respectively time, frequency and delay, $\mathbf{y}(t)$ is the $n_r \times 1$ received vector, $\mathbf{x}(t)$ is the $n_t \times 1$ transmit vector, $\mathbf{n}(t)$ is an $n_r \times 1$ additive standardized white Gaussian noise vector.

In the rest of the paper, we will only be interested in the frequency domain modelling (knowing that the impulse response matrix can be accessed through an inverse Fourier transform according to relation 1.2). We would like to provide some theoretical grounds to model the frequency response matrix $\mathbf{H}(f, t)$ based on a given state of knowledge. In other words, knowing only certain things related to the channel (Directions of Arrival (DoA), Directions of Departure (DoD), bandwidth, center frequency, number of transmitting and receiving antennas, number of chairs in the room...), how to attribute a joint probability distribution to the entries $h_{ij}(f, t)$ of the matrix:

$$\mathbf{H}_{n_r \times n_t}(f, t) = \begin{pmatrix} h_{11}(f,t) & \ldots\ldots & h_{1n_t}(f,t) \\ \vdots & \ldots\ldots & \vdots \\ \vdots & \ldots\ldots & \vdots \\ h_{n_r1}(f,t) & \ldots\ldots & h_{n_rn_t}(f,t) \end{pmatrix} \tag{1.3}$$

This question can be answered in light of the Bayesian probability theory and the principle of maximum entropy. Bayesian probability theory has led to a profound theoretical understanding of various scientific areas [11, 12, 13, 14, 15, 16, 17, 18] and has shown the potential of entropy as a measure of our degree of knowledge when encountering a new problem. The principle of maximum entropy[3] is at present the clearest theoretical justification in conducting scientific inference: we do not need a model, entropy maximization creates a model for us out of the information available. Choosing the distribution with greatest entropy avoids the arbitrary introduction or assumption of information that is not available[4]. Bayesian probability theory improves on maximum entropy by expressing some prior knowledge on the model and estimating the parameters of the model.

As we will emphasize all along this paper, channel modelling is not a science representing reality but only our knowledge of reality as thoroughly stated by Jaynes in [20]. It answers in particular the following question: based on a given state of knowledge (usually brought by raw data or prior information), what is the best model one can make? This is, of course, a vague question since there is no strict definition of what is meant by best. But what do we mean then by best? In this contribution, our aim is to derive a model which is adequate with our state of knowledge. We need a measure of uncertainty which expresses the constraints of our knowledge and the desire to leave the unknown parameters to lie in an unconstrained space. To this end, many possibilities are offered to us to express our uncertainty. However, we need an information measure which is consistent (complying to certain common sense desiderata, see [21] to express these desiderata and for the derivation of entropy) and easy to manipulate: we need a general principle for translating information into probability assignment. Entropy is the measure of information that fulfills this criteria. Hence, already in 1980, Shore et al. [21] proved that the principle of maximum entropy is the correct method of inference when given new information in terms of expected values. They proved that maximizing entropy is correct in the following sense: maximizing any function but entropy will lead to inconsistencies unless that function and entropy have the same maximum[5]. The consistency argument is at the heart of scientific inference and can be expressed through the following axiom[6]:

[3] The principle of maximum entropy was first proposed by Jaynes [12, 13] as a general inference procedure although it was first used in Physics.

[4] Keynes named it judiciously the principle of indifference [19] to express our indifference in attributing prior values when no information is available.

[5] Thus, aiming for consistency, we can maximize entropy without loss of generality.

[6] The consistency property is only one of the required properties for any good calculus of plausibility statement. In fact, R.T Cox in 1946 derived three requirements known as Cox's Theorem[22]:

- Divisibility and comparability: the plausibility of of a statement is a real number between 0 (for false) and 1 (for true) and is dependent on information we have related to the statement.
- Common sense: Plausibilities should vary with the assessment of plausibilities in the model.
- Consistency: If the plausibility of a statement can be derived in two ways, the two results should be equal.

**Lemma 1.** *If the prior information* $\mathbf{I_1}$ *on which is based the channel model* $\mathbf{H_1}$ *can be equated to the prior information* $\mathbf{I_2}$ *of the channel model* $\mathbf{H_2}$ *then both models should be assigned the same probability distribution* $P(\mathbf{H}) = P(\mathbf{H_1}) = P(\mathbf{H_2})$.

Any other procedure would be inconsistent in the sense that, by changing indices 1 and 2, we could then generate a new problem in which our state of knowledge is the same but in which we are assigning different probabilities. More precisely, Shore et al. [21] formalize the maximum entropy approach based on four consistency axioms stated as follows[7]:

- Uniqueness: If one solves the same problem twice the same way then the same answer should result both times.
- Invariance: If one solves the same problem in two different coordinate systems then the same answer should result both times.
- System independence: It should not matter whether one accounts for independent information about independent systems separately in terms of different densities or together in terms of a joint density.
- Subset independence: It should not matter whether one treats an independent subset of system states in terms of a separate conditional density or in terms of the full system density.

These axioms are based on the fundamental principle that if a problem can be solved in more than one way, the results should be consistent. Given this statement in mind, the rules of probability theory should lead every person to the same unique solution, provided each person bases his model on the same information.[8]

Moreover, the success over the years of the maximum entropy approach (see Boltzmann's kinetic gas law, [23] for the estimate of a single stationary sinusoidal frequency, [14] for estimating the spectrum density of a stochastic process subject to autocorrelation constraints, [24] for estimating parameters in the context of image reconstruction and restoration problems, [25] for applying the maximum entropy principle on solar proton event peak fluxes in order to determine the least biased distribution) has shown that this information tool is the right way so far to express our uncertainty.

Let us give an example in the context of spectral estimation of the powerful feature of the maximum entropy approach which has inspired this paper. Suppose a stochastic process $x_i$ for which $p+1$ autocorrelation values are known i.e $\mathbb{E}(x_i x_{i+k}) = \tau_k, k = 0, ..., p$ for all $i$. What is the consistent model one can make of the stochastic process based only on that state of knowledge, in other words the model which makes the least assumption on the structure of the signal? The maximum entropy approach creates for us a model and shows that, based on the previous information, the stochastic process is a $p^{\text{th}}$ auto-regressive (AR) order model process of the form [14]:

$$x_i = -\sum_{k=1}^{p} a_k x_{i-k} + b_i$$

where the $b_i$ are i.i.d zero mean Gaussian distributed with variance $\sigma^2$ and $a_1, a_2, .., a_p$ are chosen to satisfy the autocorrelation constraints (through Yule-Walker equations).

In this contribution, we would like to provide guidelines for creating models from an information theoretic point of view and therefore make extensive use of the principle of maximum entropy together with the principle of consistency.

## 1.2 Some Considerations

### 1.2.1 Channel Modelling Methodology

In this contribution, we provide a methodology (already successfully used in Bayesian spectrum analysis [23, 17]) for inferring on channel models. The goal of the modelling methodology is twofold:

[7] In all the rest of the document, the consistency argument will be referred to as Axiom 1.

[8] It is noteworthy to say that if a prior distribution Q of the estimated distribution P is available in addition to the expected values constraints, then the principle of minimum cross-entropy (which generalizes maximum entropy) should be applied. The principle states that, of the distribution P that satisfy the constraints, one should choose the one which minimizes the functional:

$$D(P,Q) = \int P(x) \log\left(\frac{P(x)}{Q(x)}\right) dx$$

Minimizing cross-entropy is equivalent to maximizing entropy when the prior $Q$ is a uniform distribution. Intuitively, cross-entropy measures the amount of information necessary to change the prior $Q$ into the posterior $P$. If measured data is available, $Q$ can be estimated. However, one can only obtain a numerical form for $P$ in this case (which is not always useful for optimization purposes). Moreover, this is not a easy task for multidimensional vectors such as vec($\mathbf{H}$). As a consequence, we will always assume a uniform prior and use therefore the principle of maximum entropy.

- to define a set of rules, called hereafter *consistency axioms*, where only our state of knowledge needs to be defined.
- to use a measure of uncertainty, called hereafter *entropy*, in order to avoid the arbitrary introduction or assumption of information that is not available.

In other words, if two published papers make the same assumptions in the abstract (concrete buildings in Oslo where one avenue...), then both papers should provide the same channel model.

To achieve this goal, in all this document, the following procedure will be applied: every time we have some information on the environment (*and not make assumptions on the model*!), we will ask a question based on that the information and provide a model taking into account that information and nothing more! The resulting model and its compliance with later test measurements will justify whether the information used for modelling was adequate to characterize the environment in sufficient details. Hence, when asked the question, "what is the consistent model one can make knowing the directions of arrival, the number of scatterers, the fact that each path has zero mean and a given variance?" we will suppose that the information provided by this question is unquestionable and true i.e the propagation environment depends on fixed steering vectors, each path has effectively zero mean and a given variance. We will suppose that effectively, when waves propagate, they bounce onto scatterers and that the receiving antenna sees these ending scatterers through steering directions. Once we assume this information to be true, we will construct the model based on Bayesian tools.[9].

To explain this point of view, the author recalls an experiment made by his teacher during a tutorial explanation on the duality behavior of light: photon or wave. The teacher took two students of the class, called here A and B for

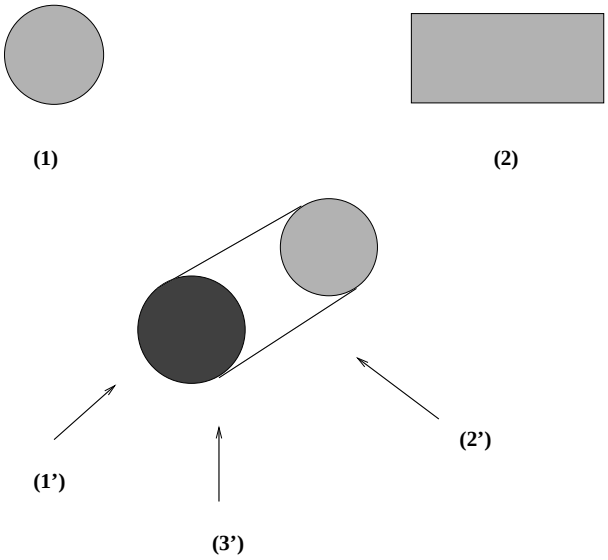



**Fig. 1.2.** Duality wave-corpuscule?

simplicity sake. To student A, he showed view (1') (see Figure 1.2) of a cylinder and to student B, he showed view (2') of the same cylinder. For A, the cylinder was a circle and for B, the cylinder was a rectangle. Who was wrong? Well, nobody. Based on the state of knowledge (1'), representing the cylinder as a circle is the best one can do. Any other representation of the cylinder would have been made on unjustified assumptions (the same applies to view (2')). Unless we have another state of knowledge (view (3')), the true nature of the object will not be found.

Our channel modelling will not pretend to seek reality but only to represent view (1') or view (2') in the most accurate way (i.e if view (1') is available then our approach should lead into representing the cylinder as a circle and not as a triangle for example). If the model fails to comply with measurements, we will not put into doubt the model but conclude that the information we had at hand to create the model was insufficient. We will take into account the failure as a new source of information and refine/change our question in order to derive a new model based on the principle of maximum entropy which complies with the measurements. This procedure will be routinely applied until the right question (and therefore the right answer) is found. When performing scientific inference, every question asked, whether right or wrong, is important. Mistakes are eagerly welcomed as they lead the path to better understand

[9] Note that in Bayesian inference, all probabilities are conditional on some hypothesis space (which is assumed to be true).

the propagation environment. Note that the approach devised here is not new and has already been used by Jaynes [20] and Jeffrey [26]. We give hereafter a summary of the modelling approach:

1. **Question selection**: the modeler asks a question based on the information available.
2. **Construct the model**: the modeler uses the principle of maximum entropy (with the constraints of the question asked) to construct the model $M_i$.
3. **Test**: (When complexity is not an issue) The modeler computes the a posteriori probability of the model and ranks the model.
4. **Return to 1.**: The outcome of the test is some "new information" evidence to keep/refine/change the question asked. Based on this information, the modeler can therefore make a new model selection.

This algorithm is iterated as many times as possible until better ranking is obtained. However, we have to alert the reader on one main point: the convergence of the previous algorithm is not at all proven. Does this mean that we have to reject the approach? we should not because our aim is to better understand the environment and by successive tests, we will discard some solutions and keep others.

We provide hereafter a brief historical example to highlight the methodology. In the context of spectrum estimation, the Schuster periodogram (also referred in the literature as the discrete Fourier transform power spectrum) is commonly used for the estimation of hidden frequencies in the data. The Schuster periodogram is defined as:

$$F(\omega) = \frac{1}{N} \mid \sum_{k=1}^{N} s_k e^{-j\omega t_k} \mid^2$$

$s_k$ is the data of length $N$ to be analyzed. In order to find the hidden frequencies in the data, the general procedure is to maximize $F(\omega)$ with respect to $\omega$ . But as in our case, one has to understand why/when to use the Schuster periodogram for frequency estimation. The Schuster periodogram answers a specific question based on a specific assumption (see the work of Bretthorst [17]). In fact, it answers the following question: "what is the optimal frequency estimator for a data set which contains a **single stationary sinusoidal frequency** in the presence of Gaussian white noise?" From the standpoint of Bayesian probability, the discrete Fourier Transform power spectrum answers a specific question about single (and not two or three....) stationary sinusoidal frequency estimation. Given this state of knowledge, the periodogram will consider everything in the data that cannot be fit to a single sinusoid to be noise and will therefore, if other frequencies are present, misestimate them. However, if the periodogram does not succeed in estimating multiple frequencies, the periodogram is not to blame but only the question asked! One has to devise a new model (a model maybe based on a two stationary sinusoidal frequencies?). This new model selection will lead to a new frequency estimator in order to take into account the structure of what was considered to be noise. This routine is repeated and each time, the models can be ranked to determine the right number of frequencies.

### 1.2.2 Information and Complexity

In the introduction, we have recalled the work of Shore et al. [21] which shows that maximizing entropy leads to consistent solutions. However, incorporating information in the entropy criteria which is not given in terms of expected values is not an easy task. In particular, how does one incorporate information on the fact that the room has four walls and two chairs? In this case, we will not maximize entropy based only on the information we have (expected values and number of chairs and walls): we will maximize entropy based on the expected values and a structured form of the channel matrix (which is more than the information we have since the chairs and walls are not constraint equations in the entropy criteria). This ad-hoc procedure will be used because it is extremely difficult to incorporate knowledge on physical considerations (number of chairs, type of room...) in the entropy criteria. Each time this ad-hoc procedure is used, we will verify that although we maximize entropy under a structured constraint, we remain consistent. Multiple iterations of this procedure will refine the structured form of the channel until the modeler obtains a consistent structured models that maximizes entropy.

A question the reader could ask is whether we should take into account all the information provided, in other words, what information is useful? We should of course consider all the available information but there is a compromise to be made in terms of model complexity. Each information added will not have the same effect on the channel model and might as well more complicate the model for nothing rather than bring useful insight on the behavior of the propagation environment. To assume further information by putting some additional structure would not lead to incorrect predictions: however, if the predictions achieved with or without the details are equivalent, then this

means that the details may exist but are irrelevant for the understanding of our model[10]. As a typical example, when conducting iterative decoding analysis [27], Gaussian models of priors are often sufficient to represent our information. Inferring on other moments and deriving the true probabilities will only complicate the results and not yield a better understanding.

## 1.3 Gaussian i.i.d Channel Model

### 1.3.1 Finite Energy Case

In this section, we give a precise justification on why and when the Gaussian i.i.d model should be used. We recall the general model:

$$\mathbf{y} = \sqrt{\frac{\rho}{n_t}}\mathbf{H}\mathbf{x} + \mathbf{n}$$

Imagine now that the modeler is in a situation where it has no measurements and no knowledge where the transmission took place. The only thing the modeler knows is that the channel carries some energy $E$, in other words, $\frac{1}{n_r n_t}\mathbb{E}\left(\sum_{i=1}^{n_r}\sum_{j=1}^{n_t} \mid h_{ij} \mid^2\right) = E$. Knowing only this information, the modeler is faced with the following question: what is the consistent model one can make knowing only the energy $E$ (but not the correlation even though it may exist) ? In other words, based on the fact that:

$$\int d\mathbf{H} \sum_{i=1}^{n_r}\sum_{j=1}^{n_t} \mid h_{ij} \mid^2 P(\mathbf{H}) = n_t n_r E \qquad \text{(Finite energy)} \tag{1.4}$$

$$\int dP(\mathbf{H}) = 1 \qquad (\text{P(H) is a probability distribution}) \tag{1.5}$$

What distribution $P(\mathbf{H})$[11] should the modeler assign to the channel? The modeler would like to derive the most general model complying with those constraints, in other words the one which maximizes our uncertainty while being certain of the energy. This statement can simply be expressed if one tries to maximize the following expression using Lagrange multipliers with respect to $P$:

$$L(P) = -\int d\mathbf{H} P(\mathbf{H}) \log P(\mathbf{H}) + \gamma \sum_{i=1}^{n_r}\sum_{j=1}^{n_t}[E - \int d\mathbf{H} \mid h_{ij} \mid^2 P(\mathbf{H})] + \beta\left[1 - \int d\mathbf{H} P(\mathbf{H})\right]$$

If we derive $L(P)$ with respect to $P$, we get:

$$\frac{dL(P)}{dP} = -1 - \log P(\mathbf{H}) - \gamma \sum_{i=1}^{n_r}\sum_{j=1}^{n_t} \mid h_{ij} \mid^2 - \beta = 0$$

then this yields:

$$\begin{aligned} P(\mathbf{H}) &= e^{-(\beta + \gamma \sum_{i=1}^{n_r}\sum_{j=1}^{n_t} |h_{ij}|^2 + 1)} \\ &= e^{-(\beta+1)} \prod_{i=1}^{n_r}\prod_{j=1}^{n_t} \exp -(\gamma \mid h_{ij} \mid^2) \\ &= \prod_{i=1}^{n_r}\prod_{j=1}^{n_t} P(h_{ij}) \end{aligned}$$

[10] Limiting one's information is a general procedure that can be applied to many other fields. As a matter of fact, the principle "one can know less but understand more" seems the only reasonable way to still conduct research considering the huge amount of papers published each year.

[11] It is important to note that we are concerned with $P(\mathbf{H} \mid I)$ where $I$ represents the general background knowledge (here the variance) used to formulate the problem. However, for simplicity sake, $P(\mathbf{H} \mid I)$ will be denoted $P(\mathbf{H})$.

with

$$P(h_{ij}) = e^{-(\gamma |h_{ij}|^2 + \frac{\beta+1}{n_r n_t})}.$$

One of the most important conclusions of the maximum entropy principle is that while we have only assumed the variance, these assumptions imply independent entries since the joint probability distribution $P(\mathbf{H})$ simplifies into products of $P(h_{ij})$. Therefore, based on the previous state of knowledge, the only maximizer of the entropy is the i.i.d one. This does not mean that we have supposed independence in the model. In the generalized $L(P)$ expression, there is no constraint on the independence. Another surprising result is that the distribution achieved is Gaussian. Once again, gaussianity is not an assumption but a consequence of the fact that the channel has finite energy. The previous distribution is the least informative probability density function that is consistent with the previous state of knowledge. When only the variance of the channel paths are known (but not the frequency bandwidth, nor knowledge of how waves propagate, nor the fact that scatterers exist...) then the only consistent model one can make is the Gaussian i.i.d model.
In order to fully derive $P(\mathbf{H})$, we need to calculate the coefficients $\beta$ and $\gamma$. The coefficients are solutions of the following constraint equations:

$$\int d\mathbf{H} \sum_{i=1}^{n_r} \sum_{j=1}^{n_t} \mid h_{ij} \mid^2 P(\mathbf{H}) = n_t n_r E$$

$$\int d\mathbf{H} P(\mathbf{H}) = 1$$

Solving the previous equations yields the following probability distribution:

$$P(\mathbf{H}) = \frac{1}{(\pi E)^{n_r n_t}} \exp\{-\sum_{i=1}^{n_r} \sum_{j=1}^{n_t} \frac{\mid h_{ij} \mid^2}{E}\}$$

Of course, if one has any additional knowledge, then this information should be integrated in the $L(P)$ criteria and would lead to a different result.
As a typical example, suppose that the modeler knows that the frequency paths have different variances such as $\mathbb{E}(\mid h_{ij} \mid^2) = E_{ij}$. Using the same methodology, it can be shown that :

$$P(\mathbf{H}) = \prod_{i=1}^{n_r} \prod_{j=1}^{n_t} P(h_{ij})$$

with $P(h_{ij}) = \frac{1}{\pi E_{ij}} e^{-\frac{|h_{ij}|^2}{E_{ij}}}$. The principle of maximum entropy still attributes independent Gaussian entries to the channel matrix but with different variances.
Suppose now that the modeler knows that the path $h_{pk}$ has a mean equal to $\mathbb{E}(h_{pk}) = m_{pk}$ and variance $\mathbb{E}(\mid h_{pk} - m_{pk} \mid^2) = E_{pk}$, all the other paths having different variances (but nothing is said about the mean). Using as before the same methodology, we show that:

$$P(\mathbf{H}) = \prod_{i=1}^{n_r} \prod_{j=1}^{n_t} P(h_{ij})$$

with for all $\{i, j, (i,j) \neq (p,k)\}$ $P(h_{ij}) = \frac{1}{\pi E_{ij}} e^{-\frac{|h_{ij}|^2}{E_{ij}}}$ and $P(h_{pk}) = \frac{1}{\pi E_{pk}} e^{-\frac{|h_{pk} - m_{pk}|^2}{E_{pk}}}$. Once again, different but still independent Gaussian distributions are attributed to the MIMO channel matrix.

The previous examples can be extended and applied whenever a modeler has some new source of information **in terms of expected values** on the propagation environment[12]. In the general case, if $N$ constraints are given on the expected values of certain functions $\int g_i(\mathbf{H}) P(\mathbf{H}) d\mathbf{H} = \alpha_i$ for $i = 1...N$, then the principle of maximum entropy attributes the following distribution [28]:

$$P(\mathbf{H}) = e^{-(1 + \lambda + \sum_{i=1}^{N} \lambda_i g_i(\mathbf{H}))}$$

where the values of $\lambda$ and $\lambda_i$ (for $i = 1..N$) can be obtained by solving the constraint equations.

[12] The case where information is not given in terms of expected values is treated afterwards.

Although these conclusions are widely known in the Bayesian community, the author is surprised that many MIMO channel papers begin with: "let us assume a $n_r \times n_t$ matrix with Gaussian i.i.d entries...". No assumptions on the model should be made. Only the state of knowledge should be clearly stated at the beginning of each paper and the conclusion of the maximum entropy approach can be straightforwardly used.[13]

As a matter of fact, the Gaussian i.i.d model should not be "thrown" away but be extensively used whenever our information on the propagation conditions is scarce (we don't know in what environment we are transmitting our signal i.e the frequency, the bandwidth, WLAN scenario, we do not know what performance measure we target...)[14].

### 1.3.2 Finite Energy unknown

We will consider a case similar to the previous section where the modeler is in a situation where it has no measurements and no knowledge where the transmission took place. The modeler does know that the channel carries some energy E but is not aware of its value.

In the case where the modeler knows the value of $E$, we have shown that:

$$P(\mathbf{H} \mid E) = \frac{1}{(\pi E)^{n_r n_t}} \exp\{-\sum_{i=1}^{n_r}\sum_{j=1}^{n_t} \frac{\mid h_{ij} \mid^2}{E}\}$$

In general, when $E$ is unknown, the probability distribution is derived according to:

$$\begin{aligned} P(\mathbf{H}) &= \int P(\mathbf{H}, E) dE \\ &= \int P(\mathbf{H} \mid E) P(E) dE \end{aligned}$$

and is consistent with the case where $E$ is known i.e $P(E) = \delta(E - E_0)$:

$$P(\mathbf{H}) = \frac{1}{(\pi E_0)^{n_r n_t}} \exp\{-\sum_{i=1}^{n_r}\sum_{j=1}^{n_t} \frac{\mid h_{ij} \mid^2}{E_0}\}$$

In the case were the energy $E$ is unknown, one has to determine $P(E)$. $E$ is a positive variance parameter and the channel can not carry more energy than what is transmitted (i.e $E \leq E_{\max}$) . This is merely the sole knowledge the modeler has about $E$ on which the modeler has to derive a prior distribution[15].

In this case, using maximum entropy arguments, one can derive $P(E)$:

$$P(E) = \frac{1}{E_{\max}} \qquad\qquad 0 \leq E \leq E_{\max}$$

As a consequence,

$$P(\mathbf{H}) = \int_0^{E_{\max}} \frac{1}{(\pi E)^{n_r n_t}} \exp\{-\sum_{i=1}^{n_r}\sum_{j=1}^{n_t} \frac{\mid h_{ij} \mid^2}{E}\} dE$$

With the change of variables $u = \frac{1}{E}$, we obtain:

$$P(\mathbf{H}) = \frac{1}{E_{\max} \pi^{n_r n_t}} \int_{\frac{1}{E_{\max}}}^{\infty} u^{n_r n_t - 2} e^{-\sum_{i=1}^{n_r}\sum_{j=1}^{n_t} |h_{ij}|^2 u} du$$

[13] "Normality is not an assumption of physical fact at all. It is a valid description of our state of information", Jaynes.

[14] In "The Role of Entropy in Wave Propagation" [29], Franceschetti et al. show that the probability laws that describe electromagnetic magnetic waves are simply maximum entropy distributions with appropriate moment constraints. They suggest that in the case of dense lattices, where the inter-obstacle hitting distance is small compared to the distance traveled, the relevant metric is non-Euclidean whereas in sparse lattices, the relevant metric becomes Euclidean as propagation is not constrained along the axis directions.

[15] Jeffrey [26] already in 1939 proposed a way to handle this issue based on invariance properties and consistency axioms. He suggested that a proper way to express incomplete ignorance of a continuous variable known to be positive is to assign uniform prior probability to its logarithm, in other words: $P(E) \propto \frac{1}{E}$. However, the distribution is improper and one can not therefore marginalize with this distribution.

Note that the distribution is invariant by unitary transformations, is not Gaussian and moreover the entries are not independent when the modeler has no knowledge on the amount of energy carried by the channel. This point is critical and shows the effect of the lack of information on the exact energy[16].

In the case $n_t = 1$ and $n_r = 2$, we obtain:

$$P(\mathbf{H}) = \frac{1}{E_{\max}\pi^2 \sum_{i=1}^{2} \mid h_{i1} \mid^2} e^{-\frac{\sum_{i=1}^{2} |h_{i1}|^2}{E_{\max}}}$$

### 1.3.3 Correlation matrix unknown

Suppose now that the modeler knows that correlation exists between the entries of the channel matrix $\mathbf{H}$ but is not aware of the value of the correlation matrix $\mathbf{Q} = \mathbb{E}(\mathrm{vec}(\mathbf{H})\mathrm{vec}(\mathbf{H})^H)$. What consistent distribution should the modeler attribute to the channel based only on that knowledge?

To answer this question, suppose that the correlation matrix $\mathbf{Q} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^H$ is known ($\mathbf{V} = [\mathbf{v}_1, ...\mathbf{v}_{n_r n_t}]$ is a $n_r n_t \times n_r n_t$ unitary matrix whereas $\mathbf{\Lambda}$ is a $n_r n_t \times n_r n_t$ diagonal matrix $\mathbf{\Lambda} = \mathrm{diag}(\lambda_1, ..., \lambda_{n_r n_t})$ with $\lambda_i \geq 0$ for $1 \leq i \leq n_r n_t$).

Using the maximum entropy principle, one can easily show that:

$$P(\mathbf{H} \mid \mathbf{V}, \mathbf{\Lambda}) = \frac{1}{\prod_{i=1}^{n_r n_t} \pi\lambda_i} \exp\{\sum_{i=1}^{n_r n_t} \frac{\mid \mathbf{v_i}^H \mathrm{vec}(\mathbf{H}) \mid^2}{\lambda_i}\}$$

The channel distribution can be obtained:

$$\begin{aligned} P(\mathbf{H}) &= \int P(\mathbf{H}, \mathbf{V}, \mathbf{\Lambda}) d\mathbf{V} d\mathbf{\Lambda} \\ &= \int P(\mathbf{H} \mid \mathbf{V}, \mathbf{\Lambda}) P(\mathbf{V}, \mathbf{\Lambda}) d\mathbf{V} d\mathbf{\Lambda} \end{aligned}$$

If the correlation matrix is perfectly known, then $P(\mathbf{V}, \mathbf{\Lambda}) = \delta(\mathbf{V} - \mathbf{V}^0)\delta(\mathbf{\Lambda} - \mathbf{\Lambda}^0)$ and

$$P(\mathbf{H}) = \frac{1}{\prod_{i=1}^{n_r n_t} \pi {\lambda^0}_i} \exp\{\sum_{i=1}^{n_r n_t} \frac{\mid \mathbf{v^0_i}^H \mathrm{vec}(\mathbf{H}) \mid^2}{{\lambda^0}_i}\}$$

In the case were the correlation matrix $\mathbf{Q}$ is unknown, one has to determine $P(\mathbf{V}, \mathbf{\Lambda}) = P(\mathbf{\Lambda} \mid \mathbf{V})P(\mathbf{V})$. This is the problem of constructing an ignorance prior corresponding to ignorance of both scale (up to some constraints proper to our problem) and rotation. The a priori distribution can be derived as well as the joint probability distribution using tools from statistical physics. Due to limited space, the result is not provided but can be found in the recent work of the author [30].

## 1.4 Knowledge of the directions of arrival, departure, delay, bandwidth, power: frequency selective channel model with time variance

### 1.4.1 Knowledge of the directions of arrival or departure

The modeler[17] is interested in modelling the channel over time scales over which the locations of scatterers do not not change significantly relative to the transmitter or receiver. This is equivalent to considering time scales over which the channel statistics do not change significantly. However, the channel realizations do vary over such time scales.

[16] In general, closed form solutions of the distributions do not exist. In this case, a powerful tool for approximate Bayesian inference that uses Markov Chain Monte Carlo to compute marginal posterior distributions of interest can be used through WinBUGS (http://www.mrc-bsu.cam.ac.uk/bugs/welcome.shtml.)

[17] We treat in this section thoroughly the directions of arrival model and show how the directions of departure model can be easily obtained from the latter case.

Imagine that the modeler is in a situation where it knows the energy carried by the channel (nothing is known about the mean)[18]. Moreover, the modeller knows from electromagnetic theory that when a wave propagates from a scatterer to the receiving antennas, the signal can be written in an exponential form

$$\mathbf{s}(t,\mathbf{d}) = \mathbf{s_0}\, e^{j(\mathbf{k}^T\mathbf{d}-2\pi ft)} \tag{1.6}$$

which is the plane wave solution of the Maxwell equations in free non-dispersive space for wave vector $\mathbf{k} \in \Re^{2\times 1}$ and location vector $\mathbf{d} \in \Re^{2\times 1}$. The reader must note that other solutions to the Maxwell equations exist and therefore the modeler is making an important restriction. The direction of the vector $\mathbf{s_0}$ gives us knowledge on the polarization of the wave while the direction of the wave vector $\mathbf{k}$ gives us knowledge on the direction of propagation. The phase of the signal results in $\phi = \mathbf{k}^T\mathbf{d}$. The modeler considers for simplicity sake that the scatterers and the antennas lie in the same plane. The modeler makes use of the knowledge that the steering vector is known up to a multiplicative complex constant that is the same for all antennas.

Although correlation might exist between the scatterers, the modeler is not aware of such a thing. Based on this state of knowledge, the modeler wants to derive a model which takes into account all the previous constraints while leaving as many degrees of freedom as possible to the other parameters (since the modeler does not want to introduce unjustified information). In other words, based on the fact that:

$$\mathbf{H} = \frac{1}{\sqrt{s_r}} \begin{pmatrix} e^{j\phi_{1,1}} & \dots & e^{j\phi_{1,s_r}} \\ \vdots & \ddots & \vdots \\ e^{j\phi_{n_r,1}} & \dots & e^{j\phi_{n_r,s_r}} \end{pmatrix} \mathbf{\Theta}_{s_r\times n_t}$$

what distribution should the modeler attribute to $\mathbf{\Theta}_{s_r\times n_t}$? $\mathbf{H}$ is equal to $\frac{1}{\sqrt{s_r}}\mathbf{\Phi\Theta}$, $\phi_{i,j} = \mathbf{k}.\mathbf{r}_{i,j}$ and $\mathbf{r}_{i,j}$ is the distance between the receiving antenna $i$ and receiving scatterer $j$ and $\mathbf{\Phi}$is a $n_r \times s_r$ matrix ($s_r$ is the number of scatterers) which represents the directions of arrival from randomly positioned scatterers to the receiving antennas. $\mathbf{\Theta}_{s_r\times n_t}$ is an $s_r \times n_t$ matrix which represents the scattering environment between the transmitting antennas and the scatterers (see Figure 1.3).

The consistency argument (see Proposition 1) states that if the DoA (Directions of Arrival) are unknown then $\mathbf{H} = \frac{1}{\sqrt{s_r}}\mathbf{\Phi}_{n_r\times s_r}\mathbf{\Theta}_{s_r\times n_t}$ should be assigned an i.i.d Gaussian distribution since the modeler is in the same state of knowledge as before where it only knew the variance.

Based on the previous remarks, let us now derive the distribution of $\mathbf{\Theta}_{s_r\times n_t}$. The probability distribution $P(\mathbf{H})$ is given by:

$$P(\mathbf{H}) = \int P(\mathbf{\Phi\Theta} \mid \mathbf{\Phi}, s_r) P(\mathbf{\Phi} \mid s_r) P(s_r) ds_r d\mathbf{\Phi}$$

- When $\mathbf{\Phi}$ and $s_r$ are known, then $P(\mathbf{\Phi} \mid s_r) = \delta(\mathbf{\Phi} - \mathbf{\Phi}^0)$ and $P(s_r) = \delta(s_r - {s_r}^0)$. Therefore $P(\mathbf{H}) = P(\mathbf{\Phi}^0\mathbf{\Theta})$.
- When $\mathbf{\Phi}$ and $s_r$ are unknown: the probability distribution of the frequency path $h_{ij}$ is:

$$P(h_{ij}) = \int P\left(h_{ij} \mid \mathbf{\Phi}, s_r\right) P(\mathbf{\Phi} \mid s_r) P(s_r) d\mathbf{\Phi} ds_r \tag{1.7}$$

In the case when $P(\mathbf{\Phi} \mid s_r)$ and $P(s_r)$ are unknown, the consistency argument states that:

- The $\mathbf{\Theta}_{s_r\times n_t}$ matrix is such as each $h_{ij}$ is zero mean Gaussian.
- The $\mathbf{\Theta}_{s_r\times n_t}$ matrix is such as $\mathbb{E}(h_{ij}{h_{mn}}^*) = \delta_{im}\delta_{jn}$ (since $h_{ij}$ is Gaussian, decorrelation is equivalent to independence).

In this case, the following result holds:

**Proposition 1.** *$\mathbf{\Theta}_{s_r\times n_t}$ i.i.d. zero mean Gaussian with unit variance is solution of the consistency argument and maximizes entropy.*

**Proof:** Since $\mathbf{\Phi}$ is unknown, the principle of maximum entropy attributes independent uniformly distributed angles to each entry $\phi_{ij}$:

$$P(\phi_{ij}) = \frac{1}{2\pi}\mathbf{1}_{[0,2\pi]}.$$

[18] The case where the paths have different non-zero means can be treated the same way.

Let us show that $\mathbf{\Theta}_{s_r \times n_t}$ i.i.d zero mean with variance 1 is solution of the consistency argument.

Since $h_{ij} = \frac{1}{\sqrt{s_r}} \sum_{k=1}^{s_r} \theta_{kj} e^{j\phi_{ik}}$ then $P(h_{ij} \mid \mathbf{\Phi}, s_r) = N(0, \frac{1}{s_r} \sum_{k=1}^{s_r} \mid e^{j\phi_{ik}} \mid^2 = 1) = \frac{1}{\sqrt{2\pi}} e^{-\frac{|h_{ij}|^2}{2}}$ and therefore $h_{ij}$ is zero mean Gaussian since:

$$\begin{aligned} P(h_{ij}) &= \int P\left(h_{ij} \mid \mathbf{\Phi}, s_r\right) P(\mathbf{\Phi} \mid s_r) P(s_r) d\mathbf{\Phi} ds_r \\ &= \int \frac{1}{\sqrt{2\pi}} e^{-\frac{|h_{ij}|^2}{2}} P(\mathbf{\Phi} \mid s_r) P(s_r) d\mathbf{\Phi} ds_r \\ &= \frac{1}{\sqrt{2\pi}} e^{-\frac{|h_{ij}|^2}{2}} \int P(\mathbf{\Phi} \mid s_r) P(s_r) d\mathbf{\Phi} ds_r \\ &= \frac{1}{\sqrt{2\pi}} e^{-\frac{|h_{ij}|^2}{2}} \end{aligned}$$

Moreover, we have:

$$\begin{aligned} \mathbb{E}(h_{ij} h_{mn}^*) &= \mathbb{E}_{\mathbf{\Theta},\mathbf{\Phi}} \left( \frac{1}{\sqrt{s_r}} \sum_{k=1}^{s_r} \theta_{kj} e^{j\phi_{ik}} \frac{1}{\sqrt{s_r}} \sum_{l=1}^{s_r} \theta_{ln}^* e^{-j\phi_{ml}} \right) \\ &= \frac{1}{s_r} \sum_{k=1}^{s_r} \sum_{l=1}^{s_r} \mathbb{E}_{\mathbf{\Theta}} (\theta_{kj} \theta_{ln}^*) \mathbb{E}_{\mathbf{\Phi}} (e^{j\phi_{ik} - j\phi_{ml}}) \\ &= \frac{1}{s_r} \sum_{k=1}^{s_r} \sum_{l=1}^{s_r} \delta_{kl} \delta_{jn} \mathbb{E}_{\mathbf{\Phi}} (e^{j\phi_{ik} - j\phi_{ml}}) \\ &= \delta_{jn} \frac{1}{s_r} \sum_{k=1}^{s_r} \mathbb{E}_{\mathbf{\Phi}} (e^{j\phi_{ik} - j\phi_{mk}}) \\ &= \delta_{jn} \delta_{im} \end{aligned}$$

which proves that $\mathbf{H}$ is i.i.d Gaussian for unknown angles.

One interesting point of the maximum entropy approach is that while we have not assumed uncorrelated scattering, the above methodology will automatically assign a model with uncorrelated scatterers in order to have as many degrees of freedom as possible. But this does not mean that correlation is not taken into account. The model in fact leaves free degrees for correlation to exist or not. The maximum entropy approach is appealing in the sense that if correlated scattering is given as a prior knowledge, then it can be immediately integrated in the channel modelling approach (as a constraint on the covariance matrix for example). Note also that in this model, the entries of $\mathbf{H}$ are correlated for general DoA's.

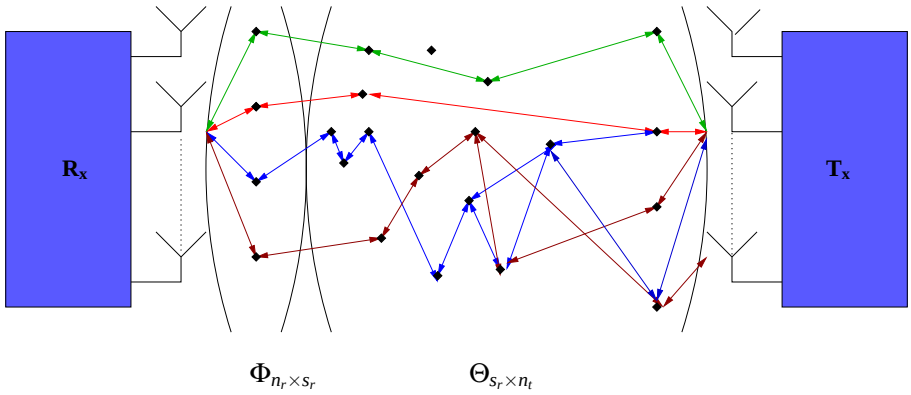



**Fig. 1.3.** Directions of arrival based model.

Suppose now that the modeler assumes that the different steering vectors have different amplitudes $\sqrt{P_i^T}$. What distribution should the modeler attribute to the matrix $\Theta_{s_r \times n_t}$ in the following representation:

$$\mathbf{H} = \frac{1}{\sqrt{s_r}} \begin{pmatrix} e^{j\phi_{1,1}} & \dots & e^{j\phi_{1,s_r}} \\ \vdots & \ddots & \vdots \\ e^{j\phi_{n_r,1}} & \dots & e^{j\phi_{n_r,s_r}} \end{pmatrix} \begin{pmatrix} \sqrt{P_1^r} & 0 & \dots \\ 0 & \ddots & 0 \\ \vdots & 0 & \sqrt{P_{s_r}^r} \end{pmatrix} \mathbf{\Theta}_{s_r \times n_t} ?$$

**Proposition 2.** *$\mathbf{\Theta}_{s_r \times n_t}$ i.i.d Gaussian with variance 1 is solution of the consistency argument and maximizes entropy*

**Proof:** We will not go into the details as the proof is a particular case of the proof of Proposition 3.

### 1.4.2 Knowledge of the Directions of Arrival and Departure

The modeler is now interested in deriving a consistent double directional model i.e taking into account simultaneously the directions of arrival and the directions of departure. The motivation of such an approach lies in the fact that when a single bounce on a scatterer occurs, the direction of arrival and departure are deterministically related by Descartes laws and therefore the distribution of the channel matrix depends on the joint DoA-DoD spectrum. The modeler assumes as a state of knowledge the directions of departure from the transmitting antennas to the set of transmitting scatterers $(1 \dots s_t)$. The modeler also assumes as a state of knowledge the directions of arrival from the set of receiving scatterers $(1 \dots s_r)$ to the receiving antennas. The modeler also has some knowledge that the steering directions have different powers. However, the modeler has no knowledge of what happens in between. The set $(1 \dots s_t)$ and $(1 \dots s_r)$ may be equal, $(1 \dots s_t)$ may be included in $(1 \dots s_r)$ or there may be no relation between the two. The modeler also knows that the channel carries some energy. Based on this state of knowledge, what is the consistent model the modeler can make of $\mathbf{H}$

$$\mathbf{H} = \frac{1}{\sqrt{s_r s_t}} \begin{pmatrix} e^{j\phi_{1,1}} & \dots & e^{j\phi_{1,s_r}} \\ \vdots & \ddots & \vdots \\ e^{j\phi_{n_r,1}} & \dots & e^{j\phi_{n_r,s_r}} \end{pmatrix} \begin{pmatrix} \sqrt{P_1^r} & 0 & \dots \\ 0 & \ddots & 0 \\ \vdots & 0 & \sqrt{P_{s_r}^r} \end{pmatrix}$$
$$\mathbf{\Theta}_{s_r \times s_t} \begin{pmatrix} \sqrt{P_1^t} & 0 & \dots \\ 0 & \ddots & 0 \\ \vdots & 0 & \sqrt{P_{s_t}^t} \end{pmatrix} \begin{pmatrix} e^{j\psi_{1,1}} & \dots & e^{j\psi_{1,n_t}} \\ \vdots & \ddots & \vdots \\ e^{j\psi_{s_t,1}} & \dots & e^{j\psi_{s_t,n_t}} \end{pmatrix} ?$$

In other words, how to model $\mathbf{\Theta}_{s_r \times s_t}$? As previously stated, the modeler must comply with the following constraints:

- The channel has a certain energy.
- Consistency argument: If the DoD and DoA are unknown then $\frac{1}{\sqrt{s_r s_t}} \mathbf{\Phi}_{n_r \times s_r} \mathbf{P}^{\mathbf{r}\frac{1}{2}} \mathbf{\Theta}_{s_r \times s_t} \mathbf{P}^{\mathbf{t}\frac{1}{2}} \mathbf{\Psi}_{s_t \times n_t}$ should be assigned an i.i.d zero mean Gaussian distribution.

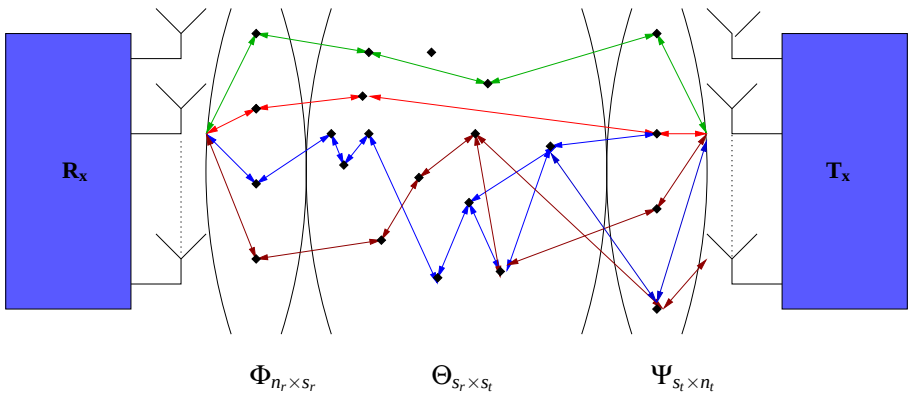



**Fig. 1.4.** Double directional based model.

Let us now determine the distribution of $\mathbf{\Theta}_{s_r \times s_t}$. The probability distribution of $P(\mathbf{H})$ is given by:

$$P(\mathbf{H}) = \int P(\mathbf{\Phi}\mathbf{P^r}^{\frac{1}{2}}\mathbf{\Theta}\mathbf{P^t}^{\frac{1}{2}}\mathbf{\Psi} \mid \mathbf{\Phi}, \mathbf{\Psi}, \mathbf{P^r}, \mathbf{P^t}, s_r, s_t) \\ P(\mathbf{\Psi}, \mathbf{\Phi} \mid s_r, s_t) P(\mathbf{P^r}, \mathbf{P^t} \mid s_t, s_r) \\ P(s_t, s_r) ds_r ds_t d\mathbf{P^r} d\mathbf{P^t} d\mathbf{\Psi} d\mathbf{\Phi}$$

- When $\mathbf{\Psi}, \mathbf{\Phi}, s_r, s_t, \mathbf{P^r}, \mathbf{P^t}$ are known: $P(\mathbf{\Phi\Psi} \mid s_r, s_t) = \delta(\mathbf{\Phi} - \mathbf{\Phi}^0)\delta(\mathbf{\Psi} - \mathbf{\Psi}^0)$, $P(s_t, s_r) = \delta(s_r - {s^0}_r)\delta(s_t - {s^0}_t)$,$P(\mathbf{P^r}, \mathbf{P^t} \mid s_r, s_t) = \delta(\mathbf{P^r} - \mathbf{P^{0}}^r)\delta(\mathbf{P^t} - \mathbf{P^{0}}^t)$ and

$$P(\mathbf{H}) = P(\mathbf{\Phi}^0 \mathbf{P^{0}}^{r\,\frac{1}{2}} \mathbf{\Theta} \mathbf{P^{0}}^{t\,\frac{1}{2}} \mathbf{\Psi}^0)$$

- Suppose now that $\mathbf{\Psi}, \mathbf{\Phi}, s_r, s_t$ are unknown, then each entry $h_{ij}$ of $\mathbf{H}$ must have an i.i.d zero mean Gaussian distribution. In this case, the following result holds:

  **Proposition 3.** *$\mathbf{\Theta}_{s_r \times s_t}$ i.i.d zero mean Gaussian with variance 1 is solution of the consistency argument and maximizes entropy.*

  **Proof**: Let us show that $\mathbf{\Theta}_{s_r \times s_t}$ i.i.d zero mean Gaussian with variance 1 is solution of the consistency argument and maximizes entropy. Since $\mathbf{\Phi}$ and $\mathbf{\Psi}$ are unknown, the principle of maximum entropy attributes i.i.d uniform distributed angles over $2\pi$ to the entries $\phi_{ij}$ and $\psi_{ij}$. In this case, if one chooses $\theta_{p,k}$ to be i.i.d zero mean Gaussian with variance 1 and knowing that $h_{ij} = \frac{1}{\sqrt{s_t s_r}} \sum_{k=1}^{s_t} \sum_{p=1}^{s_r} \theta_{pk} \sqrt{{P_k}^t} \sqrt{{P_p}^r} e^{j\psi_{kj}} e^{j\phi_{ip}}$, then: $P\left(h_{ij} \mid \mathbf{\Psi}, \mathbf{\Phi}, s_r, s_t\right) = N(0, \frac{1}{s_t s_r} \sum_{p=1}^{s_r} \sum_{k=1}^{s_t} \mid \sqrt{{P_p}^r} e^{j\phi_{ip}} \sqrt{{P_k}^t} e^{j\psi_{kj}} \mid^2 = 1) = \frac{1}{\sqrt{2\pi}} e^{-\frac{|h_{ij}|^2}{2}}$ (since $\frac{1}{s_r} \sum_{k=1}^{s_r} {P_k}^t = 1$ and $\frac{1}{s_t} \sum_{p=1}^{s_t} {P_p}^t = 1$ (due to power normalization as we assume the energy known). Therefore

$$\begin{aligned} P(h_{ij}) &= \int \frac{1}{\sqrt{2\pi}} e^{-\frac{|h_{ij}|^2}{2}} P(\mathbf{\Phi}, \mathbf{\Psi} \mid s_t, s_r) P(\mathbf{P^r}, \mathbf{P^t} \mid s_t, s_r) P(s_t, s_r) d\mathbf{\Phi} d\mathbf{\Psi} \\ &\quad d\mathbf{P^r} d\mathbf{P^t} ds_t ds_r \\ &= \frac{1}{\sqrt{2\pi}} e^{-\frac{|h_{ij}|^2}{2}} \int P(\mathbf{\Phi}, \mathbf{\Psi} \mid s_t, s_r) P(\mathbf{P^r}, \mathbf{P^t} \mid s_t, s_r) P(s_t, s_r) d\mathbf{\Phi} d\mathbf{\Psi} d\mathbf{P^r} d\mathbf{P^t} ds_t ds_r \\ &= \frac{1}{\sqrt{2\pi}} e^{-\frac{|h_{ij}|^2}{2}} \end{aligned}$$

  Moreover, we have :

$$\begin{aligned} \mathbb{E}_{\mathbf{\Phi},\mathbf{\Psi},\mathbf{\Theta}}(h_{ij} h^*_{mn}) &= \frac{1}{s_t s_r} \sum_{k=1}^{s_t} \sum_{p=1}^{s_r} \sum_{r=1}^{s_t} \sum_{l=1}^{s_r} \mathbb{E}_{\mathbf{\Theta}}(\theta_{pk} \theta^*_{lr}) \mathbb{E}_{\Psi}(e^{-j\psi_{rn} + j\psi_{kj}}) \mathbb{E}_{\Phi}(e^{-j\phi_{ml} + j\phi_{ip}}) \\ &\quad \sqrt{{P_k}^t} \sqrt{{P_r}^t} \sqrt{{P_p}^r} \sqrt{{P_l}^r} \\ &= \frac{1}{s_t s_r} \sum_{k=1}^{s_t} \sum_{p=1}^{s_r} \sum_{r=1}^{s_t} \sum_{l=1}^{s_r} \delta_{pl} \delta_{kr} \mathbb{E}_{\Psi}(e^{-j\psi_{rn} + j\psi_{kj}}) \mathbb{E}_{\Phi}(e^{-j\phi_{ml} + j\phi_{ip}}) \\ &\quad \sqrt{{P_k}^t} \sqrt{{P_r}^t} \sqrt{{P_p}^r} \sqrt{{P_l}^r} \\ &= \frac{1}{s_t s_r} \sum_{k=1}^{s_t} \sum_{p=1}^{s_r} \mathbb{E}_{\Psi}(e^{-j\psi_{kn} + j\psi_{kj}}) \mathbb{E}_{\Phi}(e^{-j\phi_{mp} + j\phi_{ip}}) {P_k}^t {P_p}^r \\ &= \delta_{im} \delta_{jn} \frac{1}{s_t s_r} \sum_{k=1}^{s_t} \sum_{p=1}^{s_r} {P_k}^t {P_p}^r \\ &= \delta_{im} \delta_{jn} \end{aligned}$$

which proves that $\mathbf{\Theta}_{s_r \times s_t}$ is solution of the consistency argument. Once again, instead of saying that this model represents a rich scattering environment, it should be more correct to say that the model makes allowance for every case that could be present to happen since we have imposed no constraints besides the energy.

### 1.4.3 Considering more features

The modeler wants to derive a consistent model taking into account the direction of arrivals and respective power profile, directions of departure and respective power profile, delay, Doppler effect. As a starting point, the modeler assumes that the position of the transmitter and receiver changes in time. However, the scattering environment (the buildings, trees,...) does not change and stays in the same position during the transmission. Let $\mathbf{v_t}$ and $\mathbf{v_r}$ be respectively the vector speed of the transmitter and the receiver with respect to a terrestrial reference (see Figure 1.5). Let $\mathbf{s}_{ij}^t$ be the signal between the transmitting antenna $i$ and the first scatterer $j$. Assuming that the signal can be written in an exponential form (plane wave solution of the Maxwell equations) then:

$$\begin{aligned}\mathbf{s}_{ij}^t(t) &= \mathbf{s_0} e^{j(\mathbf{k}_{ij}^t(\mathbf{v_t}t+\mathbf{d_{ij}})+2\pi f_c t)} \\ &= \mathbf{s_0} e^{j2\pi(\frac{f_c \mathbf{u}_{ij}^t \mathbf{v_t}}{c}t+f_c t)} e^{j\psi_{ij}}\end{aligned}$$

Here, $f_c$ is the carrier frequency, $\mathbf{d}_{ij}$ is the initial vector distance between antenna $i$ and scatterer $j$ ($\psi_{ij} = \mathbf{k}_{ij}^t.\mathbf{d}_{ij}$ is the scalar product between vector $\mathbf{k}_{ij}^t$ and vector $\mathbf{d}_{ij}$), $\mathbf{k}_{ij}^t$ is such as $\mathbf{k}_{ij}^t = \frac{2\pi}{\lambda}\mathbf{u}_{ij}^t = \frac{2\pi f_c}{c}\mathbf{u}_{ij}^t$. The quantity $\frac{1}{2\pi}\mathbf{k}_{ij}^t\mathbf{v_t}$ represents the Doppler effect.

In the same vein, if we define $\mathbf{s}_{ij}^r(t)$ as the signal between the receiving antenna $j$ and the scatterer $i$, then:

$$s_{ij}^r(t) = \mathbf{s_0} e^{j(2\pi(\frac{f_c \mathbf{v_r} \mathbf{u}_{ij}^r}{c}t+f_c t))} e^{j\phi_{ij}}$$

In all the following, the modeler supposes as a state of knowledge the following parameters:

- speed $\mathbf{v_r}$.
- speed $\mathbf{v_t}$.
- the angle of departure from the transmitting antenna to the scatterers $\psi_{ij}$ and power $P_j^t$.
- the angle of arrival from the scatterers to the receiving antenna $\phi_{ij}$ and power $P_j^r$.

The modeler has however no knowledge of what happens in between except the fact that a signal going from a steering vector of departure $j$ to a steering vector of arrival $i$ has a certain delay $\tau_{ij}$ due to possible single bounce or multiple bounces on different objects. The modeler also knows that objects do not move between the two sets of scatterers. The $s_r \times s_t$ delay matrix linking each DoA and DoD has the following structure:

$$\mathbf{D}_{s_r \times s_t}(f) = \begin{pmatrix} e^{-j2\pi f \tau_{1,1}} & \dots & e^{-j2\pi f \tau_{1,s_t}} \\ \vdots & \ddots & \vdots \\ e^{-j2\pi f \tau_{s_r,1}} & \dots & e^{-j2\pi f \tau_{s_r,s_t}} \end{pmatrix}$$

The modeler also supposes as a given state of knowledge the fact that each path $h_{ij}$ of matrix $\mathbf{H}$ has a certain power. Based on this state of knowledge, the modeler wants to model the $s_r \times s_t$ matrix $\mathbf{\Theta}_{s_r \times s_t}$ in the following representation:

$$\begin{aligned}\mathbf{H}(f,t) = \frac{1}{\sqrt{s_r s_t}} &\begin{pmatrix} e^{j(\phi_{1,1}+2\pi\frac{f\mathbf{u}_{11}^r\mathbf{v_r}}{c}t)} & \dots & e^{j(\phi_{1,s}+2\pi\frac{f\mathbf{u}_{1s}^r\mathbf{v_r}}{c}t)} \\ \vdots & \ddots & \vdots \\ e^{j(\phi_{r,1}+2\pi\frac{f\mathbf{u_{r1}}^r\mathbf{v_r}}{c}t)} & \dots & e^{j(\phi_{r,s}+2\pi\frac{f\mathbf{u}_{rs}^r\mathbf{v_r}}{c}t)} \end{pmatrix} \begin{pmatrix} \sqrt{P_1^r} & 0 & \dots \\ 0 & \ddots & 0 \\ \vdots & 0 & \sqrt{P_{s_r}^r} \end{pmatrix} \\ &\mathbf{\Theta}_{s_r \times s_t} \bigodot \mathbf{D}_{s_r \times s_t}(f) \\ &\begin{pmatrix} \sqrt{P_1^t} & 0 & \dots \\ 0 & \ddots & 0 \\ \vdots & 0 & \sqrt{P_{s_t}^t} \end{pmatrix} \begin{pmatrix} e^{j(\psi_{1,1}+2\pi\frac{f\mathbf{u}_{11}^t\mathbf{v_t}}{c}t)} & \dots & e^{j(\psi_{1,n_t}+2\pi\frac{f\mathbf{u}_{1n_t}^t\mathbf{v_t}}{c}t)} \\ \vdots & \ddots & \vdots \\ e^{j(\psi_{s_1,1}+2\pi\frac{f\mathbf{u}_{s_1 1}^t\mathbf{v_t}}{c}t)} & \dots & e^{j(\psi_{s_1,n_t}+2\pi\frac{f\mathbf{u}_{s_1 n_t}^t\mathbf{v_t}}{c}t)} \end{pmatrix}\end{aligned}$$

$\bigodot$ represents the Hadamard product defined as $c_{ij} = a_{ij}b_{ij}$ for a product matrix $\mathbf{C} = \mathbf{A} \bigodot \mathbf{B}$. As previously stated, one has to comply with the following constraints:

- Each entry of $\mathbf{H}(f,t)$ has a certain energy.

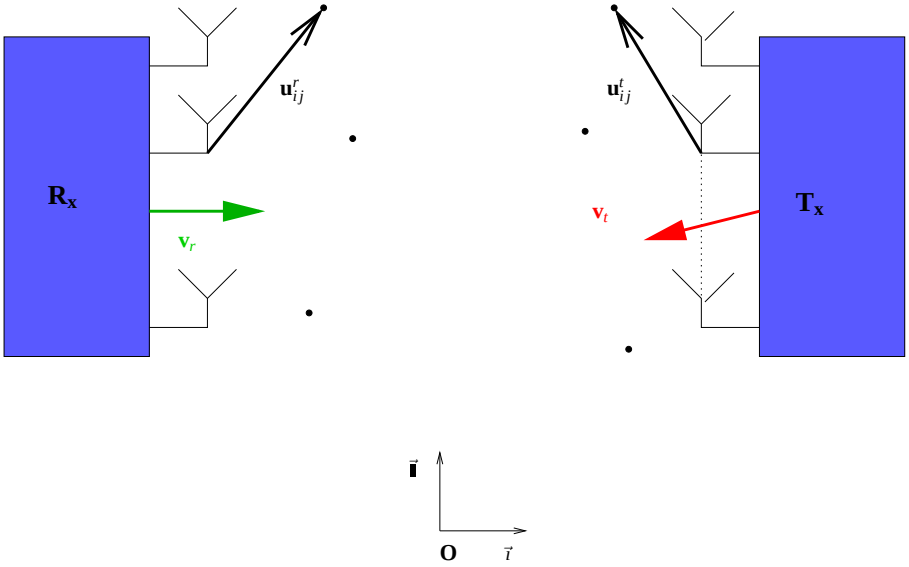



**Fig. 1.5.** Moving antennas.

- Consistency argument: if the DoA, DoD, powers, the delays, the Doppler effects are unknown then matrix $\mathbf{H}$ should be assigned an i.i.d Gaussian distribution.

**Proposition 4.** $\mathbf{\Theta}_{s_r \times s_t}$ *i.i.d zero mean Gaussian with variance 1 is solution of the consistency argument and maximizes entropy.*[19]

**Proof:** We will not go into the details but only provide the guidelines of the proof. First, remark that if $\mathbf{\Phi}$ and $\mathbf{\Psi}$ are unknown, then the principle of maximum entropy attributes i.i.d uniform distribution to the angles $\phi_{ij}$ and $\psi_{ij}$. But what probability distribution should the modeler attribute to the delays and the Doppler effects when no information is available?

- **Delays**: The modeler knows that there is, due to measurements performed in the area, a maximum possible delay for the information to go from the transmitter to the receiver $\tau_{\max}$. The principle of maximum entropy attributes therefore a uniform distribution to all the delays $\tau_{ij}$ such as $P(\tau_{ij}) = \frac{1}{\tau_{\max}}$ with $\tau_{ij} \in [0, \tau_{\max}]$
- **Doppler effect**: The modeler knows that the speed of the transmitter and receiver can not exceed a certain limit $v_{\text{limit}}$ (in the least favorable case, $v_{\text{limit}}$ would be equal to the speed of light) but if the transmission occurs in a city, the usual car speed limit can be taken as an upper bound. In this case, the speed $v_t$ and $v_r$ have also a uniform distribution such as $P(v_t) = P(v_r) = \frac{1}{v_{\text{limit}}}$. Moreover, if $\mathbf{v}_t = v_t \cos(\alpha_t)\boldsymbol{\imath} + v_t \sin(\alpha_t)\boldsymbol{\jmath}$, $\mathbf{v}_r = v_r \cos(\alpha_r)\boldsymbol{\imath} + v_r \sin(\alpha_r)\boldsymbol{\jmath}$, $\mathbf{u^t}_{ij} = \cos(\beta^t{}_{ij})\boldsymbol{\imath} + \sin(\beta^t{}_{ij})\boldsymbol{\jmath}$ and $\mathbf{u^r}_{ij} = \cos(\beta^r{}_{ij})\boldsymbol{\imath} + \sin(\beta^r{}_{ij})\boldsymbol{\jmath}$, the modeler will attribute a uniform distribution over $2\pi$ to the angles $\alpha_t$, $\alpha_r$,$\beta^t{}_{ij}$ and $\beta^r{}_{ij}$.

With all these probability distributions derived and using the same methodology as in the narrowband (in terms of frequency selectivity) MIMO model proof, one can easily show that $\mathbf{\Theta}_{s_r \times s_t}$ i.i.d Gaussian is solution of the consistency argument and maximizes entropy.

Note that in the case $f = 0$, $\mathbf{v_t} = 0$ and $\mathbf{v_r} = 0$, the same model as the narrowband model is obtained. If more information is available on correlation or different variances of frequency paths, then this information can be incorporated in the matrix $\mathbf{D}_{s_r \times s_t}$, also known as the channel pattern mask [31]. Note that in the case of a ULA (Uniform Linear Array) geometry and in the Fourier directions, we have $\mathbf{u}^r_{ij} = \mathbf{u}^r_j$ (any column of matrix $\mathbf{\Phi}$ has a given direction) and $\mathbf{u}^t_{ij} = \mathbf{u}^t_i$ (any line of matrix $\mathbf{\Psi}$ has a given direction). Therefore, the channel model simplifies to:

[19] Why does normality always appear in our models? Well, the answer is quite simple. In all this paper, we have always limited ourselves to the second moment of the channel. If more moments are available, then normal distributions would not appear in general.

$$\mathbf{H}(f,t) = \frac{1}{\sqrt{s_r s_t}} \begin{pmatrix} 1 & \dots & 1 \\ \vdots & \ddots & \vdots \\ e^{j2\pi \frac{d(n_r-1)\sin(\phi_1)}{\lambda}} & \dots & e^{j2\pi \frac{d(n_r-1)\sin(\phi_{s_r})}{\lambda}} \end{pmatrix} \mathbf{\Theta}_{s_r \times s_t} \bigodot \mathbf{D}_{s_r \times s_t}(f,t) \begin{pmatrix} 1 \dots & e^{j2\pi \frac{d(n_t-1)\sin(\psi_1)}{\lambda}} \\ \vdots \ddots & \vdots \\ 1 \dots & e^{j2\pi \frac{d(n_t-1)\sin(\psi_{s_t})}{\lambda}} \end{pmatrix}$$

In this case, the pattern mask $\mathbf{D}_{s_r \times s_t}$ has the following form:

$$\mathbf{D}_{s_r \times s_t}(f,t) = \begin{pmatrix} \sqrt{P_1^r}\sqrt{P_1^t} e^{-j2\pi f \tau_{1,1}} e^{j2\pi \frac{ft}{c}(\mathbf{u}_1^r \mathbf{v}_r + \mathbf{u}_1^t \mathbf{v}_t)} & \dots & \sqrt{P_1^r}\sqrt{P_{s_t}^t} e^{-j2\pi f \tau_{1,s_t}} e^{j2\pi \frac{ft}{c}(\mathbf{u}_1^r \mathbf{v}_r + \mathbf{u}_{s_t}^t \mathbf{v}_t)} \\ \vdots & \ddots & \vdots \\ \sqrt{P_{s_r}^r}\sqrt{P_1^t} e^{-j2\pi f \tau_{s_r,1}} e^{j2\pi \frac{ft}{c}(\mathbf{u}_{s_r}^r \mathbf{v}_r + \mathbf{u}_1^t \mathbf{v}_t)} & \dots & \sqrt{P_{s_r}^r}\sqrt{P_{s_t}^t} e^{-j2\pi f \tau_{s_r,s_t}} e^{j2\pi \frac{ft}{c}(\mathbf{u}_s^r \mathbf{v}_r + \mathbf{u}_{s_t}^t \mathbf{v}_t)} \end{pmatrix}$$

Although we take into account many parameters, the final model is quite simple. It is the product of three matrices: Matrices $\mathbf{\Phi}$ and $\mathbf{\Psi}$ taking into account the directions of arrival and departure; matrix $\mathbf{\Theta}_{s_r \times s_t} \bigodot \mathbf{D}_{s_r \times s_t}$ which is an independent Gaussian matrix with different variances. The frequency selectivity of the channel is therefore taken into account in the phase of each entry of the matrix $\mathbf{\Theta}_{s_r \times s_t} \bigodot \mathbf{D}_{s_r \times s_t}(f,t)$.

**Remark**: In the case of a one antenna system link ($n_r = 1$ and $n_t = 1$), we obtain:

$$\begin{aligned} \mathbf{H}(f,t) = & \frac{1}{\sqrt{s_r s_t}} \left[ e^{j(\phi_1 + 2\pi \frac{f \mathbf{u}_1^r \mathbf{v_r}}{c} t)} \dots e^{j(\phi_{s_r} + 2\pi \frac{f \mathbf{u}_{s_r}^r \mathbf{v_r}}{c} t)} \right] \begin{pmatrix} \sqrt{P_1^r} & 0 & \dots \\ 0 & \ddots & 0 \\ \vdots & 0 & \sqrt{P_{s_r}^r} \end{pmatrix} \\ & \mathbf{\Theta}_{s_r \times s_t} \bigodot \mathbf{D}_{s_r \times s_t}(f) \begin{pmatrix} \sqrt{P_1^t} & 0 & \dots \\ 0 & \ddots & 0 \\ \vdots & 0 & \sqrt{P_{s_t}^t} \end{pmatrix} \begin{bmatrix} e^{j(\psi_1 + 2\pi \frac{f \mathbf{u}_1^t \mathbf{v_t}}{c} t)} \\ \vdots \\ e^{j(\psi_{s_t} + 2\pi \frac{f \mathbf{u}_{s_t}^t \mathbf{v_t}}{c} t)} \end{bmatrix} \\ = & \frac{1}{\sqrt{s_r s_t}} \left[ \sum_{k=1}^{s_r} \theta_{k,1} \sqrt{P_k^r} e^{j(\phi_k + 2\pi \frac{f \mathbf{u}_k^r \mathbf{v_r}}{c} t)} e^{-j2\pi f \tau_{k,1}} \dots \sum_{k=1}^{s_r} \theta_{k,s_R} \sqrt{P_k^r} e^{j(\phi_k + 2\pi \frac{f \mathbf{u}_k^r \mathbf{v_r}}{c} t)} e^{-j2\pi f \tau_{k,s_r}} \right] \\ & \begin{pmatrix} \sqrt{P_1^t} & 0 & \dots \\ 0 & \ddots & 0 \\ \vdots & 0 & \sqrt{P_{s_t}^t} \end{pmatrix} \begin{pmatrix} e^{j(\psi_1 + 2\pi \frac{f \mathbf{u}_1^t \mathbf{v_t}}{c} t)} \\ \vdots \\ e^{j(\psi_{s_t} + 2\pi \frac{f \mathbf{u}_{s_t}^t \mathbf{v_t}}{c} t)} \end{pmatrix} \\ = & \sum_{l=1}^{s_t} \sum_{k=1}^{s_r} \rho_{k,l} e^{j2\pi \xi_{k,l} t} e^{-j2\pi f \tau_{k,l}} \end{aligned}$$

where $\rho_{k,l}$ ($\rho_{k,l} = \frac{1}{\sqrt{s_r s_t}} \theta_{k,l} \sqrt{P_k^r}\sqrt{P_l^t} e^{j(\phi_k + \psi_l)}$) are independent Gaussian variable with zero mean and variance $\mathbb{E}(\mid \rho_{k,l} \mid^2) = \frac{1}{s_r s_t} P_k^r P_l^t$, $\xi_{k,l} = \frac{f}{c}(\mathbf{u}_k^r \mathbf{v_r} - \mathbf{u}_l^t \mathbf{v_t})$ are the doppler effect and $\tau_{k,l}$ are the delays. This previous result is a generalization of the SISO (Single Input Single Output) wireless model in the case of multifold scattering with the power profile taken into account.

## 1.5 Discussion

### 1.5.1 Müller's Model

In a paper "A Random Matrix Model of Communication via Antenna Arrays" [32], Müller develops a channel model based on the product of two random matrices:

$$\mathbf{H} = \mathbf{\Phi A \Theta}$$

where $\mathbf{\Phi}$ and $\mathbf{\Theta}$ are two random matrices with zero mean unit variance i.i.d entries and $\mathbf{A}$ is a diagonal matrix (representing the attenuations). This model is intended to represent the fact that each signal bounces off a scattering object exactly once. $\mathbf{\Phi}$ represents the steering directions from the scatterers to the receiving antennas while $\mathbf{\Theta}$ represents the

steering directions from the transmitting antennas to the scatterers. Measurements in [32] confirmed the model quite accurately. Should we conclude that signals in day to day life bounce only once on the scattering objects?

With the maximum entropy approach developed in this contribution, new insights can be given on this model and explanations can be provided on why Müller's model works so well. In the maximum entropy framework, Müller's model can be seen as either:

- a DoA based model with random directions i.e matrix $\mathbf{\Phi}$ with different powers (represented by matrix $\mathbf{A}$) for each angle of arrival. In fact, the signal can bounce freely several times from the transmitting antennas to the final scatterers (matrix $\mathbf{\Theta}$). Contrary to past belief, this model takes into account multi-fold scattering and answers the following question from a maximum entropy standpoint: what is the consistent model when the state of knowledge is limited to:
  - Random directions scattering at the receiving side.
  - Each steering vector at the receiving side has a certain power.
  - Each frequency path has a given variance.
- a corresponding DoD based model with random directions i.e matrix $\mathbf{\Theta}$ with different powers (represented by matrix $\mathbf{A}$) for each angle of departure. The model permits also in this case the signal to bounce several times from the scatterers to the receiving antennas. From a maximum entropy standpoint, the model answers the following question: what is the consistent model when the state of knowledge is limited to:
  - Random directions scattering at the transmitting side.
  - Each steering vector at the transmitting side has a certain power.
  - Each frequency has zero mean and a certain variance.
- DoA-DoD based model with random directions where the following question is answered: What is the consistent model when the state of knowledge is limited to:
  - Random directions scattering at the receiving side.
  - Random directions scattering at the transmitting side.
  - Each angle of arrival is linked to one angle of departure.

As one can see, Müller's model is broad enough to include several maximum entropy directional models and this fact explains why the model complies so accurately with the measurements performed in [33]

### 1.5.2 Sayeed's Model

In a paper "Deconstructing Multi-antenna Fading Channels" [34], Sayeed proposes a virtual representation of the channel. The model is the following:

$$\mathbf{H} = \mathbf{A}_{n_r}\mathbf{S}{\mathbf{A}_{n_t}}^H$$

Matrices $\mathbf{A}_{n_r}$ and $\mathbf{A}_{n_t}$ are discrete Fourier matrices and $\mathbf{S}$ is a $n_r \times n_t$ matrix which represents the contribution of each of the fixed DoA's and DoD's. The representation is virtual in the sense that it does not represent the real directions but only the contribution of the channel to those fixed directions. The model is somewhat a projection of the real steering directions onto a Fourier basis. Sayeed's model is quite appealing in terms of simplicity and analysis (it corresponds to the Maxent model on Fourier directions). In this case, also, we can revisit Sayeed's model in light of our framework. We can show that every time, Sayeed's model answers a specific question based on a given assumption.

- Suppose matrix $\mathbf{S}$ has i.i.d zero mean Gaussian entries then Sayeed's model answers the following question: what is the consistent model for a ULA when the modeler knows that the channel carries some energy, the DoA and DoD are on Fourier directions but one does not know what happens in between.
- Suppose now that matrix $\mathbf{S}$ has a certain correlation structure then Sayeed's model answers the following question: what is the consistent model for a ULA when the modeler knows that the channel carries some energy, the DoA and DoD are on Fourier directions but assumes that the paths in between have a certain correlation.

As one can see, Sayeed's model has a simple interpretation in the maximum entropy framework: it considers a ULA geometry with Fourier directions each time. Although it may seem strange that Sayeed limits himself to Fourier directions, we do have an explanation for this fact. In his paper [31], Sayeed was mostly interested in the capacity scaling of MIMO channels and not the joint distribution of the elements. From that perspective, only the statistics of the uncorrelated scatterers is of interest since they are the ones which scale the mutual information. The correlated scatterers have very small effect on the information. In this respect, we must admit that Sayeed's intuition is quite impressive. However, the entropy framework is not limited to the ULA case (for which the Fourier vector approach is valid) and can be used for any kind of antenna and field approximation. One of the great features of the maximum

entropy (which is not immediate in Sayeed's representation) approach is the quite simplicity for translating any additional physical information into probability assignment in the model. A one to one mapping between information and model representation is possible. With the maximum entropy approach, every new information on the environment can be straightforwardly incorporated and the models are consistent: adding or retrieving information takes us one step forward or back but always in a consistent way. The models are somewhat like Russian dolls, imbricated one into the other.

### 1.5.3 The "Kronecker" model

In a paper "Capacity Scaling in MIMO Wireless Systems Under Correlated fading", Chuah et al. study the following Kronecker [20] model:

$$\mathbf{H} = \mathbf{R}_{n_r}{}^{\frac{1}{2}} \mathbf{\Theta} \mathbf{R}_{n_t}{}^{\frac{1}{2}}$$

Here, $\mathbf{\Theta}$ is an $n_r \times n_t$ i.i.d zero mean Gaussian matrix, $\mathbf{R}_{n_r}{}^{\frac{1}{2}}$ is an $n_r \times n_r$ receiving correlation matrix while $\mathbf{R}_{n_t}{}^{\frac{1}{2}}$ is a $n_t \times n_t$ transmitting correlation matrix. The correlation is supposed to decrease sufficiently fast so that $\mathbf{R}_{n_r}$ and $\mathbf{R}_{n_t}$ have a Toeplitz band structure. Using a software tool (Wireless System Engineering [37]), they demonstrate the validity of the model. Quite remarkably, although designed to take into account receiving and transmitting correlation, the model developed in the paper falls within the double directional framework. Indeed, since $\mathbf{R}_{n_r}$ and $\mathbf{R}_{n_t}$ are band Toeplitz then these matrices are asymptotically diagonalized in a Fourier basis

$$\mathbf{R}_{n_r} \sim F_{n_r} \Lambda_{n_r} F_{n_r}^H$$

and

$$\mathbf{R}_{n_t} \sim F_{n_t} \Lambda_{n_t} F_{n_t}^H.$$

$F_{n_r}$ and $F_{n_t}$ are Fourier matrices while $\Lambda_{n_r}$ and $\Lambda_{n_t}$ represent the eigenvalue matrices of $\mathbf{R}_{n_r}$ and $\mathbf{R}_{n_t}$.

Therefore, matrix $\mathbf{H}$ can be rewritten as:

$$\begin{aligned}
\mathbf{H} &= \mathbf{R}_{n_r}{}^{\frac{1}{2}} \mathbf{\Theta} \mathbf{R}_{n_t}{}^{\frac{1}{2}} \\
&= F_{n_r} \left( \Lambda_{n_r}{}^{\frac{1}{2}} F_{n_r}{}^H \mathbf{\Theta} F_{n_t} \Lambda_{n_t}{}^{\frac{1}{2}} \right) F_{n_t}{}^H \\
&= F_{n_r} \left( \mathbf{\Theta_1} \bigodot \mathbf{D}_{n_r \times n_t} \right) F_{n_t}{}^H
\end{aligned}$$

$\mathbf{\Theta_1} = \mathbf{F}_{n_r}{}^H \mathbf{\Theta} \mathbf{F}_{n_t}$ is a $n_r \times n_t$ zero mean i.i.d Gaussian matrix and $\mathbf{D}_{n_r \times n_t}$ is a pattern mask matrix defined by:

$$\mathbf{D}_{s \times s_1} = \begin{pmatrix} \lambda_{1,n_t}^{\frac{1}{2}} \lambda_{1,n_r}^{\frac{1}{2}} & \dots & \lambda_{n_t,n_t}^{\frac{1}{2}} \lambda_{1,n_r}^{\frac{1}{2}} \\ \vdots & \ddots & \vdots \\ \lambda_{1,n_t}^{\frac{1}{2}} \lambda_{n_r,n_r}^{\frac{1}{2}} & \dots & \lambda_{n_t,n_t}^{\frac{1}{2}} \lambda_{n_r,n_r}^{\frac{1}{2}} \end{pmatrix}$$

Note that this connection with the double directional model has already been reported in [31]. Here again, the previous model can be reinterpreted in light of the maximum entropy approach. The model answers the following question: what is the consistent model one can make when the DoA are uncorrelated and have respective power $\lambda_{i,n_r}$, the DoD are uncorrelated and have respective power $\lambda_{i,n_t}$, each path has zero mean and a certain variance. The model therefore confirms the double directional assumption as well as Sayeed's approach and is a particular case of the maximum entropy approach. The comments and limitations made on Sayeed's model are also valid here. **reference also [38, 39]**

### 1.5.4 The "Keyhole" Model

In [40], Gesbert et al. show that low correlation[21] is not a guarantee of high capacity: cases where the channel is rank deficient can appear while having uncorrelated entries (for example when a screen with a small keyhole is placed in between the transmitting and receiving antennas). In [42], they propose the following model for a rank one channel:

[20] The model is called a Kronecker model because $\mathbb{E}(\text{vec}(\mathbf{H})^H \text{vec}(\mathbf{H})) = \mathbf{R}_{n_r} \bigotimes \mathbf{R}_{n_t}$ is a Kronecker product. The justification of this approach relies on the fact that only immediate surroundings of the antenna array impose the correlation between array elements and have no impact on correlations observed between the elements of the array at the other end of the link. Some discussions can be found in [35, 36].

[21] "keyhole" channels are MIMO channels with uncorrelated spatial fading at the transmitter and the receiver but have a reduced channel rank (also known as uncorrelated low rank models). They were shown to arise in roof-edge diffraction scenarios [41].

$$\mathbf{H} = \mathbf{R}_{n_r}{}^{\frac{1}{2}} \mathbf{g}_r \mathbf{g}_t{}^H \mathbf{R}_{n_t}{}^{\frac{1}{2}} \tag{1.8}$$

Here, $\mathbf{R}_{n_r}{}^{\frac{1}{2}}$ is an $n_r \times n_r$ receiving correlation matrix while $\mathbf{R}_{n_t}{}^{\frac{1}{2}}$ is a $n_t \times n_t$ transmitting correlation matrix. $\mathbf{g}_r$ and $\mathbf{g}_t$ are two independent transmit and receiving Rayleigh fading vectors. Here again, this model has connections with the previous maximum entropy model:

$$\mathbf{H} = \frac{1}{\sqrt{s_r s_t}} \mathbf{\Phi}_{n_r \times s_r} \mathbf{\Theta}_{s_r \times s_t} \mathbf{\Psi}_{s_t \times n_t} \tag{1.9}$$

The Keyhole model can be either:

- A double direction model with $s_r = 1$ and $\mathbf{\Phi}_{n_r \times 1} = \mathbf{R}_{n_r}{}^{\frac{1}{2}} \mathbf{g}_r$. In this case, $\mathbf{g}_t{}^H \mathbf{R}_{n_t}{}^{\frac{1}{2}} = \mathbf{\Theta}_{1 \times s_t} \mathbf{\Psi}_{s_t \times n_t}$ where $\mathbf{\Theta}_{1 \times s_t}$ is zero mean i.i.d Gaussian.
- A double direction model with $s_t = 1$ and $\mathbf{\Psi}_{1 \times n_t} = \mathbf{g}_t{}^H \mathbf{R}_{n_t}{}^{\frac{1}{2}}$. In this case, $\mathbf{R}_{n_r}{}^{\frac{1}{2}} \mathbf{g}_r = \mathit{\Phi}_{n_r \times s_r} \mathbf{\Theta}_{s_r \times 1}$ where $\mathbf{\Theta}_{s_r \times 1}$ is zero mean i.i.d Gaussian.

As one can observe, the maximum entropy model can take into account rank deficient channels.

### 1.5.5 Conclusion

After analyzing each of these models, we find that they all answer a specific question based on a given state of knowledge. All these models can be derived within the maximum entropy framework and have a simple interpretation. Moreover, each time the directional assumption appears which conjectures the correctness of the directional approach.

## 1.6 Testing the Models

In all the previous sections, we have developed several models based on different questions. But what is the right model, in other words how to choose between the set $\{M_0, M_1, ..., M_K\}$ of $K$ models (note that $M$ specifies only the type of model and not the parameters of the model)?

### 1.6.1 Bayesian Viewpoint

When judging the appropriateness of a model, Bayes[22] rules derives the posterior probability of the model. Bayes rule gives the posterior probability for the $i^{\text{th}}$ model according to: [23]

$$P(M_i \mid Y, I) = P(M_i \mid I) \frac{P(Y \mid M_i, I)}{P(Y \mid I)}$$

Y is the data (given by measurements), I is the prior information (ULA, far field scattering...). For comparing two models $M$ and $M_1$, one has to compute the ratio:

$$\frac{P(M_1 \mid Y, I)}{P(M \mid Y, I)} = \frac{P(M_1 \mid I)}{P(M \mid I)} \frac{P(Y \mid M_1, I)}{P(Y \mid M, I)}$$

If $P(M_1 \mid Y, I) > P(M \mid Y, I)$, then one will conclude that model $M_1$ is better than model $M$. Let us now try to understand each term.
The first term, crucially important, is usually forgotten by the channel modelling community: $\frac{P(M_1|I)}{P(M|I)}$. It favors one model or the other before the observation. As an example, suppose that the information $\{I =$ The scatterers are near the antennas $\}$ is given. Then if one has to compare the model $M$ (which considers ULA with far field scattering) and the model $M_1$ (assuming near field scattering ) then one should consider $\frac{P(M_1|I)}{P(M|I)} > 1$. [24]

[22] This chapter is greatly inspired by the work of Jaynes and Bretthorst who have made the following ideas clear.

[23] We use here the notations and meanings of Jaynes [20] and Jeffrey [18]: $P(M_i \mid Y, I)$ is the "probability that the model $M_i$ is true given that the data Y is equal to the true data $y$ and that the information $I$ on which is based the model is true". Every time, " $(|$" means conditional on the truth of the hypothesis $I$. In probability theory, all probabilities are conditional on some hypothesis space.

[24] The term $\frac{P(M_1|I)}{P(M|I)}$ can be seen as the revenge of the measurement field scientist over the mathematician. It shows that modelling is both an experimental and theoretical science and that the experience of the field scientist (which attributes the values of the prior probabilities) does matter.

For understanding the second term, let us analyze and compare the following two specific models: the DoA based model $M_{\text{doa}}$ and the double directional model $M_{\text{double}}$.

**Model $M_{\text{doa}}$:**

$$\mathbf{H}(f,t) = \frac{1}{\sqrt{s_r}} \Phi \left( \mathbf{\Theta} \bigodot D(t,f) \right)$$

with

$$D(t,f) = \begin{pmatrix} e^{-j2\pi f \tau_{1,1}} e^{j2\pi \frac{ft}{c}(\mathbf{u}_1^r \mathbf{v}_r)} & \dots & e^{-j2\pi f \tau_{1,n_t}} e^{j2\pi \frac{ft}{c}(\mathbf{u}_1^r \mathbf{v}_r)} \\ \vdots & \ddots & \vdots \\ e^{-j2\pi f \tau_{s_r,1}} e^{j2\pi \frac{ft}{c}(\mathbf{u}_s^r \mathbf{v}_r)} & \dots & e^{-j2\pi f \tau_{s_r,n_t}} e^{j2\pi \frac{ft}{c}(\mathbf{u}_s^r \mathbf{v}_r)} \end{pmatrix}$$

deals with the DoA model taking into account the delays, Doppler effect (we suppose that the transmitting antenna does not move but only the receiving one) for a ULA ($s$ is the number of scatterers). Let the information $I$ on which is based the model be such that the powers of the steering directions are identical and that the transmitting antennas do not move. We recall that $\mathbf{u}_i^r \mathbf{v}_r = (\cos({\beta^r}_i)\mathbf{i} + \sin({\beta^r}_i)\mathbf{j})(v_r \cos(\alpha_r)\mathbf{i} + v_r \sin(\alpha_r)\mathbf{j}) = v_r \cos({\beta^r}_i - \alpha_r)$

The set of parameters on which the model is defined is

$$p_{\text{doa}} = \{\mathbf{\Phi}, s_r, \mathbf{\emptyset}, v_r, \mathbf{\Theta}, \alpha_r, \beta^r\}$$

and the parameters lie in a subspace $S_{p_{\text{doa}}}$. We recall here the DoA based model for a given frequency:

$$\mathbf{y}(t,f) = \frac{1}{\sqrt{s_r}} \mathbf{\Phi} \left( \mathbf{\Theta} \bigodot D(t,f) \right) \mathbf{x}(f) + \mathbf{n}(f)$$

The term of interest $P(\mathbf{y} \mid M_{\text{doa}}, I)$ can be derived the following way:

$$P(\mathbf{y} \mid M_{\text{doa}}, I) = \int P(\mathbf{y}, p_{\text{doa}} \mid M_{\text{doa}}, I) dp_{\text{doa}} = \int P(\mathbf{y} \mid p_{\text{doa}}, M_{\text{doa}}, I) P(p_{\text{doa}} \mid M_{\text{doa}}, I) dp_{\text{doa}}$$

Let us derive each probability distribution separately: $P(\mathbf{y} \mid p_{\text{doa}}, M_{\text{doa}}, I) =$

$$\frac{1}{(2\pi\sigma^2)^{\frac{N_1 N_r}{2}}} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^{N} \sum_{j=1}^{N_1} \left( y(t_j,f_i) - \frac{1}{\sqrt{s_r}} \mathbf{\Phi}(\mathbf{\Theta} \bigodot D(t_j,f_i)) x(f_i) \right)^H \left( y(t_j,f_i) - \frac{1}{\sqrt{s_r}} \mathbf{\Phi}(\mathbf{\Theta} \bigodot D(t_j,f_i)) x(f_i) \right)}$$

and

$$\begin{aligned} P(p_{\text{doa}} \mid M_{\text{doa}}, I) &= P(\mathbf{\Phi}, s_r, \mathbf{\emptyset}, \mathbf{fi^r}, v_r, \alpha_r, \mathbf{\Theta} \mid M_{\text{doa}}, I) \\ &= P(\mathbf{\Phi} \mid s_r, M_{\text{doa}}, I) P(s_r \mid M_{\text{doa}}, I) P(v_r \mid M_{\text{doa}}, I) P(\mathbf{\emptyset} \mid M_{\text{doa}}, I) \\ &\quad P(\mathbf{\Theta} \mid M_{\text{doa}}, s_r, I) P(\alpha_r \mid M_{\text{doa}}, I) P(\mathbf{fi}^r \mid I, M_{\text{doa}}) \end{aligned}$$

since all the priors are taken independent in the case of uninformative priors. The values of these priors have already been provided (the proof is given in chapter 1.4.3) and only the prior on $\mathbf{\Theta}$ and $s_r$ remain to be given. We give these two priors now (and also the prior on the power although in the two models introduced for comparison, the power distribution is not needed):

- If only the mean and variance of each path is available then using maximum entropy arguments, one can show that:

$$\begin{aligned} P(\mathbf{\Theta} \mid s_r, M_{\text{doa}}, I) &= \frac{1}{(\sqrt{2\pi})^{n_t \times s_r}} e^{-\sum_{i=1}^{s_r} \sum_{j=1}^{n_t} |\theta_{i,j}|^2} \\ &= \frac{1}{(\sqrt{2\pi})^{n_t \times s_r}} e^{-\text{trace}(\mathbf{\Theta}\mathbf{\Theta}^H)} \end{aligned}$$

- How can we assign a prior probability $P(s_r \mid M_{\text{doa}}, I)$ for the unknown number of scatterers? The modeler has no knowledge if the measurements were taken in a dense area or not. The unknown number of scatterers could range from one (this prior only occurs in model that have a single bounce) up to a maximum. But what is the maximum value? There are $N \times N_1$ data values and if there were $N \times N_1$ scatterers, the data could be at most fit by placing a scatterer at each data value and adjusting the direction of arrivals. Because no additional information is available

about the number of scatterers, $N \times N_1$ may be taken as an upper bound. Using the principle of maximum entropy, one obtains a uniform distribution for the number of scatterers $P(s_r \mid M_{\text{doa}}, I) = \frac{1}{N \times N_1}$.
Note that in the general case, if one has precise available information then one has to take it into account. But how can the modeler translate the prior on the scatterers due to the fact that the room has three chairs and a lamp in the corner? This is undoubtedly a difficult task and representing that information in terms of probabilities is not straightforward. But difficult is not impossible. The fact that there are several chairs (with respect to the case where there is no chairs) is a source of information and will lead to attributing in the latter case a peaky prior shifted around a higher number of scatterers.

- **Power**: The transmitter is limited in terms of transmit power to an upper bound value $P^t_{\text{max}}$. Therefore, the principle of maximum entropy attributes a uniform distribution to the different amplitudes $P(P_i^t) = \frac{1}{P^t_{\text{max}}}$, $P_i \in [0, P^t_{\text{max}}]$. In the same vein, the receiver cannot, due to the amplifiers, process a receiving amplitude greater then $P^r_{\text{max}}$. In this case, the principle of maximum entropy attributes a uniform distribution such as $P(P_i^r) = \frac{1}{P^r_{\text{max}}}$, $P_i \in [0, P^r_{\text{max}}]$

With all the previous priors given, one can therefore compute:

$$
\begin{aligned}
P(\mathbf{y} \mid M_{\text{doa}}, I) = \int & \frac{1}{(2\pi\sigma^2)^{\frac{N_1 N r}{2}}} \\
& e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{N}\sum_{j=1}^{N_1}\left(y(t_j,f_i)-\frac{1}{\sqrt{s_r}}\mathbf{\Phi}\left(\mathbf{\Theta}\odot D(t_j,f_i)\right)x(f_i)\right)^H\left(y(t_j,f_i)-\frac{1}{\sqrt{s_r}}\mathbf{\Phi}\left(\mathbf{\Theta}\odot D(t_j,f_i)\right)x(f_i)\right)} \\
& P(\mathbf{\Phi} \mid s_r, M_{\text{doa}}, I) P(s_r \mid M_{\text{doa}}, I) P(v_r \mid M_{\text{doa}}, I) P(\alpha_r \mid M_{\text{doa}}, I) P(\mathbf{fi}^r \mid M_{\text{doa}}, I) \\
& P(\mathbf{ø} \mid M_{\text{doa}}, I) P(\mathbf{\Theta} \mid M_{\text{doa}}, I) d\mathbf{\Phi} d\mathbf{\Theta} ds_r d\mathbf{ø} dv_r d\alpha_r d\mathbf{fi^r}
\end{aligned}
$$

which gives:

$$
\begin{aligned}
P(\mathbf{y} \mid M_{\text{doa}}, I) = & \frac{1}{N \times N_1} \sum_{s_r=1}^{N \times N_1} \int_0^{2\pi} \int_0^{\infty} \int_0^{v_{\text{lim}}} \int_0^{\tau_{\text{max}}} \frac{1}{(2\pi\sigma^2)^{\frac{N_1 N r}{2}}} \prod_{i=1}^{N} \prod_{j=1}^{N_1} \\
& e^{-\frac{1}{2\sigma^2}\left(y(t_j,f_i)-\frac{1}{\sqrt{s_r}}\mathbf{\Phi}(\mathbf{\Theta}\odot D(t_j,f_i))x(f_i)\right)^H\left(y(t_j,f_i)-\frac{1}{\sqrt{s_r}}\mathbf{\Phi}(\mathbf{\Theta}\odot D(t_j,f_i))x(f_i)\right)} \\
& (\frac{1}{\tau_{\text{max}}})^{s_r \times n_t} \frac{1}{v_{\text{lim}}} (\frac{1}{2\pi})^{n_r \times s_r} \frac{1}{2\pi} (\frac{1}{2\pi})^{s_r} \\
& d\phi_{11} ... d\phi_{n_r s_r} d\theta_{11} ... d\theta_{s_r n_t} d\tau_{11} ... d\tau_{s_r n_t} dv_r d\alpha_r d\beta^r{}_1 ... d\beta^r{}_{s_r}
\end{aligned}
\tag{1.10}
$$

As one can see, the numerical integration is tedious but it is the only way to rank the models in an appropriate manner.

**Model $M_{\text{double}}$:**
Let us now derive model $M_{\text{double}}$:

$$
\mathbf{H}(f,t) = \frac{1}{\sqrt{s_r s_t}} \mathbf{\Phi} \left( \mathbf{\Theta} \bigodot D(t,f) \right) \mathbf{\Psi}
$$

with

$$
D(t,f) = \begin{pmatrix}
e^{-j2\pi f \tau_{1,1}} e^{j2\pi \frac{ft}{c}(\mathbf{u}_1^r \mathbf{v}_r)} & \dots & e^{-j2\pi f \tau_{1,s_t}} e^{j2\pi \frac{ft}{c}(\mathbf{u}_1^r \mathbf{v}_r)} \\
\vdots & \ddots & \vdots \\
e^{-j2\pi f \tau_{s_r,1}} e^{j2\pi \frac{ft}{c}(\mathbf{u}_{s_r}^r \mathbf{v}_r)} & \dots & e^{-j2\pi f \tau_{s_r,s_t}} e^{j2\pi \frac{ft}{c}(\mathbf{u}_{s_r}^r \mathbf{v}_r)}
\end{pmatrix}
$$

deals with the double directional model for which the set of parameters is

$$
p_{\text{double}} = \{\mathbf{\Phi}, s_r, \mathbf{\Psi}, s_t, \mathbf{ø}, v_r, , \alpha_r, \mathbf{fi_r \Theta}\} = \{p_{\text{doa}}, \mathbf{\Phi}, s_t\}
$$

by adding two new parameters $\mathbf{\Psi}$ and $s_t$ and going to the new subspace $S_{p_{\text{double}}}$ in such a way that $\mathbf{\Psi} = \mathbf{F}_{n_t}$ $(n_t = s_t)$ represents model $M_{\text{doa}}$. Indeed, in this case, we have:

$$\left(\boldsymbol{\Theta} \bigodot D(t,f)\right)\mathbf{F}_{n_t} = \begin{pmatrix} \sum_{i=1}^{n_t}\theta_{1i}e^{-j2\pi f\tau_{1,i}}e^{j2\pi\frac{ft}{c}(\mathbf{u}_1^r\mathbf{v}_r)} & \dots & \sum_{i=1}^{n_t}\theta_{1i}e^{-j2\pi f\tau_{1,i}}e^{j2\pi\frac{ft}{c}(\mathbf{u}_1^r\mathbf{v}_r)}e^{j2\pi\frac{(n_t-1)i}{n_t}} \\ \vdots & \ddots & \vdots \\ \sum_{i=1}^{n_t}\theta_{s_ri}e^{-j2\pi f\tau_{s_r,i}}e^{j2\pi\frac{ft}{c}(\mathbf{u}_s^r\mathbf{v}_r)} & \dots & \sum_{i=1}^{n_t}\theta_{s_ri}e^{-j2\pi f\tau_{s_r,i}}e^{j2\pi\frac{ft}{c}(\mathbf{u}_{s_r}^r\mathbf{v}_r)}e^{j2\pi\frac{(n_t-1)i}{n_t}} \end{pmatrix}$$
$$= \begin{pmatrix} \sum_{i=1}^{n_t}\theta_{1i}e^{-j2\pi f(\tau_{1,i}-\tau_{1,1})} & \dots & \sum_{i=1}^{n_t}\theta_{1i}e^{-j2\pi f(\tau_{1,i}-\tau_{1,n_t})}e^{j2\pi\frac{(n_t-1)i}{n_t}} \\ \vdots & \ddots & \vdots \\ \sum_{i=1}^{n_t}\theta_{s_ri}e^{-j2\pi f(\tau_{s_r,i}-\tau_{s_r,1})} & \dots & \sum_{i=1}^{n_t}\theta_{s_ri}e^{-j2\pi f(\tau_{s_r,i}-\tau_{s_r,n_t})}e^{j2\pi\frac{(n_t-1)i}{n_t}} \end{pmatrix} \bigodot D(t,f)$$
$$= \Theta_1 \bigodot D(t,f)$$

Where $\Theta_1$ is a matrix with i.i.d Gaussian entries.
We recall here the model for a given frequency:

$$\mathbf{y}(f,t) = \frac{1}{\sqrt{s_r s_t}}\boldsymbol{\Phi}\left(\boldsymbol{\Theta}\bigodot D(t,f)\right)\boldsymbol{\Psi}\mathbf{x}(f) + \mathbf{n}(f)$$

The same methodology applies and we have:

$$P(\mathbf{y} \mid M_{\text{double}}, I) = \int \frac{1}{(2\pi\sigma^2)^{\frac{N_1 N_r}{2}}}$$
$$e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{N}\sum_{j=1}^{N_1}\left(y(t_j,f_i)-\frac{1}{\sqrt{s_r s_t}}\boldsymbol{\Phi}\left(\boldsymbol{\Theta}\odot D(t_j,f_i)\right)\boldsymbol{\Psi}x(f_i)\right)^H\left(y(t_j,f_i)-\frac{1}{\sqrt{s_r s_t}}\boldsymbol{\Phi}\left(\boldsymbol{\Theta}\odot D(t_j,f_i)\right)\boldsymbol{\Psi}x(f_i)\right)}$$
$$P(\boldsymbol{\Phi} \mid s_r, M_{\text{double}}, I)P(s_r \mid M_{\text{double}}, I)P(\boldsymbol{\Psi} \mid s_t, M_{\text{double}}, I)P(s_t \mid M_{\text{double}}, I)$$
$$P(v_r \mid M_{\text{double}}, I)P(\alpha_r \mid M_{\text{double}}, I)P(\mathbf{fi_r} \mid M_{\text{double}}, I)P(\boldsymbol{\phi} \mid M_{\text{double}}, I)$$
$$P(\boldsymbol{\Theta} \mid M_{\text{double}}, I)d\boldsymbol{\Phi}d\boldsymbol{\Psi}d\boldsymbol{\Theta}ds_r ds_t d\boldsymbol{\phi}dv_r d\alpha_r d\mathbf{fi}^{\mathbf{r}}$$

and

$$P(\mathbf{y} \mid M_{\text{double}}, I) = (\frac{2}{N\times N_1})^2 \sum_{s_r=1}^{\frac{N\times N_1}{2}}\sum_{s_t=1}^{\frac{N\times N_1}{2}}\int_0^{2\pi}\int_0^{\infty}\int_0^{v_{\text{lim}}}\int_0^{\tau_{\max}}\frac{1}{(2\pi\sigma^2)^{\frac{N_1 N_r}{2}}}\prod_{i=1}^{N}\prod_{j=1}^{N_1}$$
$$e^{-\frac{1}{2\sigma^2}\left(y(t_j,f_i)-\frac{1}{\sqrt{s_r s_t}}\boldsymbol{\Phi}(\boldsymbol{\Theta}\odot D(t_j,f_i))\boldsymbol{\Psi}x(f_i)\right)^H\left(y(t_j,f_i)-\frac{1}{\sqrt{s_r s_t}}\boldsymbol{\Phi}(\boldsymbol{\Theta}\odot D(t_j,f_i))\boldsymbol{\Psi}x(f_i)\right)}$$
$$(\frac{1}{\tau_{\max}})^{s_r\times s_t}\frac{1}{v_{\text{lim}}}(\frac{1}{2\pi})^{n_r\times s_r}(\frac{1}{2\pi})^{s_t\times n_t}\frac{1}{2\pi}(\frac{1}{2\pi})^{s_r}$$
$$d\phi_{11}...d\phi_{n_r s_r}d\psi_{11}...d\psi_{1n_t}d\theta_{11}...d\theta_{s_r n_t}d\tau_{11}...d\tau_{s_r n_t}dv_r d\alpha_r d\beta^r{}_1...d\beta^r{}_{s_r} \tag{1.11}$$

A common problem in the modelling process is the following: suppose, when testing the models with the data, that both models $M$ and $M_1$ have the same maximum likelihood, in other words:

$$P(\mathbf{y} \mid p_{\text{doa}}{}^{\max}, M_{\text{doa}}, I) = P(\mathbf{y} \mid p_{\text{double}}{}^{\max}, M_{\text{double}}, I)$$

Which model should we choose? Hereafter, we give an example to show that Bayesian probability will choose the model with the smallest number of parameters.

First of all, we will suppose that the information $I$ available does not give a preference to model before seeing the data: $P(M_{\text{double}} \mid I) = P(M_{\text{doa}} \mid I)$.
As previously shown,

$$P(\mathbf{y} \mid M_{\text{doa}}, I) = \int P(\mathbf{y}, p_{\text{doa}} \mid M_{\text{doa}}, I)dp_{\text{doa}}$$
$$= \int P(\mathbf{y} \mid p_{\text{doa}}, M_{\text{doa}}, I)P(p_{\text{doa}} \mid M_{\text{doa}}, I)dp_{\text{doa}}$$

and

$$P(\mathbf{y} \mid M_{\text{double}}, I) = \int P(\mathbf{y}, p_{\text{double}} \mid M_{\text{double}}, I)dp_{\text{double}} \tag{1.12}$$
$$= \int P(\mathbf{y} \mid p_{\text{double}}, M_{\text{double}}, I)P(p_{\text{double}} \mid M_{\text{double}}, I)dp_{\text{double}} \tag{1.13}$$

Since

$$\begin{aligned} P(p_{\text{double}} \mid M_{\text{double}}, I) &= P([p_{\text{doa}}, \mathbf{\Psi}, s_t] \mid M_{\text{double}}, I) \\ &= P(p_{\text{doa}} \mid \mathbf{\Psi}, s_t, M_{\text{double}}, I) P(\mathbf{\Psi}, s_t \mid M_{\text{double}}, I) \end{aligned}$$

From equation (1.12), we have:

$$\begin{aligned} P(\mathbf{y} \mid M_{\text{double}}, I) = \int\int & P(\mathbf{y} \mid [p_{\text{doa}}, \mathbf{\Psi}, s_t], M_{\text{double}}, I) P(p_{\text{doa}} \mid \mathbf{\Psi}, s_t, M_{\text{double}}, I) \\ & P(\mathbf{\Psi}, s_t \mid M_{\text{double}}, I) dp_{\text{doa}} d\mathbf{\Psi} ds_t \end{aligned}$$

In the following, we will suppose that the likelihood function $P(\mathbf{y} \mid [p_{\text{doa}}, \mathbf{\Psi}, s_t], M_{\text{double}}, I)$ is peaky around the maximum likelihood region and has near zero values elsewhere. Otherwise, the measurement data Y would be useless in the sense that the data does not provide any information. Suppose now that with model $M_{\text{double}}$, the maximum likelihood $P(\mathbf{y} \mid [p_{\text{doa}}, \mathbf{\Psi}, s_t] M_{\text{double}}, I)$ occurs at a point near $\mathbf{\Psi} = \mathbf{F}_t$ and $s_t = n_t$ for the parameters $\mathbf{\Psi}$ and $s_t$ in other words $P(\mathbf{y} \mid [p_{\text{doa}}, \mathbf{\Psi}, s_t], M_{\text{double}}, I)$ is always null except for the value of $\mathbf{\Psi} = \mathbf{F}_t$ and $s_t = n_t$ then:

$$\begin{aligned} P(\mathbf{y} \mid M_{\text{double}}, I) = & \int\int\int P(\mathbf{y} \mid [p_{\text{doa}}, \mathbf{\Psi}, s_t], M_{\text{double}}, I) P(p_{\text{doa}} \mid \mathbf{\Psi}, s_t M_{\text{double}}, I) \\ & P(\mathbf{\Psi}, s_t \mid M_{\text{double}}, I) dp_{\text{doa}} d\mathbf{\Psi} ds_t \\ \approx & \int P(\mathbf{y} \mid [p_{\text{doa}}, \mathbf{\Psi} = \mathbf{F}_{n_t}, s_t = n_t], M_{\text{Double}}, I) \\ & P(p_{\text{doa}} \mid [\mathbf{\Psi} = \mathbf{F}_{n_t}, s_t = n_t] M_{\text{double}}, I) \qquad (1.14) \\ & P([\mathbf{\Psi} = \mathbf{F}_{n_t}, s_t = n_t] \mid M_{\text{double}}, I) dp_{\text{doa}} \qquad (1.15) \end{aligned}$$

One has to notice that $P(p_{\text{doa}} \mid [\mathbf{\Psi} = \mathbf{F}_{n_t}, s_t = n_t], M_{\text{double}}, I) = P(p_{\text{doa}} \mid M_{\text{doa}}, I)$ and $P(\mathbf{y} \mid [p_{\text{doa}}, \mathbf{\Psi} = \mathbf{F}_{n_t}, s_t = n_t], M_{\text{double}}, I) = P(\mathbf{y} \mid p_{\text{doa}}, M_{\text{doa}}, I)$ since both models are the same when $\mathbf{\Psi} = \mathbf{F}_{n_t}$ and $s_t = n_t$. We also have $P(\mathbf{\Psi} = \mathbf{F}_{n_t}, s_t = n_t \mid M_{\text{double}}, I) \leq 1$ (In fact, we can derive the exact value. Indeed, since we have no knowledge of the directions of arrival, $P(\mathbf{\Phi} = \mathbf{F}_{n_t}, s_t = n_t \mid M_{\text{double}}, I) = \frac{1}{(2\pi)^{n_r \times s_t}}$). Using equation (1.14), Bayesian probability shows us that:

$$\begin{aligned} P(\mathbf{y} \mid M_{\text{double}}, I) \leq & \int P(\mathbf{y} \mid [p_{\text{doa}}, \mathbf{\Psi} = \mathbf{F}_{n_t}, s_t = n_t], M_{\text{Double}}, I) \\ & P(p_{\text{doa}} \mid [\mathbf{\Psi} = \mathbf{F}_{n_t}, s_t = n_t] M_{\text{double}}, I) P([\mathbf{\Psi} = \mathbf{F}_{n_t}, s_t = n_t] \mid M_{\text{double}}, I) dp_{\text{doa}} \\ = & \int P(\mathbf{y} \mid p_{\text{doa}}, M_{\text{doa}}, I) P(p_{\text{doa}} \mid M_{\text{doa}}, I) P([\mathbf{\Psi} = \mathbf{F}_{n_t}, s_t = n_t] \mid M_{\text{double}}, I) dp_{\text{doa}} \\ \leq & \int P(\mathbf{y} \mid p_{\text{doa}}, M_{\text{doa}}, I) P(p_{\text{doa}} \mid M_{\text{doa}}, I) dp_{\text{doa}} \\ = & \int P(\mathbf{y}, p_{\text{doa}} \mid M_{\text{doa}}, I) dp_{\text{doa}} \\ & P(\mathbf{y} \mid M_{\text{doa}}, I) \end{aligned}$$

Since $M_{\text{doa}}$ has less parameters then $M_{\text{double}}$, Bayesian probability will favor the model $M_{\text{doa}}$ with less parameters and therefore shows that "the best explanation is always the simplest"[25]. It is therefore wrong to think that by increasing the number of parameters one can always find a good model: one can indeed better fit the data to the model (expression $P(\mathbf{y} \mid p_{\text{doa}}, M_{\text{doa}}, I)$) but the prior probability $P(p_{\text{doa}} \mid M_{\text{doa}}, I)$ will spread over a larger space and assign as a consequence a lower value to $P(\mathbf{y} \mid M_{\text{doa}}, I)$.

But how does the a posteriori computation compare with the usual methodology of maximizing the likelihood $P(\mathbf{y} \mid p, M, I)$?

Following [20], let us expand $\log P(\mathbf{y} \mid p, M, I)$ around the maximum likelihood point $\hat{p} = \{p^1{}_{\text{max}}, ..., p^m{}_{\text{max}}\}$

$$\log P(\mathbf{y} \mid p, M, I) = \log P(\mathbf{y} \mid p_{\text{max}}, M, I) + \frac{1}{2} \sum_{i,j=1}^{m} \frac{d^2 \log(P)}{dp^i dp^j} (p^i - p^i{}_{\text{max}})(p^j - p^j{}_{\text{max}}) + O()$$

[25] In statistical inference, this is known as Occam's razor. William of Occam was a theologian of the 14th century who wrote against the papacy in a series of treatise in which he tried to avoid many established pseudo explanations. In his terms, the logic of simplicity was stated in the following form "Causes shall not be multiplied beyond necessity" [28]. Note that Occam's razor has been extended to other fields such as metaphysics where it is interpreted as "nature prefers simplicity".

then near the peak a good approximation is a multivariate Gaussian such as:

$$P(\mathbf{y} \mid p, M, I) = P(\mathbf{y} \mid p_{\max}, M, I) e^{-\frac{1}{2}(p-p_{\max})\Delta^{-1}(p-p_{\max})}$$

with the inverse covariance matrix defined as:

$$\Delta^{-1}{}_{ij} = \left( \frac{d^2 \log(P)}{dp^i dp^j} \right)_{\pi=\pi_{\max}}$$

Therefore,

$$\begin{aligned} P(\mathbf{y} \mid M, I) &= P(\mathbf{y} \mid p_{\max}, M, I) \int e^{-\frac{1}{2}(p-p_{\max})\Delta^{-1}(p-p_{\max})} P(p \mid M, I) dp \\ &= P(\mathbf{y} \mid p_{\max}, M, I) G(M, I) \end{aligned}$$

All the tools are now provided to better understand what is happening. Suppose we want to compare two models $M$ and $M_1$. The a posteriori probability ratio for model $M$ over $M_1$ is:

$$\begin{aligned} \frac{P(M \mid \mathbf{y}, I)}{P(M_1 \mid \mathbf{y}, I)} &= \frac{P(M \mid I)}{P(M_1 \mid I)} \frac{P(\mathbf{y} \mid M, I)}{P(\mathbf{y} \mid M_1, I)} \\ &= \frac{P(M \mid I)}{P(M_1 \mid I)} \frac{P(\mathbf{y} \mid p^{\max}, M, I)}{P(\mathbf{y} \mid {p_1}^{\max}, M_1, I)} \frac{G(M, I)}{G(M_1, I)} \end{aligned}$$

In the conventional methods, $M$ is better than $M_1$ if $\frac{P(\mathbf{y}|p^{\max}M,I)}{P(\mathbf{y}|{p_1}^{\max}M_1,I)} > 1$ which is only one part of the three terms to be computed. In fact, in order to compare two models, three terms have to be calculated and the mistake persists thinking that any model $M_1$ versus $M$ is good as long as we increase the number of parameters: indeed, the fitting will get better and the ratio $\frac{P(\mathbf{y}|p^{\max},M,I)}{P(\mathbf{y}|{p_1}_{\max},M_1,I)}$ will decrease but this is only looking at one part of the problem. First of all, one has to consider $\frac{P(M|I)}{P(M_1|I)}$ and moreover $\frac{G(M,I)}{G(M_1,I)}$. This last term depends on the prior information about the internal parameters and as the number of parameters increases this term decreases due to the fact that we add more and more uninformative priors.

### 1.6.2 Conventional Methods

In the previous section, we have shown how probability theory can be used to rank the models. However, the integrals derived in equation (1.10) and equation (1.11) are not easy to compute, especially in the case of interest with a high number of antennas ($8 \times 8$) since we have to marginalize our integrals across a great number of parameters. But however difficult the problem may be, it is not a reason to hide problems and the use of other methods should be clearly explained. The reader must now know that one can rank models and that there is an optimum number of parameters when representing information. The Bayesian framework gives us an answer by comparing the a posteriori probability ratios: $\frac{P(M|\mathbf{y},I)}{P(M_1|\mathbf{y},I)}$. If one is to use other testing methods, then one has to clearly understand the limitations of these methods and justify the use of the criteria. In the following, we explain two procedures used by the channel modelling community and explain their limitations.

1- Parameter estimation methods

In this procedure, the data is cut into two parts, one for estimating the parameters, the other to validate the model incorporating the parameters.

- For estimating the parameters such as the angles of arrival, non-parametric methods such as the beamforming or the Capon method [43] can be used. In the case of parametric methods such as Music [44], Min-Norm [45] or Esprit method [46], they rely on properties of the structure of the covariance $\mathbf{R} = \mathbb{E}(\mathbf{y}\mathbf{y}^{\mathbf{H}}) = \mathbf{\Phi}\mathbf{K}\mathbf{\Phi}^{\mathbf{H}} + \boldsymbol{\sigma}^{\mathbf{2}}\mathbf{I}$ of the output signal. In this case, one has to assume that matrix $\mathbf{K}$ ($\mathbf{K} = \mathbb{E}(\mathbf{\Theta}\mathbf{\Psi}\mathbf{x}\mathbf{x}^H\mathbf{\Psi}^H\mathbf{\Theta}^H)$) has full rank.
- Once the parameters of the model have been estimated, the other set of the data is used to test the model. A mean square error is given. In general, a small mean square error is acknowledged to yield a good model and one seeks the smallest error possible.

If one is to use this procedure, one has to understand that in no way will it lead into judging the appropriateness of a model. Indeed, by adding more and more parameters to the model, one can always find a way of achieving a low mean square error by adjusting accordingly the parameters. This fact explains why some many models comply in the literature with the measurements. If the model minimizes the mean square error, then it is a **possible** candidate but the modeler can not conclude that it is a **good** candidate.

Moreover, since the testing method has no real justification, many problems arise when using it.

- How does one cut the set of data? Do we use half the data to estimate the parameters and half the data to test the model? Why not using one quarter and three quarter? In the Bayesian viewpoint, this is not at all a problem as one takes into account all the data available and does not make any unjustified transformation on the data.
- If one is to use a Music or Esprit algorithm, $\mathbf{K}$ has to be full rank. This is obviously not the case for a double directional model where the steering DoD matrix $\mathbf{\Psi}$ is not always full rank since $\mathbf{K} = \mathbb{E}(\mathbf{\Theta\Psi x x}^H \mathbf{\Psi}^H \mathbf{\Theta}^H)$.

2- Moment fitting:

Other authors [47] validate their model by finding the smallest error of a set of moments. They derive explicit theoretical formulas of the $n_{\text{th}}$ moment $m_n(f)$ of the matrix $\mathbf{H}^H(f)\mathbf{H}(f)$ and find the optimal parameters in order to minimize:

$$\frac{1}{N}\sum_{n=1}^{N} \mid \frac{m_n(f)}{\hat{m}_n(f)} - 1 \mid$$

where

$$\hat{m}_n(f) = \frac{\text{Trace}(\mathbf{H}^H(f)\mathbf{H}(f))^n}{\text{Trace}(\mathbf{H}^H(f)\mathbf{H}(f))}$$

As previously stated, many models can minimize this criteria by adding more and more parameters and one cannot obviously conclude in this case if a model is better then the other or not. Moreover, how useful is it to have a channel that fits a certain amount of moments?[26].

The previous remarks show that when the abstract of a paper asserts: “This paper finds the theoretical predictions to accurately match data obtained in a recent measurement campaign”, one has to be really cautious on the conclusions to be drawn.

## 1.7 Conclusion

Where do we stand on channel modelling?[27] This question is not simple to answer as many models have been proposed and each of them validated by measurements. Channel models are not getting better and better but they only answer different questions based on different states of knowledge[28]. The crucial point is not creating a model but asking the right question based on a given state of knowledge (raw measurement data, prior information, are we in a urban area? is it a fixed network?..). A generic method for creating models based on the principle of maximum entropy has been provided and proved to be theoretically sound. At every step, we create a model incorporating only our prior information and not more! The model achieved is broad as it complies as best it can with any case having more constraints (but at least includes the same prior constraints). The channel modelling method is summarized hereafter:

- $H(p) = \int -p\log p + \sum_i \lambda_i \{\text{prior information}\}_i$
- Argument of consistency

The consistency argument is extremely important as it shows that two channel modelling methods based on the same state of knowledge should lead to the same channel model. This fact has not always been fulfilled in the past. Our models are logical consequence of the use of the principle of maximum entropy and need not to be assumed without deeper justification. The models proposed may seem inadequate to reality for some readers: we argue as in [20] that the purpose of channel modelling is not to describe reality but only our information about reality. The model we achieve are consistent and any other representation is obviously unsound if based on the same state of knowledge. However, one must bear in mind that the less things are assumed as a priori information the greater are the chances that the model complies with any mismatched representation.

But what if the model fails to comply with measurements? The model is not to blame as it is a logic consequence of information theoretic tools [20]. With the methodology introduced, failure is greatly appreciated as it is a source of information and the maximum entropy approach is avid of information: the result of non-compliance is automatically taken into account as some new information evidence to be incorporated in the question. It only means that the question asked was not correct (double directional rather than directional for example) and should be adjusted accordingly in

[26] Note that if all the moments fit, then the criteria is sound in the sense that measures such as mutual information or SINR (which are of interest in communications) will behave similarly.

[27] This question has to be taken in light of a talk “Where do we stand on maximum entropy?” made by E.T. Jaynes in 1978 at MIT [48].

[28] This point of view is not new and the misconception persists in many other fields. Descartes, already in 1637, warned us when stating in the first lines of the French essay “Le discours de la méthode”:“ la diversité de nos opinions ne vient pas de ce que les uns sont plus raisonnables que les autres, mais seulement de ce que nous conduisons nos pensées par diverses voies, et ne considérons pas les mêmes choses”.

order to imply a new model (based on some new source of information); and as it is well known, finding the right question is almost finding the right answer.

## References


1. G.J. Foschini and M.J. Gans, "On Limits of Wireless Communications in a Fading Environment when Using Multiple Antennas," *Wireless Personal Communications*, vol. 6, pp. 311–335, 1998.
2. K. Yu and B. Ottersten, "Models for MIMO Propagation Channels: A review," *Wireless Communications and Mobile Computing*, vol. 2, pp. 653– 666, November 2002.
3. H. Ozcelik, N. Czink, and E. Bonek, "What Makes a good MIMO Channel Model," in *Proceedings of the IEEE VTC conference*, 2005.
4. E. Biglieri, J. Proakis, and S. Shamai(Shitz), "Fading channels: Information-Theoretic and Communications Aspects," *IEEE Trans. on Information Theory*, vol. 44, no. 6, pp. 2619–2692, Oct. 1998.
5. S. N Diggavi, N. Al-Dhahir, A. Stamoulis, and A. R. Calderbank, "Great Expectations: The value of Spatial Diversity in Wireless Networks," in *Proc. of the IEEE*, 219-270, Feb. 2004.
6. G. Golden, C. Foschini, R. Valenzuela, and P. Wolniansky, "Detection Algorithm and Initial Laboratory Results using V-BLAST Space-Time Communication Architecture," *Electronics Letters*, vol. 35, no. 1, pp. 14–16, Jan. 1999.
7. P.W. Wolniansky, G.J. Foschini, G.D. Golden, and R.A. Valenzuela, "V-BLAST: An Architecture for Realizing Very High Data Rates Over the Rich-Scattering Wireless Channel," in *International Symposium on Signals, Systems, and Electronics*, 1998, vol. 4, pp. 295–300.
8. I.E. Telatar, "Capacity of Multi-Antenna Gaussian Channels," Technical report, AT & T Bell Labs, 1995.
9. X. Giraud and J.C Belfiore, "Constellations Matched to the Rayleigh Fading Channel," *IEEE Trans. on Information Theory*, pp. 106–115, Jan. 1996.
10. J. Boutros and E. Viterbo, "Signal Space Diversity: a Power and Bandwidth Efficient Diversity Technique for the Rayleigh Fading Channel," *IEEE Trans. on Information Theory*, pp. 1453–1467, July 1998.
11. C. E. Shannon, "A Mathematical Theory of Communication," *The Bell Labs Technical Journal*, pp. 379–457,623–656, July-October, vol. 27 1948.
12. E. T. Jaynes, "Information Theory and Statistical Mechanics, Part 1," *Phys. Rev*, vol. 106, pp. 620–630, 1957.
13. E. T. Jaynes, "Information Theory and Statistical Mechanics, Part 2," *Phys. Rev*, vol. 108, pp. 171–190, 1957.
14. J. P. Burg, *Maximum Entropy Spectral Analysis*, Ph.D. thesis, Stanford University, 1975.
15. A. Zellner, *An Introduction to Bayesian Inference in Econometrics*, J. Wiley and Sons, New York, 2nd edition, 1971.
16. J. N Kapur, *Maximum Entropy Models in Science and Engineering*, John Wiley and Sons, Inc, New York, 1989.
17. G. L. Bretthorst, *Bayesian Spectrum Analysis and Parameter Estimation*, Ph.D. thesis, Wahsington University, St. Louis, 1987.
18. H. Jeffrey, *Theory of Probability*, Oxford University Press, London, 1939, later editions, 1948, 1961.
19. J.M Keynes, *A Treatise on Probability*, MacMillan and Co., London, 1921.
20. E. T. Jaynes, *Probability Theory: The Logic of Science*, Cambridge, 2003.
21. J.E Shore and R.W Johnson, "Axiomatic Derivation of the Principle of Maximum Entropy and The Principle of Minimum Cross-Entropy," *IEEE Trans. on Information Theory*, pp. 26–36, Jan. 1980.
22. R. T Cox, *Probability, Frequency and Reasonable Expectation*, Am. Jour. Phys., 14:1-13 edition, 1946.
23. Bretthorst G. Larry, "An Introduction to Model Selection Using Probability Theory as Logic," *in Maximum Entropy and Bayesian Methods, G. R. Heidbreder (ed), Kluwer Academic Publishers, Dordrecht the Netherlands*, pp. 1–42, 1996.
24. A. Mohammad-Djafari and G. Demoment, "Utilisation de l'Entropie dans les Problèmes de Restauration et de Reconstruction d'Images," *Traitement du Signal*, vol. 5, pp. 235–248, 1998.
25. M.A. Xapsos, G.P. Summers, and E.A. Burke, "Probability Model for Peak Fluxes of Solar Proton Events," *IEEE Transactions on Nuclear Science*, vol. 45, pp. 2948–2953, 1998.
26. H. Jeffreys, *Theory of Probability*, Oxford University Press, London, later editions, 1948, 1961 edition, 1939.
27. J. Boutros and G. Caire, "Iterative Multiuser Joint Decoding: Unified Framework and Asymptotic Analysis ," *IEEE Trans. on Information Theory*, pp. 1772–1793, July 2002.
28. T. Cover and J. Thomas, *Elements of Information Theory*, Wiley, 1991.
29. M. Franceschetti, S. Marano, and F. Palmieri, "The role of entropy in wave propagation," in *IEEE International Symposium on Information Theory*, Yokohama, Japan, July 2003.
30. M. Debbah M. Guillaud and A. L. Moustakas, "Analytical Channels," in preparation, 2006.
31. K. Liu, V. Raghavan, and A. M. Sayeed, "Capacity Scaling and Spectral Efficiency in Wideband Correlated MIMO Channels," *IEEE Trans. on Information Theory*, pp. 2504 – 2526, Oct. 2003 2003.
32. R. Müller, "A Random Matrix Model of Communication via Antenna Arrays," *IEEE Trans. on Information Theory*, pp. 2495–2506, Sep 2002.
33. R. Müller, "On the Accuracy of Modeling the Antenna Array Channel with Random Matrices ," in *International Symposium on Wireless Personal Multimedia Communications*, Aalborg, Denmark, 2001.
34. A. M. Sayeed, "Deconstructing Multiantenna Fading Channels," *IEEE Trans. on Signal Processing*, pp. 2563–2579, Oct. 2002.
35. J.P. Kermoal, L. Schumacher, K.I Pedersen, P.E. Mogensen, and F. Frederiken, "A Stochastic MIMO Radio Channel Model with Experimental Validation," *IEEE Journal on Selected Areas in Communications*, pp. 1211–1225, vol. 20, no. 6 2002.

36. D. Chizhik, J. Lingand P.W. Wolnianski, R. A. Valenzuela, N. Costa, and K. Huber, "Multiple-Input Multiple Output Measurements and Modeling in Manhattan," *IEEE Journal on Selected Areas in Communications*, vol. 21, no. 3 2002.
37. S.J. Fortune, D.H. Gay, B.W. Kernighan, O. Landron, R.A Valenzuela, and M. H. Wright, "WiSE Design of Indoor Wireless Systems: Practical Computation and Optimization," *IEEE Comput. Sci. Eng.*, vol. 2, pp. 58–68, Mar. 1995.
38. H. Ozcelik, M. Herdin, W. J. Weichselberg, and E. Bonek, "Deficiencies of "Kronecker" MIMO Radio Channel Model," *IEE Electronics Letters*, vol. 39, no. 16, pp. 1209–1210, Aug. 2003.
39. T.S. Pollock, "Correlation Modelling in MIMO Systems: When can we Kronecker?," *Australian Communications Theory Workshop*, vol. Newcastle, NSW, no. Australia, pp. 149–153, 2004.
40. D. Gesbert, H. Bölcskei, D. Gore, and A. Paulraj, "MIMO Wireless Channels: Capacity and Performance Prediction," *GLOBECOM conference records*, vol. 2, pp. 1083–1088, 2000.
41. D. Chizhik, G.J Foschini, M. J. Gans, and R. A. Valenzuela, "Keyholes, Correlations and Capacities of Multielement Transmit and Receive Antennas," *IEEE Trans. on Wireless Communications*, vol. 1, no. 2, pp. 361–368, Apr. 2002.
42. D. Gesbert, H. Bolcskei, D.A Gore, and A.J Paulraj, "Outdoor MIMO Wireless Channels: Models and Performance Prediction," *IEEE Trans. on Communications*, vol. 50, no. 12, pp. 1926–1934, Dec. 2002.
43. J. Capon, "High-Resolution Frequency Wavenumber Spectrum Analysis," *Proceedings of the IEEE*, vol. 8, no. 57, pp. 1408–1418, 1969.
44. R.O Schmidt, "Multiple Emitter Location and Signal Parameter Estimation," *Proceedings of RADC: Spectral Estimation Workshop, Rome*, pp. 243–258, 1979.
45. R. Kumaresan and D. W Trufts, "Estimating the angles of arrival of multiple plane waves," *IEEE Transactions on Aerospace and Electronic Systems*, pp. 134–139, 1983.
46. R. Roy and T. Kailath, "ESPRIT-Estimation of Signal Parameters via Rotational invariance techniques," *IEEE Transactions on Signal Processing*, vol. 7, pp. 984–995, 1989.
47. R. Müller and H. Hofstetter, "Confirmation of Random Matrix Model for the Antenna Array Channel by Indoor Measurements," in *IEEE Antennas and Propagation Society International Symposium, vol.1, pp.472-475*, Boston, Ma. USA, June 2001.
48. E. T. Jaynes, "Where Do We Stand on Maximum Entropy?," *in The Maximum Entropy Formalism,R. D. Levine and M. Tribus (eds.), M. I. T. Press, Cambridge, MA,*, p. 15, 1978.

# 2

# Unsupervised Signal Processing: Concepts, Applications and Trends

Ricardo Suyama[1], Leonardo T. Duarte[1], Aline Neves[1], Rafael Ferrari[1], Romis R. F. Attux[1], Charles C. Cavalcante[2], Cynthia Junqueira[1,3], and João Marcos T. Romano[1]

[1] Laboratory of Signal Processing for Communications (DSPCom) - FEEC - Unicamp, Brazil
{rsuyama, ltduarte, aline, rferrari, romisri, cynthia, romano}@decom.fee.unicamp.br
[2] Wireless Telecommunications Research Group, Federal University of Ceará, Brazil
charles@gtel.ufc.br
[3] General-Command of Aerospace Technology (CTA/IAE), Brazil
cynthia@iae.cta.br

## 2.1 Introduction

A recurrent necessity in signal processing is that of extracting or restoring information from a corrupted version thereof. This fundamental requirement is embodied in the problems of blind equalization and blind source separation (BSS), on which it can be said that the theory of unsupervised adaptive filtering is based. This theory encompasses models and tools that are very general in the sense that a wide variety of problems can be fit into their framework: communications, biomedical, pattern recognition, econometrics and image processing, among several others. This fact in itself is indicative of how important is the research on the topic.

Interestingly enough, the development of the theory subjacent to these most interrelated problems took place along different lines: while most techniques for blind equalization were conceived in the context of a classical SISO (single-input /single-output) model, BSS evolved basically under the aegis of formulations of a purely spatial character. Two decades of efforts led these fields of research into a significant degree of maturity; nevertheless, the possibility of comparing the development of these branches and taking advantage of the potential synergy between them, which could be decisive to enrich and facilitate their development, still remains to be fully exploited.

Therefore, the objective of this tutorial is twofold: in the first part, the idea is to provide an overview of both problems, setting a basic theoretical foundation that will allow a non-expert audience to understand them; in the second part, some important recent developments, theoretical connections and modern trends are analyzed, which include:

- The relationships between blind equalization criteria and between blind and supervised equalization criteria.
- The problem of blind nonlinear equalization based on prediction-error filters.
- Blind source separation in nonlinear, convolutive and underdetermined contexts.
- An investigation of the relationships between blind equalization and blind source separation.
- The use of advanced mathematical tools in the study of source separation techniques, such as algebraic methods.
- The application of evolutionary optimization techniques to the problems of equalization and source separation.

Finally, it is important to remark that this two-part tutorial was conceived under the auspices of two hopes:

- That it be useful to those who are not familiarized with the problems of equalization and source separation, which justifies the inclusion of a first part devoted to their foundations.
- That it be attractive to experts in both areas, which explains the structure of the second part, in which modern (and even open) topics are introduced.

The tutorial is organized as follows. Firstly, we expose the basics of the problem of SISO channel equalization. Afterwards, in Section 2.3, we review some concepts about multichannel equalization. In Section 2.4, we present the fundamentals of the blind source separation problem. The second part of this document begins in Section 2.5, in which some novel aspects concerning the problem of equalization are addressed. In Section 2.6, we provide an overview about particular models in the BSS problem. A discussion about some relationships between the problems of equalization and blind source separation is presented in Section 2.7. In Section 2.8, we highlight some modern trends in the field of unsupervised signal processing.

## 2.2 Equalization/Deconvolution: Single-Input / Single-Output (SISO) Systems

The objective of a communication system is to provide the means whereby information may be properly interchanged. The process as a whole is conceived as having two poles - a group of transmitters and a group of receivers - which are interconnected by a channel, as shown in Fig. 2.1.

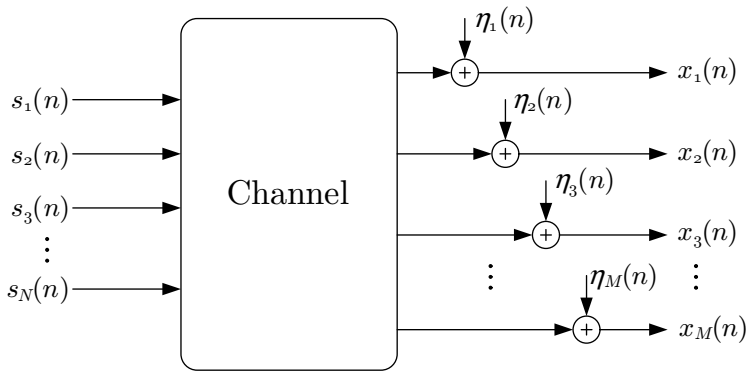



**Fig. 2.1.** MIMO channel with $N$ transmitters and $M$ received signals.

In Fig. 2.1, there are $N$ transmitted signals and $M$ received signals, a condition that characterizes a multiple-input / multiple-output (MIMO) system. From this broad notion, it is possible to reach some relevant particular cases [1]:

- SIMO (single-input / multiple-output) systems, when $N = 1$ and $M > 1$;
- MISO (multiple-input / single-output) systems, when $N > 1$ and $M = 1$;
- SISO (single-input / single-output) systems, when $N = 1$ and $M = 1$.

Initially, we will be exclusively concerned with systems belonging to the last class, particularly due to their conceptual simplicity and historical precedence. In sections 2.3 and 2.4, systems with multiple inputs and/or multiple outputs will be studied in more detail.

### 2.2.1 The channel and its effects

In general, the channel through which the information is interchanged possesses features that are responsible for modifying the transmitted signal. These distortions, if not properly dealt with, may cause the received message to be reconstructed with an unacceptable degree of imprecision.

In order to illustrate this point, let us assume that the communication channel is modeled as a linear and time-invariant (LTI) system whose output is added to a stochastic process that represents the thermal noise in order to form the received signal $x(n)$:

$$x(n) = \sum_{k=-\infty}^{\infty} h(k)s(n-k) + \eta(n), \tag{2.1}$$

where $h(n)$ is the channel impulse response and $\eta(n)$ is the additive noise. If we rearrange the terms of (2.1) in order to emphasize the presence of the transmitted signal $s(n)$[4], we will get to:

$$x\left(n\right) = h\left(0\right)s\left(n\right) + \sum_{k=-\infty, k\neq 0}^{\infty} h\left(k\right)s\left(n-k\right) + \eta(n). \tag{2.2}$$

It is noticeable that the received signal is formed by a combination of three terms: the message of interest $s(n)$, the additive noise, of whose presence we were already aware, and a term

$$\sum_{k=-\infty, k\neq 0}^{\infty} h\left(k\right)s\left(n-k\right), \tag{2.3}$$

which corresponds to a superposition of delayed versions of the transmitted signal. This term is the mathematical expression of the so-called intersymbol interference (ISI), a major concern in digital band-limited systems [2, 3].

[4] We will treat the non-delayed signal $s(n)$ exclusively for the sake of simplicity, since, as it will become clearer in the following, any equalization delay would be acceptable.

### 2.2.2 The idea of equalization

The model we have just exposed, which is the SISO channel model we will adopt in the course of the entire work, unveils the existence of two noxious effects that are potentially capable of degrading communication: ISI and noise. Now, it is time for us to analyze a classical approach to the problem of mitigating them[5]: the use of an equalizer.

An equalizer can be defined as a specially tailored filter whose objective is to process the received signal in order to generate an output signal as close as possible to the original message. In mathematical terms, the goal of an equalizer is to obtain

$$y\left(n\right) = \alpha s\left(n-d\right), \tag{2.4}$$

where $y(n)$ is the equalizer output, $\alpha$ is a constant and $d$ is the so-called equalization delay.

Since the equalizer is a filter, and, moreover, it is up to us the task of carrying out its project, there are three questions that must be answered: 1) what filtering structure will be adopted? 2) What criterion will guide the choice of its free parameters? 3) How will the optimal solution be found? In this section, we will give a classical answer to the first question: the equalizer is assumed to be a linear and time-invariant (LTI) filter with finite impulse response (FIR). The rest of our discussion will be devoted to the remaining questions.

#### The Zero-forcing criterion

If we ignore for a moment the existence of additive noise, we will be faced with a scenario in which channel and equalizer form a cascade of two LTI systems. Let us assume that the transfer function of the channel is $H(z)$ and that the transfer function of equalizer is $W(z)$. Under these circumstances,

$$Y(z) = H(z)W(z)S(z). \tag{2.5}$$

Applying the z-transform to the ideal condition (2.4), we get to:

$$Y(z) = \alpha z^{-d} S(z). \tag{2.6}$$

By replacing (2.6) into (2.5), we find the equalizer transfer function that produces the ideal condition:

$$W(z) = \frac{\alpha z^{-d}}{H(z)}. \tag{2.7}$$

This transfer function shows that the equalizer must be, in a certain sense, the inverse of the channel - a quite satisfactory result in intuitive terms. In the time domain, the above condition leads to $w(n) * h(n) = [\,0 \cdots 0\,1\,0 \cdots 0]$, which means that the combined impulse response is forced to assume a value equal to zero in all instants, except $n = d$, wherefore the solution expressed in (2.7) is referred to as zero-forcing (ZF) [1]. The ZF solution is very important from the analytical standpoint, but, in practice, it suffers from two drawbacks: 1) in the most common scenario - that in which both channel and equalizer are FIR filters - the solution is unattainable, and 2) it does not take noise into account, which may be disastrous in some cases.

#### The Wiener criterion

The Wiener criterion is arguably the cornerstone of the entire optimal filter theory. In the context of equalization, its goal is identical to that of the ZF criterion: to seek the ideal condition presented in (2.4). However, contrarily to the ZF approach, the Wiener criterion adopts a perspective that has two strong points: 1) it encompasses the idea of proximity between a given case and the ideal one, and 2) it employs a statistical measure. These positive aspects allow the application of the Wiener criterion to a wide range of structures and, in addition to that, eliminate any restrictions to the presence of randomness or noise.

The Wiener cost function is given by:

$$J_{Wiener} = E\left[\left|s\left(n-d\right) - y\left(n\right)\right|^2\right] = E\left[\left|e\left(n\right)\right|^2\right], \tag{2.8}$$

where $e(n)$ represents the error between the output signal and the desired response. The reader should notice that $J_{Wiener}$ is simply a measure of the distance between the two members of (2.4) in the ideal case in which $\alpha = 1$,

[5] Notwithstanding the fact that an equalizer is able to mitigate noise-related effects, it is more commonly associated with the task of reducing ISI. Accordingly, our analysis will focus on the latter sort of distortion.

which attests the similarity between this approach and the ZF solution. Since the equalizer is assumed to be a FIR filter, it is straightforward to obtain the optimal parameter vector, i.e., the Wiener solution:

$$\mathbf{w} = \mathbf{R}^{-1}\mathbf{p}, \tag{2.9}$$

where $\mathbf{R}$ is the correlation matrix of $x(n)$, and $\mathbf{p}$ is the cross-correlation vector between the received and transmitted signals. Thus, the optimal parameter vector depends on the correlation structure of the received signal and on the correlation between the received signal and the transmitted message (which is essential, since the filter will attempt to "build the desired signal from the received signal"). In practice, this solution is usually searched in an iterative basis with the help of the LMS (Least Mean Square) or the RLS (Recursive Least Squares) algorithm [1].

The Wiener approach is a simple and elegant solution to the equalization problem. However, it should not be overlooked that it is a solution founded on an implicit assumption: that it is possible to have access, at the receiver, to samples of the desired signal. This hypothesis, which explains why the Wiener criterion belongs to the class of supervised paradigms, may be a serious problem in a number of real-world applications. This limitation is the *raison d'être* of the so-called unsupervised or blind techniques, which will be the object of our subsequent discussions.

**Blind Equalization via Linear Prediction**

Let us analyze the problem we have just raised to find suitable blind equalization criteria. A classical "first solution" to this problem is to use a linear prediction-error filter in the role of equalizer. In order to understand this proposal, we should consider the prediction[6] task in itself.

The problem of predicting the future of a time series is commonly formulated as that of finding a filter that minimizes the Wiener-like cost function:

$$J_{Prediction} = E\left[\left|x\left(n+1\right) - y\left(n\right)\right|^2\right] = E\left[\left|e_{Prediction}\left(n\right)\right|^2\right], \tag{2.10}$$

where $y(n) = f\left[x(n), x(n-1), \cdots, x(n-k)\right]$, a function of past samples of the received signal. If the filtering structure is assumed to be linear, its optimal parameters can be calculated via the direct application of (2.9). In this case, the prediction error can be defined as being:

$$e_{Prediction} = x\left(n+1\right) - y\left(n\right). \tag{2.11}$$

The prediction-error filter is simply a device whose input is the received signal $x(n)$ and whose output is the prediction error $e_{Prediction}(n)$.

At this point, it is important to highlight two properties of prediction-error filters [4, 1]:

- A prediction-error filter is, at least in the limit of a structure with a sufficient number of parameters, a whitening filter. In other words, the output of a PEF tends to be an uncorrelated random process.
- A (forward) prediction-error filter is a minimum-phase device.

The first of these properties is essential from the standpoint of blind equalization. Since, as a rule, the transmitted signal $s(n)$ is considered to be formed by i.i.d. samples, it is tempting to imagine the possibility of using a whitening filter such as the PEF as an equalizer tuned to invert the action of the communication channel by eliminating the correlation it introduces. However, the reader should notice that to produce uncorrelated samples is different from producing independent samples [5]. This essential limitation, which will play a key role in many parts of this tutorial, dramatically reduces the scope of application of a blind equalizer based on second-order statistics (SOS). The use of SOS simply is not enough to recover the character of independence that is inherent to the transmitted signal in a general scenario. As a matter of fact, PEF are effective only in the context of minimum-phase channels, which is in consonance with the second property presented above [4].

**Blind Equalization: the Benveniste-Goursat-Rouget and Shalvi-Weinstein theorems**

Since the use of second-order statistics is not enough to produce an unsupervised paradigm that is both sound and general, it is time for us to look for more solid bases. This solidity is provided by a pair of results on which the entire blind equalization theory is founded: the Benveniste-Goursat-Ruget (BGR) and the Shalvi-Weinstein (SW) theorems. In order to understand these results, let us firstly present their fundamental hypotheses [6, 7]:

[6] In this section, we will consider, without loss of generality, that the term "prediction" refers to forward prediction.

- The transmitted signal is composed of nongaussian i.i.d. samples;
- The channel and the equalizer are LTI filters, and the additive noise is negligible;
- It is possible to reach a condition of perfect inversion (i.e. a ZF condition).

The first two assumptions are responsible for establishing a scenario in which the transmitted samples are independent and nongaussian[7] and the channel is conceived as a device that generates intersymbol interference. The last condition indicates that the theorems will emphasize the statistical aspects of blind equalization in an ideal (from the structural standpoint) case.

Having thus set the scene, we are ready to state and discuss the Benveniste-Goursat-Ruget (BGR) theorem [6]:

**Theorem 1 (Benveniste-Goursat-Ruget).** *Under the previously defined conditions, if the probability density functions of the transmitted signal and of the equalizer output are equal, then the ZF solution is necessarily attained.*

The most important aspect of the BGR theorem is the perspective it opens of expressing a ZF (i.e. an ideal) configuration without making use of the channel impulse response and of an error signal: the rule is to seek the set of parameters that engenders an output signal whose pdf matches the pdf of the transmitted signal. This result is crucial, since it discloses the possibility of obtaining an efficient equalizer with the sole aid of statistical properties. Moreover, since knowledge of a pdf implies knowledge of all of its moments [5], we have overcome the SOS limitation that characterizes the prediction-based approach described in the last section. Finally, the demand for nongaussian signals can be understood in a straightforward manner: since a linearly filtered Gaussian process remains Gaussian, the idea of matching pdf's would be, in such a case, restricted to power normalization.

The BGR theorem was responsible for demonstrating the viability of the idea of blind equalization. However, a decade later, Shalvi and Weinstein demonstrated that the condition of equality between the pdf's of the transmitted and estimated signal, albeit solid, was excessively stringent. Under the same conditions, the authors arrive at a blind expression of the ZF condition with the help of a less restrictive amount of information concerning the involved signals: their cumulants.

Cumulants are statistical measures derived from the characteristic function [5]. We denote an order $(p, q)$ cumulant as $C_{p,q}^{y(n)}$. Until third order, i.e., $p+q \leq 3$, the cumulants are equal to the moments of a random variable. Thus, $C_{1,1}^{y(n)}$, e.g., for a zero mean signal, is equal to its variance. The forth order cumulant, $C_{2,2}^{y(n)}$, is called kurtosis. Its definition based on moments is given by:

$$C_{2,2}^{y(n)} = K\left[y\left(n\right)\right] = E\left[\left|y\left(n\right)\right|^4\right] - 2E^2\left[\left|y\left(n\right)\right|^2\right] - \left|E\left[y^2\right]\right|^2 \tag{2.12}$$

With this concepts defined, we are now in position to state the Shalvi-Weinstein theorem [7]:

**Theorem 2 (Shalvi-Weinstein).** *Under conditions 1, 2 and 3, if* $E\left\{\left|\, s(n)\,\right|^2\right\} = E\left\{\left|\, y(n)\,\right|^2\right\}$ *and* $\left|K\left[y(n)\right]\right| = \left|K\left[s(n)\right]\right|$*, then the ZF solution is necessarily attained.*

This result shows that, in order to express the ZF condition in statistical terms, it is not necessary to match all the moments of the two signals of interest: it suffices to take into account the second-order moment and the fourth-order information brought by the kurtosis. In intuitive terms, this means that, after a sort of power normalization, a higher-order statistic can be used to conclude the task of finding an effective equalizer.

### Blind Equalization Criteria

The BGR and SW theorems are results that demonstrate the effectiveness of the idea of blind equalization and indicate what type of statistical entity may be employed to counterbalance the channel effects. Thus, when it comes to building equalization criteria and algorithms, the theorems are, in essence, two interesting possible starting points. There are techniques, such as the decision-directed, Sato and Godard criteria, which were proposed without making explicit use of the framework established by theorems; conversely, there is a family of algorithms associated with the Shalvi-Weinstein theorem. The objective of this section is to expose the approaches that form, in a certain sense, the class of the most widely employed blind equalization solutions. Later, we will have more to say on the properties of these techniques, as well as on their relationships.

[7] The nongaussianity restriction is, as the reader will notice, recurrent in blind equalization and source separation.

*The Decision-directed (DD) approach*

The origins of the decision-directed (DD) approach [8] are related to the need for simple practical solutions that characterized the development of the field of digital communications. Essentially, the DD criterion can be thought of as a sort of "modified Wiener criterion" in which the desired signal is replaced by the estimate provided by the decision-device. Mathematically, its cost function is defined as:

$$J_{DD} = E\left[\left|dec\left(y\left(n-d\right)\right) - y\left(n\right)\right|^2\right], \tag{2.13}$$

where $dec\left(\cdot\right)$ denotes the decision function. Calculating the gradient of this function with respect to the parameters of the equalizer and applying the usual stochastic approximation, we obtain the decision-directed algorithm[8, 1].

The idea behind this proposal becomes clearer if we imagine that it is possible to have access to samples of the transmitted signal during a relatively small training period, and that, afterwards, the process of adaptation must proceed in a blind fashion. If the supervised stage is capable of engendering a situation in which the equalizer "opens the eye pattern", it is viable to assume that the message reconstructed at the decision-device output will be accurate enough to guide the blind adaptation process. A modification of this idea gives rise to the Sato algorithm [9], which, in the context of a system endowed with a multilevel modulation (e.g. 8-PAM), attempts to recover exclusively the most significant bit via a DD-like scheme.

*Godard - Constant Modulus algorithm*

The Godard algorithm, first proposed by Dominique Godard in 1980 [10], is probably the most studied blind equalization algorithm. It is based on the following criterion:

$$J_{Godard} = E\left[\left|y\left(n\right)\right|^p - R_p\right]^q, \tag{2.14}$$

where $R_p$ is a constant that depends on the kind of modulation employed in the transmission. This cost function is based on the notion of minimizing the equalizer output dispersion around a constant value dependent on the characteristics of the adopted modulation. An important property of the Godard criterion is that it is "phase-independent", which is useful to dissociated the equalization task from that of carrier recovery. By far, the most common choice for $p$ and $q$ is $p = q = 2$, in which case (2.14) is called constant modulus (CM) criterion, and the associated stochastic gradient algorithm is known as the constant modulus algorithm (CMA)[10, 11].

*Shalvi-Weinstein : The Super Exponential Algorithm*

Another important family of algorithms used in blind equalization is directly obtained from the Shalvi-Weinstein theorem. Let the equalizer output be expressed as

$$y(n) = \sum_{k=-\infty}^{\infty} c(k)x(n-k), \tag{2.15}$$

where $c(n)$ is the combined channel-equalizer response $c(n) = w(n) * h(n)$. It is possible to show that the relashionship between the $(p, q)$ order cumulants of $s(n)$ and $y(n)$ is given by

$$\left|C_{p,q}^{y(n)}\right| = \left|C_{p,q}^{s(n)}\right| \left|\sum_k \left|c\left(k\right)\right|^4\right| \leq \left|C_{p,q}^{s(n)}\right| \left|\sum_k \left|c\left(k\right)\right|^2\right|. \tag{2.16}$$

Following Theorem 2, the first condition for obtaining a ZF solution is that $E\left\{\mid s(n) \mid^2\right\} = E\left\{\mid y(n) \mid^2\right\}$. If that is the case, from (2.16) we can conclude[8] that $\sum_k \left|c\left(k\right)\right|^2 = 1$, which means that the last inequality will become an equality only if the vector of the combined channel-equalizer response with a single nonzero element with magnitude equal to one (i.e., a ZF solution).

Based on (2.16) and on the discussion above, the criterion proposed by Shalvi and Weinstein [7] is to maximize $\left|C_{p,p}^{y(n)}\right|$, with $p + q > 2$ (higher-order cumulant), subject to the power restriction. The adopted values for $p$ and $q$ are $p = 2$, $q = 1$ (third-order cumulant) and $p = 2$, $q = 2$ , which results in the kurtosis. The latter is used more often because the third-order cumulants of symmetric distributions are zero.

[8] Remembering that $E\left\{\mid s(n) \mid^2\right\} = C_{1,1}^{s(n)}$ and $E\left\{\mid y(n) \mid^2\right\} = C_{1,1}^{y(n)}$.

The power restriction may also be substituted by a normalization what gives rise to the following criterion: maximization of $\frac{\left|C_{p,p}^{y(n)}\right|}{\left(C_{1,1}^{y(n)}\right)^{\frac{p+q}{2}}}$. This criterion is used for the deduction of the super-exponential algorithm, also proposed by Shalvi and Weinstein [12].

## 2.3 Multichannel and Multiuser Systems

Section 2.2 was dedicated to presenting the fundamentals of the blind equalization of SISO systems. Henceforth, we shall discuss some aspects involving the generalization of this problem to the cases of SIMO and MIMO systems. In a first moment, we explain how the SIMO framework allows us to surmount a structural problem present in the SISO context. Afterwards, we turn our attention to the MIMO case and briefly expose the tasks of spatial filtering and multiuser detection.

### 2.3.1 Equalization of a SIMO Channel and Fractionally-Spaced Equalizers

In Section 2.2, we remarked that, in a practical SISO scenario, it is impossible to attain the ZF condition, since the action of an FIR channel cannot be completely inverted using an FIR equalizer. Naturally, a sufficiently large equalizer can be, in theory, employed to reach a desired degree of accuracy in the process of inversion. However, this approach can be quite unpractical if one also looks for a receiver that be as simple as possible, which is indeed fundamental in communications. Fortunately, as we will show in the following, it may be possible, in the context of SIMO systems, to reach a ZF condition by using a set of FIR filters.

This can be achieved, for instance, via the notion of oversampling. In order to illustrate the idea, which is simply that of sampling the received signal with a rate $P$ times higher than the baud rate $1/T$, let us firstly consider the expression of the received signal, in the absence of noise, for the classical SISO channel discussed in Section 2.2

$$x(nT) = \sum_{k=-\infty}^{\infty} h((n-k)T)s(k). \tag{2.17}$$

Supposing that the new sampling rate is $T_s = T/P$, we may write

$$x(nT_s) = x(n\frac{T}{P}) = \sum_{k=-\infty}^{\infty} h(\frac{nT}{P} - kT)s(k). \tag{2.18}$$

It is possible to show that the samples associated with the *p*-th sequence obey the general form

$$x_p(nT) = x(nT + \frac{pT}{P}) = x((kP+p)T_s) \tag{2.19}$$

From (2.18), the following the expression

$$h_p(nT) = h(t)\Big|_{t=T_0+nT+\frac{p-1}{P}},\ p = 0, \ldots, (P-1) \tag{2.20}$$

can be interpreted as the channel discrete-time impulse response associated with the *p*-th sequence. Therefore, it is possible to view the oversampled signal as being formed by the concatenation of baud-spaced samples filtered by $P$ of these subchannels, as shown in Fig. 2.2.

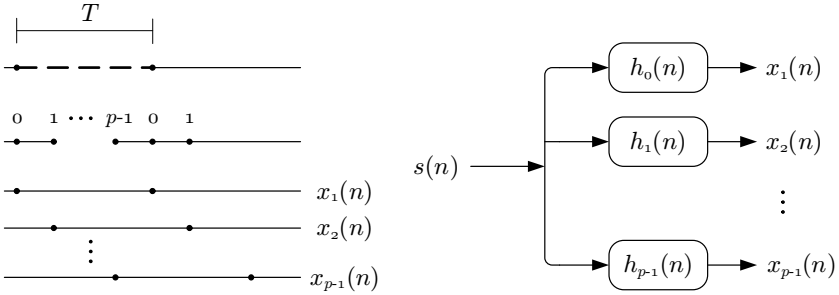



**Fig. 2.2.** SIMO channel/equalizer.

Under these circumstances, a straightforward solution is a FIR filter that process the oversampled sequence, i.e., a fractionally-spaced equalizer. This device can be understood as a set of baud-spaced subequalizers, as shown in Fig. 2.3.

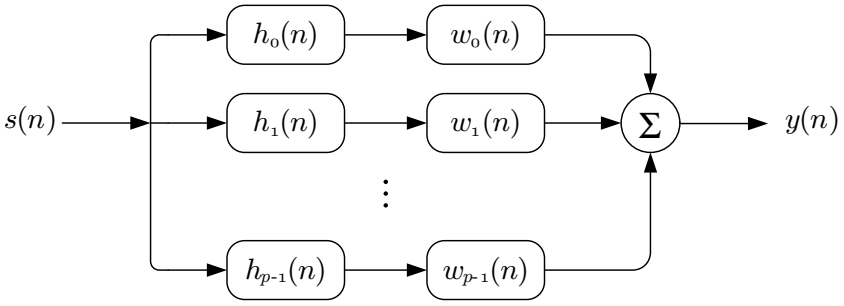



**Fig. 2.3.** SIMO channel with $P$ subchannels.

Notwithstanding the solidity of these ideas, a relevant question can be asked: under what conditions can a ZF solution be attained in the above described framework? The key to answering it is the so-called Bézout's identity[9], which can be enunciated in simple terms as: if the $P$ subchannels have no roots in common, it is possible to obtain a set of sub-equalizers such that a perfect equalization condition is reached, i.e.,

$$W^H(z)H(z) = 1, \tag{2.21}$$

where the set of transfer functions of the $P$ sub-channels by is represented by the vector of polynomials $H(z) = [H_1(z), \ldots, H_P(z)]^T$. Analogously, each element of the vector of polynomials $W(z) = [W_1(z), \ldots, W_P(z)]^T$ corresponds to a transfer function associated with one sub-equalizer. In intuitive terms, this identity simply reveals that, if we wish to invert the channel by passing from a SISO model to a SIMO model, it is important that the oversampling process bring "non-redundant" information to the scene (which justifies the demand that there be no zeros in common).

Fractionally-spaced equalizers can be adapted via the same sort of criteria discussed in Section 2.2. For the cases in which the Bézout's identity is valid, and, consequently, a ZF solution is attained, the convergence analysis of algorithms like the CMA resembles the efforts of Godard and Foschini, both in terms of mathematical tractability and in the nonexistence of local minima [14, 15].

### 2.3.2 MIMO Channels, Antenna Arrays and Multiuser Detection

After having discussed the essential aspects of SISO and SIMO equalization, it is time for us to turn our attention to the most general case: that of MIMO systems. The MIMO models of both channel and equalizer can be described, in simple terms, as a combination of multiple SIMO vector equalizers [15], as shown in Fig. 2.4.

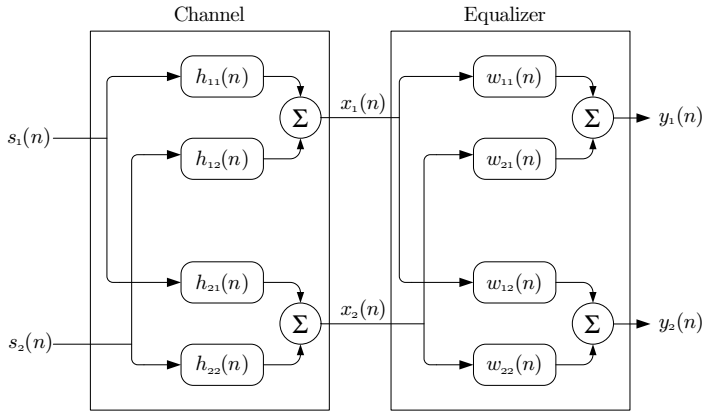



**Fig. 2.4.** Example of a MIMO channel.

[9] It is possible to generalize the Bezout's identity as follows: $W^H(z)H(z) = z^{-d}$ [13].

For this model, the channel output can be defined as

$$\mathbf{x}\left(n\right) = \mathbf{H}\left(n\right) * \mathbf{s}\left(n\right) = \sum_{k} \mathbf{H}\left(k\right) \mathbf{s}\left(n-k\right) \tag{2.22}$$

where $\mathbf{H}\left(n\right)$ represents the channel response. Analogously, the equalizer output is

$$\mathbf{y}\left(n\right) = \mathbf{W}\left(n\right) * \mathbf{x}\left(n\right) = \sum_{k} \mathbf{W}\left(k\right) \mathbf{x}\left(n-k\right) \tag{2.23}$$

Given (2.22) and (2.23), it is possible to define the ideal (ZF) equalization condition as:

$$\mathbf{W}(z)\mathbf{H}(z) = \mathbf{P}\mathbf{D}(z) \tag{2.24}$$

where $\mathbf{W}(z)$ and $\mathbf{H}(z)$ represent the z-transform of, respectively, $\mathbf{W}(n)$ and $\mathbf{H}(n)$, $\mathbf{P}$ represents a permutation matrix, and $\mathbf{D}(z)$ a diagonal matrix, whose elements are arbitrary filters[10].

The possibility of reaching the ZF condition is, in this case, related to the rank of the frequency-response matrix of the channel [15]. As it was the case in the previous sections, the analysis of blind algorithms like the CMA will be heavily dependent on the attainability of an ideal solution.

### Adaptive Antenna Arrays

A classical particular case of the presented MIMO model emerges when $M$ transmitted signals are captured by an array of $N$ antennas, each one capable of modifying the gain and the phase of an incident waveform. The most common situation is that in which the array is composed of sensors that are equally spaced and linearly disposed, as shown in Fig. 2.5.

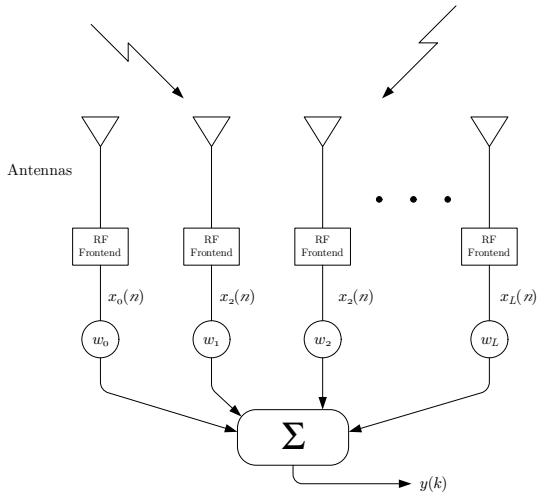



**Fig. 2.5.** Linear antenna array.

In this context, the received signal consists of a purely spatial superposition of transmitted waveforms:

$$\mathbf{x}(n) = \mathbf{H}\mathbf{s}(n) \tag{2.25}$$

where $\mathbf{H}$ is a matrix that contains all relevant information about the directions of arrival (DOAs) of the incident signals and their relationship with the geometry of the array [1].

Several algorithms can be used to adapt the coefficients of an antenna array. The Wiener approach and all blind techniques discussed in Section 2.2 can be, in theory, applied. Furthermore, there are some solutions that were specifically developed to operate in the environment established by (2.25), such as the Applebaum [16], Frost [17] and Resende [18]algorithms.

[10] It is interesting that the reader observe the similarity between (2.24) and the perfect separation condition that will be presented in section 2.4.

Finally, it is important to cite another related problem, that of multiuser detection, which is typically formulated in terms of a MIMO channel whose input is formed by several digitally-modulated signals that must be separated at the receiver. The study of the solutions to this task is an interesting bridge between the worlds of equalization, which we have analyzed so far, and of source separation, which is the domain we are about to enter [19].

## 2.4 Blind Source Separation

After describing the basics of the multichannel problem, we are now in a position to introduce the problem of blind source separation (BSS). The origin of this subject dates back to the work of Hérault, Jutten and Ans [20], who addressed the biological problem of modeling the coding related to the muscle contraction phenomena. Since then, BSS has been attracting a lot of attention from many research communities, mainly due its generical formulation, that broaden the range of applications in which the tools developed under this framework can be applied. For instance, typical BSS problems can be found in biomedical signal processing [21], telecommunications [19] and audio processing [22]. Furthermore, there are many others applications of ICA [23] (Independent Component Analysis), a data analysis technique that is closely related to BSS. In the following, we shall present the BSS basics, as well the ICA-based solutions to such problem.

Let $\mathbf{s}(n) = [s_1(n), s_2(n), \ldots, s_N(n)]^T$ denote $N$ source signals and $\mathbf{x}(n) = [x_1(n), x_2(n), \ldots, x_M(n)]^T$ be the $M$ mixtures of these sources, i.e.,

$$\mathbf{x}(n) = \mathcal{F}(\mathbf{s}(n)), \tag{2.26}$$

where the mapping $\mathcal{F}(\cdot)$ models the action of a mixture system. The model also represents the general case of a MIMO system, as discussed in Section 2.3. Typically, it is assumed that the sources are statistically independent. The aim of a BSS technique is to recover the source signals by adjusting the parameters of a separating system based solely on the observed samples of the mixtures. Mathematically, the recovering process is expressed by

$$\mathbf{y}(n) = \mathcal{G}(\mathbf{x}(n)), \tag{2.27}$$

where $\mathbf{y}(n) = [y_1(n), y_2(n), \ldots, y_N(n)]^T$ denotes the recovered sources and $\mathcal{G}(\cdot)$ models the action of the separating system. This situation is depicted in Fig. 2.6.

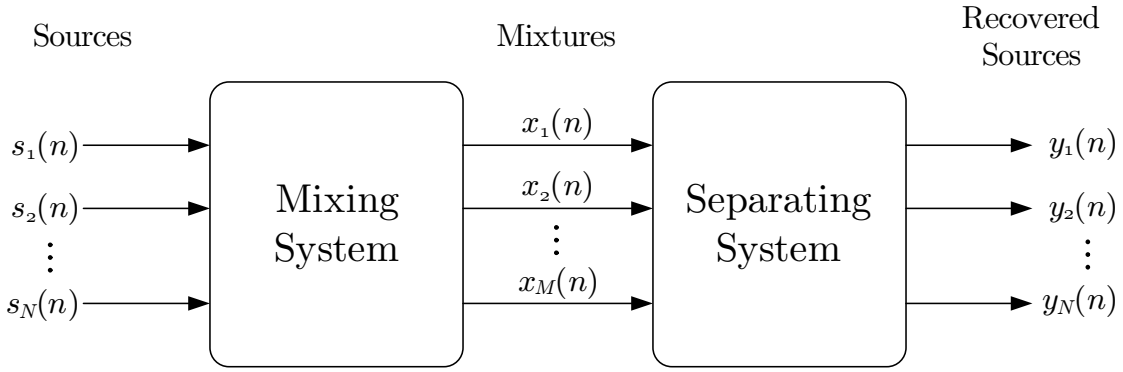



**Fig. 2.6.** The problem of blind source separation.

The simplest form of the BSS problem takes place when the mixture process is modeled as a linear and instantaneous system. Also, it is usually assumed that the number of sensors is equal to the number of sources ($N = M$). Therefore, the mixture processing in this situation can be described as follows[11]:

$$\mathbf{x} = \mathbf{A}\mathbf{s}, \tag{2.28}$$

where the mixing matrix $\mathbf{A}$ is of dimension $N \times N$. It is quite natural that the structure of the separating system in this case be given by

$$\mathbf{y} = \mathbf{W}\mathbf{x}, \tag{2.29}$$

[11] For the sake of notation simplicity, the time index $n$ will be omitted here and in the subsequent parts of this section.

where $\mathbf{W}$ is a $N \times N$ matrix. Ideally, it is expected that the separating system invert the action of the mixture process, i.e., that $\mathbf{W} = \mathbf{A}^{-1}$. Nevertheless, it would be equally satisfactory if one could recover a scaled version of the source vector, or even a permutation of it. These conditions can be expressed in mathematical terms as

$$\mathbf{W} = \mathbf{DPA}^{-1}, \tag{2.30}$$

where $\mathbf{P}$ and $\mathbf{D}$ denote a permutation and an invertible diagonal matrix, respectively. Thus, any valid BSS criterion should lead to a solution in accordance to (2.30). In the sequel, we shall discuss the crucial point of adjusting the matrix $\mathbf{W}$ so that this desirable condition be attained.

### 2.4.1 Second-Order Statistics Approach

As remarked in [24], until the 1980's it was usual, in the signal processing community, to merge the concepts of statistical independence and uncorrelatedness between random variables. This was a consequence of the widespread use of Gaussian models of signal, for which case the notions of independence and uncorrelatedness are equivalent. In this context, a first idea [20] to adjust the matrix $\mathbf{W}$ relied on the following principle: given that the sources are mutually independent, and, thus, uncorrelated, the mixing system can be viewed as a correlator, that is, a system that correlates the input signals. Therefore, it would be possible to separate the sources by guiding the adjustment of $\mathbf{W}$ in order that the estimates of the sources be uncorrelated too, in a kind of uncorrelatedness recovering[12].

This procedure is usually called whitening and can be performed by the classical data analysis technique namely Principal Component Analysis (PCA) [23]. Unfortunately, it was firstly indicated in [20] that the second-order approach does not guarantee a proper separation. Actually, the obtained solution for the separating system in this case is given by

$$\mathbf{W} = \mathbf{QA}^{-1} \tag{2.31}$$

where $\mathbf{Q}$ corresponds to $N \times N$ orthogonal (rotation) matrix. In fact, since the effect of an orthogonal linear transformation is to rotate the input vector, the mixtures $\mathbf{y} = \mathbf{Qs}$ differ from the source in a spacial phase.

Despite its incapability to accomplish the separation task, the whitening approach is commonly utilized as a preprocessing stage in BSS. After the conduction of such stage, the adaptation of the separating system is reduced to the problem of finding the orthogonal matrix $\mathbf{Q}$. This simplification improves the performance of many BSS techniques and, for this reason, it is said that the whitening can be solve a half of the BSS problem. Moreover, there are some BSS algorithms in which the realization of a whitening stage is a necessary condition for a proper operation.

### 2.4.2 Independent Component Analysis

In the seminal work of Hérault et al [20], a learning rule for the matrix $\mathbf{W}$ based on the "uncorrelatedness principle" was derived. Evidently, for the reasons discussed above, this algorithm fails to separate the sources and, in face of this limitation, they proposed a modification in this strategy. By introducing nonlinear functions in the obtained learning rule, they performed the adaptation of $\mathbf{W}$ in order to produce elements of $\mathbf{y}$ that are nonlinearly uncorrelated rather than just uncorrelated. This modification is the essence of the Hérault-Jutten algorithm, and was a major breakthrough in the BSS problem, not only because of its efficiency, but because it was the first technique to introduce higher-order statistics (HOS) in the problem[13].

It was Pierre Comon [25] who clarified this idea by formalizing the concept of ICA. In contrast with the whitening approach, the main idea in ICA is to adjust the matrix $\mathbf{W}$ so that the elements of $\mathbf{y}$ be mutually statistically independent. Based on Darmois' theorem, Comon has shown [25] that if this condition and the following ones

- There is at most one Gaussian source;
- Matrix $\mathbf{A}$ is invertible;

hold, then $\mathbf{W}$ takes the form presented in (2.30); in other words, to recover the independence implies source separation up to scaling and permutation ambiguities.

A major point in the ICA theory concerns the conception of measures that quantify the notion of statistical independence. Aiming to derive the guidelines to the development of such measures, Comon evoked the idea of contrast function, i.e., a function $\Psi(\mathbf{y})$ that must satisfy the following conditions

[12] Interestingly enough, this rather intuitive idea was also employed in the blind equalization problem, as discussed in Section 2.2, when the linear prediction approach was introduced.

[13] In [23], for instance, it is shown the use of a nonlinear correlation measure implicitly brings information about the HOS of the signals.

- $\Psi(\mathbf{y})$ should be invariant to permutations and scaling of the vector $\mathbf{y}$, i.e.:

$$\Psi(\mathbf{y}) = \Psi(\mathbf{DPy}), \tag{2.32}$$

  for any permutation matrix $\mathbf{P}$ and diagonal matrix $\mathbf{D}$;
- When the elements of $\mathbf{y}$ are statistically independent, the following expression should be valid

$$\Psi(\mathbf{y}) \leq \Psi(\mathbf{Ay}), \tag{2.33}$$

  for any invertible matrix $\mathbf{A}$.

From these conditions, it sounds natural to define the goal of the ICA approach as being the minimization, with respect to $\mathbf{W}$, of a contrast function.

A possible candidate to contrast function is the concept of mutual information, which is defined, for a random vector $\mathbf{y}$ of $N$ elements, as

$$I(\mathbf{y}) = \sum_{i=1}^{N} H(y_i) - H(\mathbf{y}), \tag{2.34}$$

where $H(\cdot)$ denotes the Shannon entropy [26]. A fundamental property of mutual information is that it is always non-negative, being zero if and only if the elements of $\mathbf{y}$ are statistically independent. Hence, one may conclude that such measure satisfies the conditions described above, and, as consequence, that it is a valid criterion to guide the optimization of $\mathbf{W}$. Nonetheless, the development of techniques in the spirit of this formulation may be extremely difficult, because it becomes necessary to introduce entropy estimation stages, which may demand a high computational effort. Fortunately, there are other means to carry out the separation task in which no such estimation is needed, as it will be seen in the following sections.

**Maximization of Nongaussianity and the FastICA Algorithm**

A well-established approach in the ICA theory is based on the nongaussianity principle [27], which can be understood through its connection with the central limit theorem (CLT) [5]. Since, in summary, the CLT states that a sum of several independent random variables tends toward a Gaussian distribution, it is expected that each mixture signal, which is a linear combination of sources, be "more Gaussian" than the sources themselves. Taking this observation into account, a straightforward strategy to perform BSS is to adjust the separating system in order to maximize the nongaussianity of its outputs.

Despite the simplicity of the above justification, the nongaussianity approach is solidly and closely related to the idea of minimizing the mutual information, as shown in [28]. This fact becomes clear after some algebraic manipulation of (2.29) and (2.35), which leads to

$$I(\mathbf{y}) = \sum_{i=1}^{N} H(y_i) - H(\mathbf{x}) - \log|\det(\mathbf{W})|. \tag{2.35}$$

From this expression, one may notice that $H(\mathbf{x})$ does not depend on the parameters of the separating system, and, therefore, can be ignored in the optimization task. Furthermore, when the matrix $\mathbf{W}$ is restricted to be orthogonal[14], and the variance of $\mathbf{y}$ is forced to be constant, the last term of (2.35) is also constant during the optimization procedure, which permits us to conclude that the minimization of the mutual information, in this case, is equivalent to the minimization of the marginal entropies of $\mathbf{y}$. Besides, from information theory, it is well-known [26] that the Gaussian distribution is the one with maximum entropy over all distributions with the same variance. Therefore, the maximization of nongaussianity is equivalent to the minimization of the marginal entropies, hence, to the minimization of mutual information - tacitly, the nongaussianity approach also follows the guideline proposed by Comon: to recover the property of statistical independence.

Naturally, as in the independence-based criteria, we must have a quantitative measure of gaussianity in the nongaussianity-based approach. One possibility is to conceive a criterion based on the kurtosis of the elements of $\mathbf{y}$, since this cumulant is equal to zero for the Gaussian distribution. For instance, the maximization of the sum of the absolute values of these kurtosis is in accordance with the nongaussianity principle. Nevertheless, there is a serious drawback related to the sensibility of kurtosis to outliers, which may deteriorate the performance of the algorithms derived from this approach [23].

[14] In the nongaussianty approach, a whitening pre-processing stage is mandatory, since the solution must be restricted to an orthogonal matrix.

In order to overcome this limitation, an alternative approach based on the concept of negentropy was developed. In accordance with the nongaussianity approach, the negentropy of a random variable $y$, which is defined as

$$J(y) = H(y_g) - H(y), \tag{2.36}$$

where $y_g$ is a Gaussian random variable with the same variance of $y$. The nice property of this quantity is that it always assumes a non-negative value, being zero only for a Gaussian distribution. Hence, the maximization of negentropy would lead to a variable $y$ that is the least gaussian as possible. At a first glance, we notice that the evaluation of negentropy requires probability density estimation and evaluation of marginal entropies, as in the mutual information approach. Fortunately, it is possible to resort to the following approximation [23]

$$J(y) \propto \left[E\{G(y)\} - E\{(G(y_g))\}\right]^2 \tag{2.37}$$

where $G(\cdot)$ is a nonquadratic function.

The most celebrated technique for BSS of linear mixtures, the FastICA algorithm [29], was conceived in the light of the nongaussianity approach. In particular, its version for negentropy maximization has been applied in many real problems. The widespread use of this algorithm can be justified by two remarkable features. The first one is its fast convergence speed, and the second is related to the possibility of separating sources in a kind of serial processing. While the former characteristic is due to the adoption of a Newton's method-based algorithm, the latter is a consequence of the nongaussianity criterion be exclusively based on marginal measures of the vector $\mathbf{y}$.

In mathematical terms, the FastICA algorithm for negentropy maximization is described as follows

$$\begin{aligned} \mathbf{w}_i &\leftarrow E\{\mathbf{x}G'(\mathbf{w}_i^T\mathbf{x})\} - E\{G''(\mathbf{w}_i^T\mathbf{x})\}\mathbf{w}_i \\ \mathbf{w}_i &\leftarrow \mathbf{w}_i \,/\, \parallel \mathbf{w}_i \parallel, \end{aligned} \tag{2.38}$$

where $\mathbf{w}_i^T$ denotes the *i*-th column of $\mathbf{W}$. The extraction of one source can be readily obtained from this learning rule. After that, if the extraction of more than one source is desired, this one-unit algorithm should be performed again. However, it becomes mandatory in this case that there be a strategy capable of avoiding the recovery of a source that was already obtained. The answer to this problem is founded on the orthogonality of $\mathbf{W}$. Actually, since the ideal solution for this matrix lies on the space of orthogonal matrices, it expected that the columns of $\mathbf{W}$ also be orthogonal when the ideal solution is reached. Therefore, by applying a mechanism that guarantees the orthogonality of its columns, the difficulty can be overcome. There are two methods to accomplish this task. In the first one, called deflation, orthogonalization is done in a serial fashion, that is, having estimated the first column, the second one should be orthogonal with respect to it, and so on. On the other hand, in the symmetric orthogonalization approach, the orthogonalization of $\mathbf{W}$ is conducted at the end of each iteration of the several one-unit instances.

**Infomax and Maximum-Likelihood Approaches**

The Infomax criterion [30] is one of the most relevant information-theoretic unsupervised learning rules for neural networks. In order to describe how this idea is applied in the context of BSS, let us consider the neural network presented in Fig. 2.7. The Infomax-based training relies on the maximization of the information transferred between $\mathbf{x}$ and $\mathbf{z}$, which, for the BSS model, culminates in the maximization of the joint entropy of $\mathbf{z}$. After some algebraic manipulations, this idea can be expressed by the following optimization problem

$$\max_{\mathbf{W}} H(\mathbf{z}) \equiv \max_{\mathbf{W}} E\{\sum_{i=1}^{N} \log(g_i^{'}(\mathbf{w}_i\mathbf{x}))\} + \log(|\det(\mathbf{W})|), \tag{2.39}$$

where $g_i(\cdot)$ corresponds to the *i*-th activation function.

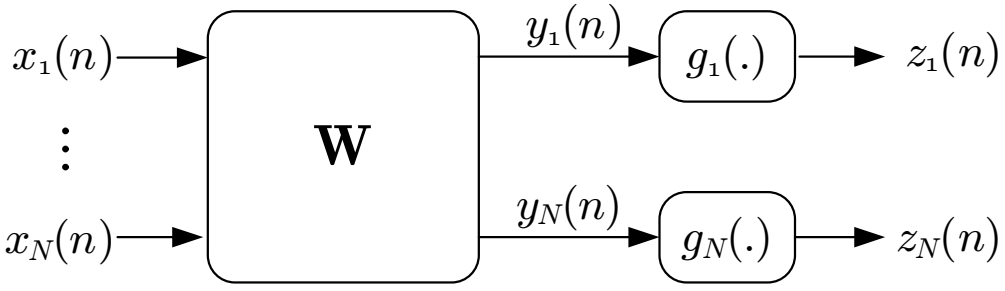



**Fig. 2.7.** Separating system in the Infomax criterion.

It is also possible to interpret the Infomax scheme under the aegis of the Maximum Likelihood (ML) estimator of the linear BSS model. In [31], Cardoso showed that these approaches are equivalent when the activation functions are defined in consonance with the cumulative distributions of the sources. More precisely, the expression (2.39) corresponds to the likelihood function for the BSS model when $g_i^{'}(\cdot)$ represents the cumulative of the *i*-th source. As a matter of fact, it seems more reasonable to conceive the Infomax as a sort of approximation of the ML approach, as no knowledge about the distributions of the sources is assumed in the BSS problem. Interestingly enough, several works [32, 33] point out that the separation task can be performed even when rough approximations of the cumulative distributions are considered, which attests the theoretical solidness of the Infomax criterion.

A very illustrative interpretation of the Infomax and ML approaches arises when the expression (2.39) is rewritten in the following terms [31]

$$\max_{\mathbf{W}} H(\mathbf{z}) \equiv \min_{\mathbf{W}} D(p_{\mathbf{y}}(\mathbf{y}), p_{\mathbf{s}}(\mathbf{s})) + constant \tag{2.40}$$

where $D(p_{\mathbf{y}}(\mathbf{y}), p_{\mathbf{s}}(\mathbf{s}))$ correspond to the divergence of Kulback-Leibler (KL) between the distributions of the sources[15], $\mathbf{s}$, and of their estimates $\mathbf{y}$. Given that the divergence of KL expresses an idea of distance between distributions, we may interpret the criterion (2.39) as a distribution matching strategy, i.e., the aim in this paradigm is to adjust the separating matrix $\mathbf{W}$ in order to obtain an output distribution as close as possible to the hypothesized one.

The first efficient solution to the BSS problem was conceived from the Infomax idea[16]. By performing the optimization of (2.39) through the application of the steepest descent method, Bell and Sejnowski [30] derived an algorithm that is extremely simple to implement, and yet permits the separation of a great number of sources. Mathematically, the Bell-Sejnowski (BS) algorithm is described by the following learning rule

$$\mathbf{W} \leftarrow \mathbf{W} + \mu\{E\{\mathbf{G}(\mathbf{W}\mathbf{x})\mathbf{x}^T\} + (\mathbf{W}^T)^{-1}\}, \tag{2.41}$$

where $\mu$ corresponds to the step size and $\mathbf{G}(\cdot) = [G_1(\cdot) \ldots G_N(\cdot)]$ is a vector of functions such that $G_i(x) = d\log(g_i^{'}(x))/dx$. If the expectation is omitted in this expression, one readily obtains the online version of the BS algorithm.

Another solution to optimize (2.39) is founded on a variant of the gradient method called natural or relative gradient [34, 35]. The principle behind the optimization of a cost function $J(\mathbf{W})$ via this strategy is to seek at each iteration, an increment $\delta$ so that $J(\mathbf{W} + \delta \cdot \mathbf{W})$ be maximized. This idea is in contrast to the classical definition of gradient, in which one searches for an increment $\triangle$ so that $J(\mathbf{W} + \triangle)$ be maximized. As pointed out by Amari [35], due to the Riemannian structure of the parameter space associated with the matrix $\mathbf{W}$, the steepest direction is not given by the ordinary gradient, but instead by the natural gradient. Accordingly, he proposed the following optimization strategy:

$$\mathbf{W} \leftarrow \mathbf{W} + \mu(\mathbf{I} + E\{\mathbf{G}(\mathbf{y})\mathbf{y}^T\})\mathbf{W}. \tag{2.42}$$

It is especially worthy of attention that this learning rule does not demand the inversion of $\mathbf{W}$ at each iteration, thus being more efficient than the BS algorithm.

**Nonlinear PCA**

As discussed in Section 2.4.1, source separation cannot be performed with the sole use of second-order statistics, which means that it is impossible to accomplish such task by in the PCA framework. Besides, we have also mentioned that it is possible to add higher-order statistics to the problem by introducing nonlinear functions in second-order based criteria, as it was done in the Hérault-Jutten algorithm. As a matter of fact, the essence of the Nonlinear PCA (NPCA) [23] approach is related to the same idea.

PCA can be defined as a technique that will seek a matrix $\widetilde{\mathbf{W}}$ such that

$$\widetilde{\mathbf{W}} = \arg\min_{\mathbf{W}} E\{\|\mathbf{x} - \sum_{i=1}^{N}(\mathbf{w}_i^T\mathbf{x})\mathbf{w}_i\|^2\}, \tag{2.43}$$

where $\mathbf{w}_i$ denotes the *i*-th column of $\mathbf{W}$ and $\mathbf{w}_i^T\mathbf{w}_j = \delta_{ij}$. In the case of the NPCA approach, the following modification is taken into account:

$$\widetilde{\mathbf{W}} = \arg\min_{\mathbf{W}} E\{\|\mathbf{x} - \sum_{i=1}^{N}(g_i(\mathbf{w}_i^T\mathbf{x}))\mathbf{w}_i\|^2\}, \tag{2.44}$$

[15] In the Infomax approach, hypothesized distributions are considered.

[16] Due to the equivalence between the Infomax and the ML approaches, the algorithms derived for one of these case can be directly applied to the other one.

where $g_i(\cdot)$ correspond to a nonlinear function. One should keep in mind that there are others nonlinear extensions of the PCA idea [23]. However, this one in special, as discussed in [36], is closely related to other criteria in BSS, such as the maximum likelihood and maximization of nongaussianity approaches.

It is equally possible to rewrite (2.44) in matrix notation:

$$J_{NPCA}(\mathbf{W}) = E\{\|\mathbf{x} - \mathbf{W}^T\mathbf{g}(\mathbf{W}\mathbf{x})\|^2\}, \tag{2.45}$$

where $\mathbf{g}(\cdot) = [g_1(\cdot) \ldots g_N(\cdot)]$. In the NPCA approach, it is quite common to perform a pre-whitening stage and, as a consequence, $\mathbf{W}$ should be an orthogonal matrix, that is to say, $\mathbf{W}\mathbf{W}^T = \mathbf{I}$.

Likewise the strategies mentioned before, a learning rule based on the steepest descent method can be derived for the optimization of (2.45), as shown in [23]. Nonetheless, the MSE formulation associated with NPCA technique has motivated the use of a recursive least squares (RLS) algorithm. In [37], a RLS-based algorithm, called PAST, was derived for the PCA problem. Its extension for the NPCA was proposed in [38]. The main attractive of such technique concerns its fast velocity of convergence in terms of iterations, which in turn is a feature inherited from the RLS algorithm.

## 2.5 Equalization: Going further

After presenting the foundations of blind equalization and blind signal separation, in this section we analyze the relationship between SW, CM and Wiener criteria and the convergence issues of the unsupervised equalization algorithms. Also, we present a blind equalization approach based on nonlinear prediction error filters in addition to the optimal predictor in the minimum mean squared error sense.

### 2.5.1 Relationships Between SW, CM and Wiener Criteria

Two of the most important blind equalization techniques, the CM and SW criteria, were proposed independently and following different principles and developments. Li and Ding, in [39], were the first to search for equivalences between them. Through an elaborate mathematical development, the authors concluded that there exists a direct correspondence between the stationary points of the two criteria. Regalia approached the problem differently, using a very simple and elegant method [40]. The first step of his line of reasoning was to rewrite both cost functions in polar coordinates, which revealed that the analyzed version of the SW cost function did not depend on the radial coordinate. In order to perform a meaningful comparison, the CM cost function was optimized with respect to the radius, resulting in the so called *reduced error surface*, $J_{CMred}(\theta)$, which also depended exclusively on the angular coordinates. The main result of this mathematical manipulation is expressed by:

$$J_{CMred}(\theta) = R_2^2\left(1 - \frac{1}{\lambda + J_{SW}(\theta)}\right) \tag{2.46}$$

where $\lambda = 3$ if $y(n)$ is real and $\lambda = 2$ if $y(n)$ is complex. It is seen through (2.46) that the reduced error surface associated with the CM criterion is a deformation of the error surface of the SW criterion. Since the operator is monotonic and order-preserving, the equilibrium points of the CM cost function are the same as those of the SW cost function. It is important to note that, even though the equilibrium points are the same, the error surfaces and the resulting algorithms are different, having different misadjustments and convergence speeds.

It is also possible to compare both criteria directly via the gradient vector of their cost functions [41]. The difference between them is given by a gain factor that, for the SW cost function, is $r^y = E\left[y^4(n)\right]/E\left[y^2(n)\right]$, and, for the CM cost function, is $r = E\left[s^4(n)\right]/E\left[s^2(n)\right]$. The importance of the power restriction in the SW criterion, given by $E\left[y^2(n)\right] = E\left[s^2(n)\right]$, is then easily seen. Only in case of a perfect equalization (ZF) solution, $r^y = r$. Moreover, both gradients have the same equilibrium points if the SW cost function is power-constrained. If this is not the case, the SW cost function presents valleys of minima that contain the CM minima [41].

Having established the equivalence between the CM and SW criteria, another interesting issue is to compare them to the Wiener criterion. Despite the supervised character of the latter approach, it represents a reference result and an optimum solution. Thus, such a comparison may help in the analysis of how good are the solutions given by the unsupervised criteria. The subject was firstly treated in 1998, in [42], considering a fractionally-spaced equalizer. Using the mean square error (MSE), which is a measure derived from the Wiener (supervised) criterion, they established an upper-bound to the performance of CM equalizers, and, moreover, showed that its minima are close to Wiener solutions, the two being collinear.

Specifically for the SW criterion, [43] showed that, as the value of $p$ increases, the solutions will be closer to Wiener solutions. However, large values of $p$ are not used in practice because the optimization of higher-order statistics involves estimators with high variance.

Another comparison between the CM/SW solutions and Wiener receivers was developed in [44], based on establishing bounds for the CM error surface. In particular, considering the transmitted signal to be binary, i.e., $s(n) \in \{-1, +1\}$, an upper-bound to the CM cost function can be shown to depend on two instances of the Least Mean Fourth (LMF) criterion [44]:

$$J_{CM} \leq \sqrt{E\left[y(n) + s(n-d)\right]^4 E\left[y(n) - s(n-d)\right]^4} \tag{2.47}$$

where $d$ is a delay. Equation (2.47) shows that the CM cost function can be related to a fourth-order supervised criterion under any equalization delay. In the work, it is shown that the upper-bound is tighter in the vicinity of "good solutions", i.e., equalizers that produce a small residual error.

Considering a situation close to the ZF ideal, the authors obtain an approximate relationship between the CM and Wiener solutions:

$$\mathbf{w}_{Wiener,d} \cong \frac{\mathbf{w}_{CM}\left(\mathbf{w}_{CM}^T \mathbf{R}_x \mathbf{w}_{CM} + 1\right)}{2\mathbf{w}_{CM}^T \mathbf{p}_d} \tag{2.48}$$

It is interesting to note that (2.48) suggests the collinearity between solutions, something that was observed, albeit in a distinct context, by [42].

These results indicate that, when both channel and equalizer are FIR filter, the best Wiener solutions, i.e., those capable of effectively reducing ISI, are close to the CM minima. This conjecture, the study of which is a most relevant research topic, establishes a scenario in which the CM approach is "as solid as a blind technique can be", since, in addition to everything we have discussed, it was shown in [45] that the CMA does not converge to spurious minima. It is important to remark that, in the light of the relationship discussed above, these ideas are also applicable to the SW criterion.

In the sequel, we will show, in general lines, the main existing results and studies about the convergence of blind equalization criteria.

### 2.5.2 Convergence Issues

The first studies analyzing the convergence of blind equalization algorithms considered a rather nonrealistic scenario of doubly infinite equalizers. These filters have an infinite number of coefficients in both the causal and anticausal parts. Such a model opens the possibility of calculating the CM minima in the combined channel+equalizer response domain, which is originated by the convolution of the channel and equalizer impulse responses. It should be noted that a doubly infinite equalizer is able to exactly invert a FIR channel.

In the context of the CM criterion, this analysis started with Godard [10] and was later detailed by Foschini [46]. Both of them showed that, considering a doubly infinite equalizer, all the CM cost function minima are ZF solutions. Foschini went further by showing that the only equilibrium points of the CM criteria in the channel+equalizer space were global minima, a local maximum (the origin) and saddle points, being the last two unstable.

It was not until 1991 that Ding *et al.* showed the limitations of such approach. In [45], the authors considered a finite filter and showed that, under an infinite impulse response channel model, the constant modulus algorithm could converge to minima that did not reduce intersymbol interference. Some time later, in 1992 [47], they showed that the analysis in the combined channel+equalizer domain could only be directly extended to the equalizer parameters domain in cases for which the convolution matrix had a trivial null space. The convolution matrix, $\mathbf{H}$, is defined by writing the combined channel+equalizer response in vector form:

$$\mathbf{c} = \mathbf{H}\mathbf{w}, \tag{2.49}$$

where $\mathbf{c}$ is the combined channel+equalizer response and $\mathbf{w}$ is the vector of the equalizer coefficients. The convolution matrix will only have a trivial null space in two situations of little practical value: when the channel is simply a gain or when the equalizer is doubly infinite. This unveils the limitations of the analysis performed by Godard and Foschini.

The work by [48] is a summary of the prevailing view on the subject. It divides the equilibrium points into three classes: the class of maxima, of which the null vector is the sole member; the class of Mazo, composed of the global minima and saddle points, which resemble the non-trivial equilibrium points discovered by Foschini; and the class of Ding, formed by the local minima.

The results presented so far were based on a direct analysis of the focused cost function and of its derivatives. A new and geometrical paradigm was inaugurated by Li and Ding [39], who were the first to analyze convergence with

the aid of the so-called global minimum cones. This approach consists of defining regions in the space of the combined channel+equalizer response that contain a single ZF solution, considering the possibility that such a solution be not attained by a FIR equalizer. The authors get to a pair of very interesting results. Firstly, the normalized kurtosis,

$$K_{norm}(b) = \frac{K(b)}{\sigma_b^4}, \tag{2.50}$$

obtained for the equalizer input and output signals, has a significant influence on the final state of the equalizer. Furthermore, they provide theoretical justifications for the use of the *center-spike* initialization, firstly proposed by Foschini [46].

In [49], Li and Ray Liu conclude that there exists two distinct sets of possible local minima: those resulting from the finite equalizer length, since it is not possible to equalize a FIR channel using a FIR filter, and those inherent to cost functions that are not continuous in the combined channel+equalizer response space. The former class is the only one that is present, for example, in the CM and SW criteria, while the latter class is found, e.g., in the decision-directed and stop-and-go algorithms [50].

It is interesting to note that the existence of local minima in the CM cost function was firstly seen as a major disadvantage in comparison with the supervised and unimodal Wiener criterion. However, a modern view on the subject tends to be different for two reasons:

- The Wiener cost function is also multimodal if the equalization delay is assumed to be a free parameter.
- Cost functions like the CM seem to possess the "ability" of eliminating "bad equalization delays" from its structure.

### 2.5.3 Blind nonlinear equalization: the prediction-based approach

The need for optimal performance and the continuous systemic refinement are the main reasons behind the growing interest in nonlinear equalization. Usually, nonlinear equalizers are adapted with the aid of a pilot signal, i.e., in a supervised fashion. This is quite natural, since the usual test of structures and algorithms must be carried out in an environment as simple as possible. Furthermore, the assumption of supervised training is reasonable in some contexts and also gives rise to a more propitious scenario for optimality analysis. However, a general nonlinear filtering paradigm should not rely on supervised learning, since a reference signal may not be available in all cases. This is the motivation behind the proposal of unsupervised equalization criteria. Although criteria based on signal statistics work well on the adaptation of linear filters, it is not certain that they will ensure the correct adaptation of nonlinear filters. Actually, the BGR theorem does not hold for nonlinear filters [51]. Ironically, this kind of problem arises exactly from the great approximation potential of nonlinear structures. Therefore, it becomes imperative to look for unsupervised equalization criteria adequate to the problem of nonlinear filtering. In particular Cavalcante et al. [52] demonstrated that a prediction approach can be effective in a linear channel context.

As stated in Section 2.4.1, linear forward prediction error equalizers are able to equalize minimum phase channels[17]. However, such equalizers are not capable of equalizing nonminimum phase channels. This is not a limitation of the prediction approach but of the linear FIR structure [52, 53]. The use of nonlinear predictors, such as artificial neural networks and fuzzy systems, allows the prediction error equalizers to overcome this limitation and makes possible the use of this type of approach in almost all linear channel scenarios [51, 52, 53].

But there remains a question: what nonlinear structure is most suitable to be used as a predictor? To answer this question we must remember that the performance of a prediction-error equalizer is related to the accuracy of the predictor employed, since, the better the predictor, more ISI will be removed from the received signal. Therefore, considering the minimum mean squared error (MMSE) as the optimality criterion, the best estimate of the channel output given a vector of past samples $\mathbf{x}(n-1) = [x(n-1), \ldots, x(n-m+1)]^T$, is given by [54]

$$\tilde{x}(n) = \mathrm{E}\left[x(n) \,|\mathbf{x}(n-1)\right]. \tag{2.51}$$

Considering a linear FIR channel and additive Gaussian noise, the evaluation of (2.51) leads us to the optimal forward predictor input-output mapping [51]

$$f_{\mathrm{MMSE}}\left(\mathbf{x}(n-1)\right) = \mathbf{w}^T \mathbf{F}\left(\mathbf{x}(n-1)\right) \tag{2.52}$$

where

$$F_i\left(\mathbf{x}(n-1)\right) = \frac{\exp\left(-\frac{\|\mathbf{x}(n-1)-\mathbf{u}_i\|^2}{2\sigma_\eta^2}\right)}{\sum_{j=1}^{N_s} \exp\left(-\frac{\|\mathbf{x}(n-1)-\mathbf{u}_j\|^2}{2\sigma_\eta^2}\right)}, \tag{2.53}$$

[17] It can be shown that maximum phase channels can be equalized using linear backward prediction error filters [4].

$\mathbf{w}$ is a weight vector, $\mathbf{u}_i$, $i = 1 \ldots N_s$, are the channel state vectors, $N_s$ is the total number of channel states and $\sigma_\eta^2$ is the noise variance. The channel states are defined as the channel output vectors in the absence of noise [51, 55]. Given the channel states and the noise variance, the weight vector $\mathbf{w}$ can be determined through the minimization of the squared prediction error using linear optimization tools. It is also possible to write the weight vector as a function of the channel states [51]. The optimal backward predictor mapping is similar to the forward predictor. The differences between them are the weight vector and, of course, the input vector.

It is important to note that, like the Bayesian equalizer [56], the optimal predictor can be implemented using fuzzy systems, which allows us to use the numerous training techniques that were developed in the context of this field to obtain the optimal predictor.

The performance of the nonlinear prediction-error equalizers is related to the absolute value of the first coefficient of the channel impulse response. On the other hand, the backward prediction error equalizers have their performance associated to the absolute value of the last coefficient. The bigger the absolute value, the lower is the symbol error rate [51]. Therefore, the forward and backward prediction error equalizers generally perform better in minimum and maximum phase channels respectively. For mixed phase channels, the best performance is usually achieved using a cascade of forward and backward prediction error equalizers in order to select the channel coefficient which has the biggest absolute value [51].

## 2.6 Blind Source Separation: Going Further

Despite its simplicity, the instantaneous model considered in the early development of blind source separation techniques provides a good approximation for the mixing process in a variety of scenarios. Nevertheless, in some situations, the tools developed so far could not be directly applied, and more specific models should be used in order to design new methods. In the following, we present some other models that have been considered in the literature, as well as some recent approaches developed to recover the sources in these scenarios.

### 2.6.1 Convolutive Mixtures

The convolutive model can be viewed as an extension of the FIR model in a SISO system. Indeed, there are practical instances, such as in the separation of audio signals [23], in which the measured signals must necessarily be understood as being formed by a combination of different sources and delayed versions of them, as seen in Section 2.3. A model of this kind bears a strong resemblance with the idea of convolution, and, in the case of blind source separation, this is exactly the reason why it is usually designated by the name of convolutive mixture.

In the convolutive case, the relationship between the sources and observed signals is given by

$$\mathbf{x}\,(n) = \sum_{k=0}^{L-1} \mathbf{A}_k \cdot \mathbf{s}\,(n-k) \tag{2.54}$$

where $\mathbf{A}(k)$ denotes the mixing matrix associated with the delay $k$.

The first approach to solving the convolutive problem is a direct extension of BSS techniques developed for the instantaneous case. However, instead of adapting an unmixing matrix, a set o matrices undergo adaptation following the classical ICA framework. In this case, however, the scale and order indeterminacies are followed by a filtering ambiguity. This new indeterminacy is due to the fact that, if $s_1(n)$ and $s_2(n)$ are independent processes, so are $f_1(n) * s_1(n)$ and $f_2(n) * s_2(n)$, where $f_1(n)$ and $f_2(n)$ are arbitrary filter responses and $*$ denotes the convolution operator.

Due to its similarity with the instantaneous case, a natural approach to solving the problem is the time-domain BSS, where ICA is applied directly to the convolutive mixture model [57, 58]. In this case, a set of matrices[18] is adapted with the objective of recovering the sources. The approach achieves good separation once the algorithm converges, since the ICA algorithm correctly evaluates the independence of separated signals. However, ICA for convolutive mixtures is not as simple as ICA for instantaneous mixtures, and the process may be computationally demanding for long convolutive mixtures.

Another possible approach is the frequency-domain BSS, in which the separation of the sources is performed in the frequency domain [59, 60, 61]. Applying the Fourrier transform to Eq. (2.54) we obtain

$$\mathbf{X}\,(w) = \mathbf{A}(w) \cdot \mathbf{S}\,(w), \tag{2.55}$$

where $\mathbf{X}\,(w)$,$\mathbf{A}(w)$ and $\mathbf{S}\,(w)$ represent, respectively, the Fourrier transform of $\mathbf{x}\,(n)$, $\mathbf{A}(n)$ and $\mathbf{s}(n)$.

[18] Or, equivalently, a set of filters.

Eq. (2.55) shows that the original convolutive problem can also be expressed, in the frequency domain, as a set of infinite instantaneous mixing problems indexed by the frequency $w$, since $w$ is a continuous variable. In practice, the values of $w$ are divided into a set of intervals (bins), and, for each interval, we have a single instantaneous mixture problem, to which complex-valued ICA can be applied [59]. The merit of this approach is that the BSS algorithm becomes simple and can be applied separately to each frequency bin. Also, any complex-valued instantaneous ICA algorithm can be employed with this approach. However, the permutation ambiguity of the ICA solution becomes a serious problem. Consider that $y_{1,1}$ and $y_{1,2}$ represent the outputs obtained by the ICA algorithm applied to frequency bins $\#1$ and $\#2$, respectively. Due to the inherent ambiguity of ICA, there is no guarantee that the outputs $y_{1,1}$ and $y_{1,2}$ are, in fact, different portions of the same source. Hence, prior to the transformation of the signals back to the time-domain, it is necessary to align the permutation in each frequency bin so that a separated signal in the time domain contains frequency components from the same source signal. This problem is well-known as the permutation problem of frequency-domain BSS, and some methods were proposed to solve this indeterminacy [62, 63].

In the discussed approaches, it is considered that the knowledge about the sources is restricted to the fact that they are independent between each other, which characterizes a blind scenario. However, in some applications, some features of the sources are known a priori and can guide the design of the separating system. For instance, the sources could be not only independent between each other but also composed of i.i.d. samples, as in the case of digital communication signals. For this scenario some specific methods can be applied , including a recent proposal [64] in which a set of nonlinear prediction-error filters is used to reduce the convolutive problem to an instantaneous one.

### 2.6.2 Nonlinear Models

In the most general scenario, the mixing process may not fit the mould of a linear system, requiring that nonlinear models be used instead. In such a case, the observed samples would be given by

$$\mathbf{x} = \mathcal{F}(\mathbf{s}), \tag{2.56}$$

where $\mathcal{F}(\cdot)$ denotes a nonlinear mapping. Therefore, the solution of the Nonlinear Blind Source Separation problem (NBSS) depends on the design of a nonlinear separating system $\mathcal{G}(\cdot)$ such that $\mathbf{y} = \mathcal{G}(\mathbf{x})$ be an accurate estimate of the sources.

As mentioned in Section 2.4.2, in the linear case, source separation can be achieved through the recovery of statistical independence between the estimates, which yields several different criteria related to this principle. Unfortunately, the same approach could not be followed for the nonlinear case. It can be shown that it is possible to find nonlinear functions that provide independence between the estimated signal and, yet, do not recover the true sources. The proof for this property can be traced back to the work of Darmois in 1951, and to [65], in which it is demonstrated how to build families of nonlinearities that recover the independence but not the sources.

In a certain sense, this particular property can be understood as a direct consequence of the high flexibility of nonlinear functions, which can mix two random vectors and still render independent vectors. Therefore, one should consider strategies to restrain the structures under adaptation, which can be achieved by either incorporating some additional penalizing terms into the cost function [23], or restraining the nonlinear mixing models to a smaller class of so-called *separable models* [66, 67, 65, 68, 69], to which the ICA framework is directly applicable.

The first approach, in theory, would allow us to deal with a larger number of scenarios, since there is no explicit restriction to the type of nonlinearity involved. However, to determine this penalizing term is, in most of the cases, not so straightforward, requiring that additional information about the mixing process itself be available beforehand.

The second one, in which the range of nonlinear mixing models is constrained, led to the development of tools for specific kinds of nonlinear systems. One such model is the so-called Post Nonlinear model (PNL) [66], which is presented in more detail in the next section.

#### Blind Source Separation of Post-Nonlinear Mixtures

First introduced in [66], the PNL model consists of a linear mixing stage followed by a set of nonlinear functions, each one applied only to one output of the previous stage. This kind of structure can be useful to model scenarions in which the mixing process is of linear nature, but the sensors used to detect relevant signals present a nonlinear character. Examples can be found in communication systems, in which amplifiers are used in the receiver end [66], and, in smart sensor arrays for chemical substances and gases [70].

Mathematically, the *i*-th output of an PNL system, shown in Fig. 2.8, is given by:

$$x_i = f_i(a_{i1}s_1 + a_{i2}s_2 + \ldots + a_{iN}s_N), \tag{2.57}$$

with $f_i(\cdot)$ corresponding to a nonlinear function. In matrix form, (2.57) can be restated as

$$\mathbf{x} = \mathbf{f}(\mathbf{A}\mathbf{s}), \tag{2.58}$$

where $\mathbf{f}(\cdot) = [f_1(\cdot) \ldots f_N(\cdot)]$.

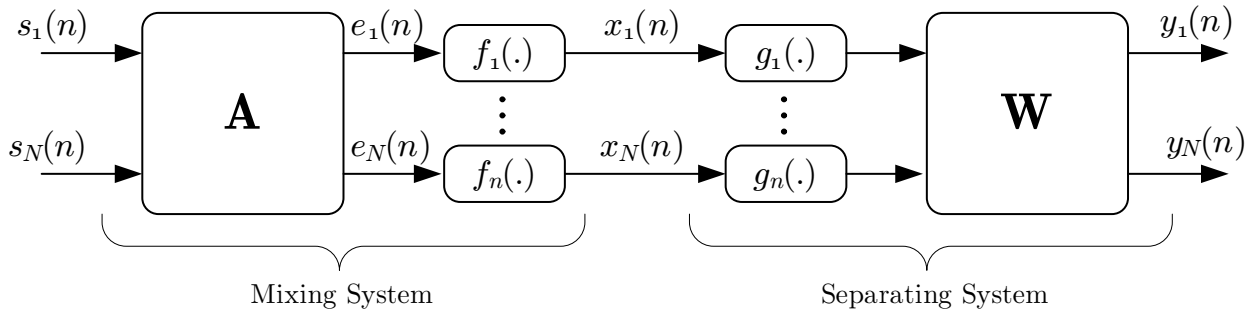



**Fig. 2.8.** Post-Nonlinear Model.

Given the structure of a PNL system, a natural choice for the separating system consists of a similar configuration: a set of nonlinearities followed by an unmixing matrix, as shown in Fig. 2.8. In mathematical terms, we have:

$$y_i = w_{i1}g_1(x_1) + w_{i2}g_2(x_2) + \ldots + w_{iN}g_N(x_N), \tag{2.59}$$

with $g_i(\cdot)$ denoting a nonlinear function. In matrix notation the unmixing process is given by

$$\mathbf{y} = \mathbf{W}\mathbf{g}(\mathbf{x}), \tag{2.60}$$

where $\mathbf{g}(\cdot) = [g_1(\cdot) \ldots g_N(\cdot)]$.

Since independence of the estimates guarantees source recovery, the next step is to adjust the free parameters in order to achieve this condition. It is important to remember, however, that in this case, not only a matrix $\mathbf{W}$, but also a set of nonlinearities, must be adapted. Considering the joint mixing/unmixing mapping

$$\mathbf{y} = \mathbf{W}\mathbf{g}(\mathbf{f}(\mathbf{A}\mathbf{s})), \tag{2.61}$$

it becomes clear that the source estimates are statistically independent when the composition $\mathbf{g} \circ \mathbf{f}$ corresponds to a vector of linear functions, and $\mathbf{WA} = \mathbf{\Lambda P}$ ($\mathbf{\Lambda}$ and $\mathbf{P}$ denote a diagonal and a permutation matrix, respectively). However, the separability of such model requires that the converse be also true.

The first formal proof for the separability of the PNL model using ICA was given in [66]. In the same paper an algorithm based on a gradient descent procedure was proposed to adapt the separating system, and the independence between the estimates was quantified using mutual information. Afterwards, other authors presented further investigations about the conditions under which the ICA approach holds in the PNL problem [69, 71], as well as new tools to solve the problem.

In [69], a geometrical approach to the problem was developed. The rationale behind the method consists of adapting separately the two portions of the separating system. For the nonlinear section, a criterion based on the geometry of the joint distribution of the observed samples was employed. After that, if the first task is fulfilled, a conventional ICA technique can be applied in the linear section of the system. Notwithstanding the possibility of separately adjusting the linear and nonlinear sections, and also using regular ICA algorithms, the method is restricted to scenarios with 2 sources.

Following the same idea of adapting the linear and nonlinear sections separately, in [72, 73, 74] strategies based on a gaussianizing process are employed. This notion can be better understood in the light of the central limit theorem, discussed in Section 2.4.2. Since each observed mixture is the sum of several independent components, it will tend to a gaussian distribution as the number of sources tend to infinity. The effect of the nonlinearity present in the mixture would then be to deviate the random variable from the gaussian distribution. Thus, in the gaussianing approach, the nonlinear section of the separating system is designed to restore the gaussian distribution of the signals. If that condition is attained, the obtained nonlinearity should be the inverse of that present in the mixing system, and the remaining linear mixture could be inverted using any reliable ICA algorithm. It is important to note that, in practical scenarios, the number of sources does not tend to infinity, and the linearly mixed signals may not present a gaussian

distribution. If that is the case, the gaussianizing process will actually introduce distortions to the observations, which may compromise the solution.

A recurrent issue found in PNL mixtures is the convergence of the methods used to adapt the separating system. Even though there is no mathematical proof for the existence of local minima, in some works [75, 69], it has been observed that gradient-based strategies could converge to non-optimal solutions, which may indicate an inherent multimodal character of the cost function. Considering this fact, some researches have proposed alternative methods to adapt the separating system, like evolutionary algorithms [76] and other heuristic search tools [77].

### 2.6.3 Underdetermined Systems

In the standard linear mixing problem, in which the mixing matrix is square, a BSS solution is reached either by adapting a separating matrix or by estimating the mixing matrix and then inverting it to obtain the separating system. The problem, however, becomes more involved when the number of sensors is smaller than the number of sources. For this kind of mixture, termed underdetermined mixture, it is no longer possible to recover the sources by simply adapting a separating matrix, and even if the mixing matrix is known, it is not easy to determine how one can estimate the sources.

The problem of underdetermined mixtures has been little addressed, in comparison with the other cases, despite its practical interest. A large number of works found in the literature dealing with underdetermined case follow a two-step procedure in order to estimate de sources: first, some method is used to estimate the mixing matrix. Then, a method, usually based on some prior knowledge an/or assumptions about the sources, is employed to estimate the original signals.

The task of estimating the mixing matrix have been tackled in a few articles [78, 79], most of then relying on multilinear algebra[19] to obtain the mixing coefficients. These methods explore some properties of the higher-order cumulant tensor[20] of the observed signals to obtain the mixing coefficients. Under some conditions [80], this tensor can be decomposed into a number of rank-1 terms, each one related to the contribution of one particular source, thus giving sufficient information to determine the mixing coefficients.

Estimating the source signals, either using the estimated or exact mixing matrix, is not a trivial task. Since the mixing matrix is not square, it is not possible to invert it and obtain the separating system, as in the regular case. In the general case, the usual approach is to apply Bayesian methods to estimate the signals. However, even in this framework, it is necessary that additional information about the sources be available. This may be reasonable in audio, or even digital communication applications, in which the pdf's of the sources are known a priori.

A third approach to deal with underdetermined mixtures heavily relies on the assumption of sparsity of the sources, or that there exist suitable linear transforms, as the Fourier transform, that can render the sources sparse [81]. In analogy to the ICA case, such approach is named Sparse Component Analysis (SCA) [82, 83, 84]. As mentioned in [81], the determination of the sparse components culminates in a clustering problem, for which several techniques can be used.

## 2.7 Blind Equalization and Blind Source Separation: Some Important Relationships

The theories of blind equalization and blind source separation, as we saw in Sections 2.2 and 2.5, evolved through different paths and from distinct motivations. Nevertheless, there are striking similarities in the manner whereby both problems are formulated, which raises a very interesting question: what kind of relationship can be established between the classical solutions belonging to these fields? This section will be devoted to analyzing three answers thereto.

### 2.7.1 Maximum Likelihood and BGR

The BGR theorem, discussed in Section 2.2, revealed that a ZF condition could be attained, under certain hypotheses, in a case in which the pdf of the equalizer output signal is identical to that of the transmitted signal. This indicates that the notion of probability density matching is solid enough to engender an unsupervised paradigm for eliminating the ISI when this is feasible from the structural standpoint.

In the maximum-likelihood BSS strategy, which was exposed in Section 2.4, it is assumed that the probability densities of the sources are known a priori, which leads naturally to the establishment of a likelihood function whose rationale is to quantify the ability of a given separating system to produce output pdfs as close as possible to those of

[19] The interested reader should see [80] for further details

[20] A tensor can intuitively be imagined as a multi-indexed array of numerical values. For instance, a matrix is a particular case of a tensor.

the sources. Alternatively, this strategy can be understood as a process of minimizing the Kullback-Leibler divergence between the pdfs of the estimated and the actual sources.

After this prologue, it is noticeable that there is an intimate connection between the BGR theorem and the ML BSS approach, since both make use of the statistical information contained in the pdf of the signals to be recovered to guide the blind adaptation process. This leads us to a relevant and general conclusion: in a linear structural context in which perfect inversion is attainable, the idea of pdf matching constitutes a solid and viable approach to recovering latent information [85].

### 2.7.2 Kurtosis and SW

In the context of SISO models, the SW theorem demonstrates that it is possible to reach a perfect equalization condition by forcing the variance of the equalizer output and its kurtosis to be equal to those of the transmitted signal. As it was shown in Section 2.2, this notion naturally leads to the SW criterion of seeking a condition in which the kurtosis is maximal. Interestingly, the idea of maximizing the kurtosis is also reached in the context of BSS, albeit through a different path. As it was discussed in Section 2.4, such approach comes from the non-gaussianity maximization principle. In brief, a clear connection can be stated from these arguments: in the SISO equalization context of the SW criterion, the channel is responsible for generating a received signal composed of multiple delayed versions of a transmitted signal. By maximizing the absolute value of the kurtosis, it is possible to obtain an estimate of one of these versions. On the other hand, in the classical BSS case, it is possible to extract multiple sources from a set of mixtures on a one-by-one basis via the same procedure. These similarities indicate two things:

- That the problem of equalization can be conceived as problem of BSS in which the multiple sources correspond to the delayed versions of the transmitted signal.
- That the ideas of matching a higher-order statistic and of maximizing a nongaussianity measure are, in essence, similar and interchangeable. Therefore, it is possible to reinterpret the SW theorem as if it were related to the goal of forcing the equalizer output to be as nongaussian as possible (which would mean that the ISI would be minimal), as well as to formulate the problem of extracting a source as one of recovering a certain level of kurtosis after the variance of the involved signals is normalized.

### 2.7.3 Nonlinear PCA and Bussgang

Another notorious relationship involves the Bussgang criteria and the idea of Nonlinear PCA. In order to understand the connection between these two approaches, let us recapitulate the cost function to be optimized in the NPCA criterion, given by

$$J_{NPCA}(\mathbf{W}) = E\{\|\mathbf{x} - \mathbf{W}^T\mathbf{g}(\mathbf{W}\mathbf{x})\|^2\}. \tag{2.62}$$

As discussed in Section 2.4, a pre-whitening stage is usually performed in this case and, as consequence, $\mathbf{W}$ is an orthogonal matrix. Taking this observation into account and after some algebraic manipulations, it is possible to rewrite (2.62) as follows

$$J_{NPCA}(\mathbf{W}) = \sum_{i=1}^{N} E\{[y_i - g_i(y_i)]^2\}. \tag{2.63}$$

As noted in [36], this expression evidencies a similarity between the NPCA and the Bussgang criteria for blind deconvolution. Therefore, there is an intimate relationship between the NPCA cost function and a sum of Bussgang-like terms.

The comparison, albeit simple, is conceptually instigating. Part I revealed that the development of the fields of equalization and source separation techniques followed a similar path - from the limitations of second-order approaches (linear prediction / PCA) to the incorporation of some kind of higher-order statistic. In the case of Bussgang techniques, higher-order information emerges from the use of a nonlinear estimator in a Wiener-like cost function. NPCA-based decomposition, on the other hand, depends on the introduction of a nonlinear mapping to transcend the correlation-based PCA framework. It is important to consider the implication of interchanging this concepts: this would allow us, for instance, to view the problem of Bussgang-based multiuser detection[86] as a classical separation problem based on NPCA and, conversely, would open the perspective of understanding the Bussgang estimators that form the core of many blind equalization techniques as being part of a nonlinear decomposition approach[21].

[21] More information on this subject can be found, for instance, in [23] and [85].

## 2.8 Modern trends

In this section, we present two research topics that we consider to be promising extensions of the concepts and applications discussed in this tutorial.

### 2.8.1 Algebraic Methods

Based on a theory proposed by [87], parameters of continuous signals as, for example, amplitude, frequency and phase of a sinusoid, can easily be estimated through the resolution of a simple system of equations. It suffices to find a differential equation, satisfied by the desired signal, whose coefficients depend on the parameters to be estimated. The algebraic and deterministic estimation methods thus developed, based on differential algebra and operational calculus, present certain desirable properties:

- Signals are treated in continuous-time.
- Signals are treated as deterministic and, for this reason, no hypotheses on their statistic characteristics are needed, including those related to noise.
- The estimation does not rely on optimization procedures, resulting in explicit formulas for the identification of the desired parameters.
- The method is robust to noise

The application of such methods in signal processing is based on modeling the channel by a rational transfer function, which, in the time domain, can be seen as a differential equation. Obtaining the Laplace transform of such equation enables the construction of a system with the same number of equations and unknowns. In the case of a deconvolution or demodulation problem, the solution of this system directly yields directly the transmitted symbols.

This method was already applied to the identification/estimation of systems (channels) modeled by rational transfer functions, to the deconvolution of BPSK and QPSK (*Binary (Quadrature) Phase Shift Keying*) signals, to the demodulation of CPM (*Continuous Phase Modulation*) signals, between others [88]. The next step would be to use it in source separation. Since undermodeled (mores sources than sensors) cases and the addition of noise continues to be opened problems, these algebraic methods could bring interesting results. A simple case of separating two sinusoids in noise, with different amplitudes, frequencies and phases, was already developed successfully [89].

### 2.8.2 Evolutionary Processing

In general, the problems of adaptive filtering are solved in accordance with a procedure that encompasses choices of:

- An appropriate filtering structure;
- A suitable criterion to guide the adjust of its free parameters;
- A search technique that allow the full potential of the combination between structure and criterion to be explored;

There are many scenarios in which a pair structure / criterion must be used that originates a multimodal optimization task in the parameter space. This commonly occurs, for instance, in the contexts of blind SISO equalization, of supervised and unsupervised nonlinear equalization and of blind separation of post-nonlinear, underdetermined and convolutive mixtures. Under these circumstances, it becomes highly attractive to look for search tools that possess not only the local search potential that characterizes the widely employed gradient techniques, but also a global search capability that avoid convergence to suboptimal solutions.

The field of evolutionary computation [90] has produced a number of solutions that meet these requirements. Evolutionary techniques are based on the implementation of mechanisms inspired by concepts drawn from theories that attempt to explain biological processes in which the idea of evolution plays a key role. In the context of signal processing applications, the most widely employed class of evolutionary techniques is that of genetic algorithms (GAs). Genetic algorithms, whose origin can be traced back to the work of John Holland [91], are population-based optimization techniques founded on the following notions:

- That of treating each possible solution to the focused problem as an individual of a population and the parameters of each solution as coded genetic information.
- That of establishing an intimate connection between the cost function to be optimized and the notion of fitness of a given individual.
- That of creating mechanisms that attempt to emulate the idea of natural selection.
- That of promoting eventual changes in the genetic patrimony of the population by means of reproduction (crossover) and mutation operators.

The first two items of this list indicate the manner whereby any given optimization problem can be "translated into evolutionary terms". The two other aspects are related to the operation of the genetic algorithm itself, which is an elegant combination between the "invisible hand" of natural selection and the processes through which the individuals suffer modifications. It is important to remark that the crossover operator is responsible for combining different solutions in a way that evokes the idea of local search, whereas the role of the mutation operator is to generate spurious modifications that significantly widen the scope of the search process. Genetic algorithms have been applied to several estimation and modeling problems [92].

Another evolutionary paradigm that has been successfully applied in the context of signal processing is that of artificial immune systems (AIS) [93]. Among the techniques belonging to this field, we highlight an artificial immune network called opt-aiNet, which was first proposed in [94]. The opt-aiNet is a search tool whose modus operandi is inspired by two theories about the process of antibody recognition in the organism of mammals:

- The theory of clonal selection and affinity maturation, which presents a view of the immune system in which defense cells become more specialized with the aid of a mechanism that combines fitness-proportional mutation and an associated selection process. This view can be understood as kind of small-scale evolution.
- The immune network theory, which conceives the immune system as a network endowed with determined "eigen-behaviors" even in the absence of antigens.

The combination of these two features is responsible for producing a technique with a remarkable balance between local and global search mechanisms, and, furthermore, capable of controlling the population size in accordance with the complexity of the search problem at hand. The opt-aiNet has been applied to variety of signal processing problems, and, in all cases, it was verified that its use was responsible for a substantial increase in the global convergence rate in comparison, for instance, with classical gradient-based approaches.

In our opinion, the use of evolutionary tools in signal processing problems is particularly justifiable in the context of scenarios characterized by a pronounced multimodal character and of applications that require a performance very close to the limit of optimality, since these tools have a computational cost that is significantly higher than that of any gradient-based technique. This tradeoff opens two perspectives:

- That evolutionary tools be widely employed in offline applications, which are common, for instance, in source separation.
- That simpler evolutionary algorithms be developed to facilitate "real world" implementations. This objective is the rationale of the so-called micro-GAs [95].

## Acknowledgments

The authors would like to thank the partial financial support from FAPESP and CNPq (Process 309830/2003-6).

## References

1. S. Haykin, *Adaptive Filter Theory*, 4th, Ed. Prentice Hall, 2001.
2. J. Proakis, *Digital Communications*, 4th ed. MacGraw-Hill, August 2000.
3. J. R. Barry, E. A. Lee, and D. G. Messerschmitt, *Digital Communication*, 3rd ed. Springer, September 2003.
4. M. Bellanger, *Adaptive Digital Filters*, 2nd ed. CRC, July 2001.
5. A. Papoulis and S. Pillai, *Probability, Random Variables and Stochastic Processes*, 4th ed. McGraw-Hill, January 2002.
6. A. Benveniste, M. Goursat, and G. Ruget, "Robust identification of a nonminimum phase system: Blind adjustment of a linear equalizer in data communications," *IEEE Trans. on Automatic Control*, vol. 25, no. 3, pp. 385–399, June 1980.
7. O. Shalvi and E. Weinstein, "New criteria for blind deconvolution of nonminimum phase systems (channels)," *IEEE Trans. on Information Theory*, vol. 36, no. 2, pp. 312–321, March 1990.
8. R. W. Lucky, "Automatic equalization for digital communication," *Bell Syst. Tech. J.*, vol. 44, pp. 547–588, 1965.
9. Y. Sato, "A method of self-recovering equalization for multilevel amplitude-modulation systems," *IEEE Transactions on Communications*, vol. 23, no. 6, pp. 679–682, June 1975.
10. D. Godard, "Self-recovering equalization and carrier tracking in two-dimensional data communication systems," *IEEE Trans. on Communications*, vol. 28, no. 11, pp. 1867–1875, 1980.
11. J. Treichler and B. Agee, "A new approach to multipath correction of constant modulus signals," *IEEE Trans. on Acoustics, Speech, and Signal Processing*, vol. 31, no. 2, pp. 459–472, April 1983.
12. O. Shalvi and E. Weinstein, "Super-exponential methods for blind deconvolution," *IEEE Trans. on Information Theory*, vol. 39, no. 2, pp. 504–519, March 1993.

13. M. S. de Castro, "Solucoes adaptativas para equalizacao multicanal," Master's thesis, UNICAMP, Campinas, Brazil, 2002.
14. J. Johnson, R., P. Schniter, T. Endres, J. Behm, D. Brown, and R. Casas, "Blind equalization using the constant modulus criterion: a review," *Proceedings of the IEEE*, vol. 86, no. 10, pp. 1927–1950, Oct. 1998.
15. Z. Ding and Y. G. Li, *Blind Equalization and Identification*. CRC, January 2001.
16. S. Applebaum, "Adaptive arrays," *IEEE Transactions on Antennas and Propagation*, vol. 24, no. 5, pp. 585–598, Sep 1976.
17. I. Frost, O.L., "An algorithm for linearly constrained adaptive array processing," *Proceedings of the IEEE*, vol. 60, no. 8, pp. 926–935, Aug. 1972.
18. L. S. Resende, J. M. T. Romano, and M. G. Bellanger, "A fast least-squares algorithm for linearly constrained adaptive filtering," *IEEE Transactions on Signal Processing*, vol. 44, no. 5, pp. 1168–1174, May 1996.
19. C. C. Cavalcante, "Sobre Separação Cega de Fontes: Proposicões e Análise de Estratégias para Processamento Multi-Usuário," Tese de Doutorado, Universidade Estadual de Campinas (Brasil), abril 2004.
20. J. Hérault, C. Jutten, and B. Ans, "Détection de grandeurs primitives dans un message composite par une architecture de calcul neuromimétique en apprentissage non supervisé," *Actes du Xème Colloque GRETSI*, pp. 1017–1022, 1985.
21. T. P. Jung, S. Makeig, T. W. Lee, M. McKeown, G. Brown, A. J. Bell, and T. J. Sejnowski, "Independent Component Analysis of Biomedical Signals," in *Proceedings of the Second International Workshop on Independent Component Analysis and Blind Signal Separation, ICA 2000*, Helsinki, Finlândia, 2000, pp. 633–644.
22. N. Mitianoudis and M. E. Davies, "Audio Source Separation of Convolutive Mixtures," *IEEE Transactions on Speech and Audio Processing*, vol. 11, no. 5, pp. 489–497, 2003.
23. A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*. Wiley-Interscience, May 2001.
24. C. Jutten and A. Taleb, "Source Separation: From Dusk till Dawn," in *Proceedings of the Second International Workshop on Independent Component Analysis and Blind Signal Separation, ICA 2000*, Helsinki, Finlândia, 2000, pp. 15–26.
25. P. Comon, "Independent Component Analysis, a New Concept?" *Signal Processing*, vol. 36, no. 6, pp. 287–314, 1994.
26. T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. John Wiley and Sons, Inc, 1991.
27. N. Delfosse and P. Loubaton, "Adaptive blind separation of independent sources: a deflation approach," *Signal Processing*, vol. 45, pp. 59–83, 1995.
28. T.-W. Lee, M. Girolami, A. J. Bell, and T. J. Sejnowski, "A unifying information-theoretic framework for independent component analysis," *Computers and Mathematics with Applications*, vol. 39, no. 11, pp. 1–21, 2000.
29. A. Hyvärinen, "Fast and robust fixed-point algorithmsfor independent component analysis," *IEEE Transactions on Neural Networks*, vol. 10, no. 3, pp. 626–634, 1999.
30. A. J. Bell and T. J. Sejnowski, "An information maximisation approach to blind separation and blind deconvolution," Computational Neurobiology Laboratory, The Salk Institute, Tech. Rep., Fevereiro 1995.
31. J. F. Cardoso, "Infomax and Maximum Likelihood of Blind Source Separation," *IEEE Signal Processing Letters*, vol. 4, no. 4, pp. 112–114, 1997.
32. T. W. Lee, M. Girolami, and T. J. Sejnowski, "Independent Component Analysis using an Extended Infomax Algorithm for Mixed Subgaussian and Supergaussian Sources," *Neural Computation*, vol. 11, pp. 417–441, 1999.
33. J. F. Cardoso, "On the stability of some source separation algorithms," in *Proc. of the IEEE Workshop on Neural Networks for Signal Processing*, Cambridge, Inglaterra, 1998.
34. J. F. Cardoso and B. H. Laheld, "Equivariant adaptive source separation," *IEEE Transactions on Signal Processing*, vol. 44, no. 12, pp. 3017–3030, 1996.
35. S.-I. Amari, "Natural gradient works efficiently in learning," *Neural Computation*, vol. 10, pp. 251–276, 1998.
36. J. Karhunen, P. Pajunen, and E. Oja, "The Nonlinear PCA Criterion in Blind Source Separation: Relations with Other Approaches," *Neurocomputing*, vol. 22, pp. 5–20, 1998.
37. B. Yang, "Projection Approximation Subspace Tracking," *IEEE Transactions on Signal Processing*, vol. 43, no. 1, pp. 95–107, 1995.
38. J. Karhunen and P. Pajunen, "Blind Source Separation Using Least-Squares Type Adaptive Algorithms," in *Proc. IEEE 1997 Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP'97)*, Munique, Alemanha, abril 1997, pp. 3361–3364.
39. Y. Li and Z. Ding, "Convergence analysis of finite length blind adaptive equalizers," *IEEE Trans. on Signal Processing*, vol. 43, no. 9, pp. 2121–2129, 1995.
40. P. Regalia, "On the equivalence between the Godard and Shalvi Weinstein schemes of blind equalization," *Signal Processing*, vol. 73, no. 1-2, pp. 185–190, 1999.
41. A. Neves, "Uma abordagem unificada para algoritmos de equalização autodidata," Tese de Mestrado, Unicamp, 2001.
42. H. Zeng, L. Tong, and C. Johnson, "Relationships between the constant modulus and wiener receivers," *IEEE Trans. on Information Theory*, vol. 44, no. 4, pp. 1523–1538, 1998.
43. P. Regalia and M. Mboup, "Undermodeled equalization: A characterization of stationary points for a family of blind criteria," *IEEE Trans. on Signal Processing*, vol. 47, no. 3, pp. 760–770, 1999.
44. R. Suyama, R. R. F. Attux, J. M. T. Romano, and M. Bellanger, "Relations entre les critères du module constant et de wiener," *19e Colloque GRETSI - Paris - França*, 2003.
45. Z. Ding, R. Kennedy, B. Anderson, and C. Jonhson, "Ill- convergence of Godard blind equalizers in data communication systems," *IEEE Trans. on Communications*, vol. 39, no. 9, pp. 1313–1327, 1991.
46. G. Foschini, "Equalization without altering or detecting data," *Bell Syst. Tech. J.*, vol. 64, no. 8, pp. 1885–1911, 1985.
47. Z. Ding, C. Jonhson, and R. Kennedy, "On the (non)existence of undesirable equilibria of Godard blind equalizers," *IEEE Trans. on Signal Processing*, vol. 40, no. 10, pp. 2425–2432, 1992.

48. C. Johnson and B. Anderson, "Godard blind equalizer error surface characteristics: White, zero-mean, binary source case," *International Journal of Adaptive Control and Signal Processing*, no. 9, pp. 301–324, 1995.
49. Y. Li, K. J. R. Liu, and Z. Ding, "Length- and cost-dependent local minima of unconstrained blind channel equalizers," *IEEE Transactions on Signal Processing*, vol. 44, no. 11, pp. 2726–2735, Nov. 1996.
50. G. Picchi and G. Prati, "Blind equalization and carrier recovery using a "stop-and-go" decision-directed algorithm," *IEEE Trans. on Communications*, vol. COM-35, pp. 877–887, 1987.
51. R. Ferrari, "Equalização de canais de comunicação digital baseada em filtros fuzzy," Tese de Mestrado, Unicamp, 2005.
52. C. C. Cavalcante, J. R. M. Filho, B. Dorizzi, and J. C. M. Mota, "Unsupervised channel equalization using fuzzy prediction-error filters," in *Proceedings of IEEE Neural Networks for Signal Processing*, Sydney, Australia, 2000.
53. R. Ferrari, C. M. Panazio, R. R. F. Attux, C. C. Cavalcante, L. N. de Castro, F. J. V. Zuben, and J. M. T. Romano, "Unsupervised channel equalization using fuzzy prediction-error filters," in *Proceedings of IEEE Neural Networks for Signal Processing (NNSP'03)*, Paris, France, 2003, pp. 1680–1686.
54. H. L. V. Trees, *Detection, Estimation and Linear Modulation Theory*. Wiley, 2001.
55. S. Chen, B. Mulgrew, and P. M. Grant, "A clustering technique for digital communications channel equalization using radial basis function networks," *IEEE Trans. on Neural Networks*, vol. 4, no. 4, pp. 570–579, julho 1993.
56. S. K. Patra and B. Mulgrew, "Efficient architecture for bayesian equalization using fuzzy filter," *IEEE Trans. on Circuits and Systems - II*, vol. 45, no. 7, pp. 812–820, July 1998.
57. K. Matsuoka and S. Nakashima, "Minimal distortion principle for blind source separation," in *Proc. ICA 2001*, December 2001, pp. 722–727.
58. S. C. Douglas and X. Sun, "Convolutive blind separation of speech mixtures using the natural gradient," *Speech Communication*, vol. 39, pp. 65–78, 2003.
59. P. Smaragdis, "Blind separation of convolved mixtures in the frequency domain," *Neurocomputing*, vol. 22, pp. 21–34, 1998.
60. S. Ikeda and N. Murata, "A method of ICA in time-frequency domain," in *Proceedings of International Workshop on Independent Component Analysis and Blind Signal Separation*, Aussions, France, January 1999, pp. 365–371.
61. Y. Zhou and B. Xu, "Blind source separation in frequency domain," *Signal Processing*, vol. 83, no. 9, pp. 2037–2046, September 2003.
62. A. Dapena and L. Castedo, "Blind Source Separation in the Frequency Domain: A Novel Solution to the Amplitude and the Permutation Indeterminacies," in *Proc. IWANN 2001*, Granada, Spain, 2001, pp. 603–610.
63. H. Sawada, R. Mukai, S. Araki, and S. Makino, "A robust and precise method for solving the permutation problem of frequency-domain blind source separation," in *Proc. ICA2003*, 2003, pp. 505–510.
64. R. Suyama, L. T. Duarte, R. Ferrari, L. E. P. Rangel, R. R. de Faissol Attux, C. C. Cavalcante, F. J. Von Zuben, and J. M. T. Romano, "A nonlinear-prediction approach to the blind separation of convolutive mixtures," *EURASIP Journal on Applied Signal Processing*, vol. Accepted for publication, 2006.
65. A. Hyvärinen and P. Pajunen, "Nonlinear Independent Component Analysis: Existence and Uniqueness Results," *Neural Networks*, vol. 3, pp. 429–439, 1999.
66. A. Taleb and C. Jutten, "Source separation in post-nonlinear mixtures," *IEEE Transactions on Signal Processing*, vol. 47, no. 10, pp. 2807–2820, 1999.
67. A. Taleb, "A generic framework for blind source separation in structured nonlinear models," *IEEE Transactions on Signal Processing*, vol. 50, no. 8, pp. 1819–1930, 2002.
68. J. Eriksson and V. Koivunen, "Blind identifiability of class of nonlinear instantaneous ICA models," in *Proceedings of the XI European Signal Processing Conference (EUSIPCO 2002)*, vol. 2, Toulouse, France, setembro 2002, pp. 7–10.
69. M. Babaie-Zadeh, "On Blind Source Separation in Convolutive and Nonlinear Mixtures," Tese de Doutorado, Institut National Polytechnique de Grenoble (França), setembro 2004.
70. S. Bermejo, C. Jutten, and J. Cabestany, "ISFET Source Separation: Foundations and Techniques," *Sensors and Actuators B - Chemical*, vol. 13, pp. 222–233, 2006.
71. S. Achard and C. Jutten, "Identifiability of Post-Nonlinear Mixtures," *IEEE Signal Processing Letters*, vol. 12, no. 5, pp. 423–426, 2005.
72. J. Solé-Casals, M. Babaie-Zadeh, C. Jutten, and D.-T. Pham, "Improving Algorithm Speed in PNL Mixture Separation and Wiener System Inversion," in *Proc. of the 4th Int. Symp. on Independent Component Analysis and Blind Signal Separation*, Nara, Japão, 2003.
73. K. Zhang and L.-W. Chan, "Extended Gaussianization Method for Blind Separation of Post-Nonlinear Mixtures," *Neural Computation*, vol. 17, no. 2, pp. 425–452, 2005.
74. S. S. Chen and R. A. Gopinath, "Gaussianization," in *Proc. Neural Information Processing Systems (NIPS)*, Denver, EUA, 2000, pp. 423–429.
75. S. Achard, "Mesures de Dépendance pour La Séparation Aveugle de Sources," Tese de Doutorado, Universite Joseph Fourier (França), setembro 2003.
76. L. Duarte, R. Suyama, R. R. de Faissol Attux, F. J. V. Zuben, and J. Romano, *Blind Source Separation of Post-nonlinear Mixtures Using Evolutionary Computation and Order Statistics*, Feb. 2006, vol. 3889.
77. F. Rojas, C. G. Puntonet, M. Rodríguez-Álvarez, I. Rojas, and R. Martín-Clemente, "Blind Source Separation in Post-Nonlinear Mixtures Using Competitive Learning, Simulated Annealing, and a Genetic Algorithm," *IEEE Transactions on Systems, Man and Cybernetics - Part C: Applications and Reviews*, vol. 34, no. 4, pp. 407–416, 2004.
78. L. D. Lathauwer, B. D. Moor, and J. Vanderwalle, "Ica algorithms for 3 sources and 2 sensors," in *Proceedings of IEEE Signal Processing Workshop on Higher-Order Statistics (HOS'99)*, Caesarea, Israel, June 1999, pp. 116–120.

79. L. D. Lathauwer, B. D. Moor, J. Vandewalle, and J.-F. Cardoso, "Independent component analysis of largely underdetermined mixtures," in *Proceedings of the 4th Internationa Symposium on Independent Component Analysis and Blind Signal Separation (ICA2003)*, Nara, Japan, Abril 2003, pp. 29–33.
80. L. D. Lathauwer, B. D. Moor, and J. Vanderwalle, "A multilinear singular value decomposition," *SIAM Journal on Matrix Analysis and Applications*, vol. 21, no. 4, pp. 1253–1278, April 2000.
81. P. Bofill and M. Zibulevsky, "Underdetermined blind source separationnext term using sparse representations," *Signal Processing*, vol. 81, no. 11, pp. 2353–2362, November 2001.
82. P. Georgiev and A. Cichoki, "Sparse component analysis of overcomplete mixtures by improved basis pursuit method," in *Proceedings of the 2004 International Symposium on Circuits and Systems, 2004. ISCAS '04.*, vol. 5, 23-26 May 2004, pp. V–37–V–40Vol.5.
83. S. Amari, "Independent component analysis, sparse component analysis and non-negative matrix factorization," in *Proc. Nonlinear Signal and Image Processing, 2005. NSIP 2005*, 18-20 May 2005, pp. viii–viii.
84. P. Georgiev, F. Theis, and A. Cichocki, "Sparse component analysis and blind source separation of underdetermined mixtures," *IEEE Transactions on Neural Networks*, vol. 16, no. 4, pp. 992–996, July 2005.
85. E. Kofidis, "Blind source separation: Fundamentals and recent advances," Mini-curso no XIX Simpósio Brasileiro de Telecomunicações (SBrT2001), Tech. Rep., 2001.
86. C. B. Papadias and A. J. Paulraj, "A constant modulus algorithm for multiuser signal separation in presence of delay spread using antenna arrays," *IEEE Signal Processing Letters*, vol. 4, no. 6, pp. 178–181, June 1997.
87. M. Fliess and H. Sira-Ramírez, "An algebraic framework for linear identification," *ESAIM Control Optim. Calculus Variations*, no. 9, pp. 151–168, 2003.
88. A. Neves, "Identification algébrique et déterministe de signaux et systèmes à temps continu: Application à des problèmes de communication numérique," Ph.D. dissertation, Université René Descartes - Paris V, 2005.
89. H. Sira-Ramirez, J. R. Trapero, and V. Feliu-Batlle, "Frequency identification in the noisy sum of two sinusoidal signals," in *Proc. Mathematical Theory of Networks and Systems (MTNS)*, Kyoto, Japan, 2006.
90. D. E. Goldberg, *Genetic Algorithms in Search, Optimization and Machine Learning*. Reading, MA, USA: Addison-Wesley, 1989.
91. J. Holland, *Adaptation in Natural and Artificial Systems*. Ann Arbor, USA: University of Michigan Press, 1975.
92. A. Neubauer, "Non-linear adaptive filters based on genetic algorithms with applications to digital signal processing," in *IEEE International Conference on Evolutionary Computation*, vol. 2, 29 Nov.-1 Dec. 1995, pp. 527–532vol.2.
93. L. N. de Castro and J. I. Timmis, *Artificial Immune Systems: A New Computational Intelligence Approach*. London: Springer-Verlag, September 2002.
94. L. N. de Castro and J. Timmis, "An artificial immune network for multimodal function optimization," in *Proceedings of the 2002 Congress on Evolutionary Computation*, vol. 1, 12-17 May 2002, pp. 699–704.
95. K. Krishnakumar, "Micro-genetic algorithms for stationary and non-stationary function optimization," in *SPIE's Intelligent Control and Adaptive Systems Conf.*, 1989.

# 3

# Tensor Decompositions and Applications to Wireless Communication Systems

André L. F. de Almeida[1], Gérard Favier[1] and João C. M. Mota[2]

[1] I3S Laboratory/CNRS, University of Nice-Sophia Antipolis, France, {lima,favier}@i3s.unice.fr
[2] Wireless Telecommunication Research Group, Federal University of Ceará, Brazil, mota@deti.ufc.br

## 3.1 Introduction

The theory of tensors is a branch of linear algebra, called *multilinear algebra*. The word "tensor" was first introduced by William R. Hamilton in 1846, to denote what is now called *modulus*. In 1899, Woldemar Voigt was the first who used this word in its current meaning. The first tensor notations and developments were done by Gregorio Ricci-Curbastro around 1890 in the context of differential geometry. A broader acceptance of tensor calculus was achieved with the introduction of Einstein's theory of general relativity, around 1915.

A different way of viewing and treating a tensor was developed between 1960 and 1970, when the attention was given to the analysis, factorization or decomposition of third-order tensors. Tucker [1], Harshman [2], Carroll and Chang [3] and Kruskal [4] are the first "players" in the development of *tensor decompositions*, which can be seen as extensions of matrix decompositions to higher-orders. Among them, two types have been extensively studied in the literature, while being focus of several applications in different domains. These are the Parallel Factor (PARAFAC) decomposition [2, 5], also known as Canonical Decomposition (CANDECOMP) [3], and the Tucker3 decomposition, which is a generalization of principal component analysis to higher orders [1]. A solid contribution to the area of multilinear algebra and tensor decompositions was given by De Lathauwer in 1997 [6], who generalized the concept of matrix Singular Value Decomposition (SVD) to tensors of higher-orders, with applications to blind source separation problems.

Around 2000, Sidiropoulos introduced tensor decompositions as a powerful modeling and signal processing tool for array processing problems in the context of wireless communications. This interdisciplinary research field has gained more and more attention, and several recent contributions have proposed tensor models for a wide variety of signal processing problems, such as blind multiuser detection for wireless communication systems including Code Division Multiple Access (CDMA) and Orthogonal Frequency Division Multiplexing (OFDM) systems [7, 8, 9, 10, 11, 12, 13], high-resolution direction finding [14], blind beamforming and equalization [15, 16, 17], multipath parameter estimation [18, 19, 20], and multiantenna space-time coding [17, 21, 22, 23, 24], among others.

From a wireless communications point of view, the fact that the received signal is, for example, a third-order tensor, means that each received signal sample to be processed is associated with a given coordinate in a three-dimensional space, i.e., it is modeled by three indices, each one of which characterizing a particular type of systematic variation on the received signal. In such a three-dimensional space, each *axis* of the received signal tensor can be interpreted as a particular form of signal *diversity*. In most of cases, two of these three axes account for *space* and *time* diversity dimensions. The *space* dimension corresponds to the number of receiver antennae while the *time* dimension corresponds to the number of symbols processed at the receiver. The *extra* (third) dimension generally depends on a particular type of processing that is done either at the transmitter or at the receiver or both. For instance, in a CDMA system this third diversity dimension is the *spreading* dimension which appears due to the use of a spreading code at the transmitter. Similarly, the use of temporal oversampling at the output of each receive antenna or multicarrier modulation such as OFDM at the transmitter also creates such an extra dimension to the receiver signal, which is called *oversampling* and *frequency* dimensions. In multiantenna coding problems, the code *redundancy* also plays the role of the third signal dimension. More attention is generally given to decompositions of third-order tensors or *three-way arrays*, since this is the case in most of the wireless communication applications encountered in practice.

The rest of this chapter is divided as follows. In Section 2, a background on fundamental aspects of tensor decompositions is given. The focus is on Tucker3, PARAFAC and constrained Block-PARAFAC decompositions, which are the most important ones in the present context. This section also introduces some notations and useful properties. In Section 3, some recently developed applications of tensor decompositions to wireless communications are presented.

The focus is on the use of tensor decompositions for a unified modeling of oversampled, DS/CDMA and OFDM systems, for the design of new multiantenna coding techniques, and for the estimation of wireless channel parameters. The applications are illustrated by some simulation results. The chapter is concluded in Section 4.

## 3.2 Background on Tensor Decompositions

This section is focused on the decomposition of higher-order tensors (multi-way arrays). Tensor decompositions, also referred as *multi-way factor analysis* is an area of the multilinear algebra that characterizes a tensor as a linear combination of *outer product factors*. Depending on the considered approach, tensor decompositions can be viewed as generalizations of Principal Component Analysis (PCA) or Singular Value Decomposition (SVD) to higher-orders. The analysis of a tensor in terms of its factors is useful in problems where a *multilinear mixture* of different factors or contributions must be identified from measured data. In the context of wireless communication signal processing, the computation of an observed data tensor decomposition allows to separate the signals transmitted by different sources. This is exactly the goal of several signal processing problems that are addressed in this chapter.

In the following, some tensor decompositions are presented. More attention is given to decompositions of third-order tensors or *three-way arrays*, since this will be the case in most of the applications encountered in this work. Therefore, each considered decomposition is first discussed assuming the third-order case. This will simplify their exposition and understanding.

*Notations and Properties*: Some notations and properties are now defined. $\mathbf{A}^T, \mathbf{A}^{-1}$ and $\mathbf{A}^\dagger$ stand for transpose, inverse and pseudo-inverse of $\mathbf{A}$, respectively. The operator $Diag(\mathbf{a})$ forms a diagonal matrix from its vector argument; $BlockDiag(\mathbf{A}_1 \cdots \mathbf{A}_N)$ constructs a block-diagonal matrix of $N$ blocks from its argument matrices; $D_i(\mathbf{A})$ forms a diagonal matrix holding the $i$-th row of $\mathbf{A}$ on its main diagonal; $vec(\cdot)$ stacks the columns of its matrix argument in a vector; $\otimes$ and $\diamond$ denote the Knonecker product and the Khatri-Rao product, respectively:

$$\mathbf{A} \diamond \mathbf{B} = [\mathbf{A}_{\cdot 1} \otimes \mathbf{B}_{\cdot 1}, \ldots, \mathbf{A}_{\cdot R} \otimes \mathbf{B}_{\cdot R}],$$

where $\mathbf{A} = [\mathbf{A}_{\cdot 1} \ldots \mathbf{A}_{\cdot R}] \in \mathbb{C}^{I \times R}$ and $\mathbf{B} = [\mathbf{B}_{\cdot 1} \ldots \mathbf{B}_{\cdot R}] \in \mathbb{C}^{J \times R}$. We make use of the following property of the Khatri-Rao product:

$$(\mathbf{A} \diamond \mathbf{B})^T (\mathbf{A} \diamond \mathbf{B}) = (\mathbf{A}^T \mathbf{A}) \odot (\mathbf{B}^T \mathbf{B}), \tag{3.1}$$

where $\odot$ is the Schur-Hadamard (element-wise) product. Given $\mathbf{A} \in \mathbb{C}^{I \times R}$, $\mathbf{B} \in \mathbb{C}^{J \times S}$, $\mathbf{C} \in \mathbb{C}^{R \times P}$ and $\mathbf{D} \in \mathbb{C}^{S \times P}$, we have:

$$(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \diamond \mathbf{D}) = \mathbf{AC} \diamond \mathbf{BD}, \tag{3.2}$$

and

$$(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = \mathbf{AC} \otimes \mathbf{BD}. \tag{3.3}$$

In this paper, scalars are denoted by lower-case letters $(a, b, \ldots, \alpha, \beta, \ldots)$, vectors are written as boldface lower-case letters $(\mathbf{a}, \mathbf{b}, \ldots)$, matrices correspond to boldface capitals $(\mathbf{A}, \mathbf{B}, \ldots)$, and tensors are written as calligraphic letters $(\mathcal{A}, \mathcal{B}, \ldots)$.

### 3.2.1 Tucker3

The Tucker3 decomposition was proposed by L. Tucker in the sixties [1]. It can be seen as an extension of bilinear factor analysis to third-order tensors. The Tucker3 decomposition is also a common name to denote the SVD of a third-order tensor [6]. The Tucker3 decomposition is general in the sense that it incorporates most of the other third-order tensor decompositions as special cases. The Tucker3 decomposition of a tensor $\mathcal{X} \in \mathbb{C}^{I \times J \times K}$ can be written in scalar form as:

$$x_{i,j,k} = \sum_{p=1}^{P} \sum_{q=1}^{Q} \sum_{r=1}^{R} a_{i,p} b_{j,q} c_{k,r} g_{p,q,r}, \tag{3.4}$$

where $a_{i,p} = [\mathbf{A}]_{i,p}$, $b_{j,q} = [\mathbf{B}]_{j,q}$ and $c_{k,r} = [\mathbf{C}]_{k,r}$ are scalar components of *three factor matrices* $\mathbf{A} \in \mathbb{C}^{I \times P}$, $\mathbf{B} \in \mathbb{C}^{J \times Q}$ and $\mathbf{C} \in \mathbb{C}^{K \times R}$ respectively, and $g_{p,q,r} = [\mathcal{G}]_{p,q,r}$ is a scalar component of the *core tensor* $\mathcal{G} \in \mathbb{C}^{P \times Q \times R}$.

It can be concluded from (3.4), that a Tucker3-decomposed tensor is equal to a linear combination (or weighted sum) of $PQR$ outer products, where the coefficient (or weighting factor) of each outer product term is the corresponding scalar component of the core tensor. We call $P$ as the number of factors in the first mode of the tensor $\mathcal{X}$. Similarly, $Q$ and $R$ denote the number of factors in the second and third modes of $\mathcal{X}$. The Tucker3 decomposition can

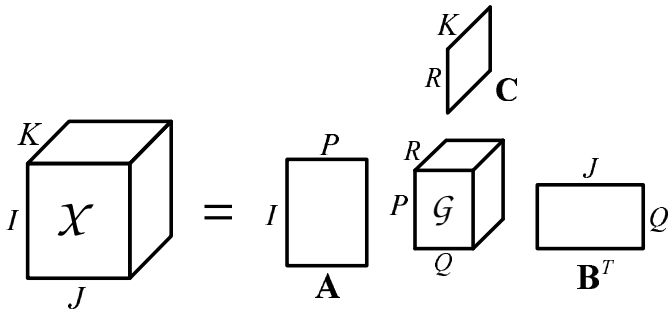



**Fig. 3.1.** The Tucker3 decomposition.

be referred as a tensor decomposition that allows *interactions* among factors across the three modes of the tensor [2]. An illustration of the Tucker3 decomposition is given in Fig. 3.1. The Tucker3 decomposition can be stated without using the triple sum and resorting to the $n$-mode product notation.

**Definition 1.** *($n$-mode product) The $n$-mode product of a tensor $\mathcal{X} \in \mathbb{C}^{I_1 \times I_2 \times \cdots \times I_N}$ and a matrix $\mathbf{A} \in \mathbb{C}^{J_n \times I_n}$, denoted by $\mathcal{X} \times_n \mathbf{A}$ is specified as:*

$$[\mathcal{X} \times_n \mathbf{A}]_{i_1,i_2,\ldots,i_{n-1},j_n,i_{n+1},\ldots,i_N} = \sum_{i_n=1}^{I_n} x_{i_1,i_2,\ldots,i_{n-1},i_n,i_{n+1},\ldots,i_N} a_{j_n,i_n} \tag{3.5}$$

*The result of an $n$-mode product is a tensor of the same order, but with a new $n$-th dimension $J_n$. It is straightforward to define $n$-mode products acting on more than one mode. Successive $n$-mode products are associative, i.e., for two matrices $\mathbf{A} \in \mathbb{C}^{J_n \times I_n}$ and $\mathbf{B} \in \mathbb{C}^{J_m \times I_m}$, we have that [6]:*

$$(\mathcal{X} \times_n \mathbf{A}) \times_m \mathbf{B} = (\mathcal{X} \times_m \mathbf{B}) \times_n \mathbf{A}. \tag{3.6}$$

Taking this definition into account, we can alternatively express the Tucker3 decomposition as:

$$\mathcal{X} = \mathcal{G} \times_1 \mathbf{A} \times_2 \mathbf{B} \times_3 \mathbf{C}. \tag{3.7}$$

Alternatively, we can state the Tucker3 decomposition using a *matrix-slice notation*. This notation characterizes the tensor by a set of parallel matrix-slices that are obtained by "slicing" the tensor in a given "direction". Each matrix-slice is obtained by fixing one index of a given *mode* and varying the two indices of the other two modes. For a third-order tensor there are three possible slicing directions. We call $\mathbf{X}_{i\cdot\cdot} \in \mathbb{C}^{J \times K}$ the $i$-th *first-mode slice*, $\mathbf{X}_{\cdot j\cdot} \in \mathbb{C}^{K \times I}$ the $j$-th *second-mode slice* and $\mathbf{X}_{\cdot\cdot k} \in \mathbb{C}^{I \times J}$ the $k$-th *third-mode slice*. In order to obtain the matrix-slice notation for the Tucker3 decomposition, we rewrite (3.4) as:

$$x_{i,j,k} = \sum_{q=1}^{Q} \sum_{r=1}^{R} b_{j,q} c_{k,r} \left( \sum_{p=1}^{P} a_{i,p} g_{p,q,r} \right), \tag{3.8}$$

and define an "equivalent" (first-mode combined) core as:

$$u_{i,q,r}^{(a)} = \sum_{p=1}^{P} a_{i,p} g_{p,q,r} = [\mathcal{G} \times_1 \mathbf{A}]_{i,q,r}. \tag{3.9}$$

The $i$-th matrix slice $\mathbf{X}_{i\cdot\cdot}$, $i = 1, \ldots, I$, is given by:

$$\mathbf{X}_{i\cdot\cdot} = \mathbf{B}\mathbf{U}_{i\cdot\cdot}^{(a)}\mathbf{C}^T, \quad i = 1, \ldots I, \tag{3.10}$$

where $\mathbf{U}_{i\cdot\cdot}^{(a)}$ is the $i$-th first-mode matrix-slice of the transformed core tensor $\mathcal{U}^{(a)} \in \mathbb{C}^{I \times Q \times R}$. The other two matrix-slice notations are obtained in a similar way, by changing the order of the summations in (3.4) and defining

$$u_{p,j,r}^{(b)} = \sum_{q=1}^{Q} b_{j,q} g_{p,q,r}, \quad u_{p,q,k}^{(c)} = \sum_{r=1}^{R} c_{k,r} g_{p,q,r}, \tag{3.11}$$

as scalar component of the transformed core tensors $\mathcal{U}^{(b)} = [\mathcal{G} \times_2 \mathbf{B}] \in \mathbb{C}^{P\times J\times R}$ and $\mathcal{U}^{(c)} = [\mathcal{G} \times_3 \mathbf{C}] \in \mathbb{C}^{P\times Q\times K}$. This leads to:

$$\mathbf{X}_{\cdot j\cdot} = \mathbf{C}\mathbf{U}^{(b)}_{\cdot j\cdot}\mathbf{A}^T, \quad j = 1, \ldots J, \tag{3.12}$$

and

$$\mathbf{X}_{\cdot\cdot k} = \mathbf{A}\mathbf{U}^{(c)}_{\cdot\cdot k}\mathbf{B}^T, \quad k = 1, \ldots K. \tag{3.13}$$

Let $\mathbf{X}_1 \in \mathbb{C}^{KI\times J}$, $\mathbf{X}_2 \in \mathbb{C}^{IJ\times K}$ and $\mathbf{X}_3 \in \mathbb{C}^{JK\times I}$ be the first- second- and third-mode matrix unfoldings of $\mathcal{X}$. These matrices are defined as:

$$\mathbf{X}_1 = \begin{bmatrix} \mathbf{X}_{\cdot\cdot 1} \\ \vdots \\ \mathbf{X}_{\cdot\cdot K} \end{bmatrix}, \quad \mathbf{X}_2 = \begin{bmatrix} \mathbf{X}_{\cdot 1\cdot} \\ \vdots \\ \mathbf{X}_{\cdot J\cdot} \end{bmatrix}, \quad \mathbf{X}_3 = \begin{bmatrix} \mathbf{X}_{1\cdot\cdot} \\ \vdots \\ \mathbf{X}_{I\cdot\cdot} \end{bmatrix},$$

It can be shown from (3.10), (3.12) and (3.13) that $\mathbf{X}_1$, $\mathbf{X}_2$ and $\mathbf{X}_3$ can be expressed as:

$$\mathbf{X}_1 = (\mathbf{C}\otimes\mathbf{A})\mathbf{G}_1\mathbf{B}^T, \quad \mathbf{X}_2 = (\mathbf{B}\otimes\mathbf{C})\mathbf{G}_2\mathbf{A}^T, \quad \mathbf{X}_3 = (\mathbf{A}\otimes\mathbf{B})\mathbf{G}_3\mathbf{C}^T, \tag{3.14}$$

where $\mathbf{G}_1 \in \mathbb{C}^{RP\times Q}$, $\mathbf{G}_2 \in \mathbb{C}^{PQ\times R}$ and $\mathbf{G}_3 \in \mathbb{C}^{QR\times P}$ are unfolded matrices of the core tensor $\mathcal{G}$, which are constructed in the same fashion as (3.14), i.e.:

$$\mathbf{G}_1 = \begin{bmatrix} \mathbf{G}_{\cdot\cdot 1} \\ \vdots \\ \mathbf{G}_{\cdot\cdot R} \end{bmatrix}, \quad \mathbf{G}_2 = \begin{bmatrix} \mathbf{G}_{\cdot 1\cdot} \\ \vdots \\ \mathbf{G}_{\cdot Q\cdot} \end{bmatrix}, \quad \mathbf{G}_3 = \begin{bmatrix} \mathbf{G}_{1\cdot\cdot} \\ \vdots \\ \mathbf{G}_{P\cdot\cdot} \end{bmatrix}. \tag{3.15}$$

Each one of the three matrix unfolding representations in (3.14) are different rearrangements of the same information contained in the tensor $\mathcal{X}$.

The Tucker3 decomposition is not unique, since there are infinite solutions for the factor matrices and for the core tensor leading to the same tensor $\mathcal{X}$. In other words, the Tucker3 decomposition allows *free rotation* or linear transformations on the three factor matrices (provided that the inverse of these transformations is applied to the core tensor) without affecting the reconstructed tensor $\mathcal{X}$. In order to see this, let us define nonsingular matrices $\mathbf{T}_a \in \mathbb{C}^{P\times P}$, $\mathbf{T}_b \in \mathbb{C}^{Q\times Q}$ and $\mathbf{T}_c \in \mathbb{C}^{R\times R}$. Considering the unfolded matrix $\mathbf{X}_1$ and using property (3.3), we have that:

$$\begin{aligned}\mathbf{X}_1 &= (\mathbf{C}\mathbf{T}_c\mathbf{T}_c^{-1}\otimes\mathbf{A}\mathbf{T}_a\mathbf{T}_a^{-1})\mathbf{G}_1(\mathbf{B}\mathbf{T}_b\mathbf{T}_b^{-1})^T \\ &= [(\mathbf{C}\mathbf{T}_c)\otimes(\mathbf{A}\mathbf{T}_a)]\left[(\mathbf{T}_c^{-1}\otimes\mathbf{T}_a^{-1})\mathbf{G}_1\mathbf{T}_b^{-T}\right](\mathbf{B}\mathbf{T}_b)^T,\end{aligned}$$

i.e.,

$$\mathbf{X}_1 = (\mathbf{C}'\otimes\mathbf{A}')\mathbf{G}_1'\mathbf{B}'^T, \tag{3.16}$$

where $\mathbf{A}' = \mathbf{A}\mathbf{T}_a$, $\mathbf{B}' = \mathbf{B}\mathbf{T}_b$ and $\mathbf{C}' = \mathbf{C}\mathbf{T}_c$ are transformed factor matrices and $\mathbf{G}_1' = (\mathbf{T}_c^{-1}\otimes\mathbf{T}_a^{-1})\mathbf{G}_1\mathbf{T}_b^{-T}$ is a transformed Tucker3 core. Equation (3.16) means that we have an infinite number of matrices $\mathbf{A}'$, $\mathbf{B}'$, $\mathbf{C}'$ and $\mathbf{G}_1'$ giving rise to the same matrix $\mathbf{X}_1$. This fact clearly states the general nonuniqueness of the Tucker3 decomposition. Complete uniqueness of the factor matrices and the core tensor of a Tucker3 decomposition is only possible in some special cases, where at least two factor matrices have some special structure that allows a unique determination of the transformation matrices. It has been proved that partial uniqueness (i.e., uniqueness of at least some factors of the model) may exist in cases where the Tucker3 core tensor is constrained to have several elements equal to 0 [25].

**Special cases: Tucker2 and Tucker1**

Consider a Tucker3 decomposition and rewrite (3.4) as:

$$x_{i,j,k} = \sum_{p=1}^{P}\sum_{q=1}^{Q} a_{i,p}b_{j,q}\left(\sum_{r=1}^{R} c_{k,r}g_{p,q,r}\right)$$

or equivalently,

$$x_{i,j,k} = \sum_{p=1}^{P}\sum_{q=1}^{Q} a_{i,p}b_{j,q}h_{p,q,k}, \tag{3.17}$$

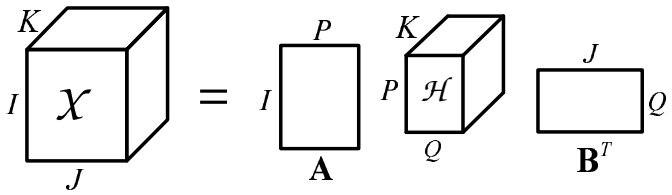



**Fig. 3.2.** The Tucker2 decomposition.

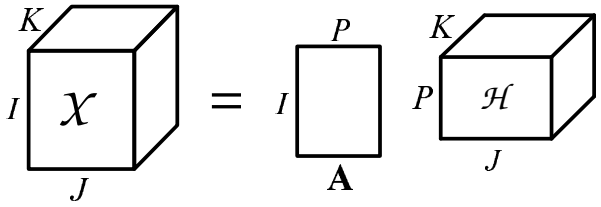



**Fig. 3.3.** The Tucker1 decomposition.

where $c_{k,r}$ has been absorbed in the core $g_{p,q,r}$, giving rise to an equivalent core $h_{p,q,k}$ i.e., $h_{p,q,k} = [\mathcal{G} \times_3 \mathbf{C}]_{p,q,k}$. Equation (3.17) is the scalar notation of an equivalent Tucker2 decomposition. Note that the Tucker2 decomposition is simpler than its Tucker3 equivalent, since the number of outer product terms has been reduced to $PQ$. A Tucker2 decomposition also arises from a Tucker3 one when one of the factor matrices, say $\mathbf{C}$, is equal to the identity matrix. The matrix-slice and unfolding notations for the Tucker2 case can be easily obtained from (3.10), (3.12), (3.13) and (3.14) by setting $\mathbf{C} = \mathbf{I}_K$ and $\mathcal{G} = \mathcal{H} \in \mathbb{C}^{P \times Q \times K}$. The Tucker2 decomposition is illustrated in Fig. 3.2.

Now, let us rewrite (3.4) as:

$$x_{i,j,k} = \sum_{p=1}^{P} a_{i,p} \left( \sum_{q=1}^{Q} \sum_{r=1}^{R} b_{j,q} c_{k,r} g_{p,q,r} \right),$$

i.e.,

$$x_{i,j,k} = \sum_{p=1}^{P} a_{i,p} h_{p,j,k}, \tag{3.18}$$

where both $b_{j,q}$ and $c_{k,r}$ have been absorbed in the core $g_{p,q,r}$, resulting in another core $h_{p,j,k}$ i.e., $h_{p,j,k} = [\mathcal{G} \times_2 \mathbf{B} \times_3 \mathbf{C}]_{p,j,k}$. Equation (3.18) is the scalar notation of a Tucker1 decomposition. A Tucker1 decomposition also arises from a Tucker3 one when two factor matrices, say $\mathbf{B}$ and $\mathbf{C}$, are equal to identity matrices. The matrix-slice and unfolding notations for the Tucker1 case are obtained from (3.10), (3.12), (3.13) and (3.14) by setting $\mathbf{B} = \mathbf{I}_J$, $\mathbf{C} = \mathbf{I}_K$ and $\mathcal{G} = \mathcal{H} \in \mathbb{C}^{P \times J \times K}$. Figure 3.3 illustrates the Tucker1 decomposition.

### 3.2.2 Parallel Factors (PARAFAC)

The PARAllel FACtors (PARAFAC) decomposition, also known as CANonical DECOMPosition (CANDECOMP), was independently developed by Harshman [2] and Carol & Chang [3] in the seventies. It is also known by the acronym CP (Candecomp-Parafac). For a third-order tensor, it is a decomposition in a sum of *triple products* or *triads*. PARAFAC can be equivalently stated as a decomposition of a three-way array in a sum of rank-1 tensors. The PARAFAC decomposition of a tensor $\mathcal{X} \in \mathbb{C}^{I \times J \times K}$ has the following scalar form:

$$x_{i,j,k} = \sum_{q=1}^{Q} a_{i,q} b_{j,q} c_{k,q}, \tag{3.19}$$

where $a_{i,q} = [\mathbf{A}]_{i,q}$, $b_{j,q} = [\mathbf{B}]_{j,q}$ and $c_{k,q} = [\mathbf{C}]_{k,q}$ are scalar components of factor matrices $\mathbf{A} \in \mathbb{C}^{I \times Q}$, $\mathbf{B} \in \mathbb{C}^{J \times Q}$ and $\mathbf{C} \in \mathbb{C}^{K \times Q}$ respectively. $Q$ is the number of factors, also known as the *rank* of the decomposition. The columns of the first-, second- and third-factor matrices $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$ are respectively called first-, second- and third-mode *factor loadings*. Other synonyms for the columns of $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$ are *loading patterns* or *loading vectors*. In Fig. 3.4, a third-order PARAFAC decomposition is depicted. From the three possible slicing directions, we get the following matrix

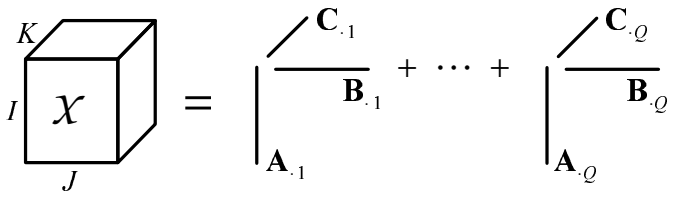



**Fig. 3.4.** The third-order PARAFAC decomposition.

writings for the PARAFAC decomposition:

$$\begin{aligned}\mathbf{X}_{i\cdot\cdot} &= \mathbf{B}D_i(\mathbf{A})\mathbf{C}^T,\\ \mathbf{X}_{\cdot j\cdot} &= \mathbf{C}D_j(\mathbf{B})\mathbf{A}^T,\\ \mathbf{X}_{\cdot\cdot k} &= \mathbf{A}D_k(\mathbf{C})\mathbf{B}^T.\end{aligned} \tag{3.20}$$

By stacking row-wise the first-, second- and third-mode slices we have:

$$\begin{aligned}\mathbf{X}_1 &= \begin{bmatrix}\mathbf{X}_{\cdot\cdot 1}\\ \vdots \\ \mathbf{X}_{\cdot\cdot K}\end{bmatrix} = \begin{bmatrix}\mathbf{A}D_1(\mathbf{C})\\ \vdots \\ \mathbf{A}D_K(\mathbf{C})\end{bmatrix}\mathbf{B}^T = (\mathbf{C}\diamond\mathbf{A})\mathbf{B}^T,\\ \mathbf{X}_2 &= \begin{bmatrix}\mathbf{X}_{\cdot 1\cdot}\\ \vdots \\ \mathbf{X}_{\cdot J\cdot}\end{bmatrix} = \begin{bmatrix}\mathbf{C}D_1(\mathbf{B})\\ \vdots \\ \mathbf{C}D_J(\mathbf{B})\end{bmatrix}\mathbf{A}^T = (\mathbf{B}\diamond\mathbf{C})\mathbf{A}^T,\\ \mathbf{X}_3 &= \begin{bmatrix}\mathbf{X}_{1\cdot\cdot}\\ \vdots \\ \mathbf{X}_{I\cdot\cdot}\end{bmatrix} = \begin{bmatrix}\mathbf{B}D_1(\mathbf{A})\\ \vdots \\ \mathbf{B}D_I(\mathbf{A})\end{bmatrix}\mathbf{C}^T = (\mathbf{A}\diamond\mathbf{B})\mathbf{C}^T.\end{aligned} \tag{3.21}$$

One of the most interesting properties of PARAFAC is its uniqueness. Contrarily to bilinear (matrix) decompositions, which are in general not unique for ranks greater than one (rank-one matrices are unique up to a scalar factor), the PARAFAC decomposition of tensors of rank greater than one can be unique up to scaling and permutation of factors.

The first uniqueness studies of the PARAFAC decomposition were done in the seventies by Harshman [2, 5]. The deepest formal uniqueness proof was provided by Kruskal in [4]. Kruskal derived *sufficient* conditions for uniqueness of third-order PARAFAC decompositions of real-valued tensors. Around two decades later, Sidiropoulos *et al* [7] extended Kruskal condition to complex-valued tensors. Sidiropoulos & Bro [26] further generalized Kruskal's uniqueness condition to $N$-th order tensors. In [27], Sidiropoulos and ten Berge showed that Kruskal's condition is not only sufficient but also *necessary* for $Q \in \{2, 3\}$. Further PARAFAC uniqueness issues were addressed by Jiang & Sidiropoulos in [28], who derived necessary and sufficient conditions for uniqueness of the so-called *restricted* PARAFAC decomposition (i.e., those where at least one factor matrix is full column-rank).

The study of the PARAFAC uniqueness condition is based on the fundamental concept of $k$-rank (*Kruskal*-rank), which is more restricted than the usual concept of matrix rank. The $k$-rank concept was proposed by Kruskal in his seminal paper [4], although the term "*Kruskal*-rank" was first used by Harshman and Lundy [29]. The $k$-rank concept has been extensively used as a key concept for stating PARAFAC uniqueness.

**Definition 2.** *($k$-rank): The rank of $\mathbf{A} \in \mathbb{C}^{I\times Q}$, denoted by $r_{\mathbf{A}}$, is equal to $r$* iff $\mathbf{A}$ *contains* at least *a collection of $r$ linearly independent columns but no collection of $r+1$ linearly independent columns. The* Kruskal-*rank (or $k$-rank) of $\mathbf{A}$ is the maximum number $k$ such that* every *set of $k$ columns of $\mathbf{A}$ is linearly independent. Note that the $k$-rank is always less than or equal to the rank, and we have that $k_{\mathbf{A}} \leq r_{\mathbf{A}} \leq min(I, Q), \forall\mathbf{A}$.*

Consider the set of $I$ matrix-slices $\mathbf{X}_{i\cdot\cdot} = \mathbf{B}D_i(\mathbf{A})\mathbf{C}^T, i = 1, \cdots, I$, defined in (3.20). If

$$k_{\mathbf{A}} + k_{\mathbf{B}} + k_{\mathbf{C}} \geq 2Q + 2, \tag{3.22}$$

the matrices $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$ are unique up to common permutation and (complex) scaling of their columns [4, 7]. This means that any matrices $\mathbf{A}'$, $\mathbf{B}'$ and $\mathbf{C}'$ satisfying (3.20) are linked to $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$ by:

$$\mathbf{A}' = \mathbf{A}\boldsymbol{\Pi}\boldsymbol{\Delta}_1, \quad \mathbf{B}' = \mathbf{B}\boldsymbol{\Pi}\boldsymbol{\Delta}_2, \quad \mathbf{C}' = \mathbf{C}\boldsymbol{\Pi}\boldsymbol{\Delta}_3, \tag{3.23}$$

where $\boldsymbol{\Pi}$ is a permutation matrix and $\boldsymbol{\Delta}_1$, $\boldsymbol{\Delta}_2$ and $\boldsymbol{\Delta}_3$ are diagonal matrices satisfying the condition

$$\boldsymbol{\Delta}_1\boldsymbol{\Delta}_2\boldsymbol{\Delta}_3 = \mathbf{I}_Q. \tag{3.24}$$

Condition (3.22) is also necessary if $Q \leq 3$. However, ten Berge & Sidiropoulos provided in [27] a simple counter-example to the necessity of (3.22). They also claimed that the uniqueness of PARAFAC depends on the particular joint pattern of zeros in the factor matrices. This was better explained and clarified by Jiang & Sidiropoulos in [28]. They provided both necessary and sufficient uniqueness conditions for PARAFAC, when one factor matrix is full column-rank. This justified the examples shown in [27]. When the elements of $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$ are drawn from a continuous distribution, the $k$-rank and the conventional rank coincides, and the PARAFAC uniqueness condition (3.22) can be equivalently stated as [7]:

$$min(I, Q) + min(J, Q) + min(K, Q) \geq 2Q + 2. \tag{3.25}$$

Condition (3.25) is sufficient but not always necessary, as shown in [27]. For general PARAFAC decompositions with $Q > 1$, *two necessary conditions* for uniqueness are [30]:

$$min(r_{\mathbf{A}\diamond\mathbf{B}}, r_{\mathbf{B}\diamond\mathbf{C}}, r_{\mathbf{C}\diamond\mathbf{A}}) = Q \quad \text{and} \quad min(k_{\mathbf{A}}, k_{\mathbf{B}}, k_{\mathbf{C}}) \geq 2. \tag{3.26}$$

This condition can be stated alternatively. Note that $r_{\mathbf{A}\diamond\mathbf{B}} \leq min(IJ, Q)$, $r_{\mathbf{B}\diamond\mathbf{C}} \leq min(JK, Q)$ and $r_{\mathbf{C}\diamond\mathbf{A}} \leq min(KI, Q)$. Note also that the meaning of $k_{\mathbf{A}} \geq 2$ is that matrix $\mathbf{A}$ has no proportional columns (otherwise $k_{\mathbf{A}} = 1$, according to the definition of $k$-rank). It is thus equivalent to state that, uniqueness is possible only if: i) *the product of any two dimensions of the tensor is at least equal to the number of factors* and ii) *none of the three factor matrices is allowed to have a pair of proportional columns*.

It was proven in [28] that condition (3.26) is necessary and sufficient, provided that one factor matrix is full column-rank. Assuming for instance, that $\mathbf{C}$ is full column-rank i.e., $r_{\mathbf{C}} = Q$, it is easily checked that condition (3.26), can be equivalently stated as:

$$r_{\mathbf{A}\diamond\mathbf{B}} = Q \quad \text{and} \quad min(k_{\mathbf{A}}, k_{\mathbf{B}}) \geq 2, \tag{3.27}$$

which means that PARAFAC is unique *iff* the Khatri-Rao product $\mathbf{A} \diamond \mathbf{B}$ is full column-rank and ii) neither $\mathbf{A}$ nor $\mathbf{B}$ has a pair of proportional columns. In [28], it was also proposed an equivalent necessary and sufficient condition, which is valid for general PARAFAC decompositions and is also easier to verify than (3.26) and (3.27). PARAFAC models where at least one factor matrix is full column-rank are very frequently encountered in the applications we will consider. Hence, the aforementioned conditions have its relevance to our context.

### 3.2.3 Constrained Block-PARAFAC decomposition

Let $\mathcal{X} \in \mathbb{C}^{I_1 \times I_2 \times I_3}$ be a third-order tensor and define two sets of matrices $\{\mathbf{A}^{(q)}\} \in \mathbb{C}^{I_1 \times R_1^{(q)}}$ and $\{\mathbf{B}^{(q)}\} \in \mathbb{C}^{I_2 \times R_2^{(q)}}$ with typical elements $a^{(q)}_{i_1, r_1^{(q)}} = [\mathbf{A}^{(q)}]_{i_1, r_1^{(q)}}$, $b^{(q)}_{i_2, r_2^{(q)}} = [\mathbf{B}^{(q)}]_{i_2, r_2^{(q)}}$, and a set of $Q$ third-order tensors $\{\mathcal{C}^{(q)}\} \in \mathbb{C}^{R_1^{(q)} \times R_2^{(q)} \times I_3}$ with typical element $c^{(q)}_{r_1^{(q)}, r_2^{(q)}, i_3}$. We consider the following decomposition of $\mathcal{X}$:

$$x_{i_1,i_2,i_3} = \sum_{q=1}^{Q} \sum_{r_1^{(q)}=1}^{R_1^{(q)}} \sum_{r_2^{(q)}=1}^{R_2^{(q)}} a^{(q)}_{i_1, r_1^{(q)}}\, b^{(q)}_{i_2, r_2^{(q)}}\, c^{(q)}_{r_1^{(q)}, r_2^{(q)}, i_3}. \tag{3.28}$$

The tensor $x_{i_1,i_2,i_3}$ is decomposed in the form of a sum of $Q$ trilinear blocks, every block being itself the sum of $R_1^{(q)} R_2^{(q)}$ triple products. Note that $a^{(q)}_{i_1, r_1^{(q)}}$, $i_1 = 1, \ldots, I_1$, contributes $R_2^{(q)}$ times and $b^{(q)}_{i_2, r_2^{(q)}}$, $i_2 = 1, \ldots, I_2$, contributes $R_1^{(q)}$ times and $c^{(q)}_{r_1^{(q)}, r_2^{(q)}, i_3}$ contributes $R_1^{(q)} R_2^{(q)}$ times to the composition of the full tensor. This tensor decomposition constitutes a generalization of the PARAFAC decomposition with constrained structure originally proposed in [13, 12]. This generalization was also studied in [17]. It can be viewed as a particular case of the block tensor decomposition introduced in [31].

Define a set of $Q$ matrices $\{\mathbf{C}^{(1)}, \ldots, \mathbf{C}^{(Q)}\} \in \mathbb{C}^{I_3 \times R_1^{(q)} R_2^{(q)}}$ in the following way:

$$[\mathbf{C}^{(q)}]_{i_3, (r_1^{(q)}-1)R_2^{(q)} + r_2^{(q)}} = c^{(q)}_{r_1^{(q)}, r_2^{(q)}, i_3}, \quad q = 1, \ldots, Q,$$

where the $q$-th matrix $\mathbf{C}^{(q)}$ is linked to tensor $\mathcal{C}^{(q)}$ by:

$$\mathbf{C}^{(q)} = [vec(\mathbf{C}_{..1}^{(q)T}) \cdots vec(\mathbf{C}_{..I_3}^{(q)T})]^T,$$

$\mathbf{C}_{..i_3}^{(q)} \in \mathbb{C}^{R_1^{(q)} \times R_2^{(q)}}$ being the $i_3$-th matrix slice of $\mathcal{C}^{(q)}$, $i_3 = 1, \ldots, I_3$. The matrix slices $\mathbf{X}_{..i_3} \in \mathbb{C}^{I_1 \times I_2}$ of $\mathcal{X}$ can be expressed as [13]:

$$\mathbf{X}_{..i_3} = \sum_{q=1}^{Q} (\mathbf{A}^{(q)} \otimes \mathbf{1}_{R_2^{(q)}}^T) D_{i_3}(\mathbf{C}^{(q)}) (\mathbf{1}_{R_1^{(q)}}^T \otimes \mathbf{B}^{(q)})^T, \tag{3.29}$$

where $\mathbf{1}_{R_1^{(q)}}$ and $\mathbf{1}_{R_2^{(q)}}$ are "all one" vectors of dimensions $R_1^{(q)} \times 1$ and $R_2^{(q)} \times 1$, respectively.

Using property (3.3), let us make use of the following equivalences:

$$\mathbf{A}^{(q)} \otimes \mathbf{1}_{R_2^{(q)}}^T = (\mathbf{A}^{(q)} \otimes 1)(\mathbf{I}_{R_1^{(q)}} \otimes \mathbf{1}_{R_2^{(q)}}^T) = \mathbf{A}^{(q)}(\mathbf{I}_{R_1^{(q)}} \otimes \mathbf{1}_{R_2^{(q)}}^T) = \mathbf{A}^{(q)} \boldsymbol{\Psi}^{(q)},$$

and

$$\mathbf{1}_{R_1^{(q)}}^T \otimes \mathbf{B}^{(q)} = (1 \otimes \mathbf{B}^{(q)})(\mathbf{1}_{R_1^{(q)}}^T \otimes \mathbf{I}_{R_2^{(q)}}) = \mathbf{B}^{(q)}(\mathbf{1}_{R_1^{(q)}}^T \otimes \mathbf{I}_{R_2^{(q)}}) = \mathbf{B}^{(q)} \boldsymbol{\Phi}^{(q)},$$

where

$$\boldsymbol{\Psi}^{(q)} = \mathbf{I}_{R_1^{(q)}} \otimes \mathbf{1}_{R_2^{(q)}}^T, \qquad \boldsymbol{\Phi}^{(q)} = \mathbf{1}_{R_1^{(q)}}^T \otimes \mathbf{I}_{R_2^{(q)}} \tag{3.30}$$

are *constraint matrices* that model interactions or linear combinations of factors of different modes within the $q$-th block. They have dimensions $R_1^{(q)} \times R_1^{(q)} R_2^{(q)}$ and $R_2^{(q)} \times R_1^{(q)} R_2^{(q)}$, respectively. These definitions allow us to rewrite (3.29) as:

$$\mathbf{X}_{..i_3} = \sum_{q=1}^{Q} \mathbf{A}^{(q)} \boldsymbol{\Psi}^{(q)} D_{i_3}(\mathbf{C}^{(q)}) (\mathbf{B}^{(q)} \boldsymbol{\Phi}^{(q)})^T. \tag{3.31}$$

Now, let us define the following block-matrices:

$$\begin{aligned} \mathbf{A} &= [\mathbf{A}^{(1)}, \ldots, \mathbf{A}^{(Q)}] \in \mathbb{C}^{I_1 \times R_1} \\ \mathbf{B} &= [\mathbf{B}^{(1)}, \ldots, \mathbf{B}^{(Q)}] \in \mathbb{C}^{I_2 \times R_2} \\ \mathbf{C} &= [\mathbf{C}^{(1)}, \ldots, \mathbf{C}^{(Q)}] \in \mathbb{C}^{I_3 \times R_3}, \end{aligned}$$

where we have defined

$$R_1 = \sum_{q=1}^{Q} R_1^{(q)}, \quad R_2 = \sum_{q=1}^{Q} R_2^{(q)}, \quad R_3 = \sum_{q=1}^{Q} R_1^{(q)} R_2^{(q)}. \tag{3.32}$$

Define also *block-diagonal constraint matrices* as:

$$\begin{aligned} \boldsymbol{\Psi} &= BlockDiag(\boldsymbol{\Psi}^{(1)} \cdots \boldsymbol{\Psi}^{(Q)}) \quad (R_1 \times R_3) \\ \boldsymbol{\Phi} &= BlockDiag(\boldsymbol{\Phi}^{(1)} \cdots \boldsymbol{\Phi}^{(Q)}) \quad (R_2 \times R_3). \end{aligned} \tag{3.33}$$

Taking these definitions into account, (3.31) can be expressed in a compact matrix form as:

$$\mathbf{X}_{..i_3} = \mathbf{A} \boldsymbol{\Psi} D_{i_3}(\mathbf{C}) (\mathbf{B} \boldsymbol{\Phi})^T. \tag{3.34}$$

By comparing (3.20) and (3.34), we deduce the following correspondences: $\mathbf{A} \rightarrow \mathbf{A}\boldsymbol{\Psi}$, $\mathbf{B} \rightarrow \mathbf{B}\boldsymbol{\Phi}$ and $\mathbf{C} \rightarrow \mathbf{C}$. Hence, by analogy with (3.21), $\mathbf{X}_{i=1,2,3}$ can be written as:

$$\mathbf{X}_1 = (\mathbf{C} \diamond \mathbf{A}\boldsymbol{\Psi})(\mathbf{B}\boldsymbol{\Phi})^T, \quad \mathbf{X}_2 = (\mathbf{B}\boldsymbol{\Phi} \diamond \mathbf{C})(\mathbf{A}\boldsymbol{\Psi})^T, \quad \mathbf{X}_3 = (\mathbf{A}\boldsymbol{\Psi} \diamond \mathbf{B}\boldsymbol{\Phi})\mathbf{C}^T. \tag{3.35}$$

The demonstration of (3.35) is provided in [13].

From the previous equations and definitions, the following can be said about this tensor decomposition:

1. It is the factorization of a third-order tensor in a sum of $Q$ structured/constrained PARAFAC blocks, everyone of them being a function of three *component matrices* $\mathbf{A}^{(q)}, \mathbf{B}^{(q)}$ and $\mathbf{C}^{(q)}$. Each component matrix models the variation of the tensor data along one dimension or *mode*.
2. Within the same constrained PARAFAC block, it is permitted that columns of different component matrices are linearly combined to generate the tensor data. The term *interaction* is used to denote such a linear combination.

3. The interaction patterns within a block are modeled by the *constraint matrices* $\boldsymbol{\Psi}^{(q)}$ and $\boldsymbol{\Phi}^{(q)}$, which may differ from block to block.
4. When the computation of the decomposition is performed, the term *between-block resolution* is a synonym of *separability* of the $Q$ blocks, while the term *within-block uniqueness* stands for a unique determination of the three component matrices of the corresponding block (up to permutation and scaling). It depends on the particular interaction structure of each block.

Note that (3.34) can be interpreted as a structured/constrained PARAFAC model [2, 3] with augmented component matrices $\mathbf{A}\boldsymbol{\Psi}$, $\mathbf{B}\boldsymbol{\Phi}$ and $\mathbf{C}$. It is also worth noting that the within-block structure of constrained Block-PARAFAC is similar to that of the PARALIND (PARAllel profiles with LINear Dependencies) model recently proposed in [32, 33]. The proposed approach is more general, since it considers multiple constrained blocks, each one with its own interaction structure.

The constrained Block-PARAFAC model can also be linked to Tucker2 and Tucker3 analysis [1, 34]. From the scalar notation (3.28), it is easier to conclude that this model naturally takes the form of a sort of "Block-Tucker2" model with a variable within-block interaction structure. It can also be linked in some sense to constrained Tucker3 decompositions [35, 27]. Further details can be found in [17].

**Theorem 1.** *Consider the set of matrix-slices* $\mathbf{X}_{\cdot\cdot i_3}$, $i_3 = 1, \ldots, I_3$. *Assume that the component matrices* $\mathbf{A}^{(q)} \in \mathbb{C}^{I_1 \times R_1^{(q)}}$, $\mathbf{B}^{(q)} \in \mathbb{C}^{I_2 \times R_2^{(q)}}$ *and* $\mathbf{C}^{(q)} \in \mathbb{C}^{I_3 \times R_1^{(q)} R_2^{(q)}}$ *are full-column rank,* $q = 1, \ldots, Q$. *If every set* $\{\mathbf{A}^{(1)}, \ldots, \mathbf{A}^{(Q)}\}$, $\{\mathbf{B}^{(1)}, \ldots, \mathbf{B}^{(Q)}\}$ *and* $\{\mathbf{C}^{(1)}, \ldots, \mathbf{C}^{(Q)}\}$ *is linear independent, the following condition is necessary for uniqueness of this decomposition:*

$$I_1 I_2 \geq R_3, \quad I_1 I_3 \geq R_2, \quad I_2 I_3 \geq R_1, \tag{3.36}$$

*where* $R_1, R_2, R_3$ *are defined in (3.32), between-block uniqueness/resolution is achieved, (i.e., the computation of the model separates the* $Q$ *blocks), and there are non-singular block-diagonal matrices*

$$\begin{aligned} \mathbf{T}_a &= BlockDiag(\mathbf{T}_a^{(1)} \cdots \mathbf{T}_a^{(Q)}), \\ \mathbf{T}_b &= BlockDiag(\mathbf{T}_b^{(1)} \cdots \mathbf{T}_b^{(Q)}), \\ \mathbf{T}_c &= BlockDiag(\mathbf{T}_c^{(1)} \cdots \mathbf{T}_c^{(Q)}), \end{aligned} \tag{3.37}$$

*satisfying*

$$(\mathbf{T}_a^{(q)} \otimes \mathbf{T}_b^{(q)})^{-1} = \mathbf{T}_c^{(q)T}, \quad q = 1, \ldots, Q, \tag{3.38}$$

*such that* $\overline{\mathbf{A}} = \mathbf{A}\mathbf{T}_a$, $\overline{\mathbf{B}} = \mathbf{B}\mathbf{T}_b$ *and* $\overline{\mathbf{C}} = \mathbf{C}\mathbf{T}_c$ *give rise to the same matrices* $\{\mathbf{X}\}_{i=1,2,3}$, *up to permutation ambiguities.*

The proof is provided in [17].

The conditions (3.36) can be stated in a simpler form. Define $\overline{R}_1 = max(R_1^{(1)}, \ldots, R_1^{(Q)})$ and $\overline{R}_2 = max(R_2^{(1)}, \ldots, R_2^{(Q)})$. Condition (3.36) can be rewritten as:

$$I_1 I_2 \geq Q\overline{R}_1\overline{R}_2, \quad I_1 I_3 \geq Q\overline{R}_2, \quad I_2 I_3 \geq Q\overline{R}_1,$$

which is equivalent to:

$$min\left(\lfloor \frac{I_1 I_2}{\overline{R}_1 \overline{R}_2} \rfloor, \lfloor \frac{I_1 I_3}{\overline{R}_2} \rfloor, \lfloor \frac{I_2 I_3}{\overline{R}_1} \rfloor\right) \geq Q, \tag{3.39}$$

where $\lfloor \cdot \rfloor$ stands for the greatest integer number that is smaller than its argument. This condition guarantees between-block uniqueness of constrained Block-PARAFAC.

The block-diagonal structure of $\mathbf{T}_a$, $\mathbf{T}_b$ and $\mathbf{T}_c$ means that rotational freedom is confined within the blocks. In other words, the constrained Block-PARAFAC model has *between-block uniqueness*. *Within-block non-uniqueness*, however, remains in the general case $R_1^{(q)} \geq 2$ and $R_2^{(q)} \geq 2$. Note however, that rotational freedom affecting the component matrices of a given block is constrained and obeys $\mathbf{T}_c^{(q)T} = (\mathbf{T}_a^{(q)} \otimes \mathbf{T}_b^{(q)})^{-1}$, which imposes some uniqueness constraints on the constrained Block-PARAFAC model that are not as strong as those of unconstrained Tucker3 models. For example, complete within-block uniqueness is restored if $\mathbf{C}$ is known. Otherwise, if $\mathbf{C}$ has some special structure (e.g. Block-diagonal, Toeplitz, etc) one can enforce the underlying structure during the computation of the model, in order to restrict within-block non-uniqueness.

## 3.3 Applications of Tensor Decompositions to Wireless Communications

In several wireless communication systems, the received signal is *multidimensional* in nature and can be interpreted as a tensor, although it is not always treated as such. A common approach for modeling and processing the received signal is based on matrix (2D) models. The usually considered dimensions are *space* and *time* [36]. The space dimension is generally due to the use of an antenna array at the receiver, and contains multiple (phase-shifted or uncorrelated) copies of the received signal. On the other hand, the time dimension is due to the fact that the received signal is processed as a block containing consecutive symbol-spaced samples. Regarding the blind recovery of information, blind algorithms generally take special (problem-specific) structural properties of the transmitted signals into account such as orthogonality, finite-alphabet, constant-modulus or cyclostationarity in order to overcome the non-uniqueness of matrix decompositions and successfully perform multiuser signal separation and equalization [36, 37, 38].

From a signal processing perspective, treating the received signal as a 3D tensor makes possible to simultaneously exploit the multiple forms of "diversity" inherent to it for a signal recovery purpose. In a seminal paper [7], it was shown that a mixture of DS/CDMA signals has trilinear structure and can be modeled as a third-order (3D) tensor. By formulating the received signal model as a PARAFAC model, the authors derived a new receiver for blind multiuser detection/separation in DS/CDMA systems, which exploits the identifiability properties of the PARAFAC decomposition for separating/detecting possibly more users than antennae. The work of [7] was constrained to flat-fading wireless channels with no intersymbol interference. In [8], the PARAFAC modeling approach of [7] was extended to frequency-selective channels with large delay spread, still in the context of DS/CDMA systems. A two-step receiver, different from that of [7], was proposed for multiuser separation and equalization. A different tensor modeling for the blind separation of DS/CDMA signals in the frequency-selective case was proposed in [10], where a generalized PARAFAC model was developed. This model was also shown to be valid for signal separation in Multiple-Input Multiple-Output (MIMO) DS/CDMA systems [9].

The usefulness of tensor decompositions in signal processing is not limited to the separation of DS/CDMA signals. PARAFAC analysis have also been proposed for other signal processing problems such as multiple-invariance sensor array processing [14], multidimensional harmonic retrieval [39] blind beamforming with oversampling at the receiver [15], multiantenna space-time coding [21] and spatial signature estimation [18].

This chapter provides a survey of some applications of tensor modeling to wireless signal processing. The considered applications are: i) unified modeling of oversampled, DS/CDMA and OFDM systems, ii) multiantenna coding techniques, and iii) estimation of wireless channel parameters.

*Previous works and new contributions*

In several works [7, 8, 10, 9], the problem of blind multiuser detection for the DS/CDMA system has been modeled using tensor decompositions. The main difference between [7] and [8] is that the first does not take a frequency-selective channel model into account. Both [8] and [10] adopt a frequency-selective channel model with inter-symbol interference. For the same system model, [8] proposes a sort of PARAFAC model with linear dependencies, while [10] proposes a generalized PARAFAC model and introduces new identifiability conditions for this model, which can be seen as a generalization of those of [7]. Both works assume that each user contributes with a single path to the received signal, which is the case when the multipath scattering is in the far field of the receive antenna array. Concerning the DS/CDMA system, our contribution is to further generalize [8] and [10] by considering that each user contributes with a finite number of multipaths. This generalization is based on a new block-tensor decomposition with constrained structure called constrained Block-PARAFAC [12, 17]. This tensor decomposition is also used for modeling oversampled and OFDM wireless communication systems. A unified perspective is adopted here for modeling these systems using a tensor approach.

Our contributions also concern the problem of multiantenna coding for MIMO systems. [21] proposed a blind Khatri-Rao space-time code, which relies on a PARAFAC model for the received signal. In this context, we propose a more general approach, which exhibits increased diversity gain, has more transmission flexibility, and is suited to the downlink multiuser MIMO systems. The proposed multiantenna code combines multistream multiplexing and space-time spreading and is also an application of the constrained Block-PARAFAC decomposition. This application of tensor modeling appears in [22, 23].

The later problem addressed using tensor decompositions is concerned with the estimation of multipath parameters of wireless communication channels. This problem is not new, and was addressed in several works such as [40, 41, 42]. With respect to these works, our contribution is to show that this problem can be formulated using a PARAFAC tensor model, which results in a new method for multipath channel estimation. We also consider the problem of estimating block-fading MIMO channels using a parametric tensor modeling.

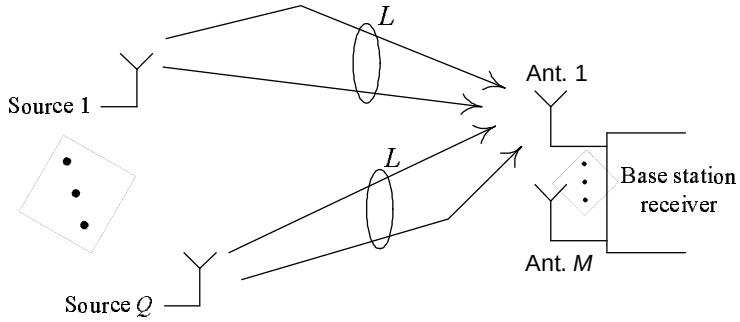



**Fig. 3.5.** Multiuser/multipath propagation scenario.

### 3.3.1 Unified modeling of oversampled, DS/CDMA and OFDM systems

A new tensor model for the received signal is presented, which unifies the received signal model of three systems: i) a temporally-oversampled system, ii) a DS/CDMA system [43] and iii) an OFDM system [44]. For all these systems an antenna array is assumed at the receiver front-end. The proposed tensor model assumes specular multipath propagation, where each user in the system contributes with a finite number of multipaths to the received signal. We show that the proposed model is subject to structural constraints in some of its component matrices, and that the same "general" tensor model is shared by the three considered systems. For each particular system, the model can be obtained from the general model by making appropriate choices in the structure/dimension of its matrix components. For the DS/CDMA system, our tensor modeling approach generalizes those of [8] and [10], which consider a special propagation model with a single path per user. It also covers the tensor model proposed in [15] as a special case, which assumes multiple paths per user but does not consider a frequency-selective channel model.

Let us consider a linear and uniformly-spaced array of $M$ antennae receiving signals from $Q$ co-channel sources/users. Assume that the signal transmitted by each user is subject to multipath propagation and arrives at the receiver via $L$ effective specular multipaths [3]. The propagation channel is assumed to be time-dispersive and it is considered that multipath delay spread exceeds the inverse of the coherence bandwidth of the system, so that fading is frequency-selective. The channel impulse response is assumed to span $K$ symbols. The discrete-time base-band representation of the signal received at the $m$-th antenna of a linear and uniformly-spaced array at the $n$-th symbol interval is given by:

$$x_m(n) = \sum_{q=1}^{Q} \sum_{l=1}^{L} \beta_l^{(q)} a_m(\theta_{lq}) \sum_{k=1}^{K} g(k-1-\tau_{lq}) s^{(q)}(n-k+1) + v_m(n),$$

where $\beta_l^{(q)}$ is the fading envelope of the $l$-th path of the $q$-th user. The term $a_m(\theta_{lq})$ is the response of the $m$-th antenna to the $l$-th path of the $q$-th user, $\theta_{lq}$ being the associated angle of arrival. Similarly, the term $\tau_{lq}$ is the propagation delay (normalized by the symbol period $T$) and the term $g(k-1-\tau_{lq})$ represents the $k$-th component of the pulse-shaping filter response. The channel impulse response has finite duration and is assumed to be zero outside the interval $[0,(K-1)T)$. Finally, $s^{(q)}(n)$ is the symbol transmitted by the $q$-th user at the $n$-th symbol interval and $v_m(n)$ is the measurement noise at the $m$-th antenna, at the $n$-th symbol interval.

#### Temporally-Oversampled System

At the output of each receiver antenna, the signal is sampled at a rate that is $P$ times the symbol rate. Due to temporal oversampling, the resolution of the pulse-shaping filter response is increased by a factor $P$, which also increases the temporal resolution of the received signal by the same factor. Such an increase in the temporal resolution can be interpreted as an incorporation of a third axis (or dimension) to the received signal, called the *oversampling dimension*. We define the $p$-th oversample associated with the $k$-th component of the pulse-shaping filter response to the $(l,q)$-th path as the following third-order tensor:

$$g_{l,k,p}^{(q)} = g(k-1+(p-1)/P-\tau_{lq}), \quad p=1,\ldots,P.$$

[3] In this model we have assumed that all the users have the same number of multipaths in order to simplify the mathematical notation and the presentation of the model.

The overall channel response, the array response and transmitted symbols are defined as:

$$h_{l,k,p}^{(q)} = \beta_l^{(q)}\, g_{l,k,p}^{(q)}, \quad a_{m,l}^{(q)} = a_m(\theta_{lq}), \quad s_{n,k}^{(q)} = s^{(q)}(n-k+1).$$

The received signal can be interpreted as a 3D tensor $\mathcal{X} \in C^{M\times N\times P}$. Its $(m,n,p)$-th scalar component can be decomposed as:

$$x_{m,n,p} = \sum_{q=1}^{Q}\sum_{l=1}^{L} a_{m,l}^{(q)} \sum_{k=1}^{K} h_{l,k,p}^{(q)}\, s_{n,k}^{(q)} + v_{m,n,p}. \tag{3.40}$$

Equation (3.40) is the *scalar notation* for the received signal tensor. It expresses the received signal in the form of summations and products involving three factors $a_{m,l}^{(q)}$, $h_{l,k,p}^{(q)}$ and $s_{n,k}^{(q)}$ associated with space, oversampling and time dimensions of the received signal, respectively. (3.40) can be alternatively written in matrix-slice form. Let $\mathbf{a}_l^{(q)} = [a_{1,l}^{(q)} a_{2,l}^{(q)} \cdots a_{M,l}^{(q)}]^T \in \mathbb{C}^M$ be the array steering vector, $\mathbf{g}_{l,k}^{(q)} = [g_{l,k,1}^{(q)} g_{l,k,2}^{(q)} \cdots g_{l,k,P}^{(q)}]^T \in \mathbb{C}^P$ be the pulse-shape response vector, and $\mathbf{s}_n^{(q)} = [s_{n,1}^{(q)} s_{n,2}^{(q)} \cdots s_{n,K}^{(q)}]^T \in \mathbb{C}^K$ be the symbol vector and $\mathbf{b}^{(q)} = [b_1^{(q)} \cdots b_L^{(q)}]^T \in \mathbb{C}^L$ be the vector of multipath gains. Let us also define:

$$\mathbf{A}^{(q)} = [\mathbf{a}_1^{(q)} \cdots \mathbf{a}_L^{(q)}] \in \mathbb{C}^{M\times L}, \quad \mathbf{G}^{(q)} = [\mathbf{g}_{1,1}^{(q)} \cdots \mathbf{g}_{l,k}^{(q)} \cdots \mathbf{g}_{L,K}^{(q)}]^T \in \mathbb{R}^{P\times LK}, \quad \mathbf{S}^{(q)} = [\mathbf{s}_1^{(q)} \cdots \mathbf{s}_N^{(q)}]^T \in \mathbb{C}^{N\times K}.$$

$\mathbf{S}^{(q)} \in \mathbb{C}^{N\times K}$ is a Toeplitz matrix with first row and column defined as $\mathbf{S}_{1\cdot}^{(q)} = [s_{1,1}^{(q)}\, 0 \cdots 0]$ and $\mathbf{S}_{\cdot 1}^{(q)} = [s_{1,1}^{(q)}\, s_{2,1}^{(q)} \cdots s_{N,1}^{(q)}]^T$ respectively. The overall temporal channel response $\mathbf{H}^{(q)}$ can be expressed as:

$$\mathbf{H}^{(q)} = \mathbf{G}^{(q)}(Diag(\mathbf{b}^{(q)}) \otimes \mathbf{I}_K) \quad \in \mathbb{C}^{P\times LK}. \tag{3.41}$$

By comparing (3.40) with (3.28), we can deduce the following correspondences:

$$(I_1, I_2, I_3, R_1^{(q)}, R_2^{(q)}, \mathbf{A}^{(q)}, \mathbf{B}^{(q)}, \mathbf{C}^{(q)}) \rightarrow (M, N, P, L, K, \mathbf{A}^{(q)}, \mathbf{S}^{(q)}, \mathbf{H}^{(q)}) \tag{3.42}$$

Fixing the oversampling dimension, and using the analogy with (3.31) and the correspondences (3.42), the space-time slice $\mathbf{X}_{\cdot\cdot p} \in \mathbb{C}^{M\times N}$, $p = 1, \ldots, P$, can be factored as:

$$\mathbf{X}_{\cdot\cdot p} = \sum_{q=1}^{Q} \mathbf{A}^{(q)} \boldsymbol{\Psi} D_p(\mathbf{H}^{(q)}) (\mathbf{S}^{(q)} \boldsymbol{\Phi})^T + \mathbf{V}_{\cdot\cdot p}, \quad p = 1, \ldots, P, \tag{3.43}$$

where $\boldsymbol{\Psi} = \mathbf{I}_L \otimes \mathbf{1}_K^T$ and $\boldsymbol{\Phi} = \mathbf{1}_L^T \otimes \mathbf{I}_K$ are the constraint matrices of the tensor model.

**DS/CDMA System**

At the transmitter, each symbol is spread by a signature (spreading code) sequence of length $J$ with period $T_c = T/J$, where $T$ is the symbol period. The spreading sequence associated with the $q$-th user is denoted by $\mathbf{c}^{(q)} = [c_1^{(q)} c_2^{(q)} \cdots c_J^{(q)}]^T \in \mathbb{C}^J$. As a result of spreading operation, each symbol to be transmitted is converted into *J chips*. The chip-sampled pulse-shape response to the $(l,q)$-th path is defined in scalar form as:

$$g_{l,k,j}^{(q)} = g(k - 1 + (j-1)/J - \tau_{lq}), \quad j = 1, \ldots, J.$$

A scalar component of the overall channel response, i.e. including path gains, is defined as $h_{l,k,j}^{(q)} = \beta_l^{(q)} g_{l,k,j}^{(q)}$. The received signal can be interpreted as a 3D tensor $\mathcal{X} \in C^{M\times N\times J}$ and its $(m,n,j)$-th scalar component has the following factorization:

$$x_{m,n,j} = \sum_{q=1}^{Q}\sum_{l=1}^{L} a_{m,l}^{(q)} \sum_{k=1}^{K}\sum_{j'=1}^{J} h_{l,k,j-j'}^{(q)}\, c_{j'}^{(q)}\, s_{n,k}^{(q)} + v_{m,n,j}. \tag{3.44}$$

Equation (3.44) decomposes the received signal in the form of summations and products involving four scalar factors $a_{m,l}^{(q)}$, $h_{j,l,k}^{(q)}$, $c_j^{(q)}$ and $s_{n,k}^{(q)}$. Defining

$$u_{l,k,j}^{(q)} = \sum_{j'=1}^{J} h_{l,k,j-j'}^{(q)}\, c_{j'}^{(q)}$$

as the convolution between the spreading sequence and the overall channel response, we can rewrite (3.44) as:

$$x_{m,n,j} = \sum_{q=1}^{Q} \sum_{l=1}^{L} a_{m,l}^{(q)} \sum_{k=1}^{K} u_{l,k,j}^{(q)} \, s_{n,k}^{(q)} + v_{m,n,j}. \tag{3.45}$$

Note that (3.45) is equivalent to (3.40), where $u_{l,k,j}^{(q)}$ plays the role of $h_{l,k,p}^{(q)}$ and $J \rightarrow P$. Therefore, by fixing the spreading dimension and slicing the received signal tensor $\mathcal{X} \in \mathbb{C}^{M \times N \times J}$ along the third-dimension, its $j$-th slice $\mathbf{X}_{\cdot\cdot j} \in \mathbb{C}^{M \times N}$, $j = 1, \ldots, J$, is given by:

$$\mathbf{X}_{\cdot\cdot j} = \sum_{q=1}^{Q} \mathbf{A}^{(q)} \boldsymbol{\Psi} D_j(\mathbf{U}^{(q)}) (\mathbf{S}^{(q)} \boldsymbol{\Phi})^T + \mathbf{V}_{\cdot\cdot j}, \quad j = 1, \ldots, J, \tag{3.46}$$

where $\mathbf{U}^{(q)}$ is defined as:

$$\mathbf{U}^{(q)} = \mathbf{C}^{(q)} \mathbf{H}^{(q)} \in \mathbb{C}^{J \times KL}, \tag{3.47}$$

$\mathbf{C}^{(q)} \in \mathbb{C}^{J \times J}$ being a Toeplitz matrix with first row and column defined as $\mathbf{C}_{1\cdot}^{(q)} = [c_1^{(q)} \, 0 \cdots 0]$ and $\mathbf{C}_{\cdot 1}^{(q)} = [c_1^{(q)} \, c_2^{(q)} \cdots c_J^{(q)}]^T$ respectively, and $\mathbf{H}^{(q)}$ being analogous to that defined in (3.41), where $J$ replaces $P$.

**OFDM System**

In an OFDM system, the symbol sequence to be transmitted is organized into blocks of $F$ symbols (serial-to-parallel conversion). Multicarrier modulation consists in linearly combining the $F$ symbols using an inverse $N \times N$ Fast Fourier Transform (IFFT) matrix. After the IFFT stage, a cyclic prefix (CP) of length $K$ is inserted to each OFDM block [4], before transmission. Due to the insertion of CP, the resulting OFDM block has length $F + K$. At the receiver, inverse processing is done. The CP is removed and each received OFDM block is linearly combined using an $N \times N$ Fast Fourier Transform (FFT) matrix. Thanks to the use of IFFT/FFT together with insertion/removal of the CP, it can be shown [44] that the length-$K$ convolutive channel is converted into a set of $F$ scalar channels, i.e. the overall channel at each subcarrier is frequency-flat. This means that the overall channel matrix has a diagonal frequency response.

Let us define $h_{l,f}^{(q)} = \beta_l^{(q)} g_{l,f}^{(q)}$ as a scalar component of the overall frequency channel response at the $f$-th subcarrier, associated with the $(l,q)$-th path. The scalar $s_{n,f}^{(q)}$ denotes the $f$-th symbol of the $n$-th OFDM block associated with the $q$-th user. The received signal is also a 3D tensor and its $(m,n,f)$-th scalar component can be decomposed as:

$$x_{m,n,f} = \sum_{q=1}^{Q} \sum_{l=1}^{L} a_{m,l}^{(q)} \, h_{l,f}^{(q)} \, s_{n,f}^{(q)} + v_{m,n,f}. \tag{3.48}$$

As it has been done for the previous systems, we are interested in representing the received signal tensor in matrix-slice notation. Let us define the vector $\mathbf{s}_n^{(q)} = [s_{n,1}^{(q)} \, s_{n,2}^{(q)} \cdots s_{n,F}^{(q)}]^T \in \mathbb{C}^F$ collecting the symbols associated with the $n$-th OFDM block, $n = 1, \ldots, N$. We define $\breve{\mathbf{G}}_l^{(q)} \in \mathbb{R}^{F \times F}$ as a *circulant channel matrix* built up from the $(l,q)$-th pulse-shape response $\mathbf{g}_l^{(q)} = [g_{l,1}^{(q)} g_{l,2}^{(q)} \cdots g_{l,K}^{(q)}]^T \in \mathbb{R}^K$. Let

$$\Omega_{cp} = \begin{bmatrix} \mathbf{0}_{K \times (F-K)} \; \mathbf{I}_K \\ \mathbf{I}_F \end{bmatrix}, \quad \overline{\Omega}_{cp} = [\mathbf{0}_{F \times K} \; \mathbf{I}_F],$$

be the matrices that represent the insertion and removal of the CP, of dimensions $(F+K) \times F$ and $F \times (F+K)$ respectively, and define a $(F+K) \times (F+K)$ Toeplitz matrix $\tilde{\mathbf{G}}_l^{(q)} = Toeplitz[\mathbf{g}_l^{(q)}]$. The $(l,q)$-th circulant matrix $\breve{\mathbf{G}}_l^{(q)}$ is given by:

$$\breve{\mathbf{G}}_l^{(q)} = \overline{\Omega}_{cp} \tilde{\mathbf{G}}_l^{(q)} \Omega_{cp}.$$

Note that the circulant structure of $\breve{\mathbf{G}}_l^{(q)} \in \mathbb{C}^{F \times F}$ is an equivalent representation of the convolution matrix $\tilde{\mathbf{G}}_l^{(q)}$, when it is pre- and post-multiplied by the matrices that represent the removal and insertion of the CP, respectively. For further details on this construction, please see [44].

[4] The CP is used to avoid interference between adjacent OFDM blocks. Its length should exceed the maximum delay spread of the channel. See [44] for further details.

Define $\breve{\mathbf{G}}^{(q)} = [\breve{\mathbf{G}}_1^{(q)} \cdots \breve{\mathbf{G}}_L^{(q)}] \in \mathbb{R}^{F \times LF}$ as a *block-circulant* matrix concatenating the $L$ circulant channel matrices associated with the $q$-th user. By analogy with (3.41), the *overall circulant channel matrix* $\breve{\mathbf{H}}^{(q)}$ can be written as:

$$\breve{\mathbf{H}}^{(q)} = \breve{\mathbf{G}}^{(q)} \left( Diag(\mathbf{b}^{(q)}) \otimes \mathbf{I}_F \right) \quad \in \mathbb{C}^{F \times LF}.$$

Taking IFFT/FFT transformations at the transmitter/receiver into account, the $q$-th user *overall frequency response channel matrix* $\boldsymbol{\Lambda}^{(q)} = [\boldsymbol{\Lambda}_1^{(q)} \cdots \boldsymbol{\Lambda}_L^{(q)}] \in \mathbb{C}^{F \times LF}$, can be expressed as:

$$\boldsymbol{\Lambda}^{(q)} = \boldsymbol{\Gamma} \breve{\mathbf{H}}^{(q)} (\mathbf{I}_L \otimes \boldsymbol{\Gamma})^H, \tag{3.49}$$

where $\boldsymbol{\Gamma} \in \mathbb{C}^{F \times F}$ is a FFT matrix the $(i,k)$-th entry of which is given by $\gamma_{i,k} = (1/\sqrt{F}) \cdot exp(-j2\pi ik/N)$, $i,k = 0,1,\ldots,N-1$. Note that $\boldsymbol{\Lambda}_l^{(q)}$ is a diagonal matrix holding the vector $\mathbf{h}_l^{(q)} = [h_{1,l}^{(q)}\, h_{2,l}^{(q)} \cdots h_{F,l}^{(q)}]^T$, on its diagonal, $l = 1,\ldots,L$, i.e., $\boldsymbol{\Lambda}_l^{(q)} = Diag(\mathbf{h}_l^{(q)})$.

Now, collect $N$ symbol blocks in $\mathbf{S}^{(q)} = [\mathbf{s}_1^{(q)} \cdots \mathbf{s}_N^{(q)}] \in \mathbb{C}^{F \times N}$. Taking these definitions into account, the $f$-th slice $\mathbf{X}_{..f} \in \mathbb{C}^{M \times N}$, $f = 1,\ldots,F$, of the received signal tensor $\mathcal{X} \in \mathbb{C}^{M \times N \times F}$ can be decomposed as:

$$\mathbf{X}_{..f} = \sum_{q=1}^{Q} \mathbf{A}^{(q)} \boldsymbol{\Psi} D_f(\boldsymbol{\Lambda}^{(q)}) (\mathbf{S}^{(q)} \boldsymbol{\Phi})^T + \mathbf{V}_{..f}, \quad f = 1,\ldots,F. \tag{3.50}$$

It is worth noting that (3.43), (3.46) and (3.50) are similar PARAFAC-based decompositions, and all of them follow (3.31). The main differences are in the structure and/or the dimension of certain component matrices, but the general tensor formulation is identical. Such similarity allows us to formalize a "general" or *unified* tensor model for the three systems. In order to achieve this, we define $I_3$ as the length of the third dimension of the received signal tensor. For the temporally-oversampled system $(I_3, R_2) = (P, K)$, for the DS/CDMA system $(I_3, R_2) = (J, K)$ and for the OFDM system $(I_3, R_2) = (F, F)$. The other parameters, which are common for all the systems are $I_1 = M$, $I_2 = N$, $R_1 = L$. In its general form, the tensor modeling of the three systems can be unified in the following expression:

$$\mathbf{X}_{..i_3} = \sum_{q=1}^{Q} \mathbf{A}^{(q)} \boldsymbol{\Psi} D_{i_3}(\mathbf{Z}^{(q)}) (\mathbf{S}^{(q)} \boldsymbol{\Phi})^T + \mathbf{V}_{..i_3}, \quad i_3 = 1,\cdots,I_3,$$

where

$$\boldsymbol{\Psi} = \mathbf{I}_L \otimes \mathbf{1}_{R_2}^T \in \mathbb{C}^{L \times LR_2}, \quad \boldsymbol{\Phi} = \mathbf{1}_L^T \otimes \mathbf{I}_{R_2} \in \mathbb{C}^{LR_2 \times R_2},$$

and $\mathbf{Z}^{(q)}$ is either $\mathbf{H}^{(q)}$ (c.f. (3.41)), or $\mathbf{U}^{(q)}$ (c.f. (3.47)) or $\boldsymbol{\Lambda}^{(q)}$ (c.f. (3.49)), depending on the considered system. Table 3.1 summarizes the unified modeling.

| | **Oversampled** | **DS/CDMA** | **OFDM** |
|---|---|---|---|
| $(I_3)$ | oversampling($P$) | spreading($J$) | frequency($F$) |
| $(R_2)$ | delay($K$) | delay($K$) | frequency($F$) |
| $\mathbf{Z}^{(q)}$ | $\mathbf{H}^{(q)}$ | $\mathbf{U}^{(q)}$ | $\boldsymbol{\Lambda}^{(q)}$ |
| dimensions | $P \times LK$ | $J \times LK$ | $F \times LF$ |
| structure | no structure | no structure | diagonal blocks |
| $\mathbf{S}^{(q)}$ | $\mathbf{S}^{(q)}$ | $\mathbf{S}^{(q)}$ | $\mathbf{S}^{(q)}$ |
| dimensions | $N \times K$ | $N \times K$ | $N \times F$ |
| structure | Toeplitz | Toeplitz | no structure |

**Table 3.1.** Summary of the unified tensor model for Oversampled, DS/CDMA an OFDM systems.

**Application to blind multiuser separation/equalization**

The blind multiuser equalization problem consists in recovering the information sequence from several users under the assumption of frequency-selective fading. In other words, a blind multiuser equalizer performs the tasks of signal separation and equalization. In this paper, we propose a tensor-based receiver performing user separation and equalization iteratively. For the temporally-oversampled system, multiuser signal separation is carried out in the 3D tensor space, exploiting *oversampling*, *time* and *space* dimensions of the received signal. Making use of the necessary uniqueness

conditions of the constrained Block-PARAFAC model, we can derive an upper bound on the maximum number of co-channel users that can be blindly separated. By analogy with (3.39), we have:

$$min\left(\lfloor\frac{MN}{LK}\rfloor,\lfloor\frac{MP}{K}\rfloor,\lfloor\frac{NP}{L}\rfloor\right)\geq Q. \tag{3.51}$$

This condition allows us to study the receiver configuration such as the required number of receive antennae $M$, number of processed symbols $N$ and oversampling factor $P$, for a given number of multipaths $L$, a given channel impulse response length $K$ and a given number $Q$ of users.

The alternating least squares (ALS) algorithm [7] is used for this purpose. Equalization is done in the 2D matrix space, where the Toeplitz structure of the symbol matrix as well as the Finite-Alphabet (FA) property of the transmitted symbols are exploited to estimate the transmitted symbols via a subspace method. The key aspect of the proposed algorithm is that multiuser signal separation (PARAFAC stage) and equalization (Subspace+FA stage) is done in an iterative way. The goal of the PARAFAC stage is to estimate three component matrices from which the PARAFAC model parameters can be determined. In turn, the goal of the subspace+FA stage is to solve the partial rotation ambiguity problem that is inherent to the proposed model as well as to estimate the transmitted symbols in the 2D space, exploiting the FA property. The FA-projected symbols are in turn used as an input to the PARAFAC stage to refine the multiuser signal separation in the 3D space. Further details of this algorithm can be found in [13].

The performance of the tensor-based blind multiuser receiver is evaluated through computer simulations. Each obtained result is an average over $R = 1000$ independent Monte Carlo runs. For each run, multipath fading gains are generated from an i.i.d. Rayleigh generator while the user symbols are generated from an i.i.d. distribution and are modulated using binary-phase shift keying (BPSK). Perfect synchronization is assumed at the receiver. In all cases, a block of $N = 100$ received samples is used in the blind estimation process. The channel impulse response follows a raised cosine pulse shape with roll-off factor 0.35. We consider $K = 2$ channel taps, $Q = 2$ users and a fixed oversampling factor $P = 8$.

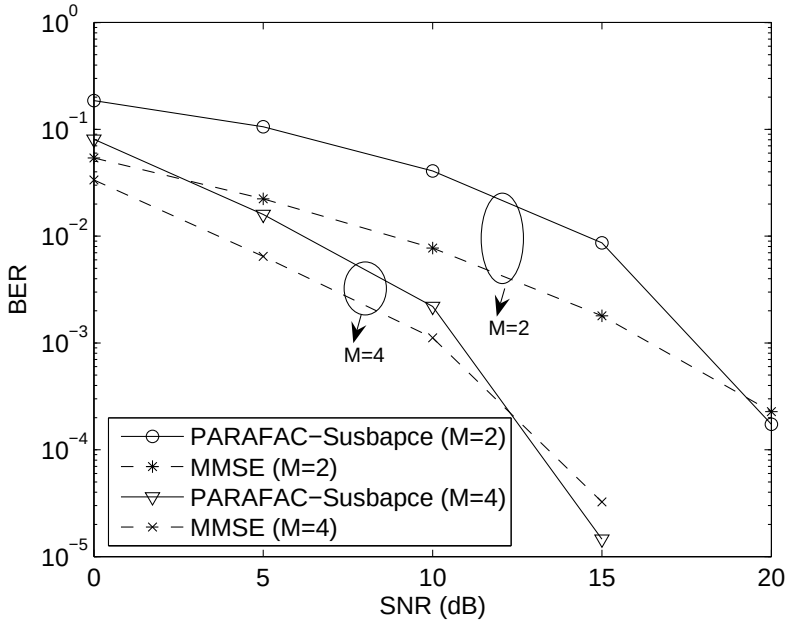



**Fig. 3.6.** Blind PARAFAC-Subspace receiver versus MMSE receiver with perfect channel knowledge.

The following simulation results consider $L = 2$ effective multipaths per user, each path being associated with a different cluster of scatterers. We consider a propagation scenario with $(\theta_{11},\theta_{21}) = (0^\circ, 20^\circ)$ for the first user and $(\theta_{12},\theta_{22}) = (0^\circ, 10^\circ)$ for the second one. Note that the first (zero-delayed) path of both users have the same angle of arrival. All multipaths have the same average power $E[\beta_{lq}\beta_{lq}^*] = 1$, $l = 1,\ldots,L$, $q = 1,\ldots,Q$.

In order to provide a performance reference for the proposed PARAFAC-Subspace receiver, we also evaluate the performance of the Minimum Mean Square Error (MMSE) receiver. In contrast to our blind iterative PARAFAC-Subspace receiver, the MMSE one assumes perfect knowledge of all the channel parameters as well as the knowledge

of the SNR. We consider $M = 2$ and $4$ receive antennae. Figure 3.6 shows that the PARAFAC-Subspace receiver has nearly the same BER vs. SNR improvement than that of MMSE with perfect channel knowledge. The gap between both receivers is smaller for $M = 4$. For example, considering a target BER of $10^{-2}$, the proposed receiver provides a loss in performance of 5dB for $M = 2$ and 3dB for $M = 4$, with respect to the MMSE receiver.

### 3.3.2 Multiantenna coding techniques

Multi-antenna coding is an effective way for combating multipath fading (i.e. random attenuations affecting the transmitted radio signals) in scattering-rich wireless propagation environments, aiming at improving the transmission reliability in wireless communication systems. This is achieved by means of space-time coding [45]. Multi-antenna techniques have also the objective of offering high spectral efficiencies, by transmitting multiple independent signals from different antennae at the same time slot and frequency band. In this case, the transmitted signals interfere each other and should be separated at the receiver [46, 47].

We propose a new multiantenna coding concept that relies on a tensor modeling of the transmitted/received signals. Let us consider a multiantenna system with $M'_T$ transmit antennae and $M_R$ receive antennae. Assume that the wireless channel is frequency-flat and constant during $N$ time-slots. Each time-slot comprises $P$ symbol periods, where $P$ is the length of the multiantenna code. In the considered multiantenna system, an antenna array of $M'_T$ transmit antennae is divided into $Q$ transmission layers of $M_T^{(q)}$ antennae each, i.e., $M_T^{(1)} + \cdots + M_T^{(Q)} = M'_T$. Each layer transmits its own symbol sequence. The $Q$ input signals can be either assigned to a single receiver/user (point-to-point transmission) or they can be assigned to $Q$ different receivers/users (point-to-multipoint transmission). The proposed modeling approach is valid for both cases, although we do not distinguish between them here.

For each layer, the input stream is demultiplexed into $R^{(q)}$ sub-sequences, which are spread over a $M_T^{(q)} \times N \times P$ grid associated with $M_T^{(q)}$ transmit antennae, $N$ time-slots and $P$ symbol periods, and then linearly combined. This is called Tensor-based Space-Time Multiplexing (TSTM). We model the TSTM structure at the $q$-th layer, as a third-order tensor $\mathcal{W}^{(q)} \in \mathbb{C}^{M_T^{(q)} \times R^{(q)} \times P}$, called *multiplexing tensor*. Note that this multiantenna transmission is similar to the multiantenna code of [48], in the sense that each symbol is fully spread over space and time dimensions. Also, from a tensor modeling point of view, it generalizes the one proposed in [21]. A general block-diagram illustration is given in Figure 3.7. The goal of the "antenna mapping" block is to associate the $M_T^{(1)}, M_T^{(2)}, \ldots M_T^{(Q)}$ signals to the $M'_T$ transmit antennae in an optimal way, in order to maximize the diversity gain. This is only relevant in scenarios where the propagation channel exhibits some spatial correlation, and when the correlation properties are known at the transmitter. The design of this block is beyond the scope of this work. Thus, for simplicity reasons, the multiantenna channel is assumed to be spatially uncorrelated, which allows us to bypass the antenna mapping block here.

The discrete-time baseband equivalent model for the received signal can be modeled using the tensor formalism as:

$$x_{m_R,n,p} = \sum_{q=1}^{Q} \sum_{m_T^{(q)}=1}^{M_T^{(q)}} \sum_{r^{(q)}=1}^{R^{(q)}} h^{(q)}_{m_R,m_T^{(q)}} s^{(q)}_{n,r^{(q)}} w^{(q)}_{m_T^{(q)},r^{(q)},p} + v_{m_R,n,p}, \tag{3.52}$$

where $x_{m_R,n,p} = [\mathcal{X}]_{m_R,n,p}$ is the received signal sample at the $m_R$-th receive antenna, and associated with the $p$-th symbol of the $n$-th time-slot, and

$$h^{(q)}_{m_R,m_T^{(q)}}, \quad s^{(q)}_{n,r^{(q)}}, \quad w^{(q)}_{m_T^{(q)},r^{(q)},p} \tag{3.53}$$

are typical elements of the channel matrix, symbol matrix and multiplexing tensor, respectively. Figure 3.8 depicts the signal transmission/reception model (in absence of noise), by focusing on the contribution of the $q$-th transmission layer. The following definitions are used in this figure:

$$c^{(q)}_{m_T^{(q)},n,p} = \sum_{r^{(q)}=1}^{R^{(q)}} w^{(q)}_{m_T^{(q)},r^{(q)},p} s^{(q)}_{n,r^{(q)}}, \quad x^{(q)}_{m_R,n,p} = \sum_{m_T^{(q)}=1}^{M_T^{(q)}} h^{(q)}_{m_R,m_T^{(q)}} c^{(q)}_{m_T^{(q)},n,p}. \tag{3.54}$$

Let

$$\mathbf{s}_n^{(q)} = [s^{(q)}_{1,n} \cdots s^{(q)}_{M_T^{(q)},n}]^T \in \mathbb{C}^{M_T^{(q)}}, \tag{3.55}$$

be the symbol vector concatenating $M_T^{(q)}$ data sub-streams, and

$$\mathbf{S}^{(q)} = [\mathbf{s}_1^{(q)} \cdots \mathbf{s}_N^{(q)}]^T \in \mathbb{C}^{N \times M_T^{(q)}}. \tag{3.56}$$

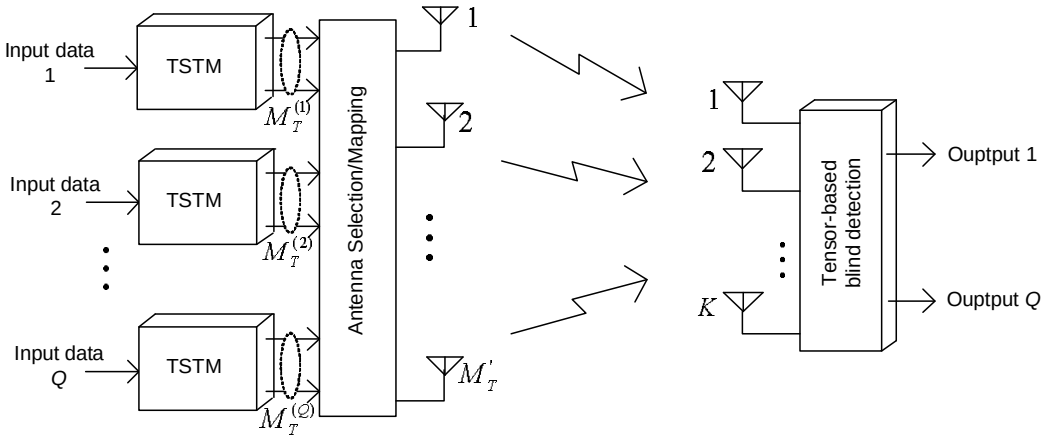



**Fig. 3.7.** Tensor-based multiantenna coding concept

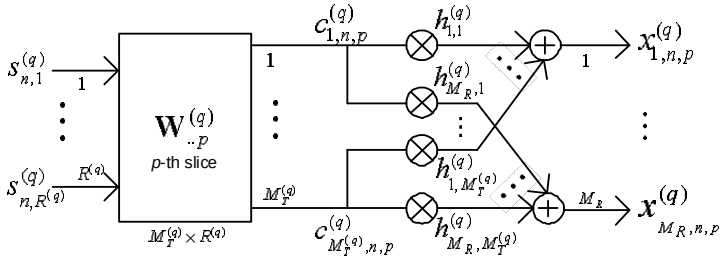



**Fig. 3.8.** Signal transmission/reception model ($q$-th layer).

The entries of $\mathbf{S}^{(q)}$ are chosen from an arbitrary signal constellation and satisfy the power constraint $E[Tr(\mathbf{S}^{(q)H}\mathbf{S}^{(q)})] = NR^{(q)}$, $q = 1, \ldots, Q$. The channel matrix $\mathbf{H} \in \mathbb{C}^{M_R \times M_T'}$ is assumed to have i.i.d. entries following a zero-mean unit-variance complex-Gaussian distribution, so that $E[Tr(\mathbf{H}\mathbf{H}^H)] = M_T' M_R$, where $Tr(\cdot)$ is the *trace* operator.

In order to write the multiantenna coding model as a constrained Block-PARAFAC model, we define a set of $Q$ multiplexing matrices $\{\mathbf{W}^{(q)}\} \in \mathbb{C}^{P \times M_T^{(q)} R^{(q)}}$ as

$$[\mathbf{W}^{(q)}]_{p,(r^{(q)}-1)M_T^{(q)}+r^{(q)}} = w^{(q)}_{m_T^{(q)},r^{(q)},p}, \quad q = 1, \ldots, Q, \tag{3.57}$$

and $\mathbf{W} = [\mathbf{W}^{(1)} \cdots \mathbf{W}^{(Q)}]$, which concatenates the $Q$ multiplexing matrices. Note that $\mathbf{W}$ is a matrix unfolding representation of the set of multiplexing tensors $\{\mathcal{W}^{(1)}, \ldots, \mathcal{W}^{(Q)}\}$. We also define $\mathbf{S} = [\mathbf{S}^{(1)} \cdots S^{(Q)}]$ as a matrix concatenating the symbol matrices for all $Q$ layers.

Taking these definitions into account, the received signal tensor can be written in constrained Block-PARAFAC form as

$$\mathbf{X}_3 = (\mathbf{H}\boldsymbol{\Psi} \diamond \mathbf{S}\boldsymbol{\Phi})\mathbf{W}^T + \mathbf{V}_3 \quad \in \mathbb{C}^{NM_R \times P}. \tag{3.58}$$

By comparing (3.52) and (3.28), we can deduce the following correspondences:

$$(I_1, I_2, I_3, R_1^{(q)}, R_2^{(q)}, \mathbf{A}^{(q)}, \mathbf{B}^{(q)}, \mathbf{C}^{(q)}) \rightarrow (M_R, N, P, M_T^{(q)}, R^{(q)}, \mathbf{H}^{(q)}, \mathbf{S}^{(q)}, \mathbf{W}^{(q)}) \tag{3.59}$$

The received signal tensor can also be expressed in two other matrix forms $\mathbf{X}_1 = (\mathbf{W} \diamond \mathbf{H}\boldsymbol{\Psi})(\mathbf{S}\boldsymbol{\Phi})^T + \mathbf{V}_1 \in \mathbb{C}^{M_R P \times N}$ and $\mathbf{X}_2 = (\mathbf{S}\boldsymbol{\Phi} \diamond \mathbf{W})(\mathbf{H}\boldsymbol{\Psi})^T + \mathbf{V}_2 \in \mathbb{C}^{PN \times M_R}$, following (3.21).

Model (3.52) covers some multiantenna coding schemes as special cases. By setting $R^{(1)} = \cdots = R^{(Q)} = 1$ and $M_T^{(1)} = \cdots = M_T^{(Q)} = 1$, it reduces to the multiantenna codes of [21], known as *Khatri-Rao Space-Time Codes*. On the other hand, for $Q = 1$ (with $R^{(1)} \geq 1$ and $M_T' > 1$), it takes the form of a *Linear Dispersion Code* [48], although the code design criterion is different here. Several codes falling between these ones are possible via the constrained Block-PARAFAC model, having different diversity-multiplexing trade-offs.

**Multiplexing structure**

The unfolded multiplexing matrix $\mathbf{W}$ is designed to satisfy the condition $\mathbf{W}\mathbf{W}^H = M'_T\mathbf{I}_P$, which is a condition required for performance maximization [21]. Defining

$$K' = \sum_{q=1}^{Q} R^{(q)} M_T^{(q)}, \tag{3.60}$$

we choose $\mathbf{W}$ as a Vandermonde matrix defined as:

$$[\mathbf{W}]_{p,k'} = e^{j\frac{2\pi}{K'}(k'-1)(p-1)},$$

where $k' = 1, \ldots, r^{(q)}m_T^{(q)}, \ldots, R^{(Q)}M_T^{(Q)}$. For achieving maximum diversity and eliminating multiuser interference, we must choose $P \geq K'$ so that $\mathbf{W}$ is a semi-unitary matrix.

**Interpretation of $\boldsymbol{\Psi}$ and $\boldsymbol{\Phi}$**

In the present context, the constraint matrices $\boldsymbol{\Psi}$ and $\boldsymbol{\Phi}$ admits an interesting physical interpretation. They can be viewed as *symbol-to-antenna loading matrices* and their structure reveals the overall joint spreading-multiplexing pattern considered at the transmitter. This means that we can construct different multistream space-time coding schemes by configuring these matrices with 1's and 0's. For example, let us consider $\overline{M}_T = 3$ transmit antennae and a transmission for $Q = 2$ users, which implies $Q = 2$ transmission layers. Assume $(M_T^{(1)}, R^{(1)}) = (2, 1)$, $(M_T^{(2)}, R^{(2)}) = (1, 3)$. From (3.30), $\boldsymbol{\Psi}$ and $\boldsymbol{\Phi}$ in this case have the following structure:

$$\boldsymbol{\Psi} = \begin{bmatrix} 1\,0\,0\,0\,0 \\ 0\,1\,0\,0\,0 \\ 0\,0\,1\,1\,1 \end{bmatrix}, \quad \boldsymbol{\Phi} = \begin{bmatrix} 1\,1\,0\,0\,0 \\ 0\,0\,1\,0\,0 \\ 0\,0\,0\,1\,0 \\ 0\,0\,0\,0\,1 \end{bmatrix}.$$

Note that both $\boldsymbol{\Psi}$ and $\boldsymbol{\Phi}$ are block-diagonal matrices with two diagonal blocks, each one associated with a given layer. Each row of $\boldsymbol{\Psi}$ defines the *spatial multiplexing factor* at each transmit antenna. More precisely, the number of 1's entries in each row of $\boldsymbol{\Psi}$ defines the number of symbols combined into each transmit antenna. Observe that the first and second rows (associated with the first layer) have only one non-zero entry, which indicates that each antenna transmits only one symbol at a time. The third row contains three non-zero entries, meaning that three symbols are simultaneously transmitted by the third antenna. Now, let us look at the structure of $\boldsymbol{\Phi}$. Again, we can identify two diagonal blocks in this matrix. Its number of rows corresponds to the total number of multiplexed data streams. Each row of $\boldsymbol{\Phi}$ defines the *spatial spreading factor* associated with each data-stream. Its first row has two non-zero entries, which means that the first data-stream is spread over the first two transmit antennae. The three other rows have only one non-zero entry, which indicates that the other data-streams are transmitted each by one transmit antenna, i.e., they are not spread in the spatial domain. The chosen spreading-multiplexing configuration can be verified in the following way:

$$\boldsymbol{\Psi}\boldsymbol{\Phi}^T = \begin{bmatrix} 1\,0\,0\,0 \\ 1\,0\,0\,0 \\ 0\,1\,1\,1 \end{bmatrix} \in \mathbb{C}^{M'_T \times K'}.$$

This matrix product reveals the spreading *versus* multiplexing pattern. Reading this matrix product for a fixed row, one can check for the number of data-streams multiplexed at a given antenna by counting the number of 1's entries in that row. On the other hand, by reading the matrix for a fixed column, one can check for the number of antennae over which a given data-stream is spread.

**Tensor-based blind detection**

We present a blind multiantenna receiver for a joint detection/decoding of the $Q$ transmitted signals. In this context, the ALS algorithm consists in fitting the constrained Block-PARAFAC model (3.58) in a Least Squares (LS) sense, by alternating between the estimations of $\mathbf{H}$ and $\mathbf{S}$. Note that the knowledge of the multiplexing structure at the receiver allows us to skip the estimation step of $\mathbf{W}$, as well as it helps us to achieve the global minimum faster (within 10-15 iterations).

| | Scheme 1 | Scheme 2 | Scheme 3 |
|---|---|---|---|
| $(M^{(1)}, R^{(1)})$ | (1,4) | (1,2) | (1,1) |
| $(M^{(2)}, R^{(2)})$ | (2,1) | (2,1) | (2,1) |

**Table 3.2.** Spreading and multiplexing parameters for the three TSTM schemes with $Q = 2$ layers.

The results are shown in terms of the average Bit-Error-Rate (BER) versus the SNR per bit. We consider $M'_T = 3$, $M_R = 2$, $N = 10$, $Q = 2$ transmission layers with $M_T^{(1)} = 2$ and $M_T^{(2)} = 1$. The code length is set to $P = R^{(1)}M_T^{(1)} + R^{(2)}M_T^{(2)}$, in order to achieve full diversity. We compare the performance of three different TSTM schemes with different number $R^{(q)}$ of multiplexed sub-sequences per layer. Table 3.2 summarizes the transmit parameters for the three schemes. Each plotted BER curve is the average over 1000 independent Monte Carlo runs. 8-Phase Shift Keying (PSK) modulation is assumed. The bit-rate of each scheme is also a performance metric, since it is a measure of the overall spectral efficiency. For the $q$-th layer, the bit-rate is given by $\frac{R^{(q)}}{P} log2(\mu)$ bits/channel use, where $\mu$ is the modulation cardinality.

According to Fig. 3.9, the BER performance of layer 2 is better than that of layer 1 for all the schemes. This is due to the fact that layer 2 spreads each sequence over two transmit antennae achieving a higher diversity gain, while layer 1 has a single antenna and no space spreading exists. Moreover, it can be seen that all the schemes achieve quite similar performance, but with different bit-rate distribution over the two layers. For first scheme, layer 1 achieves a bit-rate of 2 bits/channel use, which is four times that of layer 2. Note that the bit-rate ratio between the two layers decreases in the second scheme (layer 1 has twice the bit-rate of layer 2). In the third scheme both layers have the same bit-rate. The bit-rate values are summarized in the legend of Fig. 3.9. We call attention to the fact that constrained Block-PARAFAC modeling for multiantenna coding affords a variable bit-rate balance over different layers. This is particularly useful in multi-user (point-to-multipoint) downlink transmissions, where each layer is associated with a given user, but users have different diversity-rate requirements. Hence, the constrained Block-PARAFAC modeling approach adds some flexibility to the design of multiantenna coding schemes in a multi-user scenario.

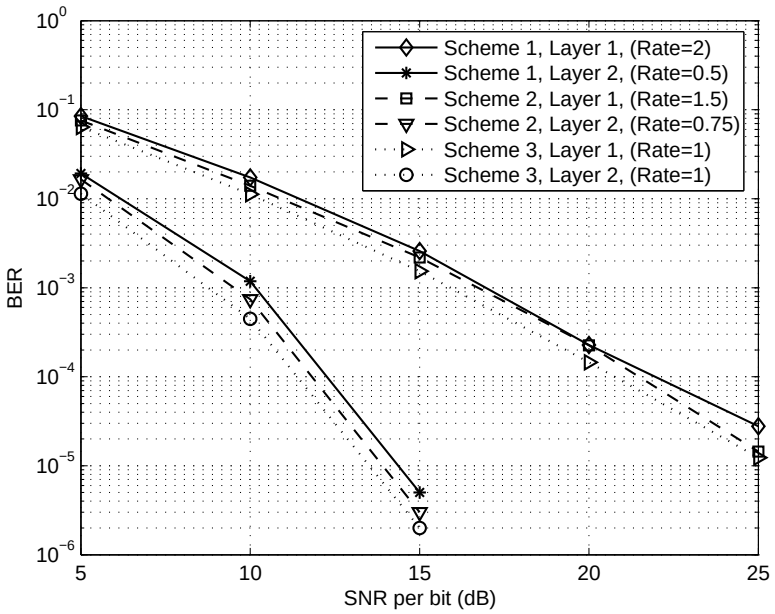



**Fig. 3.9.** Comparison of different TSTM schemes with ALS-based blind detection.

### 3.3.3 Estimation of wireless channel parameters

The issue of parametric multipath channel estimation has been exploited in several works [40, 41, 42]. Most of the approaches are based on subspace methods, which exploit shift-invariance properties and/or the knowledge of the

pulse shape function. Simultaneous estimation of angles of arrival and delays benefits from the fact that paths amplitudes are fastly-varying while angles and delays are slowly-varying over multiple transmission blocks or time-slots. In [40, 41, 42], the angles and delays are blindly-estimated using a collection of previous estimates of the space-time channel impulse response. As in [41, 42], the linear-phase variation property of the frequency domain transformed version of the known pulse shape function is exploited. Training-sequence-based space-time channel estimation methods exploiting the multislot invariance of angles and delays have been proposed recently in [49]. This problem can also be addressed using the PARAFAC decomposition, which exploits the fact that the variation of multipath amplitudes over multiple time-slots is faster than that of angles and delays. A PARAFAC model arises thanks to the use of a training sequence which is periodically extended over multiple time-slots that are jointly processed at the receiver.

We consider a wireless communication system in which a digital signal is transmitted in a specular multipath environment. The receiver is equipped with an array of $M$ antennae spaced half wavelength or closer. We focus on the case of a single-user transmission. The transmitted information symbols are organized into $J$ blocks or time-slots. Assume that the time-slots are sufficiently short so that the channel fading can be regarded as stationary over a time-interval necessary for the transmission of a whole time-slot and it varies independently from slot to slot. This is typically the case of Time Division Multiple Access (TDMA)-based systems [49]. We assume that the considered system is training-sequence-based, with the particular characteristic that it consists in reusing the training sequence: A known training sequence of $N$ symbols is periodically extended over multiple time-slots that are jointly processed at the receiver. The idea of processing multiple time-slots, based on training sequence reuse is also known as *multislot processing*.

The multipath channel model can be interpreted using the PARAFAC formalism. Let us write the slot-dependent space-time channel response in a scalar form as follows:

$$h_{m,j,k} = \sum_{l=1}^{L} a_{m,l}\beta_{j,l}g_{k,l}. \tag{3.61}$$

$h_{m,j,k}$ is interpreted here as the $(m,j,k)$-th element of a three-way array or third-order *tensor* $\mathcal{H} \in \mathbb{C}^{M\times J\times K}$. Note that $a_{m,l} = [\mathbf{A}(\boldsymbol{\theta})]_{m,l}$, $g_{k,l} = [\mathbf{G}(\boldsymbol{\tau})]_{k,l}$ and $\beta_{j,l} = [\mathbf{B}]_{j,l}$, where $\mathbf{B} \in \mathbb{C}^{J\times L}$ collects the fading amplitudes for all the slots. The three dimensions of the tensor $\mathcal{H}$ are *space*, *slot* and *time*. It is also possible to represent (3.61) as a function of matrices $\mathbf{A} = \mathbf{A}(\boldsymbol{\theta})$, $\mathbf{B}$ and $\mathbf{G} = \mathbf{G}(\boldsymbol{\tau})$. The space-time channel associated with the $j$-th slot can be regarded as the $j$-th matrix slice of the tensor $\mathcal{H}$, which is obtained by slicing the tensor along the *slot* dimension:

$$\mathbf{H}_{\cdot\, j\, \cdot} = \mathbf{A}\, D_j(\mathbf{B})\, \mathbf{G}^T,\ j = 1, \ldots, J. \tag{3.62}$$

Defining $\mathbf{H}_2 = \left[\mathbf{H}^T_{\cdot 1\, \cdot}, \ldots, \mathbf{H}^T_{\cdot J\, \cdot}\right]^T \in \mathbb{C}^{MJ\times K}$ as a matrix collecting the $J$ slices of the space-time channel, we get:

$$\mathbf{H}_2 = (\mathbf{B} \diamond \mathbf{A})\mathbf{G}^T. \tag{3.63}$$

Let $\{s(n)\}_{n=1}^{N}$ be the known training sequence. During the training period, the received baseband discrete-time signal impinging the antenna array at the $n$-th symbol period for the $j$-th slot, $\mathbf{x}_j(n) = [x_{1,j}(n), \ldots, x_{M,j}(n)]^T \in \mathbb{C}^M$ can be written as the convolution of the training sequence and the $j$-th channel response:

$$\mathbf{x}_j(n) = \sum_{k=0}^{K-1} \mathbf{h}_j(kT)s(n-k) + \mathbf{v}_j(n), \tag{3.64}$$

where $T$ is the symbol period and $\mathbf{v}_j(n)$ is the additive noise, which is assumed to be Gaussian with variance $\sigma_v^2$, irrespective of the slot. The temporal support of the channel impulse response is $(0, KT]$. A matrix model for the received signal can be obtained from (3.64):

$$\mathbf{X}_{\cdot j\cdot} = \mathbf{H}_{\cdot j\cdot}\mathbf{S}^T + \mathbf{V}_{\cdot j\cdot}, \tag{3.65}$$

where $\mathbf{X}_{\cdot j\cdot} = [\mathbf{x}_j(0), \ldots, \mathbf{x}_j(N-1)] \in \mathbb{C}^{M\times N}$, $\mathbf{H}_{\cdot j\cdot} = [\mathbf{h}_j(0), \ldots, \mathbf{h}_j(K-1)T] \in \mathbb{C}^{M\times K}$, $\mathbf{S}$ is a Toeplitz matrix formed from the training sequence, $[\mathbf{S}]_{n,k} = s(n-k)$, and $\mathbf{V}_{\cdot j\cdot} = [\mathbf{v}_j(0), \ldots, \mathbf{v}_j(N-1)] \in \mathbb{C}^{M\times N}$. By stacking of $\mathbf{X}_{\cdot 1\cdot}, \ldots, \mathbf{X}_{\cdot J\cdot}$ column-wise and using (3.63) we can express the received signal as:

$$\mathbf{X}_2 = \begin{bmatrix} \mathbf{X}_{\cdot\, 1\, \cdot} \\ \vdots \\ \mathbf{X}_{\cdot\, J\, \cdot} \end{bmatrix} = (\mathbf{B} \diamond \mathbf{A})(\mathbf{S}\mathbf{G})^T + \mathbf{V}_2. \tag{3.66}$$

We assume that the received signal is sampled with a factor $P$ equal or above the Nyquist rate, and that a Discrete Fourier Transform (DFT) is applied to the received signal at each receiver antenna. Under these assumptions, the DFT-transformed received signal matrix, denoted by $\breve{\mathbf{X}}_2$, can be approximated by [41]:

$$\breve{\mathbf{X}}_2 = (\mathbf{B} \diamond \mathbf{A})\mathbf{C}^T + \breve{\mathbf{V}}_2 \tag{3.67}$$

where $\mathbf{C} = \mathbf{S}\mathbf{G}_0\mathbf{F} \in \mathbb{C}^{N \times L}$, $\mathbf{F}$ is a Vandermonde matrix given by

$$\mathbf{F} = \begin{bmatrix} 1 & \cdots & 1 \\ \phi_1 & \cdots & \phi_L \\ \vdots & & \vdots \\ \phi_1^{(K-1)} & \cdots & \phi_L^{(K-1)} \end{bmatrix}$$
$$\phi_l = e^{-j(2\pi/K)\tau_l}, \quad l = 1, \ldots, L, \tag{3.68}$$

and $\mathbf{G}_0 = Diag(\breve{\mathbf{g}}_0)$, with $\breve{\mathbf{g}}_0 = DFT(\mathbf{g}) \in \mathbb{C}^{KP}$ containing the DFT samples of the known pulse-shape function $\mathbf{g} = [g(0)g\left(\frac{1}{P}\right) \cdots g\left(K - 1 - \frac{1}{P}\right)]$. Note that the same information contained in (3.67) can also be rearranged as $\breve{\mathbf{X}}_3 = (\mathbf{A} \diamond \mathbf{C})\mathbf{B}^T + \breve{\mathbf{V}}_3 \in \mathbb{C}^{NM \times J}$ or as $\breve{\mathbf{X}}_1 = (\mathbf{C} \diamond \mathbf{B})\mathbf{A}^T + \breve{\mathbf{V}}_1 \in \mathbb{C}^{JN \times M}$, by analogy with (3.21).

In the present context, we make the following assumptions concerning the multipath channel structure; i) the array manifold is known and the multipath signals arrive at the array at distinct angles; ii) the multipaths undergo independent fading and vary independently from slot to slot and iii) the multipaths have distinct propagation delays to the receiver. Under these assumptions, the maximum number of multipaths that can be estimated is given by the following equation:

$$\min(M, L) + \min(J, L) \geq L + 2. \tag{3.69}$$

By studying the condition (3.69), we can distinguish two cases:

1. If $J \geq L$ then $M \geq 2$ receiver antennae are sufficient for estimating angle, delay and amplitudes of the $L$ multipaths.
2. If $M \geq L$ then $J \geq 2$ slots are sufficient for estimating the set of multipath parameters.

**PARAFAC-based estimator**

The receiver algorithm for the joint estimation of angles, delays and amplitudes of the multipaths fully exploits the trilinear structure of the multipath channel model. It is based on the classical alternating least squares (ALS) algorithm for estimating the matrices $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$ from matrices $\breve{\mathbf{X}}_{i=1,2,3}$ in presence of additive Gaussian noise. After ALS convergence, *a priori* knowledge of the Vandermonde structure of $\mathbf{A}$ and $\mathbf{F}$ are exploited to eliminate the scaling ambiguity of the ALS estimation process.

The performance of the PARAFAC-based multipath parameter estimator is evaluated through computer simulations. The training sequence to be used over the $J$ slots is randomly generated at each run, following a normal distribution with unity variance. The pulse shape function is a raised cosine with roll-off factor $0.35$. The temporal support of the channel is $K = 5$. A multipath scenario with $L = 3$ paths is assumed. The angles of arrival and time delays are $\{\theta_1, \theta_2, \theta_3\} = \{-10°, 0, 20°\}$ and $\{\tau_1, \tau_2, \tau_3\} = \{0, 1.1T, 2T\}$. The paths are assumed to have the same average power. The results are averaged over 100 Monte Carlo runs. For each run, multipath fading amplitudes for the $J$ time-slots are redrawn from an i.i.d. Rayleigh generator. Random initialization is used. If convergence is not achieved within 100 iterations, we re-start the algorithm from a different initialization. It has been observed that convergence is achieved within 20-30 iterations in most of the runs. The Root Mean Square Error (RMSE) between the estimated and true matrices is used here as the evaluation metric for the estimator performance.

Figure 3.10 depicts the RMSE versus SNR for the estimation of the array (angle) and pulse shape (delay) responses, considering $M = 2$ antennae and $N = 8$ training samples. The results are shown for $J = 5$ and $10$ time-slots. It is seen that the proposed estimator exhibits a linear decrease in its RMSE as SNR increases. This is valid for both angle and delay RMSE. The performance gap between angle and delay estimation is due to the fact that the raised cosine pulse shape function is not bandlimited, which leads to some delay estimation bias. As expected, the estimator performance improves as the number of time-slots processed increases. Although not displayed in the figure, the RMSE results for the fading amplitudes are very close to those for the delay responses. Note that these performance results are achieved with fewer antennae than multipaths and with a very short training sequence, which is interesting characteristic of the proposed PARAFAC-based estimator.

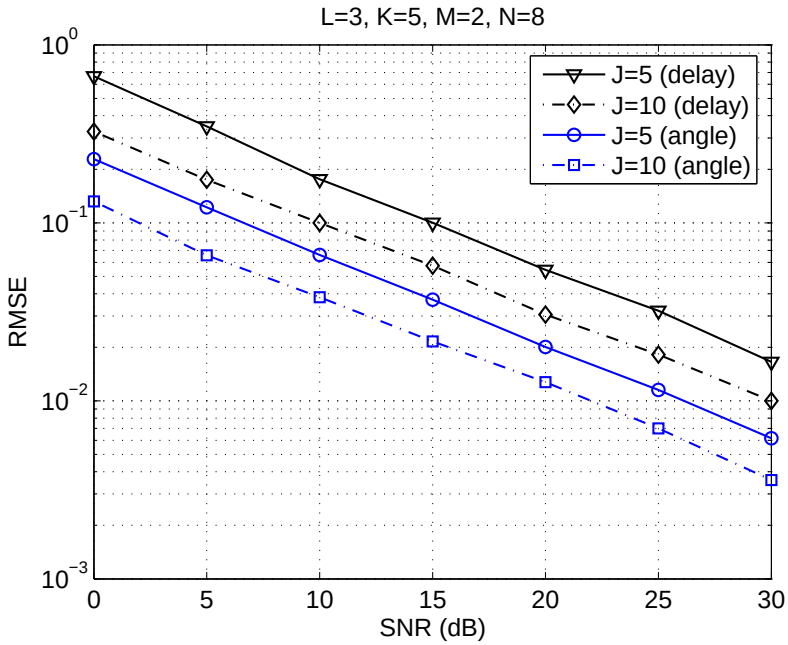



**Fig. 3.10.** RMSE versus SNR results.

**Generalization to MIMO channels**

For time-slotted MIMO systems, a "block-fading" MIMO channel model can be considered, where fading is assumed to be constant within each time-slot and changes independently from slot to slot. Different approaches for estimating MIMO channels have been recently proposed in [50, 51, 52, 53, 54]. The block-fading MIMO channel can alternatively be formulated as fourth-order PARAFAC tensor model, which can be exploited for estimating the angles of arrival/departure, delays and fading gains.

Let us consider a MIMO antenna system with $M_T$ transmit and $M_R$ receive antennae, which are closely spaced to each other, so that the far-field approximation holds. The scattering environment is such that there are $L$ dominant paths linking the transmitter to the receiver, associated with different clusters of scatterers. Each path is characterized by an Angle Of Departure (AOD) $\phi_l$, an Angle Of Arrival (AOA) $\theta_l$, a relative propagation delay $\tau_l$ and a slot-dependent fading gain $\beta_{j,l}$ for the $j$-th slot, $j = 1, \ldots, J$. The fading on the $L$ paths is uncorrelated and *slot-wise* independent (block-fading assumption). The finite support of the channel impulse response is equal to $K$ symbol periods. The array response matrices collecting the $L$ path contributions are defined as

$$\mathbf{A}^{(R)}(\boldsymbol{\theta}) = [\mathbf{a}^{(R)}(\theta_1) \cdots \mathbf{a}^{(R)}(\theta_L)] \in \mathbb{C}^{M_R \times L}, \quad \mathbf{A}^{(T)}(\boldsymbol{\phi}) = [\mathbf{a}^{(T)}(\phi_1) \cdots \mathbf{a}^{(T)}(\phi_L)] \in \mathbb{C}^{M_T \times L}.$$

A matrix collecting $L$ delayed pulse shape responses is defined as $\mathbf{G}(\boldsymbol{\tau}) = [\mathbf{g}(\tau_1) \cdots \mathbf{g}(\tau_L)] \in \mathbb{C}^{K \times L}$ and a matrix assembling the path fading gains during the $J$ slots is denoted by $\mathbf{B} = [\mathbf{b}_1 \cdots \mathbf{b}_L] \in \mathbb{C}^{J \times L}$.

The length-$N$ training sequence associated with the $m_T$-th transmit antenna is represented by a vector $\mathbf{s}_{m_T} = [s_{m_T}(1) \cdots s_{m_T}(N)]^T$. We call attention to the fact that the $M_T$ vectors $\{\mathbf{s}_1, \ldots, \mathbf{s}_{M_T}\}$ are assumed to be linearly independent but they are not necessarily constrained to be orthogonal. We also define $\mathbf{S} = [T(\mathbf{s}_1) \cdots T(\mathbf{s}_{M_T})] \in \mathbb{C}^{N \times M_T K}$ as a block-Toeplitz matrix concatenating $M_T$ Toeplitz matrices.

The MIMO channel is a fourth-order tensor $\mathcal{H} \in \mathbb{C}^{M_R \times M_T \times K \times J}$ with typical element $h_{m_R,m_T,k,j}$, and following a PARAFAC model. As a result, the collection of the multislot received signal over the $J$ training periods can be modeled as a third-order tensor $\mathcal{X} \in \mathbb{C}^{M_R \times N \times J}$ with typical element $x_{m_R,\,n,\,j}$, also following a PARAFAC model. The PARAFAC model for the channel and the received signal are given by

$$h_{m_R,m_T,k,j} = \sum_{l=1}^{L} a_{m_R}^{(R)}(\theta_l)\beta_{j,\,l}a_{m_T}^{(T)}(\phi_l)g_k(-\tau_l), \qquad x_{m_R,\,n,\,j} = \sum_{m_T=1}^{M_T}\sum_{k=1}^{K} h_{m_R,m_T,k,j}\, s_{n,\,m_T,\,k} + v_{m_R,\,n,j}, \tag{3.70}$$

with $a_{m_R}^{(R)}(\theta_l) = [\mathbf{A}^{(R)}(\boldsymbol{\theta})]_{m_R,l}$, $a_{m_T}^{(T)}(\phi_l) = [\mathbf{A}^{(T)}(\boldsymbol{\phi})]_{m_T,l}$, $g_k(-\tau_l) = g(k-1-\tau_l) = [\mathbf{G}(\boldsymbol{\tau})]_{k,l}$, $\beta_{j,\,l} = [\mathbf{B}]_{j,l}$, $s_{n,\,m_T,\,k} = [\mathbf{S}]_{n,(m_T-1)K+k}$, and $v_{m_R,\,n,j}$ represents the additive white Gaussian noise, and $T$ is the symbol period.

Let us define an unfolded matrix representation $\mathbf{H} \in \mathbb{C}^{JM_R \times M_T K}$ for the channel tensor $\mathcal{H}$ in the following manner:

$$[\mathbf{H}]_{(j-1)M_R+m_R,(m_T-1)K+k} = h_{m_R,m_T,k,j}.$$

For simplifying the notation, let us call $\mathbf{A}^{(R)} = \mathbf{A}^{(R)}(\boldsymbol{\theta})$, $\mathbf{A}^{(T)} = \mathbf{A}^{(T)}(\boldsymbol{\phi})$. and $\mathbf{G} = \mathbf{G}(\boldsymbol{\tau})$. It can be shown that $\mathbf{H}$ can be factored as a function of the MIMO channel parameters:

$$\mathbf{H} = (\mathbf{B} \diamond \mathbf{A}^{(R)})(\mathbf{A}^{(T)} \diamond \mathbf{G})^T = (\mathbf{B} \diamond \mathbf{A}^{(R)})\mathbf{T}^T, \quad \mathbf{T} = (\mathbf{A}^{(T)} \diamond \mathbf{G}) \in \mathbb{C}^{M_T K \times L}. \tag{3.71}$$

At each receive antenna, we collect $N$ samples of the received data during the training period of the $J$ slots in a matrix $\mathbf{X}_2 \in \mathbb{C}^{JM_R \times N}$, defined as $[\mathbf{X}_2]_{(j-1)M_R+m_R,n} = x_{m_R,n,j}$. Taking the definitions (3.71) into account, $\mathbf{X}_2$ can be expressed as:

$$\mathbf{X}_2 = \mathbf{H}\mathbf{S}^T = (\mathbf{B} \diamond \mathbf{A}^{(R)})\mathbf{C}^T, \quad \mathbf{C} = \mathbf{S}\mathbf{T} \in \mathbb{C}^{N \times L}. \tag{3.72}$$

$\mathbf{X}_2$ is an unfolded matrix of the received signal tensor $\mathcal{X}$ defined in (3.70), which follows a third-order PARAFAC model. The two other unfolded matrices of the received signal tensor are given by $\mathbf{X}_3 = (\mathbf{A}^{(R)} \diamond \mathbf{C})\mathbf{B}^T + \mathbf{V}_3 \in \mathbb{C}^{NM_R \times J}$ and $\mathbf{X}_1 = (\mathbf{C} \diamond \mathbf{B})\mathbf{A}^{(R)T} + \mathbf{V}_1 \in \mathbb{C}^{JN \times M_R}$, following (3.21).

According to the identifiability results of the PARAFAC model, the identifiability of $\mathbf{A}^{(R)}$, $\mathbf{B}$, and $\mathbf{C}$ is linked to the concept of $k$-rank of these matrices [14]. In our context, a sufficient condition for identifying the MIMO multipath parameters based on (3.72) can be obtained by recalling useful results on the $k$-rank of matrices having Vandermonde and Khatri-Rao product structures (see [15], Lemmas 1 and 2). Assuming that $\mathbf{S} \in \mathbb{C}^{N \times M_T K}$ is tall and full rank, a sufficient identifiability condition is

$$k_{\mathbf{A}^{(R)}} + k_{\mathbf{B}} + min(k_{\mathbf{A}^{(T)}} + k_{\mathbf{G}} - 1, L) \geq 2(L+1).$$

From the block-fading assumption and assuming distinct AOAs, AODs and delays, the following condition is sufficient for identifiability:

$$min(M_R, L) + min(J, L) + min(M_T + K - 1, L) \geq 2(L+1). \tag{3.73}$$

Column permutation is not relevant to the context of this problem, and the scaling ambiguity can be eliminated by exploiting prior knowledge of the space-time manifold structure (i.e., array responses and pulse shape function). The estimation of the MIMO multipath parameters is done in two-stages. The first one is blind, and is based on the ALS algorithm.The second one makes use of the training sequence to estimate $\mathbf{T} = (\mathbf{A}^{(T)} \diamond \mathbf{G})$ in the LS sense by solving equation $\mathbf{C} = \mathbf{S}\mathbf{T}$, i.e., $\widehat{\mathbf{T}} = \mathbf{S}^{\dagger}\widehat{\mathbf{C}}^{(conv)}$. Separated estimations of the other MIMO channel parameters $\widehat{\mathbf{A}}^{(T)}$ and $\widehat{\mathbf{G}}$ can be obtained by exploiting the Vandermonde structure of $\mathbf{A}^{(R)T}$ and the double Khatri-Rao structure of the MIMO channel defined in (3.71).

Figure 3.11 depicts the estimation performance for $\mathbf{A}^{(T)}$, $\mathbf{A}^{(R)}$ and $\mathbf{G}$, averaged over the $L$ paths and over 50 independent runs. The RMSE for the overall estimated MIMO channel $\widehat{\mathbf{H}}$ is also shown. We have assumed $M_T = 3$, $M_R = 2$, $L = 3$, and a raised cosine function with excess bandwidth 0.35 and $K = 4$. The considered multipath parameters are shown in the figure. Pseudo-random training sequences of $N = 18$ symbols were used over $J = 5$ slots. Note that the multipath parameters and the overall MIMO channel are estimated with good accuracy, and the RMSE exhibits an almost linear decrease as a function of the SNR.

## 3.4 Summary

Several phenomena in wireless communication systems, which are related to the transmitter, channel and/or receiver, can be modeled using the tensor formalism, although this is not always realized or exploited. Modeling this kind of systems using tensor decompositions or multidimensional arrays, allows one to simultaneously exploit the multiple forms of signal diversity for information recovery purposes, by allowing the use of blind processing techniques in problems of signal separation, channel estimation and equalization, among others. These are the main benefits of interpreting wireless communication signals as tensors instead of matrices. We have made a survey on tensor decompositions and some of its applications in signal processing for wireless communication systems. The chapter has first provided a theoretical background on tensor decompositions and then, some signal processing applications were studied with focus on recent results. We have shown that the constrained Block-PARAFAC decomposition can be used for unifying the tensor modeling of oversampled, DS/CDMA and OFDM systems, for performing blind multiuser equalization, as well as for the design of multiantenna coding techniques. The chapter has also applied the PARAFAC decomposition for the estimation of wireless channel parameters.

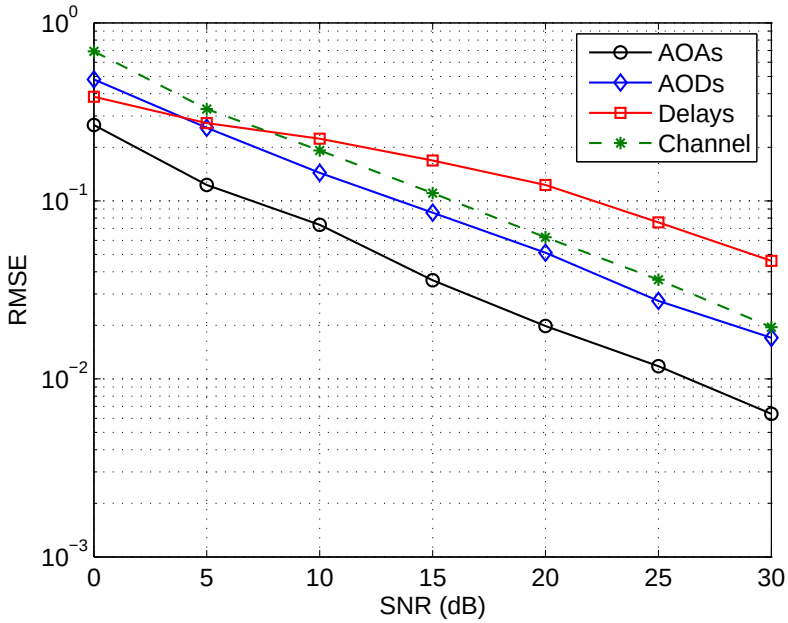



**Fig. 3.11.** RMSE performance of the PARAFAC-ALS estimator.

## References


1. L. R. Tucker, "Some mathematical notes on three-mode factor analysis," *Psychometrika*, vol. 31, pp. 279–311, 1966.
2. R. A. Harshman, "Foundations of the PARAFAC procedure: Model and conditions for an "explanatory" multi-mode factor analysis," *UCLA Working Papers in Phonetics*, vol. 16, pp. 1–84, Dec. 1970.
3. J. D. Carroll and J. Chang, "Analysis of individual differences in multidimensional scaling via an N-way generalization of "Eckart-Young" decomposition," *Psychometrika*, vol. 35, no. 3, pp. 283–319, 1970.
4. J. B. Kruskal, "Three-way arrays: Rank and uniqueness or trilinear decompositions, with applications to arithmetic complexity and statistics," *Linear Algebra Applications*, vol. 18, pp. 95–138, 1977.
5. R. A. Harshman, "Determination and proof of minimum uniqueness conditions for PARAFAC1," *UCLA Working Papers in Phonetics*, no. 22, pp. 111–117, 1972.
6. L. De Lathauwer, "Signal processing based on multilinear algebra," Ph.D. dissertation, Katholieke Univ. Leuven, Leuven, Belgium, 1997.
7. N. D. Sidiropoulos, G. B. Giannakis, and R. Bro, "Blind PARAFAC receivers for DS-CDMA systems," *IEEE Trans. Signal Processing*, vol. 48, no. 3, pp. 810–822, March 2000.
8. N. D. Sidiropoulos and G. Z. Dimic, "Blind multiuser detection in WCDMA systems with large delay spread," *IEEE Signal Processing Letters*, vol. 8, no. 3, pp. 87–89, Mar. 2001.
9. A. de Baynast, L. De Lathauwer, and B. Aazhang, "Blind PARAFAC receivers for multiple access-multiple antenna systems," in *Proc. VTC Fall*, Orlando, USA, Oct. 2003.
10. A. de Baynast and L. De Lathauwer, "Détection autodidacte pour des systèmes à accès multiple basée sur l'analyse PARAFAC," in *Proc. of XIX GRETSI Symp. Sig. Image Proc.*, Paris, France, Sep. 2003.
11. T. Jiang and N. D. Sidiropoulos, "A direct semi-blind receiver for SIMO and MIMO OFDM systems subject to frequency offset," in *Proc. SPAWC*, Rome, Italy, June 2003.
12. A. L. F. de Almeida, G. Favier, and J. C. M. Mota, "Generalized PARAFAC model for multidimensional wireless communications with application to blind multiuser equalization," in *Proc. Asilomar Conference. Sig. Syst. Comp.*, Pacific Grove, CA, October 31 - November 2 2005.
13. ——, "PARAFAC-based unified tensor modeling for wireless communication systems with application to blind multiuser equalization," *Signal Processing Elsevier, Special Issue on Tensor Signal Processing*, to appear, 2006.
14. N. D. Sidiropoulos, R. Bro, and G. B. Giannakis, "Parallel factor analysis in sensor array processing," *IEEE Trans. Signal Processing*, vol. 48, no. 8, pp. 2377–2388, Aug. 2000.
15. N. D. Sidiropoulos and X. Liu, "Identifiability results for blind beamforming in incoherent multipath with small delay spread," *IEEE Trans. Signal Processing*, vol. 49, no. 1, pp. 228–236, Jan. 2001.
16. A. L. F. de Almeida, G. Favier, and J. C. M. Mota, "PARAFAC receiver for blind multiuser equalization in wireless communication systems with temporal oversampling," in *Proc. European Signal Processing Conference (EUSIPCO)*, Antalya, Turkey, September 4-8 2005.
17. ——, "Constrained block-PARAFAC decomposition and applications in wireless communications," *SIAM Journal on Matrix Analysis (SIMAX), Special Issue on Tensor Decompositions and Applications*, submitted, 2006.

18. Y. Rong, S. A. Vorobyov, A. D. Gershman, and N. D. Sidiropoulos, "Blind spatial signature estimation via time-varying user power loading and parallel factor analysis," *IEEE Trans. Signal Processing*, vol. 53, no. 5, pp. 1697–1710, May 2005.
19. A. L. F. de Almeida, G. Favier, and J. C. M. Mota, "Multipath parameter estimation of time-varying space-time communication channels using parallel factor analysis," in *IEEE Int. Conf. Acoustics, Speech and Sig. Proc. (ICASSP)*, Toulouse, France, May 14-18 2006.
20. ——, "Estimation of frequency-selective block-fading MIMO channels using PARAFAC modeling and alternating least squares," in *Asilomar Conference. Sig. Syst. Comp.*, Pacific Grove, CA, 2006, submitted.
21. N. D. Sidiropoulos and R. Budampati, "Khatri-Rao space-time codes," *IEEE Trans. Signal Processing*, vol. 50, no. 10, pp. 2377–2388, 2002.
22. A. L. F. de Almeida, G. Favier, and J. C. M. Mota, "Space-time multiplexing codes: A tensor modeling approach," in *IEEE Int. Workshop on Sig. Proc. Advances in Wireless Commun. (SPAWC)*, Cannes, France, July 2006.
23. ——, "Tensor-based space-time multiplexing codes for MIMO-OFDM systems with blind detection," in *Proc. IEEE Int. Symp. Pers. Ind. Mob. Radio Commun. (PIMRC)*, Helsinki, Finland, September 2006.
24. ——, "Trilinear space-time-frequency codes for broadband MIMO-OFDM systems," in *International Telecom Symposium (ITS)*, Fortaleza, Brazil, September 3-6 2006.
25. J. M. F. ten Berge, "Non-triviality and identification of a constrained Tucker3 analysis," *Journal of Chemometrics*, vol. 16, pp. 609–612, 2002.
26. N. D. Sidiropoulos and R. Bro, "On the uniqueness of multilinear decomposition of N-way arrays," *Journal of Chemometrics*, vol. 14, pp. 229–239, 2000.
27. J. M. F. ten Berge and N. D. Sidiropoulos, "On uniqueness in CANDECOMP/PARAFAC," *Psychometrika*, vol. 67, pp. 399–409, 2002.
28. T. Jiang and N. D. Sidiropoulos, "Kruskal's permutation lemma and the identification of CANDECOMP/PARAFAC and bilinear models with constant modulus constraints," *IEEE Trans. Signal Processing*, vol. 52, pp. 2625–2636, Sep. 2004.
29. R. A. Harshman and M. E. Lundy, *The PARAFAC model for three-way factor analysis and multidimensional scaling*. New York: Praeger: H. G. Law, C. W. Snyder Jr., J. Hattie, and R. P. McDonald Eds., 1984.
30. N. D. Sidiropoulos, "Low-rank decomposition of multi-way arrays: A signal processing perspective," in *Invited plenary lecture at IEEE SAM*, Barcelona, Spain, July 2004.
31. L. De Lathauwer, "The decomposition of a tensor in a sum of rank-$(R_1, R_2, R_3)$ terms," in *Workshop on Tensor Decompositions and Applications*, Marseille, France, 2005.
32. R. Bro, R. A. Harshman, and N. D. Sidiropoulos, "Rank-deficient models for multi-way data," *Journal on Chemomometrics*, 2006, to be published.
33. ——, "Modeling multi-way data with linearly dependent loadings," *KVL Technical report 176*, 2005.
34. R. Bro, "Multi-way analysis in the food industry: Models, algorithms and applications," Ph.D. dissertation, University of Amsterdam, Amsterdam, 1998.
35. H. A. Kiers and A. K. Smilde, "Constrained three-mode factor analysis as a tool for parameter estimation with second-order instrumental data," *Journal of Chemometrics*, vol. 12, no. 2, pp. 125–147, Dec. 1998.
36. A. J. Paulraj and C. B. Papadias, "Space-time processing for wireless communications," *IEEE Signal Processing Mag.*, vol. 14, no. 6, pp. 49–83, Jan. 1997.
37. S. Talwar, M. Viberg, and A. J. Paulraj, "Blind separation of synchronous co-channel digital signals using an antenna array–Part I: Algorithms," *IEEE Trans. Signal Processing*, vol. 44, no. 5, pp. 1184–1197, May 1996.
38. A.-J. van der Veen, S. Talwar, and A. J. Paulraj, "A subspace approach for blind space-time signal processing," *IEEE Trans. Signal Processing*, vol. 45, no. 1, pp. 173–190, Jan 1997.
39. X. Liu and N. D. Sidiropoulos, "Almost sure identifiability of constant modulus multidimensional harmonic retrieval," *IEEE Trans. Signal Processing*, vol. 50, no. 9, pp. 2366–2368, Sep. 2002.
40. M. C. Vanderveen, C. B. Papadias, and A. Paulraj, "Joint angle and delay estimation (JADE) for multipath signals arriving at an antenna array," *IEEE Commun. Letters*, vol. 1, no. 1, pp. 12–14, Jan. 1997.
41. A.-J. van der Veen, M. C. Vanderveen, and A. Paulraj, "Joint angle and delay estimation using shift invariance techniques," *IEEE Trans. Signal Processing*, vol. 46, no. 2, pp. 405–418, Feb. 1998.
42. M. C. Vanderveen, A.-J. van der Veen, and A. Paulraj, "Estimation of multipath parameters in wireless communications," *IEEE Trans. Signal Processing*, vol. 46, no. 3, pp. 682–690, Mar. 1998.
43. J. G. Proakis, *Digital Communications*. New York: McGraw-Hill, 2001.
44. R. D. van Nee and R. Prasad, *OFDM for wireless multimedia communications*. Artech House, 2000.
45. V. Tarokh, N. Seshadri, and A. R. Calderbank, "Space-time codes for high data rate wireless communications: performance criterion and code construction," *IEEE Trans. Inf. Theory*, vol. 44, no. 2, pp. 744–765, Mar. 1998.
46. G. J. Foschini and M. J. Gans, "On limits of wireless communications when using multiple antennas," *Wireless Pers. Comm.*, vol. 6, no. 3, pp. 311–335, 1998.
47. G. D. Golden, G. J. Foschini, R. A. Valenzuela, and P. W. Wolniansky, "Detection algorithm and initial laboratory results using the V-BLAST space-time communications architecture," *Electronics Letters*, vol. 35, no. 7, pp. 14–15, 1999.
48. B. Hassibi and B. M. Hochwald, "High-rate codes that are linear in space and time," *IEEE Trans. Inf. Theory*, vol. 48, no. 7, pp. 1804–1824, July 2002.
49. O. Simeone and U. Spagnolini, "Multi-slot estimation of space-time channels," in *Proc. IEEE Int. Conf. Commun. ICC*, New York City, USA, Apr. 2002.

50. J. H. Kotecha and A. M. Sayeed, "Transmit signal design and optimal estimation of correlated MIMO channels," *IEEE Trans. Sig. Proc.*, vol. 52, no. 2, pp. 546–557, 2004.
51. O. Simeone and U. Spagnolini, "Channel estimation for block-fading MIMO channels: performance limits," in *IEEE Vehic. Tech. Conf. (VTC Spring'04)*, Milan, Italy, May 17-19 2004.
52. M. Guillaud and D. T. M. Slock, "Pathwise mimo channel modeling and estimation," in *IEEE Workshop on Sig. Proc. Advances in Wireless Commun.*, New York City, USA, June 5-8 2005.
53. J. Li, J. Conan, and S. Pierre, "Pathwise mimo channel modeling and estimation," in *Proc. Int. Symp. on Wireless Commun. Systems*, Siena, Italy, Sep. 5-7 2005, pp. 22–26.
54. A. Gershman and N. D. Sidiropoulos, *Space-time processing for MIMO communications*. England: John Wiley & Sons, 2005.

# 4

# Power Control for Wireless Networks

Raimundo A. de Oliveira Neto[1], Fabiano de S. Chaves[2], F. Rodrigo P. Cavalcanti[1], and Ricardo B. dos Santos[1]

[1] Wireless Telecommunication Research Group, Federal University of Ceará, Brazil
{neto,rodrigo,brauner}@gtel.ufc.br
[2] State University of Campinas fabiano@decom.fee.unicamp.br

## 4.1 Introduction

Power control is an essential tool in radio resource management. The transmission power affects the signal quality and the amount of interference in the co-channel links in a wireless system. Then, it is necessary to control the transmitted power so that the received power is the necessary minimum power in order to satisfy the quality of the communication and, at the same time, to prevent the generation of unnecessary interference. Moreover, Power control is essential concerning energy efficiency, since communications nodes using low power levels means longer lifetime of batteries for mobile terminals and more energy resources available for central nodes as base stations in cellular systems. Therefore, Power control serves both to manage the amount of interference in the system and to rationalize the use of energy resources, increasing the system capacity. The importance of this technique can be attested by the fact that it was standardized in third generation wireless systems and therefore requires special attention. Power control is a much researched subject. However, few works compile in a systematic way the different fronts of research in this area. Therefore, through this tutorial, we intend to provide a survey about the many facets of Power control for wireless systems. This work is organized as follows. Firstly, we present some basic definitions and the model which describes the problem and its variables in a general framework. After that, a classification of the algorithms according to the type of communication infra-structure available is given. A detailed description of the main algorithms from the literature is provided. In addition, some recent improvements on such algorithms are presented. Convergence and practical aspects are also discussed. Moreover, advanced topics such as the application of Game Theory and prediction techniques to Power control are exposed. Finally, we present the conclusions of this work.

## 4.2 Models and Basic Definitions

### Propagation Channel

The propagation channel is the physics medium which holds the propagation of the electromagnetic wave between communication antennas. When a wave travels through a propagation channel, the transmitted signal is affected by a power gain $g$. That is, if the transmitter emits a signal with power $p_t$, the received signal by the receiver will have a power given by:

$$p_r = p_t \cdot g \tag{4.1}$$

In wireless communications, most channel models assume that the power gain $g$ depends on three propagation effects: path loss, shadowing and short-term fading. The power gain is then composed by the multiplicative superposition of each of these effects. These three effects are better explained in the next subsections.

### Path Loss

The transmitted power is attenuated by a distance-dependent factor called path loss $(PL)$. One of the simplest path loss models is the free space loss. In this case, the received power $p_r$ is proportional to $1/d^2$ (Square Inverse Law) [1]. For the non-free space case, path loss is usually assumed proportional to $1/d^n$ [2], where $n$ is the so-called path-loss exponent which indicates the rate in which the path loss increases with the distance $d$. That is, the higher $n$, more

attenuating is the propagation channel. Typical values of $n$ vary from 2 to 6 [2]. It is important to emphasize that $n$ depends on environment (for instance, urban, rural, micro-cellular, height of the constructions) and frequency.

There are various other more complex and precise path-loss models. They may be empirical (e.g. Okumura-Hata [3] and [4]), semi-empirical (e.g. Cost-231 [5]) or deterministic, such as the ones based on ray-tracing [6]. However, for a given set of fixed transmission and system parameters, path loss is ultimately a function of distance. For the purpose of this tutorial, we assume the simple path loss model:

$$PL(d) = Kd^{-n} \tag{4.2}$$

where $K$ represents the perceived path loss at the reference distance of $d = 1$, which takes into account also all specific transmission and systems parameters.

**Shadowing**

Consider a transmitter and the circumference of radius $d$ with the transmitter in its center. The path loss $PL$ for all points of the circumference is the same, however power gains $g$ are different. Such phenomenon occurs due to shadowing. Shadowing provokes variations in the power gain around its mean (given by the path loss). In cellular systems, shadowing can be modeled as a log-normal random variable [2].

Among the main causes of shadowing, we can cite obstruction of the radio signals due to the large obstacles. The scale of the signal variation due to shadowing is therefore on the order of tens to hundreds of wavelengths, considering the 1-2GHz frequency band usually employed in cellular systems. For such, shadowing is also known as large-scale fading [2].

**Short-term Fading**

Short-term fading is related to fast fluctuations in the amplitude of a radio signal in a short time period or in short covered distance [2]. Its main cause is the propagation in different paths of several replicas of the transmitted signal (multi-paths). These replicas arrive in the receiver in slightly distinct instants and with random amplitude and phase. This occurs due to reflection and scattering of the signal during its propagation. The reflections occur when the wave reaches objects with dimension larger than the wavelength of the signal. On the other hand, the scattering occurs when the obstacles have the same order of magnitude of the wavelength.

When several replicas of the signal arrive in the receiver, their different phases add both constructively and destructively, generating, fast fluctuations on the amplitude of the received signal.

Another factor which influences the short-term fading is the speed of the Mobile Station (MS) as well as of the objects around it in the propagation environment. Since fading is a spatial phenomenon, the faster the MS moves, the faster it perceives the signal variations in time.

**Quality of the Signal**

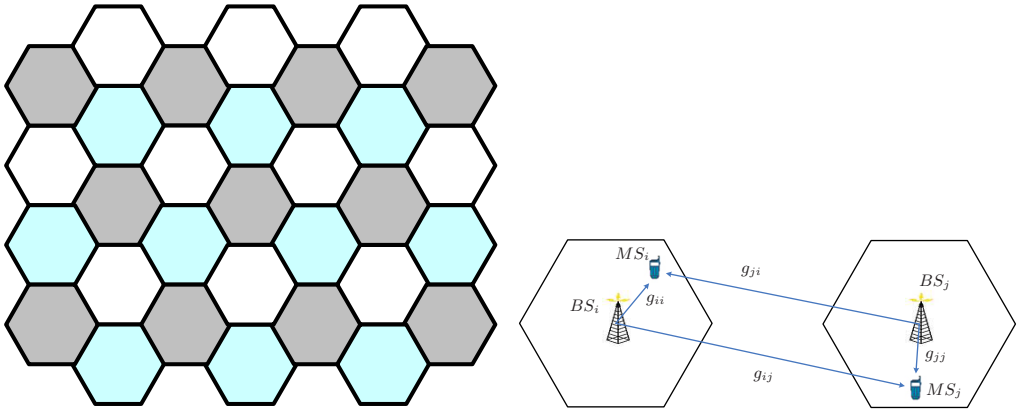
In cellular systems, adjacent cells are grouped into clusters of a suitable number $M$ of cells, as seen in figure 1(a) for $M = 3$. The entire set of available channels in the frequency pool of the system is divided into $M$ subsets of channels, and each subset is allocated to a cell in the cluster. This pattern is replicated over all clusters generating a concept called frequency reuse. Cells with the same subset of channels are called co-channel cells and cause co-channel interference, as conform illustrated in figure 1(b).

Therefore, due to the frequency reuse in the cellular systems, the propagation channels are subject to co-channel interference. The interference level affects the quality of the received signal. This influence can be measured through the Signal-to-Interference-plus-Noise Ratio (SINR). The SINR in the $ith$ link is defined as:

$$\gamma_i = \frac{g_{ii} \cdot p_i}{I_i}, \tag{4.3}$$

where $g_{ii}$, $p_i$ and $I_i$ are the power gain, transmission power and the co-channel interference plus noise in the $ith$ link, respectively. In turn, the co-channel interference plus noise in the $ith$ link is given by:

$$I_i = \sum_{j \neq i}^{M} g_{ij} \cdot p_j + \nu_i \tag{4.4}$$




(a) Frequency reuse. (b) Co-channel interference.

**Fig. 4.1.** Co-channel interference generated by the frequency reuse.

where $M$ is the number of co-channels links and $g_{ij}$ are the power gains between the Mobile Stations (MS) and the Base Stations (BS) of different links and $\nu_i$ is the noise power relative to $ith$ link. In some cases, the noise power is neglected, therefore, in this case we have a SIR (Signal-to-Interference Ratio).

It is obvious that the received signal quality depends on both the power gain and the level of co-channel interference. Power control is one of the techniques of Radio Resource Management which aims to control the quality of the signal. The values of power gains and interference serve as inputs for the formulation of algorithms, as it will be seen in further sections.

#### Classification of Power Control Algorithms

The power control algorithms presented in the next sections can be employed both in the forward link (BS to MS) and reverse link (MS to BS). They can be classified as centralized, distributed (decentralized) or semi-distributed algorithms. In centralized scheme, a centralized controller has all information about the established connections and power gains, and controls all the transmitted powers in the network [7]. Centralized power control requires extensive control signaling in the network and cannot be applied in practice. It can be used to give bounds on the performance of the decentralized algorithms. In a distributed algorithm, several decentralized controllers control only the power of one single transmitter, and the algorithm depends only on local information, such as measured the SINR or power gain of the specific user. Finally, in a semi-distributed algorithm, a compromise between the two first approaches is used, that is, control signaling among some controllers is allowed.

## 4.3 Centralized Power Control

The classical centralized power control algorithm was proposed in [8] by Zander. This algorithm assumes null noise power and therefore we consider SIRs instead of SINRs. Zander's algorithm has the objective of maximizing the minimum SIR of all co-channel links. It is possible to show that doing this is equivalent to balancing (equalizing) all SIRs of the co-channels links while maximizing this balanced SIR. This algorithm is centralized in the sense that in every moment a centralized controller has access to the entire power gain matrix and may instantaneously control all transmitted powers of the co-channel users. Therefore, the point is to find the maximum balanced SIR and the correspondent powers. In [8] this problem was solved through an inequality system and using the Perron-Frobenius Theorem [9] (see Theorem 1 below). In the following, we explain the development of the Zander algorithm.

The SIR of the $ith$ link can be expressed by:

$$\gamma_i = \frac{p_i}{\sum_{j=1}^{M} \frac{g_{ij}}{g_{ii}} \cdot p_j - p_i} = \frac{p_i}{\sum_{j=i}^{M} z_{ij} \cdot p_j - p_i}, \tag{4.5}$$

where $z_{ji} = \frac{g_{ij}}{g_{ii}}$. Let $\gamma_{min}$ be the lowest SIR of all co-channel links. Therefore, for each link $i$, we have:

$$\frac{p_i}{\sum_{j=i}^{M} z_{ij} \cdot p_j - p_i} \geq \gamma_{min}, \tag{4.6}$$

In the matrix form, this set of inequalities can be expressed as:

$$\frac{1+\gamma_{min}}{\gamma_{min}}\mathbf{p} \geq \mathbf{Zp}, \tag{4.7}$$

where $\mathbf{p} = [p_i]_{i=1,\ldots,M}$ and $\mathbf{Z} = [z_{ij}]_{i,j=1,\ldots,M}$ are the power vector and the normalized link matrix, respectively. The objective is to determine a positive power vector so that the inequality (4.7) is satisfied and $\gamma_{min}$ is maximized. The Perron-Frobenius Theorem is used to solve these problems concerning matrix inequalities. In the following, we state this theorem.

**Theorem 1.** *Given a nonnegative irreducible matrix **A**, then:*

- ***A** has exactly one real positive eigenvalue $\lambda^*$ for which the corresponding eigenvector is positive;*
- *The minimum real $\lambda$ such that the inequality $\lambda \cdot \boldsymbol{b} \geq \boldsymbol{Ab}$ which has solutions for $\boldsymbol{b} > 0$ is $\lambda = \lambda^*$;*
- *The maximum real $\lambda$ such that the inequality $\lambda \cdot \boldsymbol{b} \leq \boldsymbol{Ab}$ which has solutions for $\boldsymbol{b} > 0$ is $\lambda = \lambda^*$.*

Equation (4.7) has the same form than $\lambda \cdot \mathbf{b} \geq \mathbf{Ab}$, where:

$$\lambda = \frac{1+\gamma_{min}}{\gamma_{min}}, \tag{4.8}$$

$$\mathbf{A} = \mathbf{Z} \quad \text{and} \quad \mathbf{b} = \mathbf{p}. \tag{4.9}$$

A maximum value of $\gamma_{min}$ corresponds to a minimum value of $\lambda$. Therefore, according to the theorem, this value is $\lambda^*$ and the power vector is the corresponding eigenvector of the matrix $\mathbf{Z}$. Equation (4.8) can be rewritten substituting the minimum value $\lambda^*$:

$$\gamma_{min} = \frac{1}{\lambda^* - 1} \tag{4.10}$$

Therefore equation (4.13) gives the maximum value of the $\gamma_{min}$. However, what can we say about the other SIRs? Let's focus on the $\gamma_{max}$. For all links we can write:

$$\frac{p_i}{\sum_{j=i}^{M} z_{ij} \cdot p_j - p_i} \leq \gamma_{max}. \tag{4.11}$$

In the matrix form, we have:

$$\frac{1+\gamma_{max}}{\gamma_{max}}\mathbf{p} \leq \mathbf{Zp}. \tag{4.12}$$

According to Perron-Frobenius theorem (item 3), the solution correspondent to $\lambda^*$ and $\mathbf{p}^*$ yields a maximum value of $\frac{1+\gamma_{max}}{\gamma_{max}}$ which corresponds to a minimum value of $\gamma_{max}$ and this value is also given by:

$$\gamma_{max} = \frac{1}{\lambda^* - 1}. \tag{4.13}$$

As the solution given by eigenvector relative to eigenvalue $\lambda$ of the matrix $\mathbf{Z}$ yields in the same value for $\gamma_{min}$ and $\gamma_{max}$, we conclude that the choice of this power vector also balances the SIR of all co-channel links, while maximizing the minimum SIR.

## 4.4 Distributed Power Control

### 4.4.1 Signal-Level-Based Power Control

The Signal-Level-Based algorithm was proposed in [10]. This algorithm adjusts the transmission power $p_i(t)$ according to the power gain of the correspondent link $g_{ii}$. It does not take into account measurements of the SINR. Therefore, the algorithm can be expressed by a function of $g_{ii}$ in decibels:

$$p_{i,dB} = f(g_{ii,dB}) \tag{4.14}$$

where $p_{i,dB} = 10log_{10}(p_i)$ and $g_{ii,dB} = 10log_{10}(g_{ii})$.

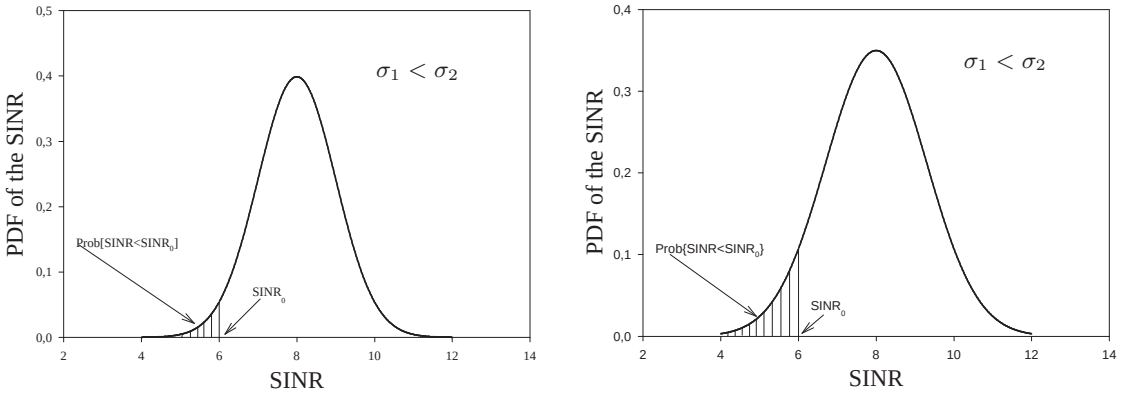
The principle of the algorithm is originated from the minimization process of the outage probability. The used minimization criterion is the Variance Minimum of the SINR. The outage probability is defined as the probability in which the SINR is below a threshold $\gamma_0$:

$$Prob[outage] = Prob[\gamma < \gamma_0] \tag{4.15}$$

Assuming that the average value of the SINR remains constant, a decreasing of the SINR variance will correspond to a decreasing of the outage probability of the system. This fact is explained qualitatively in figure 4.2. We can see that for two Probability Density Functions (PDFs) of the SINR, with same means and different standard deviations, $\sigma_1$ e $\sigma_2$, where $\sigma_1 < \sigma_2$, the PDF with lower standard deviation has lower outage probability. Therefore, the function $f(g_{ii,dB})$ in the equation (4.14) should be chosen so that the variance of the SINR is minimized. In [10], a function $f(g_{ii,dB})$ was determined which minimizes the variance of the SINR for the case of two co-channel links. Using variational analysis, the function $f(g_{ii,dB})$ which minimizes the variance of the SINR in dB is a linear function:

$$p_{i,dB} = \alpha \cdot g_{ii,dB} + \tau \tag{4.16}$$

where $\alpha$ and $\tau$ are constants. The constant $\alpha$ needs to be optimized, while $\tau$ depends on parameters of the system as the cell radio. In [10], it was shown that the optimum value $\alpha$ is $\frac{1}{2}$.




(a) Standard deviation $\sigma_1$.

(b) Standard deviation $\sigma_2$.

**Fig. 4.2.** A reduction of the SINR variance generates a reduction of the outage probability of the system.

### 4.4.2 Distributed Balancing Algorithm

Measuring all power gains, in real time, in a cellular system is impracticable, due to the signaling overheads. Therefore, a centralized power control is not desirable. Instead, we are interested in algorithms which make the power control the most independently possible of the power gains of other co-channel links. In a distributed implementation, the algorithm works only based on measurements of its own link. The algorithm assumes an iterative rule, unlike the centralized version, whose solution is given instantaneously, once the entire link matrix is known. In [11], Zander proposed a distributed version for his centralized algorithm, previously presented in section 4.3. The Zander's Distributed Power Control Algorithm is given by following iterative equation:

$$p_i(k+1) = \xi p_i(k) \left(1 + \frac{1}{\gamma_i(k)}\right) \tag{4.17}$$

where $\xi$ is a positive constant of proportionality which must be chosen adequately. This algorithm can be initialized with any positive vector $\mathbf{p}(0)$. This algorithm can be considered a distributed version of the centralized algorithm

presented in section 4.3, because the SIRs converge to the same value given by the normalized link matrix, repeated here for convenience:

$$\gamma = \frac{1}{\lambda^* - 1} \tag{4.18}$$

The power vector converges to a multiple of the eigenvector correspondent to $\lambda^*$. An important problem is that the powers are all increasing, since that the terms in equation (4.17) are positive. For the convergence of the SIRs this is not a problem, but in practical terms the powers must be kept in a limited level. The factor $\xi$ can be used to avoid the power increasing by adjusting it in each iteration $k$ according to:

$$\xi(k) = \frac{1}{|\mathbf{p}(k)|} \tag{4.19}$$

where $|\mathbf{p}(k)|$ is the norm of the power vector $\mathbf{p}(k)$.

However, one can notice that the use of equation (4.21) implies a loss of decentralization, because it would be necessary to have the signalization of the values of power in each co-channel link in order to calculate $|\mathbf{p}(k)|$.

### 4.4.3 Distributed Power Control Algorithm

Grandhi proposed in [12] an iterative algorithm similar to Zander's Distributed Power Control Algorithm, presented in section 4.4.2. Grandhi's algorithm can be expressed by:

$$p_i(k+1) = \varepsilon \frac{p_i(k)}{\gamma_i(k)} \tag{4.20}$$

where $\varepsilon$ is a positive constant of proportionality. The SIRs also converge for the value given by equation (4.18). The same problem concerning the convergence of the powers exists and can be solved through the adjustment of the factor $\varepsilon$:

$$\varepsilon(k) = \frac{1}{max(\mathbf{p}(k))} \tag{4.21}$$

where $max(\mathbf{p}(k))$ is the largest element of the power vector $\mathbf{p}(k)$. The adjustment of $\varepsilon$ would also require some coordination among the co-channel links and therefore, it would not be a completely distributed algorithm.

### 4.4.4 Autonomous SINR Balancing Algorithm

The algorithms presented so far are only based SIR (null noise power) and therefore are idealized algorithms. The Autonomous SINR Balancing Algorithm, proposed in [13] is designed to work in the presence of noise. It is originated from a differential equation, whose interpretation is the balancing of the SINRs in each link:

$$\gamma_i'(t) = -\beta[\gamma_i(t) - \gamma_t] \tag{4.22}$$

where $\gamma_t$ is the target SINR, $\gamma_i'(t)$ is the derivative of $\gamma_i(t)$ with respect to time and $\beta$ is a positive proportionality constant. According to the differential dynamics, the SINR evolves so that the SINR converges to the target SINR by a proportional amount to the offset between both. Therefore, this dynamics cannot stop unless $\gamma_i(t) = \gamma_t$. Different from the algorithms based on SIR whose resulting balanced SIR is dependent on the propagation conditions (link matrix), the Autonomous SINR Balancing Algorithm will make the SINRs converge to a pre-specific $\gamma_t$, since $\gamma_t$ is feasible. This aspect is approached in section 4.5.

Substituting equation (4.3) in equation (4.22), we have:

$$\left[\frac{g_i(t) \cdot p_i(t)}{I_i(t)}\right]' = -\beta[\gamma_i(t) - \gamma_t] \tag{4.23}$$

In a distributed implementation, the BS or MS can control only its own transmission power $p_i(t)$. The interference $I_i(t)$ and the power gain $g_i(t)$ cannot be controlled. Therefore, considering only the temporal variation of the transmission power $p_i(t)$ in equation (4.23), we have:

$$\frac{g_i(t)}{I_i(t)} \cdot p_i'(t) = -\beta[\gamma_i(t) - \gamma_t] \Rightarrow \tag{4.24}$$

$$p_i'(t) = -\frac{I_i(t)}{g_i(t)} \cdot \beta[\gamma_i(t) - \gamma_t] \tag{4.25}$$

Transforming the equation (4.25), in a difference equation, we obtain:

$$p_i(k+1) - p_i(k) = -\beta \cdot \frac{I_i(k) \cdot \gamma_i(k)}{g_i(k)} + \beta \cdot \frac{\gamma_t \cdot I_i(k)}{g_i(k)} \tag{4.26}$$

where $p_i(k)$, $g_i(k)$, $I_i(k)$ and $\gamma_i(k)$ are the transmission power, the power gain, the interference and the SINR in the iteration $k$, respectively.

Substituting $\frac{I_i(k)\cdot\gamma_i(k)}{g_i(k)}$ by $p_i(k)$ and $\frac{I_i(k)}{g_i(k)}$ by $\frac{p_i(k)}{\gamma_i(k)}$ and recombining the terms, we have finally the Autonomous SINR Balancing algorithm:

$$p_i(k+1) = p_i(k)\left[1 - \beta + \beta \cdot \frac{\gamma_t}{\gamma_i(k)}\right] \tag{4.27}$$

### 4.4.5 Unconstrained Second-Order Power Control Algorithm

The iterative algorithms proposed so far are classified as first-order algorithms in the sense that they depend on power levels of the current iteration in order to calculate the next power level. In [14], a second-order algorithm was proposed whose advantage is to produce gains in the convergence speed to the target SINR. Besides the current power, it requires the previous power level. This algorithm, called Unconstrained Second-Order Power Control (USOPC), is expressed by:

$$p_i(k+1) = \omega\frac{\gamma_t}{\gamma(k)}p_i(k) + (1-\omega)p_i(k-1), \tag{4.28}$$

where $\omega$ is the relaxation factor which must be chosen appropriately between 1 and 2. Of course, in this algorithm two initial power levels must be selected. In order to do this $p_i(0)$ can be made arbitrary while $p_i(1)$ is calculated doing $\omega = 1$ in this iteration. Note that, when $\omega = 1$, this algorithm is identical to the Autonomous SINR Balancing Algorithm with $\beta = 1$.

In order to improve the convergence, the relaxation factor can be updated in each iteration. In the simulations of [14], this new version of USOPC uses a non-stationary relaxation factor given by:

$$w(k) = 1 + \frac{1}{1.5^k}, \tag{4.29}$$

Observe that when $n$ increases, $w(k)$ tends to 1 and, therefore this algorithm behaves similarly to Autonomous SINR Balancing Algorithm with $\beta = 1$, along the iterations. Its main objective is to accelerate the converge speed of the power control.

### 4.4.6 Up-Down Algorithm

A very simple power control algorithm is the up-down algorithm also known as the fixed-step power control algorithm. At each iteration, this algorithm adjusts the transmitted power upwards or downwards by one step or keeps the power constant. The choice of this action is done based on the comparison between the actual SINR and a target SINR $\gamma_t$. The algorithm will attempt to reach a target SINR $\gamma_t$. Mathematically we can describe the algorithm in decibel (dB) scale as:

$$p_i(k+1) = \begin{cases} p_i(k) + \delta \quad [dB] & \text{if } \gamma_i < \gamma_t - \delta; \\ p_i(k) - \delta \quad [dB] & \text{if } \gamma_i > \gamma_t + \delta; \\ p_i(k) \quad [dB] & \text{otherwise .} \end{cases} \tag{4.30}$$

It can be shown that the algorithm converges once it exists a feasible solution. Practical systems such as UMTS/WCDMA (Universal Mobile Telecommunication System/Wideband Code Division Multiple Access) use this algorithm with $\delta = 1\ dB$ [15].

## 4.5 Feasibility and Convergence Aspects

In this section, we choose the Autonomous SINR Balancing Algorithm in order to explain some aspects concerning feasibility and convergence of the distributed power control.

By using the Autonomous SINR Balancing Algorithm, two questions arise. The Autonomous SINR Balancing Algorithm will get to balance the system for any target SINR $\gamma_t$? Which value of $\beta$ yields a higher convergence speed?

It is possible to show that the convergence speed is maximized when $\beta = 1$ [13]. Therefore, substituting $\beta = 1$ in equation (4.27), the Autonomous SINR Balancing Algorithm becomes:

$$p(k+1) = p(k) \cdot \frac{\rho_t}{\rho(k)} \tag{4.31}$$

Let us demonstrate the optimality of $\beta = 1$ and to establish criteria of feasibility of the power control for the case of two co-channel links, as shown in figure 4.3. The direction of communication is the reverse link, but the demonstration for the forward link is analogous.

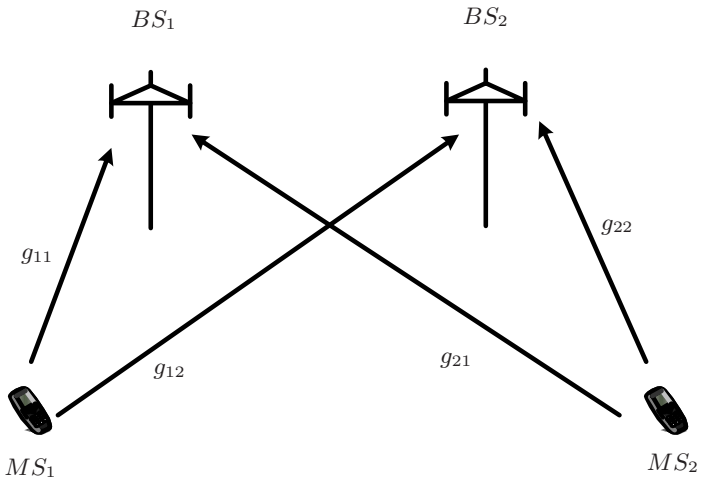



**Fig. 4.3.** Model for two co-channel links.

Firstly, let us establish a necessary and sufficient condition to check if a given target SINR is feasible. A target SINR $\gamma_t$ is feasible when the equation system has a positive solution in the variables $p_1$ and $p_2$:

$$\gamma_t = \frac{g_{11} \cdot p_1}{g_{21} \cdot p_2 + \nu_1} = \frac{g_{22} \cdot p_2}{g_{12} \cdot p_1 + \nu_2} \tag{4.32}$$

Putting the equation system in a matrix format, we have

$$\mathbf{B} \cdot \mathbf{p} = \mathbf{n} \tag{4.33}$$

where:

$\mathbf{B} = \begin{bmatrix} 1 & -\gamma_t \frac{g_{21}}{g_{11}} \\ -\gamma_t \frac{g_{12}}{g_{22}} & 1 \end{bmatrix}$,

$\mathbf{p} = \begin{bmatrix} p_1 \\ p_2 \end{bmatrix}$,

$\mathbf{n} = \begin{bmatrix} \gamma_t \frac{\nu_1}{g_{11}} \\ \gamma_t \frac{\nu_2}{g_{22}} \end{bmatrix}$.

Solving the equation system (4.33), we have:

$$p_1 = \frac{\rho_t}{1 - (\gamma_t \sigma)^2} \left( \frac{\nu_1 g_{22} + \nu_2 \gamma_t g_{21}}{g_{11} g_{22}} \right) \tag{4.34}$$

$$p_1 = \frac{\rho_t}{1-(\gamma_t\sigma)^2}\left(\frac{\nu_2 g_{11} + \nu_1\gamma_t g_{12}}{g_{11}g_{22}}\right) \tag{4.35}$$

where $\sigma = \left(\frac{g_{12}g_{21}}{g_{11}g_{22}}\right)^{1/2}$.

Therefore, the system will have positive solution when $\rho_t\sigma < 1$. Note that $\sigma$ is the absolute value of the eigenvalue of higher absolute value of the matrix $\mathbf{C} = \frac{1}{\gamma_t}(\mathbf{I}-\mathbf{B})$, where $\mathbf{I}$ is the identity matrix. This criterion of feasibility of the target SINR is also valid to *M* co-channel links [16].

Once the criterion $\gamma_t\sigma < 1$ is satisfied, let us prove that doing $\beta = 1$ in the equation (4.27), the convergence speed is maximized. Writing the equation (4.27) in the matrix form, we obtain:

$$\mathbf{p}(k+1) = \mathbf{D}\cdot\mathbf{p}(k) + \beta\mathbf{n} \tag{4.36}$$

where $\mathbf{D} = (\mathbf{I} - \beta\mathbf{B})$.

Therefore, along the iterations, we have:

$$\mathbf{p}(k) = (\mathbf{I} + \mathbf{D} + \mathbf{D}^2 + ... + \mathbf{D}^{k-2})\beta\mathbf{n} + \mathbf{D}^{k-1}\mathbf{p}(1) \tag{4.37}$$

If the absolute values of all eigenvalues of $\mathbf{D}$ are lower than 1, the series $(\mathbf{I} + \mathbf{D} + \mathbf{D}^2 + ... + \mathbf{D}^{k-2})$ will converge. The lower the absolute value of the highest absolute value eigenvalue, the higher it will be the convergence speed. Moreover, the series will converge to $(\mathbf{I}-\mathbf{D})^{-1}$ [17]. Therefore, the optimum value of $\beta$ is the one which minimizes the absolute value of the eigenvalue with highest absolute value of the matrix $\mathbf{D}$. Thus, once the absolute value of each eigenvalue of $\mathbf{D}$ are strictly lower than 1, we can affirm that:

$$\lim_{k\to\infty}\mathbf{p}(k) = (\mathbf{I}-\mathbf{D})^{-1}\beta\mathbf{n} = (\beta\mathbf{B})^{-1}\beta\mathbf{n} = \mathbf{B}^{-1}\mathbf{n} = \mathbf{p} \tag{4.38}$$

According to equation (4.38), the powers will converge to the values given by the solution of matrix equation (4.33), therefore the SINR will be balanced.

The eigenvalues of the matrix $\mathbf{D}$ for the case of two co-channel links are: $\lambda_1 = (1-\beta) + \gamma_t\beta\gamma$ and $\lambda_2 = (1-\beta) - \gamma_t\beta\sigma$. Therefore, the choice of $\beta$ which yields in a fastest convergence is $\beta = 1$. Observe that this value of $\beta$ makes $|\lambda_1| = |\lambda_2| = \gamma_t\sigma < 1$. This can be seen in the figure 4.4.

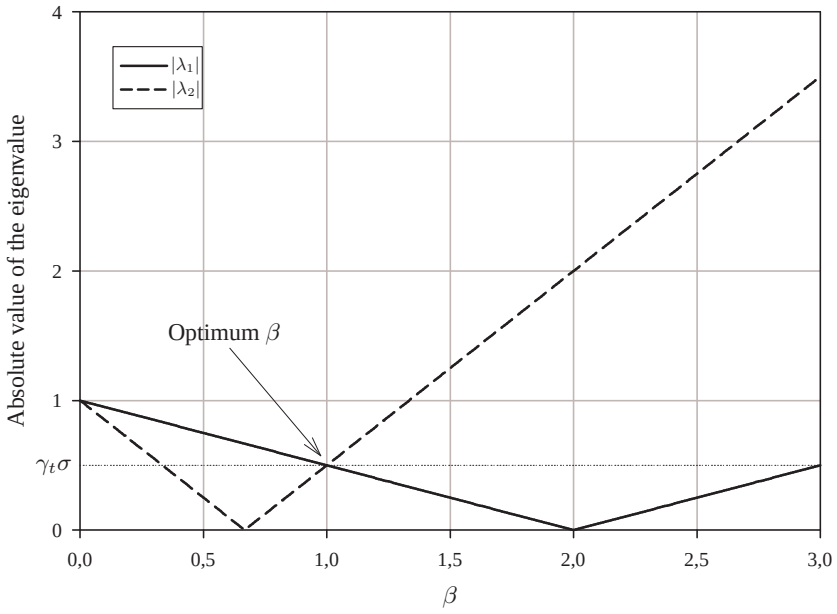



**Fig. 4.4.** Optimum value of $\beta$.

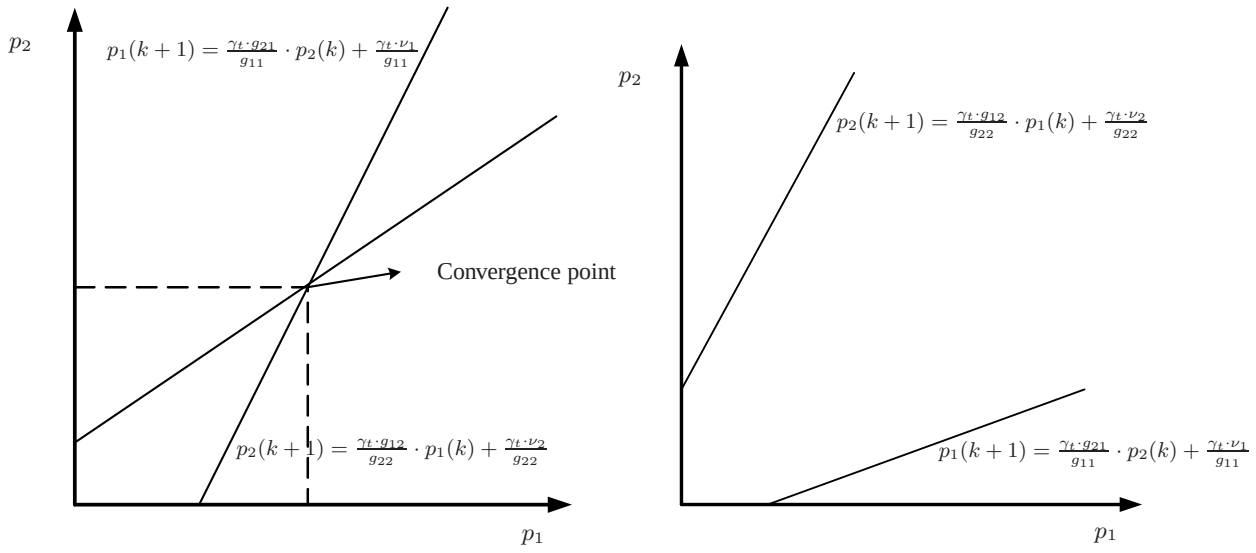
**Illustrative Example**

The following figures show illustrative examples of the (in)feasibility of the target SINR using the Autonomous SINR Balancing Algorithm with $\beta = 1$ for the case of two co-channel links. From equations (4.31), (4.3) and (4.4) it is easy to see that each transmitter will control its own power according to the equations:

$$p_1(k+1) = \frac{\gamma_t \cdot g_{21}}{g_{11}} \cdot p_2(k) + \frac{\gamma_t \cdot \nu_1}{g_{11}} \tag{4.39}$$

$$p_2(k+1) = \frac{\gamma_t \cdot g_{12}}{g_{22}} \cdot p_1(k) + \frac{\gamma_t \cdot \nu_2}{g_{22}} \tag{4.40}$$

Equations (4.39) and (4.40) represent two lines and the target SINR $\gamma_t$ will be feasible if both itself cross in the first quadrant. This will happen when $\gamma_t \sigma < 1$. In this case, the power control algorithm will get to balance the SINRs. This scenario can be seen in figure 5(a).




(a) Feasible Target SINR.

(b) Non feasible target SINR.

**Fig. 4.5.** Scenarios of feasibility of the target SINR.

Observing equations (4.39) and (4.40), we can see that the angular coefficients of the lines are influenced by three factors:

1. The target SINR $\gamma_t$;
2. The power gain $g_{ii}$ of the signal;
3. The power gain $g_{ji}$ of interfering signal.

Three situations make the angular coefficient of the lines increase, making the crossing point of figure 5(a) occur in higher powers: the target SINR $\gamma_t$ is increased or the power gain $g_{ii}$ of the signal is decreased or the power gain $g_{ji}$ of interfering signal is increased. The crossing of the lines might not even happen in the first quadrant. This situation can be seen in figure 5(b). This scenario happens when the criterion $\gamma_t \sigma < 1$ is not satisfied. In this case of infeasibility the powers would increase indefinitely and even so the power control algorithm would not balance the SINRs to the value $\gamma_t$.

## 4.6 Semi-distributed Power Control Algorithms

The algorithm presented in section 4.4.5 has the advantage of increasing the convergence speed of the algorithm. Another way of doing this is to employ semi-distributed algorithms. In [18], it was proposed a new semi-distributed

power control algorithm. This algorithm is based on the assumption that some limited control data communication between the co-channel bases is allowed. The algorithm is semi-distributed in the sense that it is run by each base station based on its local information and information sent by some of its neighbor bases. Of course, such information exchanges generate some overhead. However, the trade-off is that the new algorithm does not have any normalizing factor, as in the power control algorithms presented in sections 4.4.2 and 4.4.3 and then, no global communication is needed. Moreover, it has a nice convergence property in the sense that the minimum of the signal-to-interference ratios of all the interfering receivers converge monotonically upward to a limit.

Before the application of the algorithm, it is necessary to define the concept of network neighborhood. The network neighborhood is defined through a control data flow structure which is represented as a directed graph. This directed graph defines a static network topology where the base stations are represented by the nodes and the existence of a directed edge from base A to base B indicates that control data can be passed from A to B. A directed edge can be implemented in the practice by a microwave link or an optics fiber cable. Of course, there is a cost associated with this communication among the bases and it is desirable to minimize the cost by decreasing the number of edges of the structure.

Once that the control data flow structure is defined, the algorithm can actuate according to the following rule:

$$p_i(k+1) = \eta_i(k) p_i(k), \tag{4.41}$$

with

$$p_i(0) = p_{max} \tag{4.42}$$

and

$$\eta_i(k) = \sqrt[q]{\frac{min(\gamma_i(k), max(min_{j \in N_i} \gamma_j(k), \gamma_0))}{\gamma_i(k)}}. \tag{4.43}$$

where $N_i$ is the set of indices of base stations that send control data information to base station $i$ according to the control data flow structure, $\gamma_0$ is the minimum SIR ratio for transmission with acceptable quality, $p_{max}$ is the maximum transmission power level and $q$ is a parameter of the algorithm that control the rate of convergence. Note that the communication between the bases is made through of the coefficients $\eta_i$.

Concerning to the convergence properties, [18] states that the final SIRs depends on the maximum achievable SIR $\gamma^*$, which is given by equation (4.18). If $\gamma^* \geq \gamma_0$, then the algorithm converges to the balanced solution, that is $\gamma_i = \gamma^*$ for all base stations. If $\gamma^* < \gamma_0$, then $\gamma_i \leq \gamma_0$ for all base stations, with the strict inequality holding for at least one base station. Moreover, if the initial SIR of the $ith$ base station is larger than or equal to $\gamma_0$, then $\gamma_i = \gamma_0$.

In [19], a variation of this algorithm is presented. This new version consists of the power of each link starts from the minimum allowed power and is adjusted monotonically upward. This modification has two advantages. First, it saves user battery life. Second, the admission of a new user minimizes any sudden increase in the interference to the existing users. The disturbance to a balanced system is thus minimized.

## 4.7 Techniques of Transmitter Removal

When the system becomes congested or the propagation conditions are unfavorable, the probability of an infeasible power control solution increases. An infeasible power control solution occurs when the balanced SIR is lower than a minimum SIR $\gamma_0$ for transmission with acceptable quality. A countermeasure for dealing with infeasible situations is to remove links from the set of co-channel links. Removed links may be handed off to other channels and in an extreme case dropped.

### 4.7.1 Stepwise Removal Algorithm

In [8] it was proposed a simple procedure, called the Stepwise Removal Algorithm (SRA), for practical implementations. The SRA algorithm, one by one, removes links until the required SIR is achieved in the remaining links. This method has the advantage of requiring substantially low computational effort. SRA consists of two steps:

**Step 1**: Determine the feasible maximum SIR $\gamma^*$ from link matrix **Z** given by equation (4.18). If $\gamma^* \geq \gamma_0$ (power control is feasible), utilize the eigenvector correspondent to the eigenvalue $\lambda^*$ (theorem 1) as power vector and stop. If $\gamma^* < \gamma_0$ (power control is infeasible), execute step 2.

**Step 2**: Remove the link $k$ for which the sum of its row and column in the link matrix **Z**

$$\sum_{i=1}^{M} Z_{ki} + Z_{ik} \tag{4.44}$$

is maximized and thus forming a new square sub-matrix **Z'** with dimension $(M-1)$. Determine the new $\gamma^*$ corresponding to **Z'**. If $\gamma^* \geq \gamma_0$ (power control is feasible), utilize the corresponding eigenvector as power vector, else repeat step 2 until $\gamma^*$ becomes larger than $\gamma_0$.

Although one may think the link removal scheme is an extreme way to improve the performance of the system, we should remember that those links which do not reach $\gamma_0$ are useless while causing interference to other links. Therefore, the removal of a bad link is an effective strategy.

### 4.7.2 Stepwise Maximum Interference Removal Algorithm

In [20], it was proposed an algorithm which outperforms the SRA. This algorithm was called Stepwise Maximum-Interference Removal Algorithm (SMIRA). Different from SRA, SMIRA considers transmitting power to be an important factor in removing links, in addition to power gains of the link matrix. The idea is that the larger the transmitting power, the greater the interference it causes to MSs in the other cells. Therefore, if a link which uses a high transmitting power is removed, then it is likely that the remaining links can achieve the minimum acceptable SIR $\gamma_0$. The SMIRA is composed of two steps:

**Step 1**: Step 1 is identical to step 1 from SRA, which is determining the feasible maximum SIR $\gamma^*$ from link matrix **Z** given by equation (4.18). If $\gamma^* \geq \gamma_0$ (power control is feasible), utilize the eigenvector correspondent to the eigenvalue $\lambda^*$ (theorem 1) as power vector and stop. If $\gamma^* < \gamma_0$ (power control is infeasible), execute step 2.

**Step 2**: Calculate:

$$I_k^T = \sum_{i \neq k}^{M} Z_{ki} p_k \tag{4.45}$$

$$I_k^R = \sum_{i \neq k}^{M} Z_{ik} p_i \tag{4.46}$$

$$I_k^{max} = max(I_k^T, I_k^R) \tag{4.47}$$

and remove the link $k$ for which $I_k^{max}$ is maximum. The powers $p_k$ are the elements of eigenvector correspondent to the eigenvalue $\lambda^*$ of the link matrix **Z**. With this new link matrix **Z'**, determine the new $\gamma^*$. If $\gamma^* \geq \gamma_0$, then use the corresponding eigenvector and stop; otherwise repeat step 2.

Observe that the term "maximum interference" is used because $I_k^T$ represents the total interference caused by the mobile in cell $k$ to other cells and $I_k^R$ represents the total received interference in the cell $k$ originated from other cells. In terms of numeric complexity, the SMIRA algorithm requires more multiplications than the SRA algorithm each time a cell is removed. However, the computational complexity is dominated by finding the eigenvalue and eigenvector of the link matrix and thus these additional multiplications are negligible.

SMIRA has two variations: Stepwise Maximum Received Interference Removal Algorithm (SMRIRA) and Stepwise Maximum Transmitted Interference Removal Algorithm (SMTIRA). The difference from SMIRA procedure is the fact that the former considers $I_k^{max} = I_k^R$, while the second considers $I_k^{max} = I_k^T$.

Simulation results in [20] shows that SMIRA presents the best performance. The performances of SMRIRA and SMTIRA are similar and are better than SRA. However, SMITIRA may become worse than the SRA for high values of $\gamma_0$.

## 4.8 Techniques of SINR-target decreasing

Another direction for dealing with infeasible situations is to decrease the target SINR. In [21], instead of considering a fixed value to target SINR, a variable one is used. The variable target SINR of each transmitter depends on its transmitted power. The idea is that the more transmitted power is necessary to obtain a target SINR, the worse the propagation condition of this link is. As a consequence, this transmitter will be causing excessive interference to other co-channel links. Therefore, a good solution would be to decrease the target SINR of this link. On the other hand,

users who utilize little power to obtain its target SINR have conditions of reaching a higher value. With this in mind, [21] proposed the following equation (in dB) for the adjustment of the target SINR:

$$\gamma_t(k+1) = m \cdot p_i(k+1) + n \qquad [dB], \tag{4.48}$$

where $m$ is a negative constant and $n$ is a control parameter, both in linear scale. Considering only the effect of the power adjustment between two consecutive iterations, we have:

$$\gamma_i(k+1) = \gamma_i(k) - p_i(k) + p_i(k+1) \qquad [dB], \tag{4.49}$$

Substituting equation (4.48) in equation (4.49) yields in:

$$m \cdot p_i(k+1) + n = \gamma_i(k) - p_i(k) + p_i(k+1) \qquad [dB], \tag{4.50}$$

Rewriting equation 4.50, the strategy of power control will be given by the following algorithm:

$$p_i(k+1) = \frac{n + p_i(k) - \gamma_i(k)}{1-m} \qquad [dBm], \tag{4.51}$$

or

$$p_i(k+1) = \phi - \varphi(\gamma_i(k) - p_i(k)) \qquad [dBm], \tag{4.52}$$

where

$$\phi = \frac{n}{1-m} \quad \text{and} \quad \varphi = \frac{1}{1-m}, \tag{4.53}$$

The parameter $\phi$ controls the mean value of the transmitted power distribution and $\varphi$ controls the system's spread in transmitted power and SINR. In [21], it was shown through simulation that the optimum value for $\varphi$ is around 0.7, in terms of $10th$ percentile of the SINR.

In [22], a similar power-to-target SINR mapping is proposed including inferior and superior bounds of target SINR and a restriction of maximum transmitted power. Figure 4.6 illustrates this scheme. For $p_i \leq \underline{p_i}$, the transmitter $i$ attempts to keep a high quality connection by aiming for the maximum target SINR $\overline{\gamma}_i$. For $\overline{p}_i \leq p_i \leq p_{max}$, the transmitter $i$ aims for a minimum target $\underline{\gamma_i}$. Observe that $\overline{\gamma}_i \geq \underline{\gamma_i} \geq \gamma_0$, where $\gamma_0$ is the minimum acceptable threshold. For $\underline{p_i} \leq p_i \leq \overline{p}_i$, the transmitter $i$ aims for a target SINR that is given by a linear function of $p_i$ in dB. Note that the points $(\underline{p_i}, \overline{\gamma}_i)$ and $(\overline{p}_i, \underline{\gamma_i})$ determine uniquely this linear function.

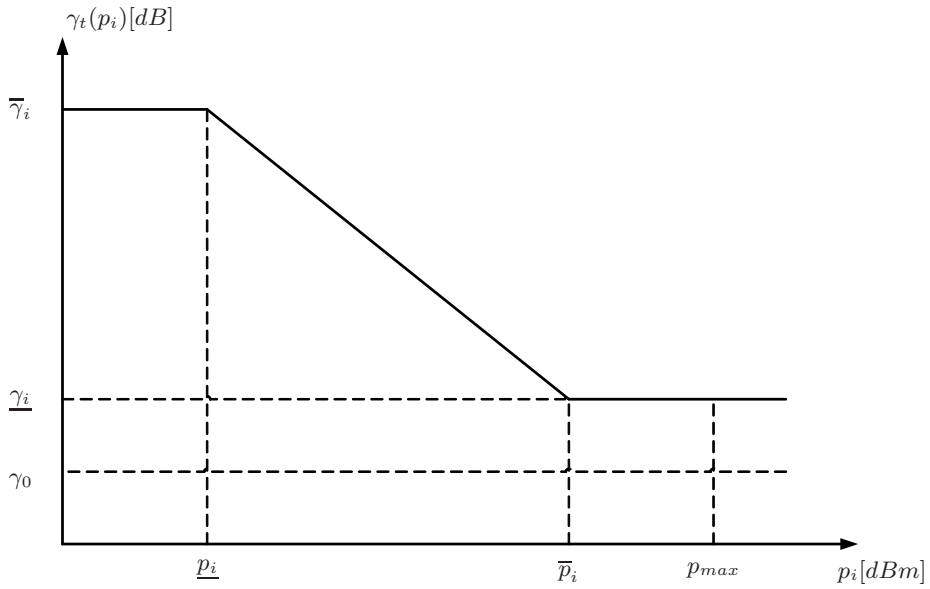



**Fig. 4.6.** Mapping of the target SINR.

## 4.9 Practical aspects of power control

### 4.9.1 Constrained Power Control Algorithms

So far, the presented algorithms can transmit without power limitation. However, this assumption is not realistic, since the maximum power of a transmitter is upper-bounded for implementation purposes. For the mobile stations, this limitation is crucial, because the battery energy is scarce. Moreover, the introduction of limits for the maximum power level restricts the impact that a single non-supported link may have over the other co-channel links. In order to consider this power limitation, [12] and [23] introduced a constraint in the algorithms given by:

$$0 \leq p_i \leq p_{max}, \tag{4.54}$$

where $p_{max}$ is the allowed maximum transmission power. Therefore, if the transmission power calculated by a power control algorithm is greater than $p_{max}$, the algorithm truncates the power transmission in $p_{max}$. A transmitter which suffers this limitation will not reach its target SINR, but this best-effort solution is better than holding a non-supported link which would cause all other link powers to increase indefinitely.

### 4.9.2 Power Control over a Discrete Power Domain

Most studies about power control assume a continuous range of transmit power levels. In practice, however, power levels are assigned from a discrete set. The convergence of most power control algorithms depends on power levels that can be controlled in a continuous domain and so the impact of discretization needs to be analyzed.

In [24] a simple method is proposed taking the "ceiling" value of the power given by the continuous algorithm. For a power value $x$, let the ceiling value $\lceil x \rceil$ be the smallest value in the discrete set of power values which is larger than or equals $x$.

For an algorithm based on received power like signal-level-based power control in section 4.4.1 is obvious that a higher transmit power will result in higher received power. So using ceiling value is a straightforward solution.

For algorithms based on a target SIR, theorem 3.1 and 3.2 in [24] grants that if $0 \leq \mathbf{p}(0) \leq \mathbf{p}^*$, where $\mathbf{p}^*$ is the optimal discrete power vector, then $\mathbf{p}(k)$ converges to $\mathbf{p}^*$.

In [24] it is stated that the ceiling approach may result in an oscillating SIR, which results in a poorer link quality and in a higher outage probability. To mitigate this problem, [25] proposes a solution that alleviate the oscillation but has a lower rate of convergence. We exploit this solution in the next paragraphs.

Consider that $p_i(k)$ is the current discrete-domain transmission power for the mobile $i$ and that $p'_i(k+1)$ is the power determined by the continuous algorithm for the next power update interval.

The algorithm in [25] is defined as

$$p_i(k+1) = \begin{cases} \lceil p'_i(k+1) \rceil & \text{if } p'_i(k+1) > p_i(k) \\ \lceil \alpha p_i(k) + (1-\alpha) p'_i(k+1) \rceil & \text{if } p'_i(k+1) \leq p_i(k) \end{cases} \tag{4.55}$$

By analyzing (4.55) we conclude that the algorithm reduces the speed of the adjustment of power when the power is increasing and permits the adjustment at full speed when the power is decreasing. That way, with a proper $\alpha$, the oscillation is reduced with little impact on the convergence of the algorithm.

### 4.9.3 Impact of Time Delays in Power Control

Closed-loop power control is the key to compensate the variations in power gains and multiple access interference. The current state of the channel, characterized by measurements (SIR, Received power, etc), is used by the algorithm to determine the transmit power in the next power control command.

It is common to control each link individually based only on local information. An outer loop provides an inner loop with a target value and the inner loop determines transmission powers or issues transmission power control commands to meet this target.

The issued power control commands use the measurement $m_i(k)$ together with corresponding target values $m_i^t(k)$ from an outer loop in the control mechanism $R_i$ to issue power control commands $s_i(k)$. These commands are sent over the radio interface. The control commands are decoded by the device $D_i$ on the transmitter side into updated transmission powers $p_i(k+1)$.

$$s_i(k) = R_i\{m_i^t(k), m_i(k)\} \tag{4.56}$$

$$\check{p}_i(k+1) = D_i\{s_i(k)\} \tag{4.57}$$

Assuming that the measurement $m_i(k)$ is a function of $p_i(k)$ given by

$$m_t(k) = p_i(k) + \beta_i(k) \qquad [dB], \tag{4.58}$$

This works well under ideal assumptions, but in real situations, time delays in the control loops reduce the performance. These delays are typically round-trip delays in the inner loop. As concluded in [26], the delays result in oscillating and possibly unstable systems.

The time delay compensation (TDC) proposed in [26] can be used to compensate for the delays. This technique can be used together with a wide range of power control algorithms. The main idea of this algorithm is consider issued commands that have not yet taken effect together with the current measurements.

Assume that issued powers or power control commands are delayed by $n_p$ samples before they take effect and a measurement is used in the control algorithm $n_m$ samples after it is collected (for example, if the measurement is sent over the radio interface). The power control algorithm itself includes a processing delay of one sample. In total, the round-trip delay is thus $n_p + n_m + 1$ power update intervals. The measurement delay $n_m$ is normally short and can be seen as part of the control algorithm processing delay $n_p$.

The actual measurement will be

$$m_i(k) = p_i(k - n_p) + \beta_i(k) \qquad [dB]. \tag{4.59}$$

However, to compensate the delay in the issued commands of the algorithm we will need the measurement when the power will be in fact used $m_i(k + n_p)$. Using (4.59) we arrive to

$$m_i(k + n_p) - m_i(k) = p_i(k) - p_i(k - n_p) + \beta_i(k - n_p) - \beta_i(k) \qquad [dB]. \tag{4.60}$$

that considering $\beta_i(k - n_p) = \beta_i(k)$ takes to the simple estimative

$$m_i(k + n_p) = m_i(k) + p_i(k) - p_i(k - n_p - n_m). \tag{4.61}$$

In practice $p_i(k)$ is not available and the device $D_i$ must be considered to determine $p_i(k)$.

Applying the TDC to the autonomous SINR balancing algorithm in section 4.4.4, for an example, with $n_p = 1$ we get

$$\gamma_i(k + 1) = \gamma_i(k) + p_i(k) - p_i(k - 1) \tag{4.62}$$
$$p_i(k + 1) = p_i(k) + (\gamma_i^t(k) - \gamma_i(k + 1)). \tag{4.63}$$

## 4.10 Power Control in WCDMA

The power control in WCDMA is based on a paradigm of three loops: Open loop, Closed loop and Outer loop. These three loops work jointly to determine the transmission power.

The open loop determine the initial transmission power. The propose of this loop is compensate for the shadowing in the first transmission. The propose of the closed loop is compensate for the fast fading. Measurements of SINR are made in the receiver and compared with a target value. After this comparison a command to increase or decrease the power is sent to the transmitter. This loop is capable to track the fast fading because of its high frequency.

The outer loop role is to compensate for errors in SINR measurements and changes in the mobility profile of the mobiles. It adjusts the target SINR of the closed loop based on measurements of the transmission errors.

The Fig. 4.7 illustrates the operation of the three loops.

## 4.11 Power Control for Multirate Systems

Modern communication systems provide different services with different transmission rates and error requirements. The transmission rate is closely related to SINR, and the SINRs can be efficiently controlled by power control. Therefore, the combined rate and power control is an interesting problem that is highlighted when multiple services are offered with different transmission rates.

Consider that the rate $r_i(t)$ is limited by

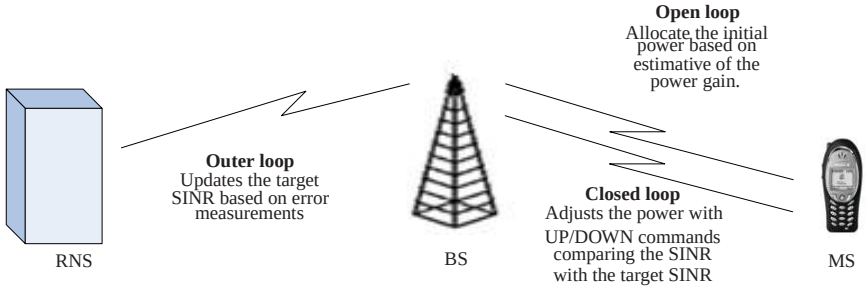



**Fig. 4.7.** Operation of the three WCDMA power control loops.

$$r_i(t) \leq f(\gamma_i(t)) \tag{4.64}$$

where $f$ is a monotonically increasing function. If we use the Shannon limit, for example, the relation between rate and SINR becomes

$$f(\gamma) = c\log(1+\gamma) \tag{4.65}$$

We will assume that the transmit power is limited and the maximum transmit power is given by the $\bar{\mathbf{p}} = (\bar{p}_1, \bar{p}_2, \ldots, \bar{p}_N)$

Define a rate vector $\mathbf{r}(\bar{\mathbf{p}}) = (r_1, r_2, \ldots, r_N)$ as instantaneously achievable if there is a positive power vector

$$\mathbf{p} = (p_1, p_2, \ldots, p_N) \leq \bar{\mathbf{p}} \tag{4.66}$$

such that

$$r_i \leq f(\gamma_i(\mathbf{p})), \forall i \tag{4.67}$$

Define a rate vector $\mathbf{r}^*(\bar{p}) = (r_1^*, r_2^*, \ldots, r_N^*)$ as achievable in the average sense if it may be expressed as

$$\mathbf{r}^* = \sum_k \alpha_k \mathbf{r}_k \tag{4.68}$$

where

$$\alpha_k \epsilon [0,1], \sum_k \alpha_k = 1 \tag{4.69}$$

where all $\mathbf{r}_k$ are instantaneously achievable rate vectors.

Assume that each link $i$ requires a minimum data rate, denoted $r_{i,min}$. Then assume that each user desires to achieve the maximum possible data rate. Then, if the service is delay-insensitive, we can consider the optimization problem

$$max \sum_{i=1}^{N} r_i^*(\bar{\mathbf{p}}) \tag{4.70}$$

subject to

$$r_i^*(\bar{\mathbf{p}}) \geq r_{i,min}, \forall i \tag{4.71}$$

Using (4.71) and (4.67) we arrive to a region of instantaneous achievable rate that, with (4.68), can be used to obtain the region of average achievable rate.

If a set of rate vector $\mathbf{r}_k$ is instantaneously achievable, it is possible to switch between rate vector, using each of them during the fraction of time $\alpha_k$ and yielding the average rate $\mathbf{r}^*(\bar{\mathbf{p}}) = (r_1^*, r_2^*, \ldots, r_N^*)$. This is important because sometimes the instantaneous achievable rate region is not convex but the average achievable rate region always is.

In [27] it is defined a class of algorithms called opportunistic algorithms. This class of algorithm uses the variability of the power gains to increase the system capacity increasing the transmission rate for links with high gains.

It proposes a power update using an iterative function

$$p(k+1) = I(p(k)) \tag{4.72}$$

where $I(p) = (I_1(p), I_2(p), \ldots, I_N(p))$.

It defines that $I(p)$ is *two-sided scalable* if

$$\left(\frac{1}{\alpha}\right) p < p' < \alpha p \quad \Rightarrow \left(\frac{1}{\alpha}\right) I(p) < I(p') < \alpha I(p) \quad \forall\, \alpha > 1 \tag{4.73}$$

The theorems that follows shows the advantage of the use of a two-sided scalable iterative function.

**Theorem 2.** *If $I(p)$ is two-sided scalable and a fixed point exists, then that fixed point is unique.*

*Proof.* Refer to [27]

**Theorem 3.** *If $I(p)$ is two-sided scalable and a fixed point $p^*$ exists, then the power vector $p(t)$ converges to $p^*$.*

*Proof.* Refer to [27]

**Theorem 4.** *Given a two-sided scalable iterative function $I(p)$ , if there exists $l, u > 0$ such that $l \leq I(p) \leq u$ for all $p$, then a fixed point exists.*

*Proof.* Refer to [27]

In some services, simply maximizing the aggregate rate like in (4.70) is not sufficient to achieve the QoS requirements. Other requirements like maximum delay must be taken into account to maintain the users satisfaction.

## 4.12 Power Control Games

Wireless communications may be generally characterized by multiple communication nodes and a limited amount of radio resources, which must be shared among the nodes. In order to assure Quality of Service (QoS) requirements, power control techniques are used. Moreover, power control is essential concerning the energy efficiency, since communications nodes using low power levels means longer lifetime of batteries for mobile terminals and more energy resources available for central nodes as base stations in cellular systems.

Distributed or decentralized power control is of special interest and importance, since it allows the use of only local information for determining a suitable transmit power [13, 21]. The decentralized power control problem may be characterized as a competition between the communication nodes. The power resources must be efficiently allocated by self-optimization, since the transmit powers of all nodes can not be jointly determined by a central controller. This indicates the adequacy of the application of noncooperative game theory concepts to the problem.

Game theory is a mathematical tool for analyzing the interaction of decision makers with conflicting objectives and for the formulation of decision strategies. The potential of game-theoretical concepts began to be explored in the distributed power control problem recently, as discussed in [28]. CDMA-like cellular systems, where power control has a critical role, have been the preferred environment of studies, with the focus mostly in best effort (or elastic) services.

### 4.12.1 Noncooperative $N$-Person Games

A game has three basic elements: a set of players, a set of possible actions for each player, and a set of objective functions mapping action profiles into real numbers. Games which involve $N$ players are called $N$-Person Games. When the players are not allowed to negotiate, i.e., they have to make decisions based only on their own objective and information, it is configured a noncooperative game. Detailed discussions about game theory can be found in [29, 30, 31].

In distributed power control games, a general model may be used to represent the wireless system as a set $\Omega$ of *N* radio links (transmitter/receiver pairs), $\Omega = \{1, ..., N\}$, where each link is affected by interference caused by all other links. In this model, the transmitters constitute the set of players and each player $j$, $j \in \Omega$, has as decision variable its own transmit power $p_j$.

Actions or decisions of players are confined to their strategy space, the set of feasible transmit power levels $P_j = [p_{j\,min}, p_{j\,max}]$. It is usually adopted a unique set of feasible transmit power for all players: $P_j = [p_{min}, p_{max}]$.

Power gain seen by the *j*th receiver is represented by $g_j$, while the correspondent interference-plus-noise power is $I_j$. Then, Signal-to-Interference-plus-Noise Ratio (SINR) perceived by *j*th receiver, $\gamma_j$, is given by:

$$\gamma_j = \frac{p_j g_j}{I_j}, \tag{4.74}$$

where $I_j$ is defined as:

$$I_j = \sum_{l=1}^{N} [p_l g_l] + \sigma^2, l \neq j, \tag{4.75}$$

with $\sigma^2$ as the average AWGN power.

The objective of each player in the optimization process is represented by its objective function, which is a function of all the decision variables, i.e., it depends on transmit powers of all players.

The decentralized power control game is naturally noncooperative, since each player makes rational decisions aiming the optimization of its objective function based only on local information. Therefore, distributed power control games are classified as noncooperative $N$-person games.

### 4.12.2 Objective Functions: Utility, Pricing and Cost

The objective function is a special element in a game, since it must represent the player interest with respect to the optimization process to be carried out. Objective functions can be classified as cost functions or utility functions. The concept of cost function refers to the expense or loss of the player as a result of its actions. Utility functions are referred to revenue or gain measures of players instead of their losses. Another modeling of losses or expenses is called pricing. The pricing works as a penalty associated to the decision made by the player. Power control algorithms in the literature differ essentially on the objective functions.

In wireless communications, objective functions are closely related to QoS. Two QoS requirements are low delay and low probability of error. QoS guaranteed services are usually associated to hard delay requirements and considerable tolerance with respect to transmission errors. This is the case of voice services. By contrast, some data services can accept some delay but are very low tolerant to errors. Such services are classified as best effort or elastic services. Different objective functions may be constructed for both types of services.

#### QoS Guaranteed Services

The establishment of a minimum SINR corresponds to a step function of the SINR as utility function [32, 33]. The system is considered unacceptable when the SINR $\gamma$ is below a target level $\gamma^t$. When SINR is above the target level, the utility or satisfaction $u$ is constant, as shown in Fig. 4.8. Then, it is implicitly assumed that there is no benefit to having SINR above the target level. This is an adequate model for QoS guaranteed services, as demonstrated for voice traffic in [34, 35], where a downlink resource (code and power) allocation problem is considered in CDMA systems.

Objective functions which represent satisfaction can be modified by including a penalty in order to avoid the unrestricted use of resources. Pricing schemes provide more efficient solutions for the power control game. In [34, 35], each player is penalized linearly with its transmit power. Then, player $j$ with transmit power $p_j$ has as total charge for service $\alpha_c + \alpha_p p_j$, where $\alpha_c$ is the price per code and $\alpha_p$ is the price per unit transmitted power. In this case, the objective function to be maximized is defined as the utility (step function) minus the charge for service. This optimization process provides two possible outcomes to each player: it remains inactive (no transmission) or uses exactly enough power to achieve the target SINR $\gamma^t$.

Another suitable objective function for QoS guaranteed services is explored in [36], in the form of a cost function of the SINR. There, SINR values far from the target SINR $\gamma^t$ are discouraged, as shown in Fig. 4.9. The minimization of the cost function (4.76), defined as the squared error between target and achieved SINRs,

$$c_j = \left(\gamma_j^t - \gamma_j\right)^2, \tag{4.76}$$

accomplishes the tracking of $\gamma_j^t$ and provides a new demonstration of the Autonomous SINR Balancing Algorithm [13].

#### Best Effort Services

Distributed power control algorithms for best effort services have been extensively studied in the context of noncooperative games [28, 32, 33, 37, 38, 39, 40, 41, 42, 43, 44, 45]. In most frameworks, utilities and pricing schemes are adopted as QoS metrics. Pricing is always associated to power consumption, while a variety of approaches is used to

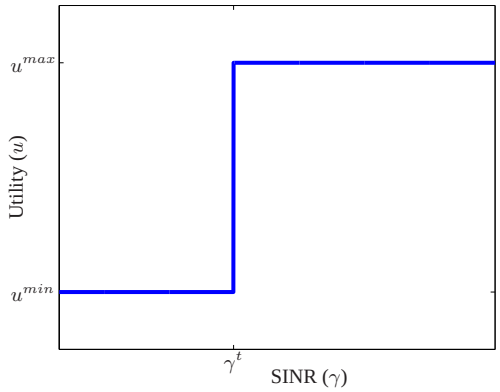



**Fig. 4.8.** Quality of service metric for QoS guaranteed services represented as a utility function.

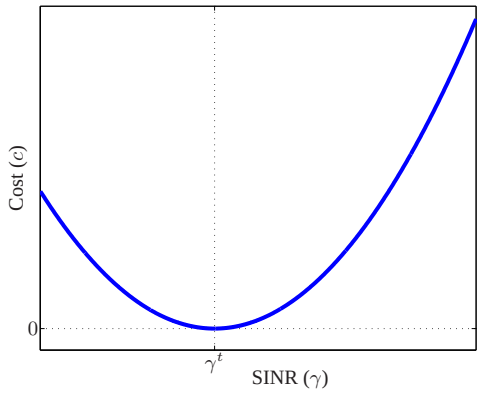



**Fig. 4.9.** Quality of service metric for QoS guaranteed services represented as a cost function.

represent some objectives as utilities. The most common objectives are: maximization of total (system) throughput and maximization of energy efficiency.

In [37], a jointly uplink power and spreading gain control problem is considered for CDMA cellular systems. An instantaneous throughput for each transmitter is used as utility function. A global optimal solution which maximizes the aggregate throughput is obtained. However, it is not considered any constraint on the spreading gain.

In [38], a framework for uplink power control in cellular systems is proposed to obtain better QoS using less power. The objective function is a concave decreasing function of power and a concave increasing function of SINR.

Utility-based algorithms with pricing, that is, algorithms which use net utilities are developed in order to improve energy efficiency [32, 33, 39, 40, 41, 42, 43]. Pricing is a monotonic increasing function with transmitted power. A net utility is composed by the difference between utility and pricing, as illustrated in Figure 4.10.

Uplink power control is developed to provide power-efficient transmission in [32, 33, 39, 40]. The utility function is related to the number of effective bits transmitted per unit of energy in a wireless data system which transmits packets containing $L$ information bits. With channel coding, the total size of each packet is $M > L$ bits and the transmission rate is $R$ bits/s. The utility function for each player $j$ is given below in bits/Joule:

$$u_j = \frac{LRf(\gamma_j)}{Mp_j}, \tag{4.77}$$

where $f(\gamma_j)$ is called efficiency function and represents the probability of correct reception as a function of the SINR $\gamma_j$.

Once the efficiency function is defined as showed in equation (4.78), with $BER_j$ denoting the binary error rate of transmitter/receiver pair $j$, the adopted utility function includes details of system lower layers, as coding and modulation.

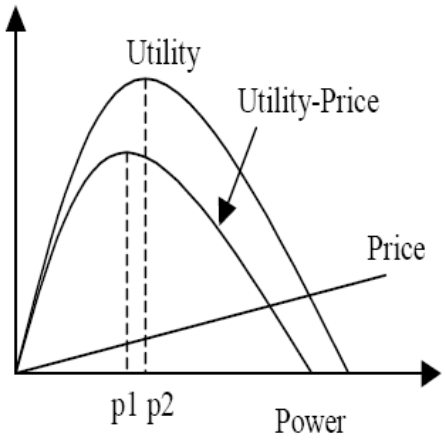



**Fig. 4.10.** Utility, pricing, and net utility.

$$f(\gamma_j) = (1 - 2BER_j)^M. \tag{4.78}$$

Such dependence on system parameters may restrict the employment of algorithms to some system configurations, as in [32, 33, 39, 40] for noncoherent frequency shift keying (FSK) modulation.

High transmit power levels as result of the utility maximization motivate the search for techniques to encourage players to transmit at lower power. An improved power control can be carried out by imposing a price on each transmission. The pricing scheme used in [32, 33, 39, 40] is a linear function of power, $\alpha_p p_j$, where $\alpha_p > 0$ is the pricing factor. Then, the net utility is given by:

$$u_j = \frac{LRf(\gamma_j)}{Mp_j} - \alpha_p p_j. \tag{4.79}$$

Other objective functions not dependent on system parameters were addressed in some works, as in [42, 41, 43]. The authors of [41] make use of the Shannon's capacity rule for AWGN channels as utility function, which corresponds to treating all interference as noise. Thus, the net utility of each player $j$ is:

$$u_j = B\log_2(1 + \gamma_j) - \alpha_p p_j, \tag{4.80}$$

where $B$ denotes the channel bandwidth in Hz.

In [42] the utility function corresponds to the effective throughput. In this work, a pricing scheme different from the pricing per unit transmit power is adopted: the pricing per unit normalized transmit power is used (normalized by total interference). Thus, the impact made by player in overall interference is more accurately measured.

The utility considered in [43] for the uplink problem is a sigmoid function of SINR. The net utility with a linear pricing is expressed in equation (4.81):

$$u_j = \frac{1}{1 + e^{-\alpha(\gamma_j - \beta)}} - \alpha_p p_j. \tag{4.81}$$

Parameters $\alpha$ and $\beta$ in the sigmoid utility can be used to tune the steepness and the center of the utility, respectively.

Cost functions are also used as objective functions for best effort services. As in [36] for QoS guaranteed services, the cost function is defined as the squared error between target and achieved SINRs in [46, 45]:

$$c_j = \left(\gamma_j^t - \gamma_j\right)^2. \tag{4.82}$$

However, for best effort services the target SINR is defined for each player in an opportunistic manner as a function of the transmit power required to achieve it: players who degenerate the performance of other players by using high power levels are designated to target low SINR levels. In this approach, restrictions (upper and lower bounds) on transmit power and QoS (SINR) requirements are considered.

Each mapping function of transmit power into target SINR with such opportunistic behavior provides a particular distributed power control algorithm [46, 45]. In [46], the target SINR is linear (in logarithmic scale) with the transmit power required to reach it, as shown in Figure 4.6. The mapping function is then expressed as:

$$\gamma^t_{j\,dB} = \alpha - \beta p_{j\,dBm}, \tag{4.83}$$

where $\alpha$ and $\beta$ are tunable parameters defined as functions of the upper and lower bounds of power and target SINR. The target SINR mapping considered in [45] is a sigmoid function of the transmit power.

Another cost function is studied in [47]. There, the cost function to be minimized is composed by the cost of the difference between target and achieved SINRs and the cost of the power consumption, as shown below in (4.84):

$$c_j = \alpha p_j + \beta \left(\gamma^t_j - \gamma_j\right)^2, \tag{4.84}$$

where $\alpha$ and $\beta$ are constant nonnegative weighting factors.

A third approach to the power control problem with the use of cost functions is suggested in [48]. There, each player $j$ uses a cost function which is linear in power and dependent on SINR in logarithmic scale, as defined in equation (4.85):

$$c_j = \alpha p_j - \beta \log_e(1 + \gamma_j), \tag{4.85}$$

where $\alpha$ and $\beta$ are the cost weights.

**Multi-Service**

The frameworks of noncooperative power control games differ essentially on the objective functions. Some of the objective functions adopted in the literature for both QoS and best effort (or elastic) services were presented. Each commented objective function is addressed to the single-service power control problem, i.e., they are suitable for QoS guaranteed services or for best effort services, exclusively.

In [44], a general framework based on the approach adopted in [46] for best effort services encloses both types of service. The cost function is the squared error between target and achieved SINRs (4.83) and the target SINR mapping function is also that illustrated in Figure 4.6. However, it is demonstrated that for a unique target SINR, i.e., for upper and lower bounds of target SINR assuming the same value, the Game-Theoretical Distributed Power Control (GT-DPC) algorithm, proposed for best effort services, becomes the Autonomous SINR Balancing Algorithm [13, 36], adequate for QoS guaranteed services.

Therefore, the GT-DPC algorithm, derived from a framework with power restrictions, where upper and lower bounds for QoS (SINR) requirements are also considered, is able to be used for best effort services only, for QoS guaranteed services only, so as for the mixed (or multi-service) scenario.

### 4.12.3 Nash Equilibrium Solutions

The transmit power that optimizes individual cost function depends on the transmit powers of all other transmitters (players). Therefore, it is necessary to determine a set of powers where each player is satisfied with its gain or with the cost that it has to pay, given the power selections of other players. Such an operating point is called equilibrium point. All commented noncooperative power control games were addressed in the point of view of the Nash equilibrium solution.

The Nash equilibrium solution (or Nash equilibrium point) is one of the most celebrated equilibrium solutions [30, 49]. Nash equilibrium (NE) concepts offer predictable and stable outcomes of a game where multiple agents with conflicting interests compete through self-optimization and reach a point where no player wishes to deviate from.

Considering the problem of minimizing a cost function, a power vector $\mathbf{p}^* = [p^*_1, ..., p^*_N]$ is a NE point of a power control game if, for each $j \in \Omega$ it holds:

$$c_j\left(p^*_j, \mathbf{p}^*_{-j}\right) \leq c_j\left(p_j, \mathbf{p}^*_{-j}\right), \tag{4.86}$$

where $\mathbf{p}^*_{-j}$ denotes the vector consisting of the elements of $\mathbf{p}^*$ other than the *j*th element. Therefore, each player is discouraged to unilaterally adopt strategies different from that defined by the NE solution, since its cost would be increased (or equivalently, in the case of utility maximization its satisfaction would be diminished).

## 4.13 Prediction for Improvements of Power Control

The employment of a power control technique in wireless networks is not trivial, since this is a multipath environment, where fast (or short-term) fading occurs. The SINR perceived in receivers depends on power gain and interference, which are influenced by fast fading. Fast fading is the phenomenon that describes the rapid amplitude fluctuations of a radio signal over a short period of time or travel distance. These rapid fluctuations cause degradation in the action of power control.

Most formulations of the power control problem do not consider the time-varying nature of power gain and interference. It is usually supposed that between two power control actuations the power gain and the interference may be considered invariant. However, some studies have been developed to the employment of prediction techniques in the power control problem.

### 4.13.1 Taylor's Series

In [50, 51, 36], a new distributed power control algorithm is presented. It differs from the Autonomous SINR Balancing Algorithm [13] in that its deduction assumes both power gain and interference to be time-varying functions and it predicts this variations through the Taylor's Series.

The expansion of a generic continuous function $f$ in Taylor's Series with the negligence of high order terms, followed by the transformation into a difference equation produces the prediction expressed below:

$$f(k+1) = 2 \cdot f(k) - f(k-1), \tag{4.87}$$

where $k$ denotes the discrete-time index.

Then, equation (4.87) is used to predict power gain and interference. The simple prediction method based on Taylor's Series is responsible for a significant performance gain of the proposed algorithm over the DPC.

In [36], the DPC algorithm is shown to correspond to a poor approximation of the Nash Equilibrium (NE) solution of a noncooperative power control game. The use of Taylor's Series for prediction of power gain and interference produces a better NE solution approximation. The proposed algorithm outperforms the DPC with respect to the guarantee of a minimum QoS, as shown in Figure 4.11 for a CDMA system in the reverse link.

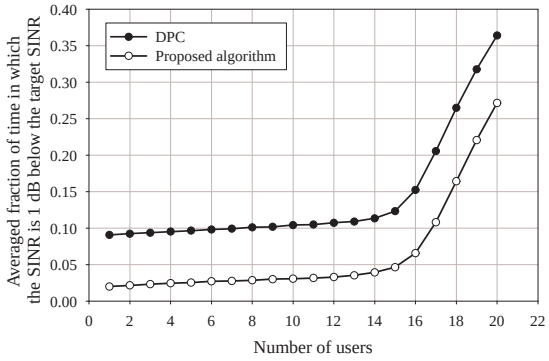



**Fig. 4.11.** Average fraction of time in which SINR is 1 dB below the target SINR for DPC and Taylor's Series based algorithm in a CDMA system.

### 4.13.2 LMS and RLS Algorithms

Predictive schemes based on both recursive-least squares (RLS) and least-mean-squares (LMS) algorithms have been investigated in the context of the power control problem for wireless networks [52, 53, 54]. Current and past samples of the received power are used as input of adaptive filters which have their tap-coefficients updated by the LMS or the RLS algorithm. Future received power is then predicted, according to the expression below:

$$\hat{p}(k+1) = \sum_{i=0}^{M-1} w_i(k)p(k-i), \tag{4.88}$$

where $\hat{p}(k+1)$ is the predicted received power at time instant $(k+1)$ and the filter is composed by $M$ tap-coefficients. The power control command decision is subsequently made based on the predicted power instead of the current estimate of the received power.

In [52], such approach is applied to the Satellite Universal Mobile Telecommunication System (S-UMTS), in order to mitigate the effects of delay. Both RLS and LMS algorithms are investigated. It is concluded that the RLS algorithm converges faster but the LMS exhibits superior tracking performance.

Prediction of future received power is also used in [53] with the goal of improving the quality of the transmission for the active users in 3G cellular systems. The proposed self-tuning predictive power control algorithm uses the RLS algorithm and performs better than the conventional power control, specially for high speed scenarios.

Similarly, in [54], adaptive filters with tap-coefficients updated by LMS and RLS algorithms are used to propose new prediction-based power control algorithms in the context of CDMA mobile radio system. However, in this case, the power gain instead of power prediction is accomplished. Simulation results show a lower outage probability when compared with the conventional power control algorithm with optimum power control threshold.

### 4.13.3 Neural Networks

A neural network based channel predictor is presented in [55]. The estimate obtained from the neural network along with the pilot assisted channel estimate is used to determine a suitable power control step. Neural network prediction based power control scheme is compared against the 3GPP based scheme in terms of Bit Error Rate (BER), average transmit power, MSE of SINR and the distribution of the SINR about the threshold. It is found that the neural network based scheme performs better at medium to high speeds.

In [56], a neural predictor for received signal power prediction in DS/CDMA systems is proposed. The results show that the optimized neural predictors can provide significant SINR gains at low speed while the improvement at high speed is clearly smaller. Other publication where power prediction is accomplished is [57]. There, a predictive reverse-link power control scheme is proposed in a CDMA mobile system. Results show that the proposed scheme outperforms the conventional power control with optimum threshold at all speed ranges considered.

### 4.13.4 Kalman Filter

The Kalman filter is widely employed in prediction tasks for improvement of power control in wireless networks, for instance [58, 59, 60]. It can be used to predict power gain, interference power and the transmit power.

In [58], Kalman filter is used to predict interference power in a TDMA wireless network, assuming that the interference signal and its measurements are corrupted by additive white Gaussian noise (AWGN). Based on the predicted interference and estimated power gain between the transmitter and receiver, transmission power is determined to achieve a desired SINR performance. Simulation results reveal that the Kalman filter method for power control provides a significant performance improvement in wireless packet networks.

Dynamic channel and power allocation algorithms are developed in [59] using the Kalman filter to provide predicted measurements for both power gain and interference power under the assumption that they are corrupted by AWGN. In [60], a power control algorithm for mobile communications that uses the Kalman filter to predict the transmit power is developed.

### 4.13.5 $H_\infty$ Filter

$H_\infty$ filtering becomes an interesting alternative to the Kalman filtering when model uncertainties are significant, since it determines upper bound to the estimation error, with no dependence on the knowledge of the disturbance statistical characteristics [61].

In [62], a SINR-based stochastic power control scheme using stochastic control theory is proposed. The measurements of SINR can be assumed to contain white noise and the stochastic power control problem is formulated as a linear stochastic discrete-time system driven by white noise. The variance minimization problem for the weighted sum of variances of SINR error and transmission power when the SINR is corrupted by AWGN is studied. In this research work, there is no assumption on channel model. Instead, a robust estimator ($H_\infty$ filter) is used to estimate the power gain. Another power control algorithm also based on the SINR error optimization is proposed in [63], with the application of an $H_\infty$ filter to predict interference power.

An alternative estimator-based approach that is also independent of the specific stochastic nature of the interference is presented in [64]. A power control algorithm which minimizes an objective function incorporating both user-centric and network-centric metrics is developed with the prediction of interference power accomplished by an $H_\infty$ filter.

## 4.14 Conclusions

Power control is an essential technique for wireless communication systems where there exists interference among the radio links. In centralized power control, the entire power channel gain matrix is available and transmission power of all transmitters may be simultaneously controlled and jointly determined. The more classical solution to this problem assumes null noise power and defines a maximum balanced SIR to be targeted by all traansmitter/receiver pairs.

Distributed (or decentralized) power control is of special interest and importance, since it allows the use of only local information. Algorithms are usually based on signal level or SINR. SINR based power control algorithms are the most common ones, with a large variety of approaches. In recent years, formulations of distributed power control as Nash noncooperative games have been intensively studied, indicating numerous interesting results.

System performance can be improved by implementing more complex techniques for power control as transmitter removal and decreasing of target SINR. However, performance of power control algorithms are limited by some aspects as time delays, discrete power domain and imperfect estimates of channel, interference and power. Prediction techniques, for instance Taylor's Series, LMS and RLS Algorithms, Neural Networks and Kalman and $H_\infty$ filtering have been increasingly used to deal with the problem of imperfect estimates.

## References

1. J. Chiller, *Mobile Communications*. Addison-Wesley, 2000.
2. T. S. Rappaport, *Wireless Communications*. Prentice-Hall, 1996.
3. T. Okumura, E. Ohmori, and K. Fukuda, "Field strengh and its variability in VHF and UHF land mobile service," *Review Electrical Communication Laboratoty*, vol. 16, pp. 825–873, 1968.
4. M. Hata, "Empirical formula for propagation loss in land mobile radio services," *IEEE Transaction on Vehicular Technology*, vol. 29, pp. 317–325, August 1980.
5. E. Damosso, *Digital Mobile Radio Towards Future Generation Systems*. European Comission, 1999.
6. J. Walfisch and H. L. Benoni, "A theoretical model for UHF propagation in urban environments," *IEEE Transaction on Antennas and Propagation*, vol. 36, pp. 1788–1796, December 1988.
7. D. M. Novakovic and M. L. Dukic, "Evolution of the power control techniques for DS-CDMA toward 3G wireless communication systems," *IEEE Communications Surveys*, p. 2-15, 2000.
8. J. Zander, "Performance of optimum tranmitter power control in cellular radio systems," *IEEE Transactions on Vehicular Technology*, vol. 41, pp. 57–62, 1992.
9. F. R. Gantmacker, *The Theory of Matrices*. Chelsea Publishing, 1960.
10. J. F. Whitehead, "Sygnal-level-based dynamic power control for co-channel interference management," *IEEE Vehicular Technology Conference*, vol. 41, pp. 499–502, 1993.
11. J. Zander, "Distributed co-channel interference control in cellular radio systems," *IEEE Transactions on Vehicular Technology*, vol. 41, pp. 305–311, 1992.
12. S. A. Grandhi, R. Vijayan, and D. J. Goodman, "Distributed power control in cellular radio systems," *IEEE Transactions on Vehicular Technology*, vol. 42, pp. 226–228, 1994.
13. G. J. Foschini and Z. Miljanic, "A simple distributed autonomous power control algorithm and its convergence," *IEEE Transactions on Vehicular Technology*, vol. 42, pp. 641–646, 1993.
14. R. Jantti and S. L. Kim, "Second-order power control with asymptotically fast convergence," *IEEE Journal Selected Areas Comunications*, vol. 18, pp. 447–457, 2000.
15. H. Holma and A. Toskala, *WCDMA for UMTS*. Wiley, 2004.
16. J. Zander and S. L. Kim, *Radio Resource Management for Wireless Networks*. Artech House Publishers, 2001.
17. G. Strang, *Linear Algebra and its Applications*. Harcourt, 1988.
18. W. S. Wong and K. H. Lam, "Distributed power balancing with a sparse information link," *IEEE Vehicular Technology Conference*, vol. 2, pp. 829–832, 1994.
19. C. W. Sung and W. S. Wong, "The convergence of an asynchronous cooperative algorithm for distributed power control in cellular systems," *IEEE Vehicular Technology Conference*, vol. 48, pp. 563–570, 1999.
20. T. H. Lee, J. C. Lin, and Y. T. Su, "Downlink power control algorithms for cellular radio systems," *IEEE Transactions on Vehicular Technology*, vol. 44, pp. 89–94, 1995.
21. M. Almgren, H. Andersson, and K. Wallstedt, "Power control in a cellular system," *IEEE Vehicular Technology Conference*, vol. 2, pp. 833–837, 1994.

22. R. D. Yates, S. Gupta, C. Rose, and S. Sohn, "Soft dropping power control," *IEEE Vehicular Technology Conference*, vol. 3, pp. 1694–1698, 1997.
23. S. A. Grandhi, J. Zander, and R. Yates, "Constrained power control," *Wireless Personal Communications*, vol. 1, pp. 257–270, 1995.
24. M. Andersin, Z. Rosberg, and J. Zander, "Distributed discrete power control in cellular PCS," *Wireless Personal Communications*, vol. 6, pp. 211–231, March 1998.
25. Y. Qian and T. Yao, "Analysis of distributed discrete power control," *Proc. International Conference on signal Processing*, vol. 2, pp. 1336–1339, August 2002.
26. F. Gunnarsson, J. Blom, and F. Gustafsson, "Power control in cellular systems subject to constraints and time delays," *Proc. IEEE Global Telecommunications Conference*, vol. 6, pp. 3645–3650, November 1998.
27. C. W. Sung and K.-K. Leung, "A generalized framework for distributed power control in wireless networks," *IEEE Transactions on Information Theory*, vol. 51, pp. 2625–2635, July 2005.
28. A. B. MacKenzie and S. B. Wicker, "Game theory and the design of self-configuring, adaptive wireless networks," *IEEE Communications Magazine*, vol. 39, pp. 126–131, November 2001.
29. R. D. Luce and H. Raiffa, *Games and Decisions*. John Wiley & Sons, 1957.
30. D. Fudenberg and J. Tirole, *Game Theory*. MIT Press, 1991.
31. T. Basar and G. J. Olsder, *Dynamic Noncooperative Game Theory*. SIAM, 1998.
32. D. Falomari, N. Mandayam, and D. Goodman, "A new framework for power control in wireless data networks: Games utility and pricing," *Proc. Allerton Conference on Communication, Control, and Computing*, vol. 1, pp. 546–555, September 1998.
33. D. Goodman and N. Mandayam, "Power control for wireless data," *IEEE Personal Communications*, vol. 7, pp. 48–54, April 2000.
34. P. Liu, P. Zhang, S. Jordan, and M. L. Honig, "Single-cell forward link power allocation using pricing in wireless networks," *IEEE Transactions Wireless Communications*, vol. 3, pp. 533–543, March 2004.
35. P. Zhang, S. Jordan, P. Liu, and M. L. Honig, "Power control of voice users using pricing in wireless networks," *Proc. SPIE ITCom Modeling and Design of Wireless Networks*, vol. 1, pp. 155–165, August 2001.
36. F. de S. Chaves, R. A. de Oliveira Neto, and F. R. P. Cavalcanti, "A game-theoretic approach to the distributed power control problem in wireless systems and the application of a simple prediction method," *Proc. Brazilian Telecommunications Symposium*, vol. 1, pp. 1–6, September 2005.
37. S. J. Oh, T. L. Olsen, and K. M. Wasserman, "Distributed power control and spreading gain allocation in CDMA data networks," *Proc. Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM)*, vol. 2, pp. 379–385, March 2000.
38. H. Ji and C.-Y. Huang, "Non-cooperative uplink power control in cellular radio systems," *Wireless Networks*, vol. 4, pp. 233–240, May 1998.
39. D. J. Goodman and N. B. Mandayam, "Network assisted power control for wireless data," *Proc. IEEE Vehicular Technology Conference (VTC)*, vol. 1, p. 1022-1026, May 2001.
40. C. U. Saraydar, N. B. Mandayam, and D. J. Goodman, "Efficient power control via pricing in wireless data networks," *IEEE Transactions on Communications*, vol. 50, pp. 291–303, February 2002.
41. S. Gunturi and F. Paganini, "Game theoretic approach to power control in cellular CDMA," *Proc. IEEE Vehicular Technology Conference (VTC)*, vol. 4, pp. 2362–2366, October 2003.
42. C. W. Sung and W. S. Wong, "A noncooperative power control game for multirate CDMA data networks," *IEEE Transactions on Wireless Communications*, vol. 2, pp. 186–194, January 2003.
43. M. Xiao, N. B. Shroff, and E. K. P. Chong, "A utility-based power-control scheme in wireless cellular systems," *IEEE/ACM Transactions on Networking*, vol. 11, pp. 210–221, April 2003.
44. F. de S. Chaves, T. F. Maciel, F. R. P. Cavalcanti, and R. A. de Oliveira Neto, "A game-theoretical distributed power control algorithm for wireless systems," *IEEE Transactions on Vehicular Technology*, submitted.
45. F. de S. Chaves, F. R. P. Cavalcanti, R. B. dos Sanstos, and R. A. de Oliveira Neto, "Opportunistic game-theoretic distributed power control algorithm for wireless communication systems," *International Journal of Communications Systems*, submitted.
46. F. de S. Chaves, V. A. de Sousa Junior, R. A. de Oliveira Neto, C. H. M. de Lima, and F. R. P. Cavalcanti, "Performance of energy efficient game theoretical-based power control algorithm in WCDMA," *Proc. IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, accepted.
47. S. Koskie, "Contributions to dynamic Nash games and applications to power control for wireless networks," *PhD Thesis, Rutgers University*, January 2003.
48. T. Alpcan, T. Basar, R. Srikant, and E. Altman, "CDMA uplink power control as a noncooperative game," *Wireless Networks*, vol. 8, p. 659-670, November 2002.
49. T. Basar and G. J. Olsder, *Dynamic Noncooperative Game Theory*. SIAM, 1998.
50. R. A. de Oliveira Neto, F. de S. Chaves, F. R. P. Cavalcanti, and T. F. Maciel, "A new distributed power control algorithm based on a simple prediction method," *Lecture Notes in Computer Science*, vol. 3124, p. 431-436, August 2004.
51. ——, "New distributed power control algorithms for mobile communications," *Journal of the Brazilian Communications Society*, vol. 20, p. 65-71, August 2005.
52. H. S. H. Gombachika, R. Tafazolli, and B. G. Evans, "A comparative study of predictive transmit power control schemes for S-UMTS," *Wireless Networks*, vol. 11, p. 215-222, May 2005.
53. I. Virtej, O. Kansanen, and H. Koivo, "Enhanced predictive fast power control for 3G systems," *Proc. Vehicular Technology Conference (VTC)*, vol. 4, pp. 2864–2868, May 2001.

54. F. C. Lau and W. M. Tam, "Novel predictive power control in a CDMA mobile radio system," *Proc. Vehicular Technology Conference (VTC)*, vol. 3, pp. 1950–1954, May 2000.
55. B. Visweswaran and T. Kiran, "Channel prediction based power control in W-CDMA systems," *Proc. 3G Mobile Communication Technologies*, vol. 471, pp. 41–45, 2000.
56. X. M. Gao, X. Z. Gao, J. M. A. Tanskanen, and S. J. Ovaska, "Power prediction in mobile communication systems using an optimal neural-network structure," *IEEE Transactions On Neural Networks*, vol. 8, pp. 1446–1455, November 1997.
57. F. C. M. Lau, "Achievable-SIR-based predictive closed-loop power control in a CDMA mobile system," *IEEE Transactions on Vehicular Technology*, vol. 51, pp. 720–728, July 2002.
58. K. K. Leung, "Novel predictive power control in a CDMA mobile radio system," *IEEE Transactions on Wireless Communications*, vol. 1, pp. 256–265, April 2002.
59. K. Shoarinejad, J. L. Speyer, and G. J. Pottie, "Integrated predictive power control and dynamic channel assignment in mobile radio systems," *IEEE Transactions on Wireless Communications*, vol. 2, pp. 976–988, September 2003.
60. T. Jiang, N. Sidiropoulos, and G. Giannakis, "Kalman filtering for power estimation in mobile communications," *IEEE Transactions on Wireless Communications*, vol. 2, p. 151-161, January 2003.
61. U. Shaked and Y. Theodor, "$H_\infty$-optimal estimation: A tutorial," *Proc. IEEE Conference on Decision and Control*, vol. 2, pp. 2278–2286, December 1992.
62. L. Qian and Z. Gajic, "Variance minimization stochastic power control in CDMA systems," *IEEE Transactions on Wireless Communications*, vol. 5, pp. 193–202, January 2006.
63. ——, "Optimal distributed power control in cellular wireless systems," *Dynamics of Continuous Discrete and Impulsive Systems*, vol. 10, pp. 537–559, 2003.
64. S. Sorooshyari and Z. Gajic, "Optimal SIR and interference-based distributed power control for wireless networks," *Proc. IEEE Conference on Decision and Control*, vol. 5, pp. 5128–5133, December 2004.

# 5

# Crosslayer Techniques - Theory and Application

Benno Zerlin and Josef A. Nossek

Institute for Circuit Theory and Signal Processing
Munich University of Technology (TUM)
Arcisstr. 21, 80290 München, Germany
{zerlin,nossek}@tum.de

## 5.1 Overview

### 5.1.1 Introducing crosslayer

Design and optimization of communication systems conventionally is based upon a vertically structured abstract system description. The different components of this model are called *layers* and their serial concatenation is known as a *protocol stack*. The different layers within a stack only employ the functionalities of their lower neighbor and export functionality exclusively to the next higher layer. Therefore the definition of interfaces between different layers provides the means to manage the complexity of systems engineering as it allows for the mutually decoupled investigation of the functionalities within different layers. A very prominent example of a system description that has received wide attention is the *open system interconnection* (OSI) model presented and commented in, e.g. [1, 2]. It classifies the functionalities of a communication system into seven layers. These OSI reference layers include at the bottom of the protocol stack the physical layer and the neighboring data link layer. The latter is comprising the media-access control (MAC) and the logical link control (LLC) sublayers. The latter part of this chapter focuses on issues within the PHY and MAC layer of mobile communication systems.

#### Motivating example

It is common practice to optimize the functionalities of different layers mutually independently. Although this approach has served well in the design of a wide number of communication systems, e.g. *second generation* (2G) mobile communication systems, it is clearly suboptimum. This lack of optimality results from several reasons, the most prominent of which are listed in the following:

- The limitation of information exchange between the layers renders complete classes of information theoretic considerations impossible. Among them are all approaches that benefit from the combination of the functionalities within different layers, e.g. channel aware scheduling.
- The components in a protocol stack do not necessarily provide a strictly monotonic mapping of inputs onto output. Hence, the optimization in a layer with respect to its output might be counterproductive in an overall sense. To a large extent this problem can be avoided by the proper design of suitable interfaces, but yet remains existent in many communication standards.
- Virtually all layers face trade-offs between different parameters. Hence an infinite number of solutions are optimum to the knowledge of the corresponding layer, i.e., they appear equivalent. But different choices might result in drastically different performance within other layers. Yet without cross-layer communication each layer has to rely on pre-defined heuristics when choosing a specific trade, which obviously results in a suboptimum system configuration. Fig. 5.1 illustrates the often addressed tradeoff between rate and error probability when transmitting data over an *additive white Gaussian noise* (AWGN) channel. For a specific setting[1] various choices for the data rate are known to result in different values for the resulting error probability. To the PHY layer all points, i.e., all PHY configurations, on one of the plotted lines are equivalent as they all require the same amount of resources. Yet, to the MAC layer, which in this example employs an *automatic repeat request* (ARQ) protocol, different configurations lead to different values for the resulting throughput, i.e., the net data rate. Including this cross-layer information when configuring the PHY layer therefore can be used to optimize the overall system performance.

[1] Different settings in this example are characterized through different signal to noise ratios and different lengths of the employed block code.

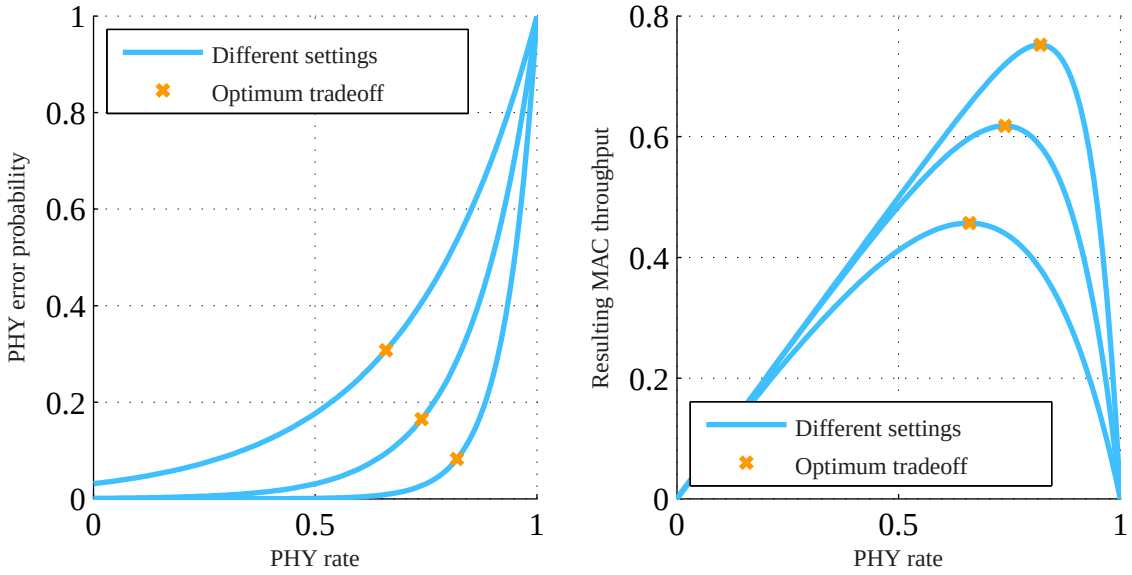



**Fig. 5.1.** Solving the rate vs. error probability tradeoff through knowledge about the used protocol, i.e., ARQ.

To overcome these imperfections the topic of cross-layer optimization has evolved. It aims at improving the interactions of different layers by introducing a limited additional information exchange and by allowing for a certain degree of cooperation between the layers. These additional degrees of freedom are used to adapt the parameters within different layers in order to optimize the mapping from inputs to outputs based on the provided cross-layer knowledge.[2]

#### Notation and assumptions

We use italic script to denote any kind of variables. Roman font is used for operators and labels. Furthermore we use lower and upper case bold face letters to denote vectors and matrices respectively. Beside the specific premises that are introduced in the system model sections of the different sections let us make some general assumptions: All occurring random processes are stationary and zero mean. We consider perfect knowledge about the statistics of the involved random variables. Hence the distributions of all random variables are known to all system units. With respect to the channel coefficients we denote this knowledge about the statistics as partial *channel state information* (CSI). Moreover, we assume that the receivers have knowledge about the instantaneous channel coefficients, i.e., has full CSI. Explicitly we do not consider the problems of obtaining the different levels of state information by means of estimation theory.

### 5.1.2 Structuring the field

There exist virtually infinitely many different problems that deserve the label cross-layer optimization. The following paragraphs shall provide three dimensions that allow to structure the field and to classify the contributions made.

#### Range

A first and almost canonical question is which layers are actually optimized by the repective scheme. We call this set of all tackled layers the range of the scheme. Typically, but of course not necessarily, cross-layer techniques jointly address problems in adjacent layers. Giving the range of a scheme automatically establishes a relationship upon virtually incomparable algorithms.

#### Modularity

Yet a certain class of cross-layer schemes[3] is not accessible to these range considerations as they merely investigate the interaction and the optimization of a modular protocol stack. These schemes might be applicable to any subset of layers

[2] Input and output of a layer are often referred to as resources and *quality of service* (QoS), respectively.
[3] In fact these were the schemes to initially come up with the term *cross-layer*.

as they propose, e.g., techniques for exchanging information among layers, for whether to optimze in a distributed way or to centralize computations etc. We refer to a perfectly modular scheme if it is scalable in the number of layers from $2$ to $\infty$.

**Orientation**

Placing the made contributions in the spectrum of published cross-layer techniques, let us differ between two fundamental approaches based upon the formulation of the respective optimization problem.[4] One part of the known cross-layer schemes aims at the cross-layer aware maximization of the output of all layers based upon their inputs. Hence the considerations start with the lowest layer and maximize its outputs subject to a set of constraints on the available resources. Doing so, the optimization regards all cross-layer information that is communicated from the upper layers. Further on the optimization makes its way upwards through the protocol stack. Layer by layer the local parameters are chosen to maximize the layer's output for the given input and under consideration of the available cross-layer information. The resulting system mode therefore employs all available resources to provide the best possible service on top of the system. Due to its directivity we refer to this approach as *bottom-up* cross-layer optimization. Complementary to that, *top-down* techniques operate on the basis of certain top-level requirements. Hence the cross-layer scheme begins with the highest layer in the stack and optimizes its parameters such that the input that is necessary to provide the required output is minimized. Again all available cross-layer information is included. The resulting minimum for the input is propagated to the next lower layer as a requirement upon corresponding output. In this manner the optimization works downwards from top until a system mode is found, that provides the required service with a minimum of overall resources. This system is then said to act at its economic maximum.

### 5.1.3 Reviewing relevant contributions

**Bottom-up**

The publications [3, 4] shows how to maximize the throughput in a multi-user system by means of optimum power allocation in PHY and MAC layer. The non-modular bottom up considerations within base upon a very abstract system that allows to transfer the resulting techniques to a variety of communication systems. Optimizing a wider span of layers [5, 6] introduces a rather modular framework for bottom up problems. It bases on utility functions, i.e., a generalized QoS expression that is implied to be monotonic with respect to the data rate on the physical link. On the background of an OFDM system with infinitesimal granularity in the frequency domain the authors derive solutions to the problem of dynamic subcarrier allocation and an adaptive power allocation scheme for QoS maximization. Moreover the results are employed to derive statements on the convexity of the feasibility regions of data rates and the global optimum of the posed cross-layer optimum. The authors in [7] moreover employ the concept of utility functions to formulate a pricing problem for the optimization of network resources in higher layers. layer description this approach leads to a multiobjective optimization problem as the core of the bottom-up cross-layer design. In [8] the authors introduce a sum throughput maximization for PHY and MAC layers in HSDPA that employs the possibility of adaptive modulation and coding and includes the ARQ protocol into this optimization. Based on a later release of HSDPA [9] combines physical layer transmit processing with a cross-layer based adaptation of the MAC modulation scheme. This approach is used to optimize the sum throughput in a cross-layer oriented scheme for HSDPA base stations.

Basing upon the analysis of application-level QoS the authors in [10] study cross-layer optimization for a variety of different 3G and 4G CDMA systems. They propose a set of protocols to serve different classes of applications with the associated QoS. Within the authors employ all available resources to maximize the system capacity. Bottom-up techniques can also be applied to video streaming services [11, 12] or gaming applications, i.e., allow for a combination of APP, MAC, and PHY layer. An example for such a bottom-up optimization that aims at video streaming applications can be found in [12]. The authors therein employ a low complexity video quality measure that can be formulated analytically. The central cross-layer optimization thereupon aims at the maximization of this QoS metric.

In the perfectly modular approach [11] the authors propose to base the layer descriptions on efficient sets of Pareto optimum points of operation. With a central monotonicity constraint on the employed layer description this approach leads to a multiobjective optimization problem as the core of the bottom-up cross-layer design.

[4] The proposed classification not only applies to cross-layer issues, but can be used for arbitrary problem statements in communications. Yet it serves especially well in this field as it allows for a general discussion of the wide variety of cross-layer contributions that have been presented in literature recently.

**Top-down**

Considering the wide span from the top of the OSI transport layer to the PHY layer the authors in [13] analyze the interconnection of system level power minimization and per user throughput on the background of lazy scheduling. A trade-off between throughput in a TCP protocol and average energy per information bit is presented and optimum scheduling policies for different sublayers are characterized. Moreover [13] proposes an efficient structure for cross-layer communication. With the modeling of voltage scaled processors, adjustable power amplifiers and variable strength coding the authors in [14] formulate the relations between range, reliability, latency and energy as the four cross-layer system parameters. The derived solutions base on an energy based description of all relevant processes. Yet the model cannot be used to formulate optimum transmission strategies, i.e., solve the underlying cross-layer optimization problem. The contributions in [15] analyze the top-down problem of minimizing the average transmit power subject to constraints on the channel capacity in the sense of [16]. Additionally [15] includes a queuing theoretic model of latency time employing Markov decision theory. The results are presented as trade-off between average transmit power and delay for a certain throughput requirement. Still the extension of these works to multi-user settings and omitting the assumption of full transmitter side CSI are open problems.

## 5.2 A generic scheme

### 5.2.1 Introducing CLARA

This Chapter introduces a generic approach to top-down cross-layer optimization of communications systems, cf. [17]. It aims at delivering a certain amount of *quality of service* (QoS) to the users, while simultaneously minimizing the required system resources.

**Notation**

To this end let us define the three relevant classes of system parameters explicitly:

- The matrix $\boldsymbol{Q} \in \mathbb{R}^{N_\mathrm{Q} \times K}$ contains the $N_\mathrm{Q}$ QoS parameters of the $K$ users that are present in the system. These QoS parameters are the only relevant interface to upper layers and sublayers or the application itself. The different service demands of the users are characterized by requirements $\boldsymbol{Q}^{(\mathrm{rq})}$ upon these QoS parameters.
- The resources of the $K$ users are denoted by a vector $\boldsymbol{P} \in \mathbb{R}^K$. Hence only a single resource parameter per user is considered.[5]
- The mode of operation $\mathcal{M}$ contains all optimization parameters that are not considered as resources.

Without loss of generality let the ordering of users in $\boldsymbol{P}$, $\mathcal{M}$ and $\boldsymbol{Q}$ coincide.

**Problem statement**

With these definitions the cross-layer problem can be formulated as the minimization of resources. For a given norm[6] in $\boldsymbol{P}$ we find the optimum resource allocation and the optimum mode of operation as the solution to the following optimization:

$$\{\boldsymbol{P}^*, \mathcal{M}^*\} = \underset{\{\boldsymbol{P}, \mathcal{M}\}}{\operatorname{argmin}} \|\boldsymbol{P}\| \quad \text{s.t.:} \quad \boldsymbol{Q} \geq \boldsymbol{Q}^{(\mathrm{rq})}, \tag{5.2}$$

[5] The extension to multiple resource parameters per user or the extension to integer resource metrics does in general not contradict the results of this Chapter. Yet the solution of the in this case matrix valued problem:

$$\boldsymbol{P}^* = \underset{\boldsymbol{P}}{\operatorname{argmin}} \|\boldsymbol{P}\| \quad \text{s.t.:} \quad \tilde{\boldsymbol{P}} \geq \tilde{\boldsymbol{P}}^{(\mathrm{rq})}, \tag{5.1}$$

which occurs to be part of the resulting cross-layer optimization algorithm can not be implied for general choices of $\boldsymbol{P} \notin \mathbb{R}^K_{+,0}$. Hence the extension to multiple resources per user is left to the specific environment where additional constraints might apply to provide a solution to (5.1).

[6] In fact any order relation can be used to define this problem. For the remainder of this thesis though we exemplarily employ the quasi order relations of different matrix norms. These norms are reflexive and transitive, i.e., a *quasi-relation,* and hence qualify in this context.

where the constraining inequalities hold component-wise. The presented technique is proven to be optimum for a wide class of systems.

To this end Section 5.2.2 introduces 3 propositions, which guarantee the applicability of the upcoming considerations. Based upon these propositions Section 5.2.3 and 5.2.3 formulate a converging iterative approach that yields the optimum solution up to arbitrary accuracy. Key to solving the problem in this wide generality is the transformation of the original program into a conditioned version that is solved in Section 5.2.3. This equivalent problem can be shown to be independent of shortterm properties. Moreover it is proven to be decoupled among the users which allows for an offline computation of the optimum mode of operation.

### 5.2.2 Formulating necessary propositions

Let the representation of the $K$ user system be given in the form:

$$\boldsymbol{Q} = \boldsymbol{\Upsilon}_{\mathfrak{M}}(\boldsymbol{P}), \quad \boldsymbol{\Upsilon}_{\mathfrak{M}} : \mathbb{R}_{+,0}^{K} \mapsto \mathbb{R}^{N_{\mathrm{Q}} \times K} \tag{5.3}$$

where the $N_{\mathrm{Q}}$ different QoS values for each of the $K$ users in $\boldsymbol{Q} \in \mathbb{R}^{N_{\mathrm{Q}} \times K}$ are defined through a mode $\mathfrak{M}$ dependent mapping of the system resources $\boldsymbol{P} \in \mathbb{R}_{+,0}^{K}$. This representation exists for all systems and for all choices of QoS and resources respectively. Yet not all of these representations can be given in closed or even in invertible explicit form. Preparing the derivation of a generic approach for cross-layer optimization of a wide class of communication systems some preconditions on $\boldsymbol{\Upsilon}_{\mathfrak{M}}$ will be made in the following paragraphs. To this end let us introduce the notation of *longterm* and *shortterm* parameters. It refers to time variant processes of the bottom layer and classifies all parameters that depend upon the instantaneous realization of these processes as shortterm parameters. Characteristics of their probability density function and variables in higher layers that are independent of these fading processes are considered longterm parameters. On this background let the following propositions hold:

**Proposition 1.** *The function $\boldsymbol{\Upsilon}_{\mathfrak{M}}$ is decomposable into three components as follows:*

$$\boldsymbol{\Upsilon}^{(1)} : \boldsymbol{P} \mapsto \tilde{\boldsymbol{P}} = \boldsymbol{\Upsilon}^{(1)}(\boldsymbol{P}), \quad \tilde{\boldsymbol{P}}, \boldsymbol{P} \in \mathbb{R}_{0,+}^{K} \tag{5.4}$$

$$\boldsymbol{\Upsilon}_{\mathfrak{M}}^{(2)} : \tilde{\boldsymbol{P}} \mapsto \tilde{\boldsymbol{Q}} = \boldsymbol{\Upsilon}_{\mathfrak{M}}^{(2)}(\tilde{\boldsymbol{P}}), \quad \tilde{\boldsymbol{Q}} \in \mathbb{R}^{N_{Q} \times K} \tag{5.5}$$

$$\boldsymbol{\Upsilon}^{(3)} : \tilde{\boldsymbol{Q}} \mapsto \boldsymbol{Q} = \boldsymbol{\Upsilon}^{(3)}(\tilde{\boldsymbol{Q}}, \boldsymbol{\pi}_{out}), \quad \boldsymbol{Q} \in \mathbb{R}^{N_{Q} \times K}. \tag{5.6}$$

*Within $\tilde{\boldsymbol{P}} = \boldsymbol{\Upsilon}^{(1)}(\boldsymbol{P})$ gives a longterm description of the fading multiple access channel. The outage probability $\boldsymbol{\pi}_{out} \in [0;1]^{K}$ is defined through the shortterm pendant $\tilde{\boldsymbol{P}}^{(st)}$ of $\tilde{\boldsymbol{P}}$ as:*

$$\boldsymbol{\pi}_{out} = Pr\left(\tilde{\boldsymbol{P}}^{(st)} < \tilde{\boldsymbol{P}}\right), \tag{5.7}$$

*where the inequality holds component wise.*

**Proposition 2.** *The functions $\boldsymbol{\Upsilon}^{(1)}$, $\boldsymbol{\Upsilon}_{\mathfrak{M}}^{(2)}$ and $\boldsymbol{\Upsilon}^{(3)}$ fulfill the following properties:*

1. *The function $\boldsymbol{\Upsilon}^{(1)}$ is strictly monotonically increasing on its diagonal[7] and is monotonically decreasing in all off-diagonal elements.*
2. *The function $\boldsymbol{\Upsilon}_{\mathfrak{M}}^{(2)}$ is diagonal.*
3. *Conditioned on $\boldsymbol{\pi}_{out}$ a solution for the unique inversion $\boldsymbol{\Upsilon}^{(3),-1}$ of $\boldsymbol{\Upsilon}^{(3)}$ exists with*

$$\boldsymbol{Q} = \boldsymbol{\Upsilon}^{(3)}\left(\boldsymbol{\Upsilon}^{(3),-1}(\boldsymbol{Q}, \boldsymbol{\pi}_{out}), \boldsymbol{\pi}_{out}\right), \tag{5.8}$$

Through the decomposability in Proposition 1 the parameters $\tilde{\boldsymbol{P}}$ and $\boldsymbol{\pi}_{\mathrm{out}}$ form a longterm description of the shortterm fading processes that is valid as long as the fading processes can be assumed stationary. For an information theoretic backup of this approach we refer to [19]. The outage probability therefore is defined as the probability that the shortterm representation $\tilde{\boldsymbol{P}}^{(\mathrm{st})}$ of $\tilde{\boldsymbol{P}}$ is smaller than $\tilde{\boldsymbol{P}}$.[8] The function $\boldsymbol{\Upsilon}_{\mathfrak{M}}^{(2)}$ gives a description of the remaining system in non-outage cases, while $\boldsymbol{\Upsilon}^{(3)}$ determines how these outage events affect the specified QoS. Later reasonings will suggest the terminology *equivalent resources* for $\tilde{\boldsymbol{P}}$ and *equivalent QoS* for $\tilde{\boldsymbol{Q}}$ respectively. Fig. 5.2 visualizes the decomposition of $\boldsymbol{\Upsilon}_{\mathfrak{M}}$. With the definition of the three representatives $\boldsymbol{\Upsilon}^{(1)}$, $\boldsymbol{\Upsilon}_{\mathfrak{M}}^{(2)}$ and $\boldsymbol{\Upsilon}^{(3)}$ the decomposable function

[7] Diagonality in this context refers to the user indices.

[8] This definition applies no matter if the strategy to adapt $\boldsymbol{P}$ operates on a longterm or a shortterm scale, i.e., no matter whether $\boldsymbol{P}$ itself is a longterm or shortterm parameter.

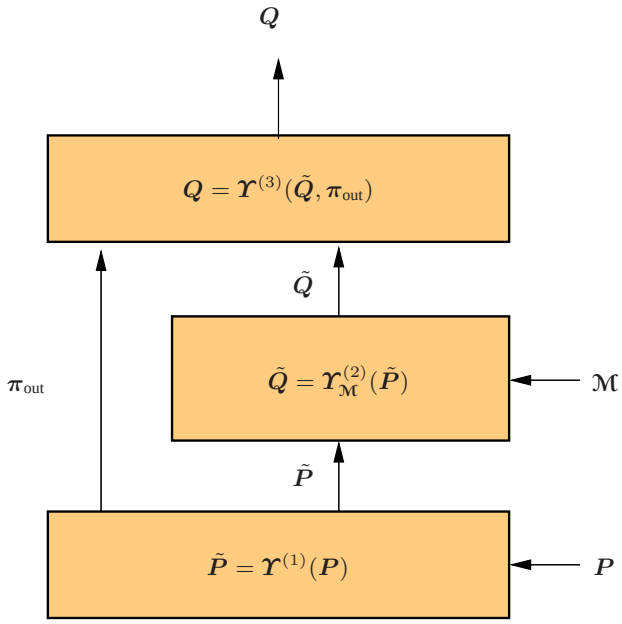



**Fig. 5.2.** Schematic representation of the system model

$\boldsymbol{\Upsilon}_{\mathfrak{M}}$ can be expressed in a concatenated form as:

$$\boldsymbol{Q} = \boldsymbol{\Upsilon}_{\mathfrak{M}}\left(\boldsymbol{P}\right) = \boldsymbol{\Upsilon}^{(3)}\left(\boldsymbol{\Upsilon}^{(2)}_{\mathfrak{M}}\left(\boldsymbol{\Upsilon}^{(1)}\left(\boldsymbol{P}\right)\right), \pi_{\text{out}}\right). \tag{5.9}$$

The components of this composition are subject to preconditions as presented in Proposition 2. According to the first clause within, the equivalent resource $\tilde{\boldsymbol{P}}$ of a user is strictly monotonically increasing with this user's resource while it is monotonically decreasing with the resources of all other users.[9] Through the second and third clause $\boldsymbol{\Upsilon}^{(1)}$ moreover is required to carry the complete regarded influence of shortterm system parameters. The required diagonal nature of $\boldsymbol{\Upsilon}^{(2)}_{\mathfrak{M}}$ translates into user wise decoupled system cores that carry the complete dependence on the mode of operation. Hence the equivalent requirements of a user $k$ in the vector $\tilde{\boldsymbol{Q}}^{(\text{rq})}_k$ must be completely determined by the equivalent scalar resource $\tilde{P}_k$ of the same user $k$. This condition is somewhat canonical as parameters that influence the performance of all users are typically classified as resources rather than as mode parameters, e.g. the transmit power in multiple access schemes. Typically the function $\boldsymbol{\Upsilon}^{(2)}_{\mathfrak{M}}$ will span the widest part of the cross-layer problem as the corresponding clause 2.2 does not make any restrictions on its diagonal elements. Clearly $\boldsymbol{\Upsilon}^{(2)}_{\mathfrak{M}}$ does not even have to be a continuous function neither does it need to be available in closed form. This fact comprises the central strength of the presented approach, because it can solve problems that are typically not accessible for conventional optimization techniques and their application in cross-layer design. The third representative $\boldsymbol{\Upsilon}^{(3)}$ includes the influence of the outage probability $\pi_{\text{out}}$ and additionally allows for an invertible non-diagonal extension of the concatenation $\boldsymbol{\Upsilon}^{(2)}_{\mathfrak{M}}\left(\boldsymbol{\Upsilon}^{(1)}\left(\boldsymbol{P}\right)\right)$. This feature allows to include upper layer resource allocation schemes, the parameters of which are not subject to the regarded cross-layer optimization themselves. Hence $\boldsymbol{\Upsilon}^{(3)}$ does not depend on the mode of operation $\mathfrak{M}$ and must be given in a description that is independent of shortterm system parameters. The extension $\boldsymbol{\Upsilon}^{(3)}$ is more important for the validity of the regarded QoS expressions than it is relevant to the optimization process with respect to $\mathfrak{M}$.

### 5.2.3 Solving the optimization

#### Conditioned optimization

The central problem of this work is to derive cross-layer resource allocation schemes for the above defined system class through the solution of the following optimization problem:

[9] The demanded property is inherent to a wide variety of multiple access schemes. It can as well be found in axiomatic approaches like [20, 21].

$$\{\boldsymbol{P}^*, \mathfrak{M}^*\} = \underset{\{\boldsymbol{P}, \mathfrak{M}\}}{\operatorname{argmin}} \|\boldsymbol{P}\| \quad \text{s.t.:} \quad \boldsymbol{Q} \geq \boldsymbol{Q}^{(\mathrm{rq})}, \quad \text{with} \quad \boldsymbol{Q} = \boldsymbol{\Upsilon}_{\mathfrak{M}}(\boldsymbol{P}). \tag{5.10}$$

To this end let us first derive the solution of this problem conditioned on the outage probability $\boldsymbol{\pi}_{\text{out}}$. Conditioning the solution to an a priori known value for $\boldsymbol{\pi}_{\text{out}}$ allows for the decomposition of the optimization task and thus shall be of a certain relevance to the optimum procedure in Section 5.2.3. Hence, the following paragraphs consider the optimization:

$$\{\boldsymbol{P}^*, \mathfrak{M}^*\} = \underset{\{\boldsymbol{P}, \mathfrak{M}\}}{\operatorname{argmin}} \|\boldsymbol{P}\| \quad \text{s.t.:} \quad \boldsymbol{\Upsilon}^{(3)}\left(\boldsymbol{\Upsilon}^{(2)}_{\mathfrak{M}}\left(\boldsymbol{\Upsilon}^{(1)}(\boldsymbol{P})\right), \boldsymbol{\pi}_{\text{out}}\right)\Big|_{\boldsymbol{\pi}_{\text{out}}} \geq \boldsymbol{Q}^{(\mathrm{rq})}. \tag{5.11}$$

As this program as well as the original task (5.10) is not accessible to conventional optimization techniques because the derivatives with respect to the possibly integer valued mode variable $\mathfrak{M}$ is not defined and $\boldsymbol{\Upsilon}^{(2)}_{\mathfrak{M}}$ additionally does not necessarily allow for a closed solution of the resulting Karush-Kuhn-Tucker conditions, let us regard the following theorem:

**Theorem 1.** *Let $\boldsymbol{\pi}_{out}$ be the outage probability that through $\boldsymbol{\Upsilon}^{(1)}$ corresponds to the equivalent resources $\tilde{\boldsymbol{P}}^{(rq)}$. Furthermore let the equivalent requirements $\tilde{\boldsymbol{Q}}^{(rq)}$ be defined through the inversion of $\boldsymbol{\Upsilon}^{(3)}$ conditioned on $\boldsymbol{\pi}_{out}$. Then the solution to the problem:*

$$\left\{\mathfrak{M}_k^*, \tilde{P}_k^{(rq)}\right\} = \underset{\{\mathfrak{M}_k, \tilde{P}_k\}}{\operatorname{argmin}} \tilde{P}_k \quad \textit{s.t.:} \quad \tilde{\boldsymbol{Q}}_k \geq \tilde{\boldsymbol{Q}}_k^{(rq)}, \tag{5.12}$$

*is independent of the requirements*[10] $\tilde{\boldsymbol{Q}}_\ell^{(rq)}$, $\forall \ell \neq k$ *and is independent of all shortterm system parameters. The union of the solutions $\mathfrak{M}_k^*$ forms an optimizer $\mathfrak{M}^*$ to the cross-layer optimization program in* (5.11)*.*

*Proof.* The first two clauses of this theorem are directly proven by Proposition 2. Due to the diagonality of $\boldsymbol{\Upsilon}^{(2)}_{\mathfrak{M}}$ the constraints are diagonal too and hence the solution for the user $k$ does not depend upon other users' equivalent QoS requirements. The same applies to the independence of the fading parameters. Key to the proof of the last clause of Theorem 1 is the monotonicity of the equivalent resources $\tilde{\boldsymbol{P}}$ with respect to all resources $\boldsymbol{P}$. This monotonicity has been part of Proposition 2 above. As the resources only through $\tilde{\boldsymbol{P}}$ and through $\boldsymbol{\pi}_{\text{out}}$ influence the QoS and furthermore the problem is conditioned on the outage probability this monotonicity suffices to show that a minimization of the resources $\boldsymbol{P}$ will inherently result in a minimization of all equivalent resources $\tilde{\boldsymbol{P}}$. Therefore the objective and the optimization with respect to $\boldsymbol{P}$ in (5.11) can be replaced by $\tilde{\boldsymbol{P}}$ without violating the validity of the solution $\mathfrak{M}_k^*$. Employing the decomposability of $\boldsymbol{\Upsilon}_{\mathfrak{M}}$ and the invertibility of $\boldsymbol{\Upsilon}^{(3)}$ the constraints of (5.11) can equivalently be expressed through a set of requirements $\tilde{\boldsymbol{Q}}^{(\mathrm{rq})}$ on the equivalent QoS $\tilde{\boldsymbol{Q}}$. This concludes the proof of the optimal nature of $\mathfrak{M}^*$ for (5.10).

The decoupled and equivalent formulation of the original program through the problems in (5.12) provides little advantage in terms of Lagrangian optimization. Still components of the optimization parameter $\mathfrak{M}_k$ are discrete, the corresponding derivatives do not exist and the Karush-Kuhn-Tucker conditions can not be applied. But the achieved decoupling among users and the gained independence of the problem from the instantaneous channel realization render the accessibility to Lagrangian methods unnecessary:

**Corollary 1.** *The equivalent requirements $\tilde{\boldsymbol{Q}}_k^{(rq)}$ uniquely determine the solution to the optimization problems in* (5.12)*. In particular the program is independent of the state of operation and therefore the solution for any equivalent QoS requirements can be obtained offline and prior to operation.*

Explicitly the optimum mode of operation $\mathfrak{M}_k^*$ can be precomputed for a sufficiently dense grid in $\tilde{\boldsymbol{Q}}_k^{(\mathrm{rq})}$ offline by sampling the $N_{\mathrm{Q}} \times K$ dimensional range of $\boldsymbol{\Upsilon}^{(2)}_{\mathfrak{M}}$. The equivalent QoS $\tilde{\boldsymbol{Q}}$ to this end is computed for an arbitrarily large but finite number of system modes and for a suitable number of values for $\tilde{\boldsymbol{P}}_k$. Storing these offline computed solutions in an $N_{\mathrm{Q}}$ dimensional database allows for the offline determination of the optimum mode of operation for grid of equivalent requirements. Each grid point defines a feasibility region through its equivalent requirements. Searching this feasibility region for the mode that provides $\tilde{\boldsymbol{Q}}_k^{(\mathrm{rq})}$ with a minimum $\tilde{P}_k^{(\mathrm{rq})}$ is chosen as the optimum mode $\mathfrak{M}_k^*$. During operation the solution to (5.12) thus can be obtained through a single table lookup. The problem (5.12) therefore can be solved very efficiently at an absolute minimum of computational cost. This makes the proposed procedure easily accessible for real-time implementations.

[10] $\tilde{\boldsymbol{Q}}_k$, $\mathfrak{M}_k$ and $\tilde{P}_k$ denote the $k$th user's portion of $\tilde{\boldsymbol{Q}}$, $\mathfrak{M}$ and $\tilde{\boldsymbol{P}}$ respectively.

**Iterative procedure**

With the above Section a low complexity solution to the conditioned problem setting (5.12) is available. The upcoming considerations in this Section now focus on optimizing the mode of operation and the resource allocation among all $K$ users employing the results form above. We target the problem of finding the mode of operation $\mathfrak{M}^*$ that fulfills a set of QoS requirements $\boldsymbol{Q}^{(\mathrm{rq})}$ with a minimum amount of resources, cf. (5.10):

$$\{\boldsymbol{P}^*, \mathfrak{M}^*\} = \underset{\{\boldsymbol{P},\mathfrak{M}\}}{\operatorname{argmin}} \|\boldsymbol{P}\| \quad \text{s.t.:} \quad \boldsymbol{Q} \geq \boldsymbol{Q}^{(\mathrm{rq})}, \quad \text{with} \quad \boldsymbol{Q} = \boldsymbol{\Upsilon}_{\mathfrak{M}}\left(\boldsymbol{P}\right), \tag{5.13}$$

where the function $\boldsymbol{\Upsilon}_{\mathfrak{M}}$ belongs to the above defined system class and fulfills the made properties introduced in Section 5.2.2. Aiming at a generic solution that applies to all representatives of the above defined class the solution of (5.13) through the calculus of Lagrangian multipliers and the Kuhn-Tucker theorem renders impossible. Moreover neither the function $\boldsymbol{\Upsilon}^{(2)}_{\mathfrak{M}}(\tilde{\boldsymbol{P}})$ nor the mode of operation $\mathfrak{M}$ necessarily is accessible to these methods.[11]

We propose an iterative framework to solve (5.13) with arbitrary accuracy. An overview of this approach is given in Fig. 5.3. The approach is based on an iterative adaptation of the outage probability $\boldsymbol{\pi}_{\text{out}}$ which in the further context will be indexed by the iteration number $i$ as $\boldsymbol{\pi}_{\text{out}}[i]$. The solutions within each iteration hence can base upon an assumption $\hat{\boldsymbol{\pi}}_{\text{out}}[i]$ on the outage probability that was obtained during the last iteration. The problem thus reduces to the conditioned optimization investigated in Section 5.2.3. Given the outage probability $\hat{\boldsymbol{\pi}}_{\text{out}}[i]$ that corresponds to the optimum vector of equivalent resources $\tilde{\boldsymbol{P}}^*$ the program reads, cf. (5.11):

$$\{\boldsymbol{P}^*, \mathfrak{M}^*\} = \underset{\{\boldsymbol{P},\mathfrak{M}\}}{\operatorname{argmin}} \|\boldsymbol{P}\| \quad \text{s.t.:} \quad \boldsymbol{\Upsilon}^{(3)}\left(\boldsymbol{\Upsilon}^{(2)}_{\mathfrak{M}}\left(\boldsymbol{\Upsilon}^{(1)}\left(\boldsymbol{P}\right)\right), \hat{\boldsymbol{\pi}}_{\text{out}}[i]\right) \geq \boldsymbol{Q}^{(\mathrm{rq})}. \tag{5.14}$$

As shown in Section 5.2.3 the propositions made can be used to decompose this optimization into three separate subproblems:

$$\tilde{\boldsymbol{Q}}^{(\mathrm{rq})} = \boldsymbol{\Upsilon}^{(3),-1}(\boldsymbol{Q}^{(\mathrm{rq})}, \hat{\boldsymbol{\pi}}_{\text{out}}[i]), \tag{5.15}$$

$$\left\{\mathfrak{M}^*_k, \tilde{P}^{(\mathrm{rq})}_k\right\} = \underset{\{\mathfrak{M}_k,\tilde{P}_k\}}{\operatorname{argmin}} \tilde{P}_k \quad \text{s.t.:} \quad \tilde{\boldsymbol{Q}}_k \geq \tilde{\boldsymbol{Q}}^{(\mathrm{rq})}_k, \forall k \tag{5.16}$$

$$\boldsymbol{P}^* = \underset{\boldsymbol{P}}{\operatorname{argmin}} \|\boldsymbol{P}\|_{\mathrm{F}} \quad \text{s.t.:} \quad \tilde{\boldsymbol{P}} \geq \tilde{\boldsymbol{P}}^{(\mathrm{rq})}. \tag{5.17}$$

For convenient reading we dropped the index $[i]$ in the notation of $\boldsymbol{P}^*$, $\tilde{\boldsymbol{P}}^{(\mathrm{rq})}$ and $\tilde{\boldsymbol{Q}}^{(\mathrm{rq})}$. While an efficient solution for the inverse of $\boldsymbol{\Upsilon}^{(3)}$ was a precondition on the regarded system class the solution to the problems (5.16) has been derived in Section 5.2.3. Moreover the solution to (5.17) for the vast majority of multiple access schemes is known or can be obtained through the monotonicity of $\boldsymbol{\Upsilon}^{(1)}$ as it was introduced in Section 5.2.2. The solution of the cross-layer optimization conditioned on $\hat{\boldsymbol{\pi}}_{\text{out}}[i]$ thus is known and can be obtained at very low computational cost. With the optimum pair of modes and resources $\{\boldsymbol{P}^*, \mathfrak{M}^*\}$ the resulting outage probability $\boldsymbol{\pi}_{\text{out}}[i]$ of the system can be determined through its definition in (5.7). Unless $\boldsymbol{\pi}_{\text{out}}[i] = \hat{\boldsymbol{\pi}}_{\text{out}}[i]$ the solutions obtained from the conditioned problem (5.16) is not a valid solution for the original task (5.13). The made assumption on $\boldsymbol{\pi}_{\text{out}}[i]$ has to be adapted and a new conditioned problem has to be solved. To this end we propose the following update rule:

$$\hat{\boldsymbol{\pi}}_{\text{out}}[i+1] = \boldsymbol{\pi}_{\text{out}}[i]. \tag{5.18}$$

The resulting structure of the iterative scheme is sketched in Fig. 5.3.

**Proof of convergence**

Through the solution of an $\hat{\boldsymbol{\pi}}_{\text{out}}[i]$ conditioned version of the original problem the cross-layer optimization can be solved through an iterative scheme as sketched in Fig. 5.3. Yet the obtained iteration is of no use if its convergence can not be proven. To this end let us state the following Theorem:

**Theorem 2.** *Let $\boldsymbol{\pi}_{out}[0] = \mathbf{0}$ be the initialization for the iterative scheme. Furthermore let $\boldsymbol{\pi}_{out}[i]$ be defined as the outage probability that results from $\mathfrak{M}^*$ and $\boldsymbol{P}^*$ as defined in* (5.14)*. Then the iteration $\hat{\boldsymbol{\pi}}_{out}[i] = \boldsymbol{\pi}_{out}[i-1]$ converges and $\boldsymbol{\pi}_{out}[i] = \boldsymbol{\pi}_{out}[i-1]$ holds with arbitrary accuracy for large $i$.*

[11] The function $\boldsymbol{\Upsilon}^{(2)}_{\mathfrak{M}}(\tilde{\boldsymbol{P}})$ in non-trivial settings usually is too complex to provide an invertible system of equations from the Karush-Kuhn-Tucker conditions, whereas the mode of operation in many relevant applications contains integer variables. Hence the derivatives involved in the conventional solution of this problem are not defined.

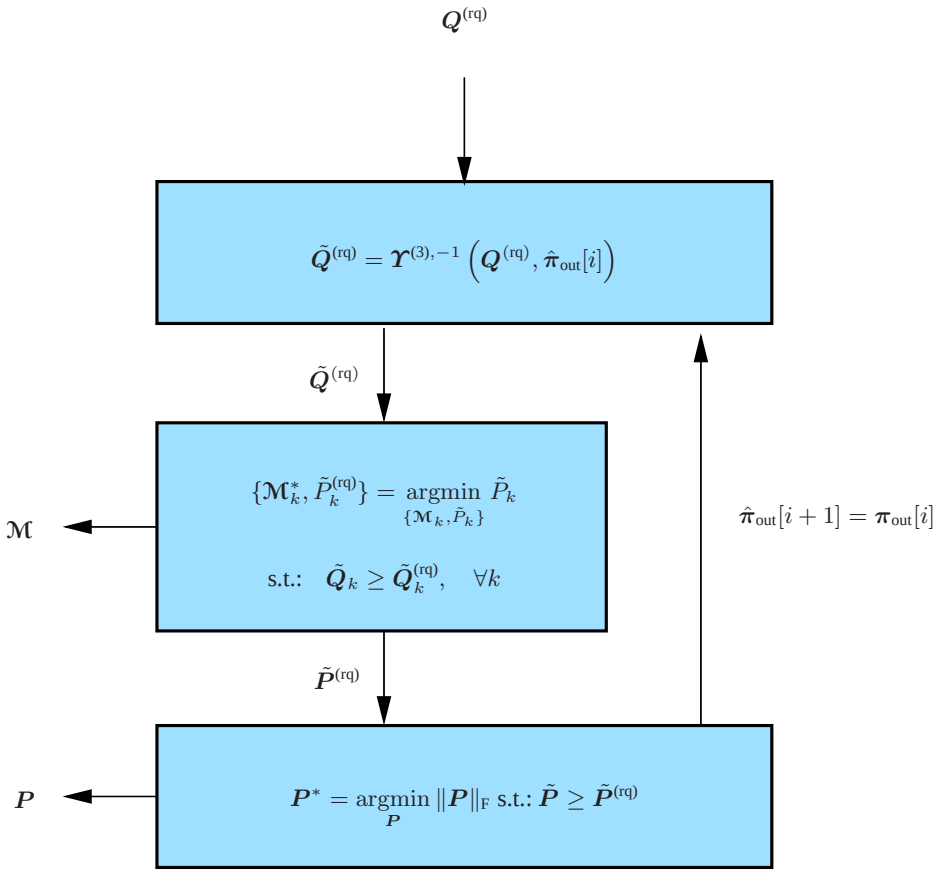



**Fig. 5.3.** Schematic representation of the iterative cross-layer optimization

*Proof.* To prove this convergence assume an arbitrary $\hat{\boldsymbol{\pi}}_{\mathrm{out}}[i] = \boldsymbol{\pi}_{\mathrm{out}}[i-1] > \hat{\boldsymbol{\pi}}_{\mathrm{out}}[i-1]$ from the interval $[0;1]^K$. By the definition of the outage probability the increase in $\hat{\boldsymbol{\pi}}_{\mathrm{out}}[i]$ necessarily results in an increase in all components of $\tilde{\boldsymbol{Q}}^{(\mathrm{rq})}$, i.e., larger equivalent requirements with respect to the iteration $i-1$. This is due to the fact that an increased outage probability will enlarge the QoS decrease in $\boldsymbol{\Upsilon}^{(3)}$. Due to the inequality nature of the constraints in (5.16) this monotonicity with respect to $\hat{\boldsymbol{\pi}}_{\mathrm{out}}[i]$ applies to the requirements on the equivalent resources $\tilde{\boldsymbol{P}}^{(\mathrm{rq})}$ as well. To this end assume that an increase in $\tilde{\boldsymbol{Q}}_k^{(\mathrm{rq})}$ from iteration $i-1$ to $i$ yields a decrease of $\tilde{P}_k^{(\mathrm{rq})}$ in the optimum solution of (5.16). Then this new value for $\tilde{P}_k^{(\mathrm{rq})}$ would provide a smaller solution to the problem in iteration $i-1$. Thus the solution in iteration $i$ can not be optimum, which is a contradiction. Hence an increase in outage probability $\hat{\boldsymbol{\pi}}_{\mathrm{out}}[i] > \hat{\boldsymbol{\pi}}_{\mathrm{out}}[i-1]$ always results in requirements $\tilde{\boldsymbol{P}}^{(\mathrm{rq})}$ that are equal to or larger than the corresponding values in the previous iteration. Due to the positive semidefinite nature of the equivalent resources and the definition of the outage probability this inherently results in $\boldsymbol{\pi}_{\mathrm{out}}[i] \geq \boldsymbol{\pi}_{\mathrm{out}}[i-1]$. Through the update $\hat{\boldsymbol{\pi}}_{\mathrm{out}}[i+1] = \boldsymbol{\pi}_{\mathrm{out}}[i]$ a single increase in $\hat{\boldsymbol{\pi}}_{\mathrm{out}}$ causes a monotonically increasing series $\hat{\boldsymbol{\pi}}_{\mathrm{out}}[i]$. Because $\boldsymbol{\pi}_{\mathrm{out}}[i]$ is positive semidefinite, the initial choice $\hat{\boldsymbol{\pi}}_{\mathrm{out}}[0] = 0$ results in $\hat{\boldsymbol{\pi}}_{\mathrm{out}}[1] \geq \hat{\boldsymbol{\pi}}_{\mathrm{out}}[0]$. In the case of equality $i = 1$ directly fulfills the condition for convergence $\boldsymbol{\pi}_{\mathrm{out}}[i] = \boldsymbol{\pi}_{\mathrm{out}}[i-1]$. In all other cases, the above derivation proves $\boldsymbol{\pi}_{\mathrm{out}}[i]$ to be an monotonically increasing sequence. Hence the proof of convergence is obtained from the bounded nature of the probability integral, i.e., $\boldsymbol{\pi}_{\mathrm{out}}[i] \in [0;1]^K$.

**Proof of optimality**

**Theorem 3.** *Let $\boldsymbol{\pi}_{out}[0] = \mathbf{0}$ be the initialization for the iterative scheme. Furthermore let $\mathcal{M}^*$ and $\boldsymbol{P}^*$ result from the fixpoint $\boldsymbol{\pi}_{out}^* = \boldsymbol{\pi}_{out}[i^*] = \hat{\boldsymbol{\pi}}_{out}[i^*]$. Then $\mathcal{M}^*$ and $\boldsymbol{P}^*$ are optimizers to* (5.11)*.*

*Proof.* Assume the existence of another fixpoint $\boldsymbol{\pi}'_{\mathrm{out}} = \boldsymbol{\pi}_{\mathrm{out}}[i'] = \hat{\boldsymbol{\pi}}_{\mathrm{out}}[i']$. First if $\boldsymbol{\pi}'_{\mathrm{out}} > \boldsymbol{\pi}^*_{\mathrm{out}}$, $\mathcal{M}'$ and $\boldsymbol{P}'$ must be suboptimum to $\mathcal{M}^*$ and $\boldsymbol{P}^*$ due to the inequality nature of the constraints. Second, $\boldsymbol{\pi}'_{\mathrm{out}} < \boldsymbol{\pi}^*_{\mathrm{out}}$ would imply the existence of $i$ such that $\hat{\pi}_{\mathrm{out}}[i] < \hat{\pi}_{\mathrm{out}}[i'] = \pi'_{\mathrm{out}} < \pi_{\mathrm{out}}[i]$. This contradicts the monotonicity of the series $\hat{\pi}_{\mathrm{out}}[i]$ as proven above and thus concludes this proof.

With the conclusion of this proof the iterative procedure of Section 5.2.3 is known to converge for all problem settings and systems that fall in the defined class. Yet this convergence does not prove the feasibility of the posed optimization task. As possible feasibility constraints on the system resources can always be included in the definition of the corresponding outage probability we propose to define the feasibility through the optimum mode of operation $\mathfrak{M}^*$. Hence a set of QoS requirements is considered feasible if for every iteration $i$ a mode of operation $\mathfrak{M}^* = \bigcup_{k=1}^{K} \mathfrak{M}_k^*$ exists among the finite number of system configurations such that:

$$\left\{\mathfrak{M}_k^*, \tilde{P}_k^{(\mathrm{rq})}\right\} = \underset{\left\{\mathfrak{M}_k, \tilde{P}_k\right\}}{\operatorname{argmin}} \tilde{P}_k \quad \text{s.t.:} \quad \tilde{\boldsymbol{Q}}_k \geq \tilde{\boldsymbol{Q}}_k^{(\mathrm{rq})}, \forall k. \tag{5.19}$$

For all feasible constellations though, the above results provide the means to solve the cross-layer problem (5.10) through the iterative consideration of $K$ decoupled longterm problems. Based upon the results from the previous iteration, these problems are all conditioned on the outage probability. They have been subject to consideration in Section 5.2.3 where the optimum was proven to be independent of the state of operation and the solution therefore was obtained through a look-up in an offline generated table. Hence the presented iterative scheme provides the means to efficiently solve the top-down cross-layer optimization program (5.2) for all systems that fulfill the Propositions 1-3.

## 5.3 Some applications

### 5.3.1 Applying CLARA to EGPRS

With the generic results obtained in the above section means are available to jointly optimize radio parameters in a wide variety of communication systems. The following subsection will demonstrate this applicability for the special case of EGPRS, an extension to the second generation standard GSM.

#### System model

The crosslayer considerations regard all instances in the PHY and MAC layer of a 2G communication system. The generic system model $\boldsymbol{Q} = \boldsymbol{\Upsilon}_{\mathfrak{M}}(\boldsymbol{P})$ is specified along

$$\{\boldsymbol{\rho}, \boldsymbol{\tau}\} = \boldsymbol{\Upsilon}_{\mathfrak{M}}(\boldsymbol{P}), \tag{5.20}$$

describing the throughput $\rho_k$ and the delay $\tau_k$ experienced by the users $k = 1, \ldots, K$. The throughput $\rho_k$ of user $k$ is defined as the average net data rate that is available on top of the MAC scheduling unit. Let $B$ be the number of information bits in a packet of length $T$ and let $f_n[n]$ be the probability, that it takes exactly $n$ time slots of length $T$ to transmit a packet error-free. Then the throughput $\rho_k$ is defined as:

$$\rho_k = \frac{1}{\mathrm{E}[n]} \frac{B}{T}. \tag{5.21}$$

On the other hand we find an outage based formulation for the $\pi_\tau$-outage delay $\tau_k$ as:

$$\tau_k = \underset{\tau'}{\operatorname{argmin}}\, \tau' \quad \text{s.t.:} \quad \sum_{n=1}^{\lfloor \tau'/T \rfloor} f_n[n] \geq 1 - \pi_\tau. \tag{5.22}$$

Hence, $\tau_k$ gives the time which in a fraction of $1 - \pi_\tau$ of all cases suffices to transmit a packet error free, i.e., the lower limit of the $\pi_\tau$ outage quantile. These QoS parameters through $\boldsymbol{\Upsilon}_{\mathfrak{M}}$ are expressed as a function of the users' transmit powers $P_k$. The mode of operation $\mathfrak{M}$ consists of the FEC code rates and the modulation alphabets of all users. To derive $\boldsymbol{\Upsilon}_{\mathfrak{M}}(\boldsymbol{P})$ we tackle the following components:

*Broadcast channels*

We employ an SINR $\gamma_k$ based description of the downlink channel along [22]. Introducing the coupling factors[12] $\nu_{k,\ell}$, the circulary symmetric complex Gaussian random variable $r_k$ with variance $\sigma_{r_k}^2$, $k = 1, \ldots, K$ and the noise power $P_\eta$ we obtain, cf. the validations in [22]:

[12] This modelling approach comprises the common assumption of adjecent channel only interference.

$$\gamma_k = \frac{r_k P_k}{\sum_{\ell=1}^{K} \nu_{k,\ell} r_k P_\ell + P_\eta}. \tag{5.23}$$

The elaborations in [23] demonstrate how this channel only model can be extended to account for different receiver structures. A central role is held by channel matched filters, i.e., rake receivers which result in $r_k$ being the sum of squared circulary symmetric complex Gaussian random variables. Their density function and distribution can be obtained through Lemmata 4.3b.1-4.3b.3 from [24]. Given knowledge about the distribution of the involved random variables allows for the computation of the probability that the fading value $\gamma_k$ falls beneath a constant threshold. This outage probability generally is defined as the integral over the set $\mathcal{H}$ of all infeasible channel realization:

$$\pi_{\text{out}} = \int_{\mathcal{H}} \prod_{i=1}^{K} f_{r_i}(r_i) \mathrm{d}r_i. \tag{5.24}$$

The latter set depends on the specifics of the employed downlink strategy, i.e., whether fast power control is employed or not. In the sequel we will assume constant power allocation. $\mathcal{H}$ hence computes along

$$\mathcal{H} = \left\{ [r_1, \ldots, r_K]^{\mathrm{T}} \,\middle|\, r_K < r_K^{(\text{th})} \right\}, \tag{5.25}$$

$$r_K^{(\text{th})} = \lim_{\varepsilon \to 0+} \frac{\left[P_\eta \mathbf{1}^{\mathrm{T}} \boldsymbol{\Psi}_t^{-1}\right]_K}{\max\left\{P_{\max} - \sum_{\ell=1}^{K-1} \left[P_\eta \mathbf{1}^{\mathrm{T}} \boldsymbol{\Psi}_t^{-1}\right]_\ell \frac{1}{r_\ell}, \epsilon\right\}} \tag{5.26}$$

Within we defined the diagonal matrix $\boldsymbol{R} = \mathrm{diag}_{k=1}^{K}(r_k)$ containing the channel gains and the coupling matrix $\boldsymbol{\Psi}_t$ as

$$[\boldsymbol{\Psi}_t]_{k,\ell} = \begin{cases} -\nu & \text{for} \quad k \neq \ell \\ \frac{\chi}{\gamma_k^{(\text{rq})} C_k} - \nu & \text{for} \quad k = \ell. \end{cases} \tag{5.27}$$

For numerical details on the computation of $\pi_{\text{out}}$ we refer to [25] and [26].

*Channel coding*

With this SINR model we can employ the noisy channel coding theorem, e.g., [27] for modeling the FEC codes:

**Theorem 4.** *For a memoryless channel with the corresponding $R_0(\gamma_k)$ value and an input alphabet $\mathcal{A}_k$ of cardinality $A_k$ there always exists a block code with block length $n$ and code rate $R_k \leq R_0(\gamma_k)$ so that with maximum likelihood decoding the error probability $\pi_{pe}$ of a code word can be expressed as:*

$$\pi_{pe} \leq 2^{-n_{\text{eq}}(R_0(\gamma_k) - R_k \mathrm{ld} A_k)}. \tag{5.28}$$

Due to the tight nature of this bound (5.53) serves as a good approximation of the block error rate. The central parameter within is the cutoff rate $R_0$ which can be obtained as the maximum of the error exponent in [27]. For the given system class and the modulation points $a_\ell$, $\ell = 1, \ldots, A_k$ the cutoff-rate $R_0$ computes to:

$$R_0(\gamma_k) = \mathrm{ld} A_k - \mathrm{ld}\left[1 + \frac{2}{A_k} \sum_{m=1}^{A_k - 1} \sum_{\ell=m+1}^{A_k} \exp\left(-\frac{1}{4} |a_\ell - a_m|^2 \gamma_k\right)\right]. \tag{5.29}$$

Unlike the more prominent capacity based approaches, the model in (5.53) includes the influences of $A_k$ and $R_k$, i.e., the mode of operation.

*Hybrid ARQ protocols*

Let us extend the above Theorem 4 to include the effects of HARQ protocols in the MAC, cf. [28]. These protocols retransmit lost packets with possibly changing parity information and decode all received instances of a packet jointly. This results in a reduction of the packet error probability with every transmission attempt. Let $m$ index these HARQ transmission attempts. Then the decrease in block error probability can be modeled through the above channel coding theorem by

- calculating the code rate $R_k[m]$ valid for this transmission as the fraction of information bits and total transmitted bits and by

- including a proportional increase of the SINR for every bit that is transmitted multiply. Hence $\gamma_k[m]$ becomes a function of the transmission number as well.

This hybrid proceeding allows to include Chase combining type HARQ protocols as well as incremental redundancy techniques. In fact EGPRS employs both schemes simultaneously. The corresponding values for $R_k[m]$ and the SINR enhancements[13] $_{\Delta}\,\gamma_k[m]$ such that

$$\gamma_k[m] = \gamma_k \; {}_{\Delta}\gamma_k[m] \tag{5.30}$$

for different modes of operation can be obtained from Table 5.1.

**Table 5.1.** Numerical details of the EGPRS model

| # | $R_k[1]$ | $R_k[2]$ | $R_k[3]$ | ${}_{\Delta}\,\gamma_k[1]$ | ${}_{\Delta}\,\gamma_k[2]$ | ${}_{\Delta}\,\gamma_k[3]$ |
|---|---|---|---|---|---|---|
| 1 | 0.51 | 0.33 | 0.33 | 1.00 | 1.31 | 1.96 |
| 2 | 0.64 | 0.33 | 0.33 | 1.00 | 1.04 | 1.56 |
| 3 | 0.83 | 0.42 | 0.33 | 1.00 | 1.00 | 1.20 |
| 4 | 0.98 | 0.49 | 0.33 | 1.00 | 1.00 | 1.00 |
| 5 | 0.37 | 0.33 | 0.33 | 1.00 | 1.80 | 2.67 |
| 6 | 0.49 | 0.33 | 0.33 | 1.00 | 1.37 | 2.04 |
| 7 | 0.75 | 0.38 | 0.33 | 1.00 | 1.00 | 1.32 |
| 8 | 0.91 | 0.46 | 0.33 | 1.00 | 1.00 | 1.10 |
| 9 | 0.99 | 0.50 | 0.33 | 1.00 | 1.00 | 1.01 |

With the adaptation of $\pi_{\text{pe}}$ we can furthermore express the probability that it takes exactly $m$ HARQ transmissions to receive a packet error-free:

$$f_m[m] = \left( \prod_{m'=1}^{m-1} \pi_{\text{pe}}\,[m'] \right) (1 - \pi_{\text{pe}}[m]) \,. \tag{5.31}$$

Employing a proof from [23] this probability can be approximated very well for large code words as $f_m[m] = \delta[m - m^*]$. This yields the probability of waiting exactly $n$ time slots for the successfull tranmission of a packet as

$$f_n[n] = \pi_{\text{out}}^{n-m^*} (1 - \pi_{\text{out}})^{m^*} \binom{n-1}{m^*-1}. \tag{5.32}$$

The compound of all these components finally allows for the expression of the throughput and the latency in the investigated system and provides the specifics of $\{\boldsymbol{\rho}, \boldsymbol{\tau}\} = \Upsilon_{\mathfrak{M}}(\boldsymbol{P})$ in a GERAN interface:

$$\rho_k = \frac{1}{\mathrm{E}[n]} \frac{B}{T} = \frac{1 - \pi_{\text{out}}}{m^*} R_k \mathrm{ld} A_k R_s, \tag{5.33}$$

$$\tau_k = \underset{\tau'}{\operatorname{argmin}}\, \tau' \quad \text{s.t.:} \quad \sum_{n=1}^{\lfloor \tau'/T \rfloor} f_n[n] \geq 1 - \pi_\tau. \tag{5.34}$$

Let $\boldsymbol{P}$, $\boldsymbol{\rho}$, and $\boldsymbol{\tau}$, denote the $K \times 1$ dimensional multi-users compounds of $P_k$, $\rho_k$, and $\tau_k$, respectively. Then the decomposition of the above model $\boldsymbol{\Upsilon}_{\mathfrak{M}} : \boldsymbol{P} \mapsto [\boldsymbol{\rho}, \boldsymbol{\tau}] = \boldsymbol{\Upsilon}_{\mathfrak{M}}(\boldsymbol{P})$ along:

$$\boldsymbol{\Upsilon}^{(1)} : \boldsymbol{P} \mapsto \boldsymbol{\gamma} = \boldsymbol{\Upsilon}^{(1)}(\boldsymbol{P}), \tag{5.35}$$

$$\boldsymbol{\Upsilon}^{(2)}_{\mathfrak{M}} : \boldsymbol{\gamma} \mapsto [\tilde{\boldsymbol{\rho}}, \tilde{\boldsymbol{\tau}}] = \boldsymbol{\Upsilon}^{(2)}_{\mathfrak{M}}(\boldsymbol{\gamma}), \tag{5.36}$$

$$\boldsymbol{\Upsilon}^{(3)} : [\tilde{\boldsymbol{\rho}}, \tilde{\boldsymbol{\tau}}] \mapsto [\boldsymbol{\rho}, \boldsymbol{\tau}] = \boldsymbol{\Upsilon}^{(3)}([\tilde{\boldsymbol{\rho}}, \tilde{\boldsymbol{\tau}}], \boldsymbol{\pi}_{\text{out}}). \tag{5.37}$$

complies with the propositions in Section 5.2.2. Hence the iterative optimization scheme, the corresponding proof of convergence, and the proof of optimality, respectively, apply.

[13] Some implementations discard all received packets after $m_{\text{max}}$ transmission attempts and reschedule the packet from the radio network control. These special cases are included in the model proposed here through a periodic (and therefore non monotonic) definition of ${}_{\Delta}\gamma[m]$.

**Optimization**

Within this Section the cross-layer optimization in HSDPA is formulated in accordance to the considerations in Section 5.2 as a resource minimization problem subject to a set of constraints on the QoS parameters throughput $\rho_k$ and delay $\tau_k$. The cross-layer optimum mode of operation $\mathfrak{M}^*$ and the corresponding optimum transmit powers $\boldsymbol{P}^*$ therefore are the solutions to the minimization program:

$$\{\mathfrak{M}^*, \boldsymbol{P}^*\} = \underset{\{\mathfrak{M},\boldsymbol{P}\}}{\operatorname{argmin}} \|\boldsymbol{P}\|_1 \quad \text{s.t.:} \begin{cases} \rho_k \geq \rho_k^{(\mathrm{rq})} \ \forall k \\ \tau_k \leq \tau_k^{(\mathrm{rq})} \ \forall k. \end{cases} \tag{5.38}$$

The nature of the optimization parameters in $\mathfrak{M}$ implicitly poses further constraints to the optimization problem, which are not stated in (5.38):

$$C_k \in \{1, 2, \ldots, 15\}, \ \forall k \tag{5.39}$$

$$A_k \in \{2, 8\}, \ \forall k \tag{5.40}$$

where $A_k = 2$ and $A_k = 8$ correspond to the modulation alphabets for GMSK and 8-PSK, respectively. The partial derivatives of the Lagrangian corresponding to (5.38) extended by (5.39) and (5.40) are not defined as $C$ and $A$ are integer numbers. Hence the Kuhn-Tucker theorem does not apply and a conventional solution of (5.38) does not exist. The upcoming sections thus employ the convergence of the iterative procedure presented in Section 5.2 to derive an efficient algorithm that solves the cross-layer problem with arbitrary accuracy along the following algorithm:

**Algorithm 1**: CLARA in EGPRS

```
Initialization: i = 0, π_out[i] = 0;
repeat
    Increment: i ← i + 1;
    Update: π̂_out[i] = π_out[i − 1];
    Equivalent requirements: [ρ̃^(rq), τ̃^(rq)] = Υ^(3),−1([ρ^(rq), τ^(rq)], π̂_out[i]);
    Mode optimization through LUT;
    Computation of π_out[i];
until π_out[i] = π̂_out[i] + ε ;
Power control;
```

**Equivalent requirements**

In the given context $\boldsymbol{\Upsilon}^{(3)}$ can be inverted by eliminating the influence of $\pi_{\text{out}}$ from throughput and delay. Hence we can find the equivalent QoS parameters as

$$\tilde{\rho}^{(\mathrm{rq})} = \frac{1}{1 - \hat{\pi}_{\text{out}}[i]} \rho^{(\mathrm{rq})}, \tag{5.41}$$

$$\tilde{\tau}^{(\mathrm{rq})} = \max_{m^*} \ Tm^* \quad \text{s.t.:} \quad \sum_{n=1}^{\lfloor \tau^{(\mathrm{rq})}/T \rfloor} f_n[n] \geq \ (1 - \pi_\tau). \tag{5.42}$$

**Decoupled mode optimization**

Referring to Corollary 1 we can find SINR requirements $\boldsymbol{\gamma}^{(\mathrm{rq})}$ that will guarantee $[\boldsymbol{\rho}^{(\mathrm{rq})}, \boldsymbol{\tau}^{(\mathrm{rq})}]$ through a single lookup-table. This database can be built from system model components, cf. 5.3.1, or may be supported through numerical simulations or even measurements. This provides the optimum solution to the $K$ problems:

$$\{\mathcal{M}_k^*, \gamma_k^{(\mathrm{rq})}\} = \underset{\{\mathcal{M}_k,\gamma_k\}}{\operatorname{argmin}} \gamma_k \quad \text{s.t.:} \begin{cases} \tilde{\rho}_k \geq \tilde{\rho}_k^{(\mathrm{rq})} \ \forall k \\ \tilde{\tau}_k \leq \tilde{\tau}_k^{(\mathrm{rq})} \ \forall k \end{cases} \tag{5.43}$$

**Resource allocation**

Last stage within an iteration is the solution of the downlink power control problem:

$$\boldsymbol{P}^* = \underset{\boldsymbol{P}}{\operatorname{argmin}}\, \mathbf{1}^{\mathrm{T}}\boldsymbol{P} \quad \text{s.t.:} \quad \gamma_k \geq \gamma_k^{(\mathrm{rq})},\ \forall k \tag{5.44}$$

It is known that the corresponding solution is completely determined by the always active constraints. Defining the diagonal matrix of channel gains $\boldsymbol{R}$ and the full-rank coupling matrix $\boldsymbol{\Psi}$ through its elements as:

$$[\boldsymbol{\Psi}]_{k,\ell} = \begin{cases} -(1-\nu) & \text{for} \quad k \neq \ell \\ \frac{\chi}{\gamma_k^{(\mathrm{rq})} C_k} - (1-\nu) & \text{for} \quad k = \ell. \end{cases} \tag{5.45}$$

we obtain the optimum power allocation for the given SINR constraints as the solution of this linear equation system:

$$\boldsymbol{R}\boldsymbol{\Psi}\boldsymbol{P} = P_\eta \mathbf{1}. \tag{5.46}$$

With these powers the framework in [26] can be employed to calculate the channel outage probability and hence conclude the iteration by updating $\hat{\boldsymbol{\pi}}_{\text{out}}[i+1] = \boldsymbol{\pi}_{\text{out}}[i]$.

This provides a low complexity algorithm to minimize the necessary transmit powers subject to a set of constraints on throughput and delay. A service request therefore can be answered in a maximum economical way, which by means of scheduling algorithms like the EQoS scheme in [29] can be transformed into increased system capacity.

**Evaluation**

The results in this section show the large potential cross-layer techniques have to enhanced 2G systems. The underlying simulation environment investigates 1000 longterm settings, each consisting of 5000 time slots of length $T$. For every longterm realization, user locations are generated randomly, assuming a uniform distribution of users in the cell. The resulting path losses, i.e., the inverse of the variances $\sigma_{r_k}^2$, are derived from the distances $d$ between MS and BS using the Hata pathloss model as introduced in, e.g., [30, 31]. For every time slot in a longterm setting, channel coefficients are generated. An industrially-deployed FEC turbo code is used, providing the ACK and NACK messages to the fully implemented HARQ protocol. Recording all relevant parameters allows the evaluations of the QoS compliance, the power savings which are possible through the proposed technique. The numerical values for the used system components are displayed in Tab. 5.2. Within the MCS indexes the available modes of operation. In addition to the

| $\nu$ | $T_1 - T_3$ | $T_4$ | $\pi_\tau$ | MCS |
|---|---|---|---|---|
| $-18\ dB$ | 10 ms | 100 ms | 0.01 | $[1-9]$ |

**Table 5.2.** Numerical parameter values used in simulations

structural settings of the model in Section 5.3.1 the preparation of a packet transmission is assumed to cause a delay of $T_i$. As the standard requires the HARQ implementation in the periodic form mentioned in Footnote 13 with $i_{\max} = 3$ every 3th transmission faces a longer delay accounting for the round trip time to the RNC and back. We assumed a receiver noise level of $-95$ dBm and a maximum transmit power of 16 W, which together with an antenna gain of 18 dBi, results in an *effective isotropic radiated power* (EIRP) of 60 dBm.

In a first step, Fig. 5.4 proves the QoS compliance of the cross-layer scheme. Serving three users with QoS demands on throughput/delay of [9 kbps / 100 ms], [18 kbps / 100 ms] and [27 kbps / 100 ms] respectively, the simulation of 1000 user settings, i.e. location of the users in the cell, resulted in 1000 values for the throughput, obtained from averaging the data rate over the evaluated 1000 channel realizations. The lines in Fig. 5.4 show a histogram of these throughput realizations, revealing a superb match of requirement and de facto measurement. The constructed system model allows the algorithm to precisely control the QoS parameters, allowing for the targeted power minimization.

**Power savings**

As the philosophy employed to derive the proposed solution was the overall minimization of transmit power, this section elaborates on stochastic descriptions of the transmit power necessary to serve the mentioned 3 users with their

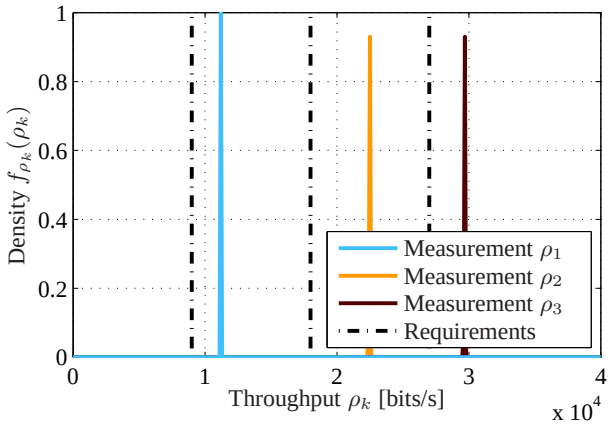



**Fig. 5.4.** Demonstration of the QoS compliance of the proposed cross-layer technique for 3 users with throughput demands of 9, 18 and 27 kbit/s respectively.

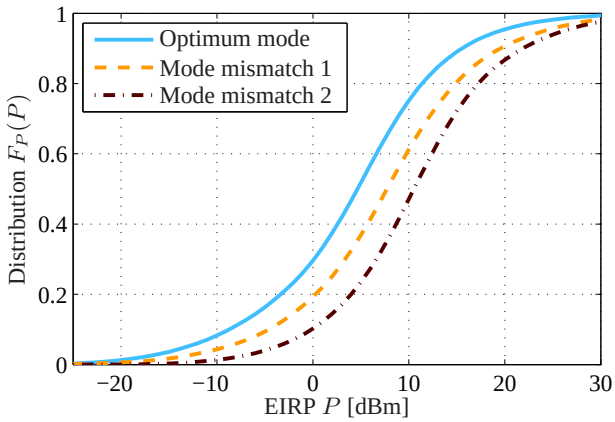



**Fig. 5.5.** Visualization of the power distributions obtained for 3 different users with different QoS requirements.

demands. Note that any power $P_k < P_{\max}$ is only possible by the above means of QoS management, i.e., top-down cross-layer optimization.

The cumulative distributions reveal that the above-discussed QoS compliance can be achieved with significantly less than maximum transmit power in a large number of cases. In $50\%$ of all time slots in the investigated example the QoS requirements can be met with only $40$ dB EIRP (cf. Fig. 5.5).

A more competitive reference can be obtained by providing the required QoS with a suboptimal mode. Exemplarily, Fig. 5.5 includes the PDFs resulting from a constant mode mismatch by selecting the mode corresponding to the next higher MCS value. The dash-dotted line in Fig. 5.5 shows the significantly larger power consumption of the resulting system configuration and thus visualizes the sensitivity of this performance measure with respect to mode selection.

### 5.3.2 Applying CLARA to HSDPA

A second exemplarily application of the proposed generic scheme shall be derived upon the background of HSDPA.

#### System model

This Section derives analytical expressions for the QoS parameters throughput $\rho_k$ and delay $\tau_k$ of user $k$. To tihs end we adopt the QoS definitions from Section 5.3.1 and define the mode of operation $\mathcal{M}_k$ of user $k$ as an element of the cartesian product:

$$\mathcal{M}_k = \{C_k, R_k, A_k\} \in \{1, 2, \ldots, 15\} \times [0; 1] \times \{4, 16\}. \tag{5.47}$$

The mode $\mathcal{M}_k$ consists of the number of employed CDMA code channels $C_k$, the code rate $R_k$ of the employed FEC code and the cardinality $A_k$ of the employed modulation alphabet $\mathcal{A}_k$. The union of all modes $\mathcal{M}_k$ is denoted as $\mathfrak{M} = \{\mathcal{M}_1, \ldots, \mathcal{M}_K\}$ and contains the modes of all users.

*Broadcast channels*

Let $K_t$ of the $K$ users share the same time slot $t$ and let the physical channel to user $k$ be modeled by frequency selective uncorrelated block Rayleigh fading. The channel coefficients are assumed constant over the period of one slot of length $T$.[14] The channel thus is characterized by an equivalent baseband complex valued impulse response of length $Q + 1$:

$$h_k[u] = \sum_{q=0}^{Q} h_{k,q}\delta[u - q], \quad h_{k,q} \in \mathbb{C}, \tag{5.48}$$

where the real and imaginary parts of the channel coefficients $h_{k,q}$ are distributed according to a zero mean Gaussian *probability density function* (PDF) of variance $\sigma^2_{k,q}$. Moreover, the channel realizations for different paths $q$, for different users and for different slots are mutually independent. We assume additive complex noise at the receiver to model intercell interference as well as thermal noise within the receiver radio components. We assume white Gaussian noise with circulary symmetric complex Gaussian distribution $\mathcal{CN}(0, P_\eta)$. Without loss of generality, we assume equal noise powers among the mobile stations.[15]

**Signal to noise and interference ratio**

In order to establish a scalable SINR metric let us first investigate the maximum available signal power at the receiver output[16]. This optimum is obtained by the *maximum ratio combining* (MRC) principle. The power gain of the resulting impulse response $(h_k[u] * g_k[u])$ is defined as $r_k = \sum_{q=0}^{Q} |h_{k,q}|^2$.[17]Then, the signal power of user $k$ in the MRC CDMA system can be written as:

$$P_{\mathrm{S}} = \chi r_k \frac{1}{C_k} P_k, \tag{5.49}$$

where $\chi$, $P_k$ and $C_k \in \mathcal{M}_k$ denote the spreading factor, the total transmit power of user $k$ and the number of different data streams. On the basis of, e.g., [36] a linear expression for the interference power along:

$$P_{\mathrm{I}} = \sum_{\ell=1}^{K_t} (1 - \nu) r_k P_\ell, \tag{5.50}$$

can be obtained by normalizing these interference power components to the product of channel gain $r_k$ and transmit power $P_\ell$.[18] Hence, the SINR expression for the $K_t$ users that are active in time slot $t$ can be obtained from e.g. [36] and is given as:

$$\gamma_k = \frac{\chi r_k \frac{1}{C_k} P_k}{\sum_{\ell=1}^{K_t} (1 - \nu) r_k P_\ell + P_\eta}. \tag{5.51}$$

The numerator includes the signal receive power accounting for the transmit power per stream $\frac{1}{C_k} P_k$, the spreading gain $\chi$ and the channel power gain $r_k$. The denominator introduces the noise power $P_\eta$ and the $\nu$ based expression for the interference through all $C_\ell$ streams $(1 - \nu) r_k P_\ell$ from user $\ell$. As the considered Walsh codes are strictly orthogonal in synchronous use, inter-symbol interference and inter-stream interference equal $(1 - \nu) r_k \frac{1}{C_\ell} P_\ell$.

[14] Standardization documents refer to this time instance as one *transmission time interval* (TTI).
[15] The extensions to different receive noise levels in all cases are straight forward.
[16] Confer with publications like [32, 33, 34] where comprehensive SINR discussions do not include the adaptive receive filter.
[17] For details on the distribution $F_r(r)$ we refer to [24, 35]
[18] In general any coupling function that fulfills the properties of positive semidefinitnes monotonicity and scalability as demanded in [20] can be used as long as the outage expressions in (5.24) can be evaluated.

**Channel outage probability**

The outage probability can be formulated through the SINR $\gamma$:

$$\pi_{\text{out}} = \Pr\left(\gamma_k < \gamma_k^{(\text{rq})}\right). \tag{5.52}$$

It can be seen from (5.51) that this probability depends upon the choice of all transmit powers $P_\ell$, $\ell = 1, \ldots, K_t$. For constant choices of $P_\ell$ the authors in [25] have derived a way to compute $\pi_{\text{out}}$ for general fading channels. For non-constant power allocation, i.e.,fast link adaptation as performed in HSDPA, the we refer to [23, 26].

*FEC coding*

Theorem 5 introduces the wording of the 'noisy channel coding theorem' as it was called in [37]:

**Theorem 5.** *For a discrete memoryless channel with the corresponding $R_0(\gamma_k)$ value and an input alphabet $\mathcal{A}_k$ of cardinality $A_k$ there always exists a block code with block length $n_{eq}$ and binary code rate $R_k \operatorname{ld} A_k \leq R_0(\gamma_k)$ in bits per CDMA channel use, so that with maximum likelihood decoding the error probability $\tilde{\pi}_{pe}$ of a code word can be expressed as:*

$$\tilde{\pi}_{pe} \leq 2^{-n_{eq}(R_0(\gamma_k) - R_k \operatorname{ld} A_k)}. \tag{5.53}$$

The theorem was formulated for rates $R_k \operatorname{ld} A_k \leq R_0(\gamma_k)$. Throughout the remainder of this work we will implicitly assume the trivial continuation $\tilde{\pi}_{\text{pe}} \leq 1$ for $R_k \operatorname{ld} A_k > R_0(\gamma_k)$. For derivations and a compact proof of the cutoff rate theorem we refer to [37, 38, 39]. For SINR $\gamma_k$ and a modulation alphabet $\mathcal{A}_k = \{a_1, \ldots, a_q\}$ of cardinality $A_k$, the results from Section 5.3.2 and the derivations in [40, 41] compute $R_0$ as:

$$R_0(\gamma_k) = \operatorname{ld} A_k - \operatorname{ld}\left[1 + \frac{2}{A_k} \sum_{m=1}^{A_k - 1} \sum_{\ell=m+1}^{A_k} \exp\left(-\frac{1}{4} |a_\ell - a_m|^2 \gamma_k\right)\right]. \tag{5.54}$$

Despite certain efforts in the literature the cutoff rate theorem has not yet been formulated explicitly for turbo-decoded convolutional codes [42, 43]. For this class of codes that is highly relevant in HSDPA settings we introduce the concept of equivalent block lengths. Let $B$ be the interleaver length between inner and outer code in a concatenated encoding system, i.e., the number of information bits per packet in a system with chip rate $R_\chi$:

$$B = R_k \operatorname{ld} A_k \, C_k \, \frac{R_\chi}{\chi} T. \tag{5.55}$$

Then the performance of the turbo decoded convolutional code in a very good approximation equals the performance of a block code with block length:

$$n_{\text{eq}} = \beta_{n_{\text{eq}}} \ln B. \tag{5.56}$$

The logarithmic dependence can be deduced from information theoretic results [44, 45]. The parameter $\beta_{n_{\text{eq}}}$ in (5.56) can be used to adapt this model to the specifics of the employed turbo code. For each block of length $n_{\text{eq}}$ Theorem 5 applies. The packet error probability therefore can be expressed by (5.53) through the Bernoulli formulation:

$$1 - \pi_{\text{pe}} = (1 - \tilde{\pi}_{\text{pe}})^{\frac{B}{n_{\text{eq}}}}. \tag{5.57}$$

Validating the derived expression Fig. 3.3 and Fig. 3.4 in [23] compare (5.57) with link level simulations of an industrially deployed turbo-code [46] and demonstrate the superb match between theoretical model and practical application. Equation (5.57) thus can be regarded a very valid model for the performance of turbo-decoded convolutional codes.

*HARQ protocols*

In general a large variety of packet combining techniques based on hard decisions [47, 48, 49] or soft channel outputs [50, 51, 52] have been proposed in the literature. For the sequel we will focus on Full Incremental Redundancy methods. Thus, the system operates with variable code rates that can be obtained through e.g., rate compatible punctured convolutional codes [52] or punctured turbo codes [53]. Let $\tilde{B}$ be the total number of bits per packet. Then the code rate in the $m$th transmission is given by:

$$R_k[m] = \frac{B}{m\tilde{B}} = \frac{1}{m} R_k[1] = \frac{1}{m} R_k. \tag{5.58}$$

Thus, the cutoff-rate theorem directly applies to the $m$th retransmission when accounting for the resulting rate of the FEC code. The packet error probability of an IR HARQ in the $m$th transmission thus can be written as:

$$\pi_{\text{pe}}[m] = 1 - \left(1 - 2^{-\beta_{n_{\text{eq}}} \ln B (R_0(\gamma_k) - \frac{1}{m} R_k \ \text{ld} A_k)}\right)^{\frac{B}{\beta_{n_{\text{eq}}} \ln B}}. \tag{5.59}$$

The above expressions (5.59) extend the formulated cutoff rate based modeling of FEC codes to the case of hybrid ARQ protocols. With these results we can now derive the probability $f_m[m]$ that it takes $m$ transmissions to decode a packet error free, i.e., the description of the error control layers that was proposed in Section 5.2. The probability $f_m[m]$ therefore is given by:

$$f_m[m] = \left(\prod_{m'=1}^{m-1} \pi_{\text{pe}}\left[m'\right]\right) \left(1 - \pi_{\text{pe}}[m]\right). \tag{5.60}$$

This distribution features an important property that gives reason to a prominent approximation:

**Lemma 1.** *For asymptotically large block lengths $\tilde{B}$ the probability $f_m[m]$ can be written as a unit impulse:*

$$\lim_{B \to \infty} f_m[m] = \delta[m - m^*] = \begin{cases} 1 \text{ for } m = m^*, \\ 0 \text{ else}. \end{cases} \tag{5.61}$$

The corresponding proof is given in detail in [23]. Due to the fast asymptotic convergence of the exponential function the limit provided in (5.61) can be used as a good approximation for $f_m[m]$ even in settings with large but finite block lengths $B$ and is considered valid for the remainder of this Section.

*MAC scheduling*

This section will present the means to include a given MAC scheduling strategy into the cross-layer system model. From the large variety of time domain scheduling algorithms we consider the three most prominent ones in their basic versions to exemplarily demonstrate the modeling approaches proposed in this section.

*Round Robin scheduling* The Round Robin approach schedules the users one after another in a predefined order. Within every round each users is granted access to the channel once. The strategy $\mathcal{S}$ containing the users that are scheduled to the time slot $t$ can be formulated as:

$$\mathcal{S}_{\text{rr}} = \left\{ k \,\middle|\, \frac{t-k}{K} \in \mathbb{N}_0 \right\}. \tag{5.62}$$

The probability of being selected by this scheme therefore is completely determined by the number of users $K$ in the system. As the scheduling decision is independent of the channel realizations so are the probabilities of being scheduled and of facing a feasible channel. Hence, the transmission probability can be written as their product and reads:

$$\pi_{\text{tx}} = \frac{1}{K}(1 - \pi_{\text{out}}). \tag{5.63}$$

*Proportional fair scheduling* Defining the constant scalar weights $\beta_{\text{pf}} \in [0; 1]$ that allow to trade multi-user gain for fairness we investigate MAC scheduling with the decision rule:

$$\mathcal{S}_{\text{pf}} = \left\{ k \,\middle|\, r_k \left(1 - \left(1 - \frac{1}{\text{E}[r_k]}\right) \beta_{\text{pf}}\right) > r_\ell \left(1 - \left(1 - \frac{1}{\text{E}[r_\ell]}\right) \beta_{\text{pf}}\right),\ \forall \ell \neq k \right\}. \tag{5.64}$$

The substitution $\beta_{\text{pf},k} = \left(1 - \left(1 - \frac{1}{\text{E}[r_k]}\right) \beta_{\text{pf}}\right)$ yields a more compact and more convenient form of the proportional fair scheduling as:

$$\mathcal{S}_{\text{pf}} = \left\{ k \,|\, r_k \beta_{\text{pf},k} > r_\ell \beta_{\text{pf},\ell},\ \forall \ell \neq k \right\}. \tag{5.65}$$

As $\beta_{\text{pf},k} r_k$ is a linear function of the random variable $r_k$ its probability density can be computed through $f_r(r)$ as $\frac{1}{\beta_{\text{pf},k}} f_{r_k}\left(\frac{r_k}{\beta_{\text{pf},k}}\right)$. The probability $\pi_{\text{s}}$ now can be written in terms of the derived density of $\beta_{\text{pf},k} r_k$:

$$\pi_{\text{s}} = \int_0^\infty \frac{1}{\beta_{\text{pf},k}} f_{r_k}\left(\frac{r_k}{\beta_{\text{pf},k}}\right) \left( \prod_{\substack{\ell=1 \\ \ell \neq k}}^{K} F_{r_\ell}\left(\frac{r_\ell}{\beta_{\text{pf},\ell}}\right) \right) \mathrm{d}r_k. \tag{5.66}$$

Hence, the transmission probability can be expressed through the product $\pi_{\text{tx}} = (1 - \pi_{\text{out}})\pi_{\text{s}}$.

With these extensive preliminaries on the exemplarily chosen MAC scheduling approaches the description of the lower layers through $f_m[m]$ in (5.60) can be used to formulate the probability $f_n[n]$ that it takes $n$ slots to transmit a packet error free. Employing the tight bound (5.61) as it was proven in Section 5.3.2 yields :

$$f_n[n] = f_{n|m^*}[n] = (1 - \pi_{\text{tx}})^{(n-m^*)} \pi_{\text{tx}}^{m^*} \binom{n-1}{m^*-1}, \quad n \leq m^*. \tag{5.67}$$

From this the QoS parameters throughput $\rho_k$ and outage delay $\tau_k$ can be obtained through the properties of the probability in (5.67). The upcoming paragraphs will derive the corresponding details. To this end the expectation of $n$ can be obtained as:

$$\begin{aligned} \mathrm{E}[n] &= \sum_{n=m^*}^{\infty} n(1 - \pi_{\text{tx}})^{(n-m^*)} \pi_{\text{tx}}^{m^*} \binom{n-1}{m^*-1}, \\ &= \frac{m^*}{\pi_{\text{tx}}}. \end{aligned} \tag{5.68}$$

The corresponding derivation through the calculus of hypergeometric functions [54] can be obtained from [55]. With this central parameter of the probability in (5.67) the definition of the throughput $\rho$ from (5.21) can be computed as:

$$\rho_k = \frac{1}{\mathrm{E}[n]} \frac{B}{T}. \tag{5.69}$$

$B$ denotes the number of information bits per packet and can be computed along (5.55) through:

$$B = R_k \, \mathrm{ld} A_k \, C_k \, \frac{R_\chi}{\chi} T. \tag{5.70}$$

With (5.22) a $f_n[n]$ based definition of the second QoS parameter delay was given. Through the expression in (5.67) the $\pi_\tau$ outage delay $\tau_k$ thus can be written as:

$$\tau_k = \underset{\tau'}{\operatorname{argmin}} \, \tau' \quad \text{s.t.:} \quad \sum_{n=1}^{\lfloor \tau'/T \rfloor} f_n[n] \geq 1 - \pi_\tau. \tag{5.71}$$

These equations conclude the derivation of the bottom-up system model. It can be easily verified that this model complies with the made propositions in Section 5.2.

**Optimization**

Within this Section the cross-layer optimization in HSDPA is formulated in accordance to the considerations in Section 5.2 as a resource minimization problem subject to a set of constraints on the QoS parameters throughput $\rho_k$ and delay $\tau_k$. The cross-layer optimum mode of operation $\boldsymbol{\mathcal{M}}^*$ and the corresponding optimum transmit powers $\boldsymbol{P}^*$ therefore are the solutions to the minimization program:

$$\{\boldsymbol{\mathcal{M}}^*, \boldsymbol{P}^*\} = \underset{\{\boldsymbol{\mathcal{M}}, \boldsymbol{P}\}}{\operatorname{argmin}} \, \|\boldsymbol{P}\|_1 \quad \text{s.t.:} \begin{cases} \rho_k \geq \rho_k^{(\text{rq})} \ \forall k \\ \tau_k \leq \tau_k^{(\text{rq})} \ \forall k. \end{cases} \tag{5.72}$$

Due CLARA compliant nature HSDPA system models as introduced above the structure from Subsection 5.3.1 can be adapted through two modifications:

- Replace $\pi_{\text{out}}[i]$ in the inversion of $\boldsymbol{\Upsilon}^{(3)}$ by $(1 - \pi_{\text{tx}}[i])$.
- Use the HSDPA model components or corresponding measurements for building the look-up-table required to solve the mode optimization in the iterative scheme.

This provides the converging optimum scheme for HSDPA.

**Scheduling**

In [56] techniques have been proposed to use the cross-layer knowledge of CLARA and her ability to control QoS for MAC scheduling purposes. These *cross-layer assisted scheduling* (CLASS) algorithms consist of a sequence of hypothetical scheduling sets with following feasibility test through CLARA. These techniques will be used in the upcoming evaluation to demontrate CLARA benefits in terms of system capacity as well.

**Evaluation**

The performed HSDPA cell level evaluations of the proposed cross-layer optimization technique shall support the theoretical elaborations from above and demonstrate the applicability of the optimization, as well as the resulting system enhancements. The underlying simulation environment investigates 1000 longterm settings, each consisting of 5000 time slots of length $T$. For every longterm realization, user locations are generated randomly, assuming a uniform distribution of users in the cell. The resulting path losses, i.e., the inverse of the variances $\sigma^2_{k,q}$, are derived from the distances $d$ between MS and BS using the Hata pathloss model as introduced in [30, 31] and referenced in [57]. For every time slot in a longterm setting, channel coefficients are generated to fulfill the made stochastic model. With the metrics described above, the scheduling algorithms are implemented and decide which user is granted access to its channel for every channel realization. An industrially-deployed FEC turbo code is used, providing the ACK and NACK messages to the fully implemented HARQ protocol. Recording all relevant parameters allows the evaluations of the QoS compliance, the power savings and the capacity improvements which are possible through the proposed technique. The numerical values for the used system components are displayed in Tab. 5.3. In addition to the structural settings of the model in Section 5.3.2 the preparation of a packet transmission is assumed to cause a delay of $T_i$. As

| $\chi$ | $\nu$ | $T_1 - T_3$ | $T_4$ | $Q$ | $_\Delta\gamma$ @ 4QAM | $_\Delta\gamma$ @ 16QAM | $\pi_\tau$ | $\beta_{n_{\text{eq}}}$ | CQI | $\beta_{\text{pf}}$ |
|---|---|---|---|---|---|---|---|---|---|---|
| 16 | 0.1 | 10 ms | 100 ms | 3 | $[3, 4.8, 6]$ dB | $[2, 3.8, 5]$ dB | 0.01 | 32 | $[1 - 30]$ | 1 |

**Table 5.3.** Numerical parameter values used in simulations

the standard requires the HARQ implementation in the periodic form mentioned in Footnote 13 with $i_{\max} = 4$ every 4th transmission faces a longer delay accounting for the round trip time to the RNC and back. We assumed a receiver noise level of $-95$ dBm and a maximum transmit power of 16 W, which together with an antenna gain of 18 dBi, results in an *effective isotropic radiated power* (EIRP) of 60 dBm.

### 5.3.3 QoS Compliance

In a first step, Fig. 5.6 proves the QoS compliance of the cross-layer scheme. Serving three users with QoS demands on throughput/delay of [200 kbps / 100 ms], [500 kbps / 100 ms] and [1 Mbps / 100 ms] respectively, the simulation of 1000 user settings, i.e., location of the users in the cell, resulted in 1000 values for the throughput, obtained from averaging the data rate over the evaluated 5000 channel realizations. The lines in Fig. 5.6 show a histogram of these throughput realizations, revealing a superb match of requirement and de facto measurement. The constructed system model allows the algorithm to precisely control the QoS parameters, allowing for the targeted power minimization.

In the vast majority of all user settings, the MT scheduling algorithm does not allow the QoS true serving of all users for which the corresponding MTS curve is missing. As the delay demands in this setting are implicitly accomplished by fulfilling the throughput demands, the corresponding curves are omitted here.

### 5.3.4 Power Advantage

Because the philosophy employed to derive the proposed solution was the overall minimization of transmit power, this section elaborates on stochastic descriptions of the transmit power necessary to serve the mentioned 3 users with their demands. Note that any power $P_k < P_{\max}$ is only possible by the above means of QoS management, i.e., top-down cross-layer optimization. Reference techniques will therefore always employ the full available transmit power. Against this background Fig. 5.7 displays the cumulative distribution of the EIRP measured during the above mentioned simulations. The cumulative distribution reveal that the above-discussed QoS compliance can be achieved with significantly less than maximum transmit power in a large number of cases. Due to the large scale of path loss, shadowing and fading effects, the resulting gains are significant. Although using the full available EIRP is a coarse

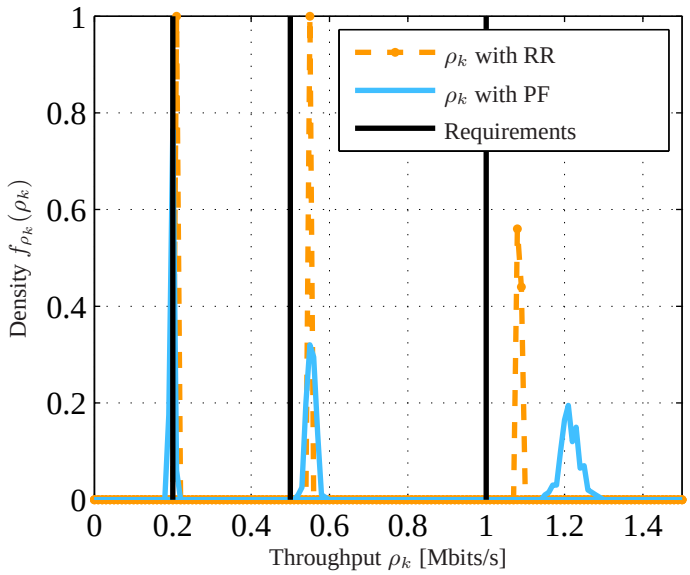



**Fig. 5.6.** Demonstration of the QoS compliance of the proposed cross-layer technique for 3 users with throughput demands of 0.2, 0.5 and 1Mbit/s respectively.

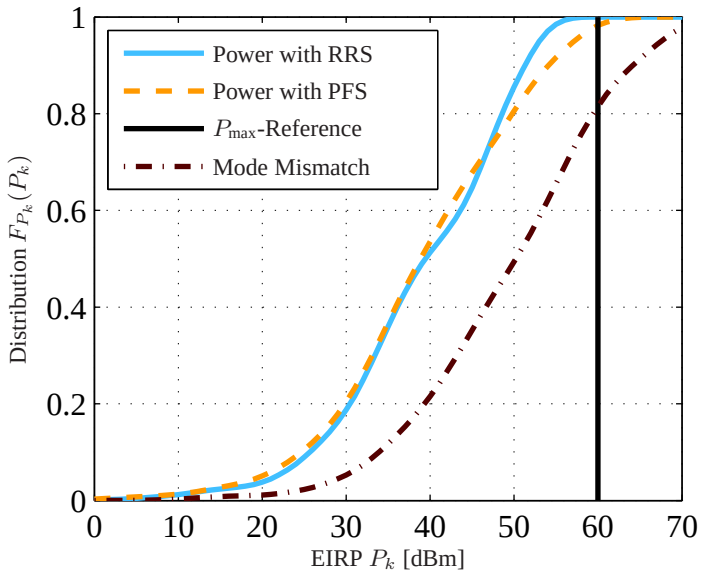



**Fig. 5.7.** Visualization of the sensitivity with respect to mode mismatches: Comparison of the EIRP over all users.

reference, it is a useful landmark, as Section 5.3.5 will confirm. However a more competitive reference can be obtained by providing the required QoS with a suboptimal mode. Exemplarily, Fig. 5.7 includes the PDF resulting from a constant mode mismatch in a Round Robin scheme. Indexing the modes with the CQI, the 'mode mismatch' reference approach does not select the optimal mode with respect to (5.72), but instead selects the mode corresponding to the next higher CQI value. The dash-dotted line in Fig. 5.7 shows the significantly larger power consumption of the resulting system configuration and thus visualizes the sensitivity of this performance measure with respect to mode selection. The increased power demand of the higher mode additionally increases the channel outage probability, which makes further power increases necessary. Through this iterative connection, the PDF shift in Fig. 5.7 significantly exceeds the SINR gap between neighboring modes. This result provides another strong argument for the necessity of cross-layer approaches, as this relation is not regarded in conventional single-layer models.

### 5.3.5 Capacity Increase

Finally let us demonstrate how the saved power can contribute to maximize the system capacity in terms of servable users. To this end Fig. 5.8 plots the distribution of the number of users that can be scheduled with different scheduling approaches comparing CLARA cross-layer methods as mentioned in Section 5.3.2 with MT and PF scheduling in their basic versions, always scheduling the first $K$ users from a prioritized queue.

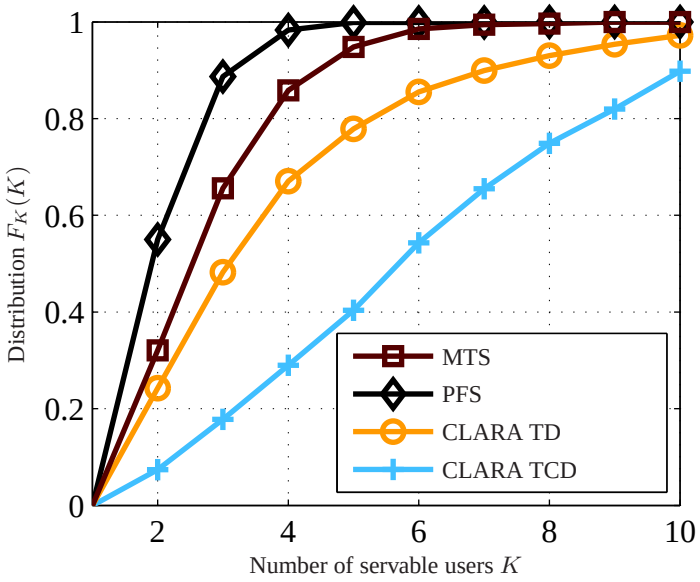



**Fig. 5.8.** Demonstration of the possible capacity gains, obtainable through cross-layer assisted scheduling.

Fig. 5.8 shows the resulting distribution obtained from the simulation of 1000 scenarios. Assuming a queue of 15 users with descending priorities, their location as well as their QoS requirements for every scenario are randomly chosen. Locations are distributed uniformly across the cell with a radius of 1.5 km, throughput requirements have uniformly distributed square root in the interval of $\sqrt{[0; 12.8]}$Mbps and the latency requirements are uniformly distributed in [20 ms; 100 ms].

With the drastic power saving obtained through the proposed optimization, many time slots allow the additional scheduling of users on free CDMA code channels. For the given traffic scenario, the top-down cross-layer optimization of the HSDPA link allows us to double the median of the system capacity, even with respect to the already optimized CLARA TD scheme. This results illustrates the enormous potential of cross-layer assisted schemes to the scheduling unit of HSDPA systems.

## 5.4 Conclusion

This chapter has introduced a generic approach for the top-down cross-layer optimization of communication systems. It bases on an iterative detection of the system's outage probability and relies on a series of table look-ups to obtain the optimum solution at an absolute minimum of computational cost. Because the initial formulation was based on an abstract set of mathematical properties the evolved scheme can be proven optimum for a wide class of systems. With the application of this scheme to EGPRS and HSDPA the latter part of this chapter has demonstrated the enormous potentials the presented technique has in these prominent state of the art systems.

## References

1. D. MacKinnon, W. McCrum, and D. Sheppard, *An Introduction to Open Systems Interconnection*. Computer Science Press, 1990, iSBN: 0716781808.

2. B. N. Jain and A. K. Agrawala, *Open Systems Interconnection, Its Architecture and Protocols*. ELSEVIER, 1990, iSBN: 0444884904.
3. D. N. C. Tse, "Optimal power allocation over parallel Gaussian channels," in *Proc. Int. Symp. Information Theory, Ulm, Germany*, June 1997.
4. ——, *Transmitter directed, multiple receiver system using path diversity to equitably maximize throughput*, May 24 1999, patent filed.
5. G. Song and Y. G. Li, "Cross-layer optimization for OFDM wireless networks - part I: theoretical framework," *IEEE Transactions on Wireless Communications*, vol. 4, no. 2, pp. 614–624, March 2005.
6. ——, "Cross-layer optimization for OFDM wireless networks - part II: algorithm development," *IEEE Transactions on Wireless Communications*, vol. 4, no. 2, pp. 625–634, March 2005.
7. J. K. MacKie-Mason and H. R. Varian, "Pricing Congestible Network Resources," *IEEE Journal on Selected Areas in Communications*, vol. 13, no. 7, pp. 1141–1149, September 1995.
8. S. A. C. Gonzalez, L. Szczecinski, "Throughput maximization of ARQ transmission protocol employing adaptive modulation and coding," in *Proc. IEEE and IEE International Conference on Telecommunications (ICT)*, August 2004, pp. 607 – 615, fortaleza Ceara, Brazil.
9. C. A.-H. J. Lopez-Vicario, "Joint transmit antenna selection and adaptive modulation in cross-layer oriented designs for HS-DPA systems," in *Proc. IEEE Statistical Array and Multi-channel Signal Processing Workshop (SAM04) Sitges*, July 2004, barcelona, Spain.
10. F. Fitzek, A. Koepsel, A. Wolisz, M. Krishnam, and M. Reisslein, "Providing application-level QoS in 3G/4G wireless systems: a comprehensive framework based on multi-rate CDMA," in *Proc. of IEEE 3Gwireless*, 2001, pp. 344 – 349, san Francisco, United States.
11. J. Brehmer and W. Utschick, "Physical layer characterization of MIMO systems by means of multiobjective optimization," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, March 2005, philadelphia, United States.
12. L.-U. Choi, M. T. Ivrlac, E. Steinbach, and J. A. Nossek, "Bottom-Up Approach to Cross-layer Design for Video Transmission over Wireless Channels (*sic*)," in *Proceedings of the IEEE 2005 Spring Semiannual Vehicular Technology Conference*, May - June 2005, stockholm, Sweden.
13. B. Bougard, S. Pollin, A. Dejonghe, F. Catthoor, and W. Dehaene, "Cross-layer power management in wireless networks and consequences on system-level architecture," *EURASIP Signal Processing Journal*, 2005.
14. R. Min and A. Chandrakasan, "A framework for energy-scalable communication in high-density wireless networks," in *ISLPED '02: Proceedings of the 2002 International Symposium on Low Power Electronics and Design*, 2002, pp. 36–41, monterey, California, USA.
15. R. A. Berry and R. G. Gallager, "Communication over fading channels with delay constraints," *IEEE Transactions on Information Theory*, vol. 48, no. 5, pp. 1135–1149, May 2002.
16. C. E. Shannon, "A Mathematical Theory of Communication," *Bell Sys. Tech. J.*, vol. 27, pp. 379–423, 623–656, 1948.
17. B. Zerlin and J. A. Nossek, "A Generic Approach to Cross-Layer Assisted Resource Allocation," March 2006, accepted for publication in the proceedings of the IEEE ITG Workshop on Smart Antennas.
18. M. T. Ivrlac, "Mimo systems - models, measures and algorithms," Ph.D. dissertation, Munich University of Technology, 2005.
19. L.H.Ozarow, S. Shamai, and A.D.Wyner, "Information theoretic considerations for cellular mobile radio," *IEEE Transactions on Vehicular Technology*, vol. 43, no. 2, pp. 359–378, May 1994.
20. R. D. Yates, "A Framework for Uplink Power Control in Cellular Radio Systems," *IEEE Journal on Selected Areas in Communications*, vol. 13, no. 7, pp. 1341–1347, 1995.
21. H. Boche, M. Schubert, S. Stanczak, and M. Wiczanowski, "An axiomatic approach to resource allocation and interference balancing," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*.
22. T. Halonen, J. Romero, and J. Melero, *GSM, GPRS and EDGE performance*. John Wiley & Sons, Ltd, 2003.
23. B. Zerlin, "CLARA - Cross-layer assisted resource allocation," Dissertation, Technische Universität München, June 2006, iSBN xxx, Shaker Verlag.
24. S. B. P. A. M. Mathai, *Quadratic Forms In Random Variables*, ser. Statistics: Textbooks and Monographs. Marcel Dekker, Inc., 1992, vol. 126.
25. Y. c. Ko, M. S. Alouini, and M. K. Simon, "Outage probability of diversity systems over generalized fading channels," *IEEE Transactions on Communications*, vol. 48, no. 11, pp. 1783–1787, November 2000.
26. B. Zerlin, J. Brehmer, and J. A. Nossek, "On the channel outage probability of multi-user time diversity systems," in *Proceedings of the 16th Annual IEEE International Symposium on Personal Indoor and Mobile Radio Communication (PIMRC)*, September 2005, berlin, Germany.
27. R. G. Gallager, "A simple derivation of the coding theorem and some applications," vol. IT-11, pp. 3–18, 1965.
28. G. Caire and D. Tunietti, "The throughput of hybrid-ARQ protocols for the Gaussian collision channel," *IEEE Transactions on Information Theory*, vol. 47, no. 5, pp. 1971–1988, July 2001.
29. D. Fernandez and H. Montes, "An Enhanced Quality of Service Method for Guaranteed Bitrae Services over Shared Channels in EGPRS Systems," in *Proc. 55th IEEE Vehicular Technology Conference (VTC Spring)*, May 2002.
30. M. Hata, "Empirical Formula for Propagation Path-Loss in Land Mobile Radio Services," in *Proceedings of the IEEE Vehicular Technology Conference*, vol. 29. IEEE, August 1980, pp. 317–325.
31. Y. Oda, K. Tsunekawa, and M. Hata, "Advanced LOS path-loss model in microcellular mobile communications," in *Proceedings of the Vehicular Technology Conference*, vol. 49, no. 6. IEEE, November 2000, pp. 2121–2125.

32. M. Schubert and H. Boche, "Iterative multiuser uplink and downlink beamforming under SINR constraints," *IEEE Transactions on Signal Processing*, vol. 53, no. 7, pp. 2324–2334, July 2005.
33. D. Catrein, L. Imhof, and R. Mathar, "Power Control, Capacity, and Duality of Up- and Downlink in Cellular CDMA Systems," *IEEE Transactions on Communications*, vol. 52, no. 10, pp. 1777–1785, 2004.
34. L. Imhof and R. Mathar, "Capacity Regions and Optimal Power Allocation for CDMA Cellular Radio," *IEEE Transactions on Information Theory*, vol. 51, no. 6, pp. 2011–2019, June 2005.
35. G. E. P. Box, "Some theorems on quadratic forms applied in the study of analysis of variance problems," *The Annals of Mathematical Statistics*, vol. 1, no. 25, pp. 290–302, 1954.
36. T. Frey, "Analyse und Vergleich von Codierverfahren für Mobilfunksysteme mit höherstufigem Codemultiplex," Ph.D. dissertation, Universität Ulm, Abteilung Informationstechnik, 2000.
37. R. G. Gallager, *Information Theory and Reliable Communication*. John Wiley and Son, 1968, iSBN: W-471-29048-3.
38. R. Mathar, *Informationstheorie*. Augustinus Buchh., Aachen, Mai 1998.
39. B. Friedrichs, *Kanalcodierung*, J. H. H. Marko, Ed. Springer Verlag, 1995, no. ISBN 3-540-58232-0.
40. M. Ivrlac, T. Kurpjuhn, C. Brunner, and W. Utschick, "Efficient use of fading correlations in MIMO systems," in *Proceedings of the 54th IEEE Vehicular Technology Conference*, 2001, pp. 2763–2767, atlantic City, NJ, USA.
41. T. Kurpjuhn and M. Ivrlac, "Information Theoretic Measures for MIMO-Channels," Lehrstuhl für Netzwerktheorie und Signalverarbeitung, Technische Universität München, Tech. Rep. TUM-LNS-TR-01-01, 2001.
42. C. Berrou, A. Glavieux, and P. Thitimajshima, "Near Shannon limit error-correcting coding and decoding: Turbo codes," in *Proceedings of the International Conference on Communications*. IEEE, March 1993, pp. 1064–1070, geneva, Switzerland.
43. S. Benedetto, D. Divsalar, and J. Hagenauer, Eds., *Special Issue on Concatenated Coding Techniques and Iterative decoding: Sailing towards Capacity*, ser. Journal on Selected Areas in Communications. IEEE, February 1998, vol. 16.
44. B. Zerlin, M. Bossert, and J. A. Nossek, "On the logarithmic dependence of the performance of Turbo-codes with respect to interleaver lengths," February and March 2005, private discussions in Karlsruhe and Munich.
45. M. T. Ivrlac, "Effective and Pragmatic Code Length of Convolutional Codes," Munich University of Technology, Institute for Network Theory and Signal Processing, Tech. Rep., 2004.
46. B. Zerlin, M. T. Ivrlač, W. Utschick, and J. A. Nossek, "Joint Optimization of Radio Parameters," TUM-NWS, Tech. Rep. 6, October 2004, final Report on the Research Cooperation JORP with the Siemens AG.
47. P. S. Sindhu, "Retransmission Error Control with Memory," *IEEE Transactions on Communications*, vol. 25, pp. 473–379, 1977.
48. G. Benelli, "An ARQ Scheme with Memory and Soft Error Detection," *IEEE Transactions on Communications*, vol. 33, pp. 285–288, 1985.
49. S. B. Wicker, "Adaptive Error Control through the Use of Diversity Combining Majority Logic Decoding in Hybrid ARQ Protocol," *IEEE Transactions on Communications*, vol. 39, pp. 380–385, 1991.
50. D. Chase, "Code Combining - A Maximum Likelihood Decoding Approach for Combining an Arbitrary Number of Noise Packets," *IEEE Transactions on Communications*, vol. 33, no. 5, pp. 385–393, May 1985.
51. B. A. Harvey and S. B. Wicker, "Packet combining system based on the Viterbi decoder," *IEEE Transactions on Communications*, vol. 42, pp. 1544–1557, 1994.
52. S. Kallel, "Analysis of type-II hybrid ARQ scheme with code combining," *IEEE Transactions on Communications*, vol. 38, pp. 1133–1137, 1990.
53. K. R. Narayanan and G. L. Stuber, "A Novel ARQ Technique Using the Turbo Coding Principle," *IEEE Communications Letters*, vol. 1, pp. 49–51, 1997.
54. P. Koev and A. Edelman, "The Efficient Evaluation of the Hypergeometric Function of a Materix Argument," may 2005, submitted to Mathematics of Computation.
55. M. Abramowitz and I. Stegun, Eds., *Handbook of Mathematical Functions*. United States Department of Commerce, 1964.
56. B. Zerlin, M. T. Ivrlač, J. A. Nossek, and A. Klein, "On cross-layer assisted resource allocation in HSDPA," in *Proceedings of the 12th International Conference on Telecommunication*. IEEE, May 2005, cape Town, South Africa.
57. G1, "TS 03.30 8.3.0 - Cell Sizes and RF Budgets," Third Generation Partnership Project, Tech. Rep., 2000.

# 6

# Power Line Communications: A Promising Communication System's Paradigm for Last Miles and Last Meters Applications

Moisés Vidal Ribeiro[1]

Electrical Circuit Department, Engineering Faculty, Federal University of Juiz de Fora, Campus Universitário, Bairro Martelos, 36 036 330, Juiz de fora MG, Brazil. E-mail: mribeiro@ieee.org
This work was supported in part by CNPq under Grant 552371/01-7 and Grant 150064/2005-5, Brazil.

## 6.1 Introduction

Recently, increasing demand for fast Internet access, new multimedia services, surveillance networks, sensor networks and simple and powerful automation and monitoring networks, the development of new and feasible signal processing techniques associated with faster and low-cost digital signal processors, as well as the deregulation of the telecommunication market have placed major emphasis on the value of investigating hostile media, such as *power line* (PL) channels for the last meters and last miles application demanding high-rate data transmissions. The use of PL channels is very promising because about 50% of all investments in the telecommunication infrastructure is needed for telecommunication access networks. Moreover, 95% of residencies and buildings around the world make use of electric circuits for energy delivery. As stated in [1]: *the Last meters and last miles access markets are still evolving very slowly due to the daunting cost of home networking and cabling installation. Pulling wires in an existing home is difficult, while it is not a solution amenable to the mass market. Most consumers are unwilling or cannot afford a large-scale residential or building rewiring, especially in the brick/concrete buildings even in developed countries. Then, apart from some exceptions, many vendors and service providers have put great focus on the so-called "no-new-wires" solutions that eliminate the need for wires pulling.*

Then, digital communication through PL channels is a very competitive alternative to a number of competing access network technologies, ranging from copper enhacement/*digital subscriber line variations* (e.g. ADSL, HDSL, VDSL), cable TV/ modem (e.g. DOCSIS), wireless solutions [2], satellite communications to *fibre to the curb* (FTTC), *fiber to the home* (FTTH), and various combinations between fibre/twisted pair (e.g. FTTC combined with VDSL) or coaxial (*hybrid fibre coaxial* - HFC).

In this regard, the aim of the research in the *power line communication* (PLC) field is to provide last miles and last meters networks as reliable and simple as possible to meet the increasing demands for bandwidth in residential, industrial, and commercial facilities in metropolitan areas as well as rural areas. In fact, the access networks are very important for network providers because of their high costs and the need for a direct access to the end users/subscribers. Therefore, the main investigations for providing it focus on communication systems able to provide remarkable improvement in terms of achievable bit-rate with as low as possible costs for applications where the use of fibre or other well-known medium is a very expensive solution. As a result, recent investigations have positioned the PLC technology as one of the greatest opportunity to develop simple and powerful communication systems with flexibility to support *narrowband* and *broadband* applications, such as surveillance, automation, telemetric, and monitoring systems, *high definition television* (HDTV), *voice over* PLC (VoPLC), fast internet access and all other kinds of multimedia services not only for residential, industrial, and commercial facilities, but also for all kinds of vehicles that make use of power lines for energy delivery.

Recent advances in PLC field have resulted in the advent of novel, impressive, and promising PLC technology (see Fig. 6.1) that will provide in the near future way to overcome the PL disadvantages and, consequently, guarantee the widespread use of PL channels for *narrowband* PLC (N-PLC, NPL or NaPLC) and *broadband* PLC (B-PLC, BPL or BoPLC) data transmissions. For instance, nowadays some companies are offering PLC modems with mean and peak bit-rates around 100 Mbps and 200 Mbps, respectively, and IEEE formed a group to discuss and propose a PLC standard. However, advanced B-PLC modems will surpass this performance because the capacity of PL channels can surpass 600 Mbps [3, 4, 5, 6, 7]. Some special schemes or solutions for coping with the following issues should be addressed [8, 9, 10]: *i*) considerable differences between power line network topologies and physical properties of cables; *ii*) hostile properties of PL channels, such as attenuation proportional to frequency and distance increases, high-power impulse noise occurrences, time-varying behavior, time- and frequency-varying impedance, strong *inter-symbol*

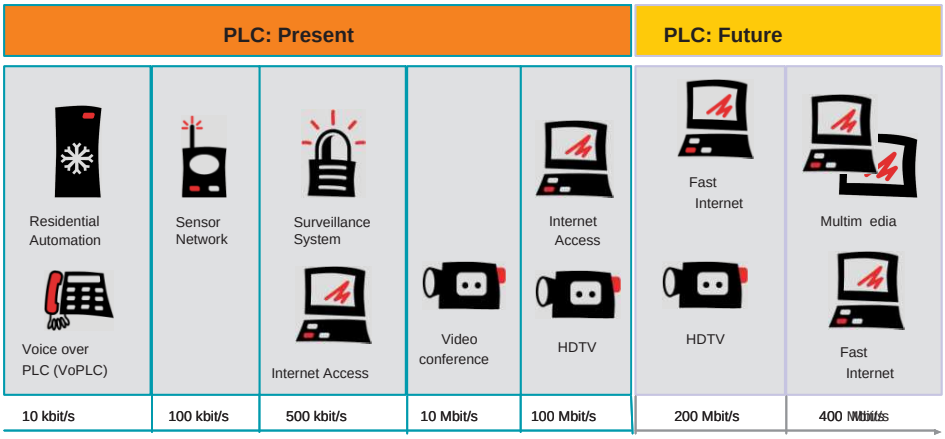



**Fig. 6.1.** PLC technology: present and future.

*interference* (ISI) effects, and frequency selectivity; *iii*) interoperability with other well-established communication systems; *iv*) electromagnetic compatibility with other well-established communication systems working in the same spectrum, *v*) climatic conditions in different parts of the world, *vi*) reliability and *quality of service* (QoS) guarantee for video and voice transmissions; *vii*) security against non-authorized access, and *viii*) different demands and needs from developed, developing, and underdeveloped countries.

To deal with these issues, novel and efficient signal processing, digital communication, and computational intelligence techniques as well as simple and powerful *digital signal processor* (DSP)-, *flexible programmable gate array* (FPGA)-, and *application-specific integrated circuit* (ASIC)-based hardware have to be researched for the development of new PLC solutions. It can lead to exciting research frontiers with very promising results for last miles and last meters applications. In fact, the used of signal processing, digital communication, and computational intelligence tools either individually or in combined form can result in reliable and powerful future generations of NaPLC and BoPLC systems that can be suited for applications in developed, developing, and underdeveloped countries. The research results will provide novel paradigms for overcoming limitations and problems that restrict the widespread use of PLC technology as a new/old wireline medium.

## 6.2 Historical Perspective

The use of PL for data transmission date back to 1838 when the first remote electricity supply metering was introduced and in 1897 when the first patent on power line signaling was proposed in the United Kingdom [11]. After that, in 1905 applications were patented in the United States, and in 1913 the first commercial production of electromechanical meter repeaters took place. Around 1920, the first *carrier transmission over powerlines* (CTP) by power system utilities began to operate on high-voltage lines in the frequency range of 15-500 kHz for handling operations management of the power supply by means of voice. After 1940, new CTP for telemetering and telecontrolling of power system was introduced. Initially, the data-rate was 50 bits/s, and later it was increased to 100 bits/s and 200 bits/s. In the past and present days the main purpose of CTP was to maintain the operability of the power supply. The CTP under favorable and unfavorable conditions can bridge up to 900 km with a transmission power of only 10 W [8].

At very beginning of 1930 *ripple carrier signaling* (RCS) started to operate on low and medium voltages power lines for power system management operations. The used frequency ranged from 125 Hz up to 3 kHz with *amplitude shift keying* (ASK) modulation technique. The frequencies close to the power frequency allow the information to flow over the transformers between medium and low voltages without the need of coupling measures. The data rates achieved by RCS was of the order of a few bits per second. Load management and automatic reconfiguration of power distribution networks were among the most important tasks performed by RCS. Among the applications provided by RCS were the management of street lights and load control.

The advancement of signal modulation technologies, digital signal processing, and error control coding have minimized the restrictions of channel imperfections, and high-speed signal transmission through power lines. As a result, bi-directional communication was developed in the late 1980s and early 1990s. The main difference between early PLC systems and modern ones is that much higher frequencies and a substantial reduction of the signal levels are used on today's power grid network for data transmission.

In 1991, the European CELENEC standard EN 50065 was launched. It makes use of the frequency range from 3 up to 148.5 kHz for NaPLC applications with a maximum signal power of 5 mW or 134 dB$\mu$V and rates up to 144 kbps over distances around 500 m. While the main member of European Union restricts the use of PLC technology by imposing lower power transmission level, the USA and Japan apply less stringent restriction on power transmission. Also, USA and Japan specify a spectrum below 500 kHz for NaPLC applications.

With the deregulation of telecommunication market in 1998, the development of alternative fixed network access systems within the last miles and last meters have pushed forward the development of advanced NaPLC and mainly BoPLC technologies. Soon after, the standard HomePlug 1.0 was introduced. It was developed by the HomePlug Alliance formed in 2000 by 13 founding members [12]. This standard is one of the most famous power line communication technologies and it supports up to a 14-Mb/s transmission rate. New HomePlug AV [12], which was finalized in 2005, is expected to have the data rate of 200 Mb/s in the *physical* (PHY) layer (100 Mb/s in the *medium access control* (MAC) layer). HomePlug AV provides both connection-oriented and contention-free services based on periodic *time-division multiple access* (TDMA) and connectionless services based on *carrier sense multiple-access/collision avoidance* (CSMA/CA) technology, which is used in HomePlug 1.0 MAC [13].

In January 2004, the *Open PLC European Research Alliance* (OPERA) [14] started an ambitious project to develop a new PLC technology generation as an alternative for access networks. Not only, entities from European countries joined this project, but also entities from countries outside Europe, such as Brazil, collaborate. The main objective of OPERA project is to perform the necessary research and development to overcome any remaining obstacles, allowing PLC operators to provide competitive PLC-based broadband services. Their final results are expected to the end of 2006.

In 2005, ES 59013 draft standard from CENELEC allocated the 1.6 MHz to 12.7 MHz band to the access systems and the 14.35 MHz to 30 MHz band for in-home ones.

After 2005, an IEEE working group started the development of a standard for MAC and PHY layers specifications for BoPLC [15]. This standard aims at the development of a standard for high speed ($>$100 Mbps at the PHY layer) communication devices through power lines, so called Broadband over Power Line devices in the frequency range between 1 MHz and 100 MHz. This standard will be used by all classes of BoPLC devices, including BoPLC devices used for the first-mile/last-mile connection ( distances $<$ 1500 m to the premise) to broadband services as well as BoPLC devices used in buildings for LANs and other data applications ( distances $<$ 100 m between devices).

Additionally, this IEEE P1901 working group is focusing on the balanced and efficient use of PL channels by all classes of BoPLC devices, defining detailed mechanisms for coexistence and interoperability between different BoPLC devices, and ensuring that desired bandwidth and quality of service may be delivered. The standard will address the necessary security questions to ensure the privacy of communications between users and allow the use of BoPLc for security sensitive services. It is also the intent of this effort to quickly progress towards a robust standard so that PLC applications may begin to impact the marketplace. The standard will also comply with EMC limits set by national regulators, so as to ensure successful coexistence with wireless and telecommunication systems.

Among the reasons for putting a great research effort on PLC system viability, the following advantages should be mentioned [9, 2, 16, 17, 18]: *i*) PLC can provide an extensive coverage for last mile and last meter applications, since the power lines are already installed almost everywhere. This is advantageous especially for substations in rural areas where there is usually no communication infrastructure available for use. while PL channel is present in more than 95% of households around the word, other broadband alternatives such as xDSL and cable modems have only reached less than 10% households, even though 60% and 15% of households in developed and developing countries are already connected to the Internet. In fact, The biggest capillarity of existing power networks makes possible the access of the Internet to most of the population and it makes possible the proper services of the electric energy utilities that depend on data communication, besides adding value to its electric assets (wires); *ii*) the communication network can be established quickly and cost-effectively because it utilizes the existing wires to carry the communication signals. Thus, PLC can offer new cost-saving methods for remotely monitoring power uses and outages as well as other NaPLC and BoPLC applications; *iii*) PLC is easy to install and use - no more wires are needed -just plug in. High transmission rates, multiple Mbps, are right now available and more than 100 Mbps are possible. Together with secure data encryption you can utilize your existing power source for all your communication needs, including TV quality video, streaming audio and broadband multimedia.

However, the following disadvantages have discouraged the investigation of PLC technology for the last few decades [19, 9, 2, 16, 17, 18, 20]: *i*) the PL are noisy environments for data communications due to several noise sources such as electronic-based equipment, electrical motors, power supplies, fluorescent lights and radio signal interferences. *ii*) past contributions present limit theoretical channel capacity for low-, medium-, and high-voltage PLs. *iii*) communications over PLs are severely restricted by the use of switches, reclosers and sectionalizers, *iv*) the attenuation and distortion of signals are immense due to the reasons such as physical topology of the power network and load impedance fluctuation over the power lines. Besides, there is significant signal attenuation at specific

frequency bands due to wave reflection at the terminal points. *v*) considering the physical and nature of PLs, security concerns have been addressed, because PLs irradiate considerable amount of *electro magnetic interference* (EMI). Then, an unauthorized person can intercept the data information carried by power cable if an appropriate encryption technique is applied. *vi*) the lack of worldwide regulation for BoPLC and NaPLC to guarantee the interoperability and low-cost PLC systems working in a shared medium with other telecommunication technology working in the same frequency range such as mobile communications, broadcasting channels and military communications.

Recent initiatives and great research efforts show that PLC is becoming more a reality than a promise. In fact, recent contributions on this field will lead to very powerful and cheap last mile and last meters solutions for medium- and low- voltages PLs of distribution networks of power systems as well as electrical circuits in residential, vehicular, and building facilities. As a result, one can predict that the new generation of PLC modem with data-rate in the PHY layer up to 400 Mbps will come.

### Case Study: PLC in Brazil

Since a long time ago, the use of power line for data communications has been restricted to the NaPLC applications, such as power system protection and relaying by utility companies from Brazil. These companies have focused only on electric energy market because their administrators are much more interested in keeping their main energy market than data service. In fact, the commercial use of PLC is a novel, promising and challenging paradigm whose advantages few administrators of electrical energy utilities are aware of. Such advantages are the increase of portfolio of service, venue, profit, and client's fidelity. In fact, the utility companies can increase considerably their profit by offering communication service for their clients and customs. Figs. 6.2 and 6.3 illustrate the use of indoor and outdoor energy distribution network for BoPLC and NaPLC applications.

NaPLC and BoPLC infrastructures for voice, video and data traffics in the low and medium voltages PLs can be provided by electrical energy utilities demanding lower investment than that required by well-known technologies such as wireline (cable TV, xDSL, fibre cables), and fixed- and mobile- wireless systems (WiMax and satellite communications) to offer the following services [22]: *i*) Fast and secure surfing on the Internet, e-commerce, e-mail and e-banking; *ii*) powerful telephone connection using the internet that offers security and good speech quality. Sendind and receiving fax messages is also possible; *iii*) smart home applications with remote maintenance and in-house control of internet enabled household appliances like refrigerators, heating systems, smoke and fire alarm systems; *iv*) surveillance systems with both visual and motion detectors that can be monitored by you and a security service; *v*) health care services; *vi*) online reading of utility meters for flexible and easy up-to-date billing; and, *vii*) easy PCs, telephones, and multimedia connection devices by simply plugging in the existing electric socket.

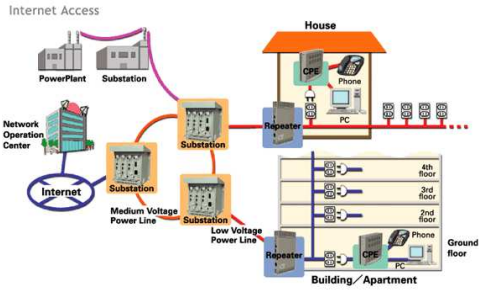



**Fig. 6.2.** Access PLC networks [126].

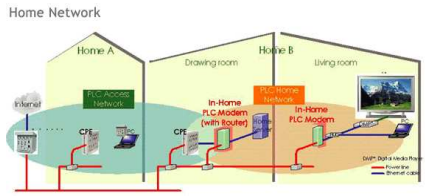



**Fig. 6.3.** Residential PLC networks [126].

As a result, the electrical energy utilities will start to dispute the last mile telecommunication market with Telcos with one remarkable advantage: *capillarity in the last mile*. Telecommunication infrastructure capillarity is the key connectivity problem to be minimized in the last mile environment and the electrical energy utilities have it available because electricity service is nearly ubiquitous (the theoretical coverage from PLC is almost 100% in most of developed countries) and have higher electricity penetration than telephony infra-structure. In a developing country, like Brazil, a social program from federal government named "*Luz para todos* (light for everyone)" [23] will guarantee electrical energy infrastructure for, at least, 98% of residencies and buildings until 2008. As a result, households in metropolitan and in rural areas will have electrical infrastructures available not only for energy delivery, but also for data communication.

In Brazil, as well as around the world, the main competitors for BoPLC are [24]: *i*) the xDSL technology with the broadband last mile technology leader being the current *Asymmetric Digital Subscriber Line* (ADSL), with average bandwidth of 2 Mbps for 6 km distance; *ii*) cable networks which have the capacity to deliver significantly higher bandwidth to the end user. It presents a relatively limited coverage area and subscribers base. To increase the number of customers, investment should be made to redesign existing networks to allow higher connection density and expand the coverage area; and, *iii*)wireless broadband technologies that have a good potential particularly in low-density areas. The key advantages are fast installation and the full mobility for the end costumers. However, market solutions depend upon the availability of sufficient spectrum and the necessary investment on transport systems and Fiber Optics backbones.

Regarding the use of BoPLC technology in the Brazilian network access telecommunication market, the following issues have to be taken into account [24]: *i*) the market is dominated by a few technology patterns. As a result, BoPLC proponents can expect to face strong barriers when trying to develop partnerships with incumbent Telcos; *ii*) many end customers do not differentiate one telecom service from another. Then, BPL in competition with a traditional telecom strategy may mean lower prices. This competition may result in lower margins for suppliers, which is not a good business model; *iii*) there are barriers to new entrants because existing players control most of the distribution channels; *iv*) according to recent investigations, Telcos are not ready to commercialize the full potential of telecom convergence. Then, BoPLC can be competitive depending the companies' strategies, even in the Telco homeland; and *v*) a small number of big organizations using few access technologies such as xDSL, HFC, wireless, FTTH and competing over cost leadership strategy.

The widespread use of PLC technology in developing countries, such as Brazil contributes to minimize the distance from people digitally inserted in the world from those digitally excluded. Additionally, a PLC standard for developing and underdeveloped countries sponsored by Brazil could support and motivate the start of high-tech companies in Brazil to produce modems, systems, and accessories for BoPLC and NaPLC. Then, the development of a so-called "Brazilian PLC standard for BoPLC and NaPLC" should be a priority for those responsible for the technological advances in access networks because NaPLC and BoPLC are great telecommunication niches for Brazil. Unfortunately, the initiatives or research efforts on this field in Brazil are far from reaching this objective. Actually, the Brazilian efforts towards PLC technology can be summarized as follows:

*i*) the use of PLC technologies developed in USA and Europe: these PLC technologies are not related to any kind of P&D for PLC system improvement as well as development. Usually, USA and European PLC system are installed in proof camp for utilities companies to verify their technical viability and in poor and distant communities to demonstrate that PLC system can help them access the digital contents through internet as well as offer other kinds of data communication services. The main motivation for this is that PLC technology is a very promising one in terms of cost and performance to digitally assist rural and isolated communities or even for those that do not have telecommunication infrastructure available. In this regard, several Brazilian companies joined the Brazilian forum so-called "Fórum Aptel de PLC/Brasil" [25], which aims at fomenting and encouraging the widespread use of PLC technology in Brazil. This forum collaborates with the European PLC forum through the OPERA project in order to verify the viability of PLC business in Brazil.

*ii*) the analysis of PHY and MAC layers of PLC system at universities: basically analysis of single- and multi-carrier digital communication systems suited for NaPLC and BoPLC applications have been analyzed. Additionally, implementation of PLC modem based on USA and European equipment is being carried out for indoor and outdoor PLC applications. It is worth mentioning that a chipset based on a OFDM system for BoPLC is under development at Federal University of Juiz de Fora, Brazil, that tries to take into account the main features of the PL channels found in Brazil. The expected results is a system capable of delivering more than 100 Mbps at the PHY layer. The reason for this hardworking development resides on the fact that the PLC technology from USA and Europe do not take into account the characteristics as well as the needs of developing countries like Brazil.

## 6.3 Channel Models

The development of a modelling method for PL channels should follow the following principles: *i*) The propagation model has to be as simple as possible and use few cable properties for each cable type; *ii*) the model input data could only be extracted from the network topology ; *iii*) the development of a measuring method is required to analyze totally unknown networks; *iv*) experimental verification of the propagation and the measuring methods have to be carried out under different conditions and compared with other methods; *v*) the propagation model and the measuring method have to be based on easily calculable/measurable quantities, related only to network terminals rather than inner (hidden) nodes within the network; and *vi*) the elements needed to make the propagation analysis software tool have to be modeled by the propagation model.

The power distribution network was not designed or even optimized for low- or high-date communication, with numerous bridges, splits, taps, branchings, etc., as well as a large amount equipment on route such as capacitor banks, relays, reclosers, sectionalizers and transformers. The unavailability of a model for a precise description of signal propagation along PL lines has conducted to pessimistic conclusions. The PL channels are impossible to be modeled *a priori* and, consequently, do not allow the superposition of effects. They have very little or no determinism embedded in it, and consequently, it is not reliable as communication medium for high-speed data transmission [20]. Despite that, research efforts for a better characterization of PL channels have been carried out around the world offering novel results that dismisses several arguments against PLC. The two well-established approaches for modeling PL channels are as follows:

*i*) *top-down*: this approach considers the PL communication channel as a black box from which the PL transfer function is estimated. Basically, channel measurements are provided in order to estimate the frequency domain parameters from the PL channel. Based on the *top-down* approach, the so-called multipath model, which is a well-known model offering the frequency response of indoor, outdoor, and broadband PL channels is given by [26]

$$H(f) = \sum_{i=1}^{P} G_i(f, d_i) A_i(f, d_i) \exp(-j2\pi f \tau_i), \tag{6.1}$$

where $G_i(f, d_i)$ refers to the weighting factor in the *i*th multi-path, $A_i(f, d_i)$ is the attenuation term produced by the *i*th multi-path, $d_i$ is the distance of the *i*th multi-path, $\tau_i$ is the delay portion in the *i*th multi-path, and *P* is the number of multi-paths. The typical attenuation of three PL channels occupying the spectrum ranging from 0.5 to 20 MHz is depicted in Fig. 6.4 in which each curve corresponds to PL channel with different distance between *receiver* (Rx) and *transmitter* (Tx). One can note that the attenuation profile depend hardly upon the frequency range as well as Tx-Rx distance. Also, deep narrow-band notches caused by multiple reflections at impedance discontinuities appears in the transfer function. These notches can be spread over the frequency range.

It is worth stressing that the multipath model is quite appropriate in the case of very simple topologies, such as a cable with a single branch and the physical reasons for the observed results (cable loss, reflection, and transmission factors) can be easily identified. In real network topologies, which are always more complicated, a back-tracing of measurement results to physical reasons will generally turn out to be impossible.

Regarding NaPLC applications, a very interesting PL channel model in the frequency range of the CELENEC standard EN 50065 is presented in [27, 28].

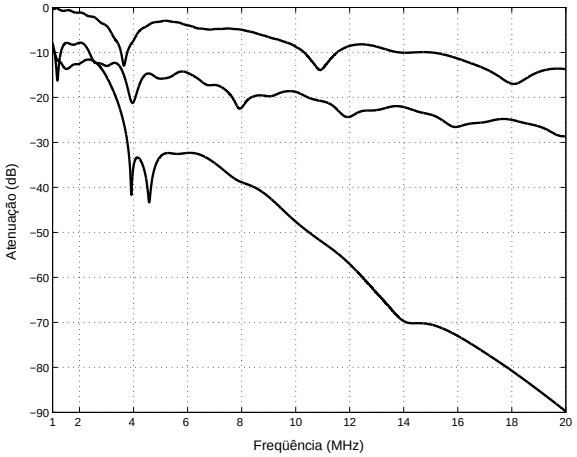



**Fig. 6.4.** Frequency attenuation of three typical PL channels.

*ii*) *bottom-up*: this approach describes the transfer function of the power line by using scattering parameters matrices or four poles impedance or admittance matrices. This model requires a complete knowledge about the components of the networks to determine the elements of the matrices. The disadvantage of such approach is the large number of parameters to be manipulated. Based on the fact that indoor PL channels are simpler than others and have their parameters well-known, the *bottom-up* approach is widely used for modeling them. Regarding this approach in [29]-[30] and [31] a deterministic model grounded on *multiconductor transmission line* (MTL) theory and modal decomposition is presented for indoor PL channels that allows one to compute *a priori* and in a deterministic fashion the transfer function of any PL link by using two-port transmission matrices and it demonstrates the viability of PL channels for BoPLC applications in indoor environments. In [6] a PL channel model based on *bottom-up* approach is introduced to precisely model overhead MV power line showing that MV cables present a theoretical channel capacity higher than 600 Mbps.

PLs connect the power generation station to a variety of customers widespread in a area. Power transmission is done using varying voltage levels and power line cables. The voltages levels are interconnected by transformers. Based on the voltage levels at which they transfer, power lines are distinguished as high-, medium- and low-voltages.

### 6.3.1 *High-Voltage* (HV) Lines

High-voltage lines connect electricity generation stations to distribution stations, bridging distances from several dozen to several hundred kilometers. The voltage levels on these lines are typically in the order of hundreds of kilovolts ($110 - 300$ kV). Voltages over 300 kV are named *extremely high voltage* (EHV). The EHV and HV levels establish a pure transmission network, since no customer premises are directly connected to the lines.

For these levels, the PLs present energy losses due to heating, leakage, and corona effects. The MV cables have been used only for narrowband communication system operating in the frequency up $500$ kHz for protections and restricted supervision and operation of power systems [8]. The limited use of these channels is because the length of HV level cables are up to 500 km and the attenuation in the HV cables increases remarkably with frequency as well as distance, the high-cost and difficulty for coupling modem PLC in HV cables. The HV lines are very good communication channels in the sense that almost no branches are observed on them and a small number of disturbances that strongly affect the broadband and narrowband communication systems are noted on them. The CTP is the only well-known communication system developed to transmit data in HV lines.

### 6.3.2 *Medium-Voltage* (MV) Lines

The MV cables connect the distribution stations to pole mounted transformers. MV level cables are used to supply electric energy to rural areas, towns, cities, individual industries and companies. The nominal voltage levels are of the order of a few kilo volts ($10 - 30$ kV) and they run over distances of the order of a few kilometers (5-25 km).

In developed countries, most of the MV level cables are underground lines, while in other countries most of these cables are overhead ones.

Over MV cables, two types of noises are dominant [32]: *i*) colored background noise and *ii*) narrowband noise. The former is the environmental noise, which is highly dependent on the weather, geography, above ground height, etc. One has to note that corona discharge [33] is a major cause of background noise, especially under humid and severe weather conditions. Therefore, a PL appears as a noise source by itself due to corona discharges on HV lines. The latter is the interference from wireless devices and services operating in the frequency range of BoPLC systems. Narrowband noise varies with time and place.

The channel capacity of MV cables are higher than the one observed in the low-voltage cables, because the former cables present lower number of branches than in the latter ones. In fact, the characteristics of the MV for broadband communication systems are better than those ones of low-voltage power lines.

MV networks can be used as a distribution network connecting a number of PLC access networks to the backbone as well as a solution for the distribution network [34, 35]. However, the use of the MV lines has been limited to N-PLC because past research has considerably underestimated the theoretical capacity of these MV lines. Recently, in [6] it was demonstrated that the theoretical channel capacity of MV lines is higher than 600 Mbps. Therefore, the MV lines are very competitive to provide last mile access for residential, comercial, and industrial facilities. In this regard, research leading to promising last mile access systems based on PLC is being developed to introduce in the near future new generations of PLC systems operating in distances up to 25 km and delivering some Gbps with a good tradeoff regarding performance and cost.

### 6.3.3 *Low-Voltage* (LV) Lines

The low-voltage lines connect pole-mounted transformers to individual households as well as appliances and electrical equipment in predial, residential, industrial and commercial electrical circuits. The voltage levels on these lines are of the order of a few hundred volts ($< 400$ V) and these run over distances of the order of a few hundred meters. The typical supplies radiuses of a feed point (pole-mounted transformers or low voltage transformer station) are from $100$ to $500\ m$ and that the number of users connected to a LV cable is between 20 and 80. Recently, the LV cables have appeared as a new frontier for last meters applications because they are in more than $95$ % of residential, commercial, and industrial facilities around the world. Additionally, one can note that all kinds of vehicles (car, airplane, ship, aircraft, etc) own LV cables for energy delivery in the *alternating current* (AC) and *direct current* (DC) forms.

Due to the fact that the attenuation of PL channels becomes considerably high as the distance from the *transmitter* (Tx) to *receiver* (Rx) and the frequency of data transmission increases, some authors have advocated that outdoor PLC system (from transformers to individual households) should occupy the frequency range from 1 up to 10 MHz for BoPLC, while the spectrum over 10 MHz should be used for indoor PLC system (inside residencies and buildings).

Although the PLC channels are time-varying, most of them can be considered time-invariant during the interval of symbol transmissions [29, 30, 31, 26]. Recent investigations are focusing on time-varying behavior of PL channels, see [36, 37, 27].

## 6.4 Additive Noise

For NaPLC applications, a good characterization of noise scenario is provided in [38] and [28]. The additive noises model in BoPLC scenarios for last mile and last meters applications as well as other environment, such as aerospace, vehicular, etc is well known from [39]. Referring to [39], the noise in last miles or last meters low-, medium- and high-voltages networks is expressed by

$$v(n) = v_{bkgr}(t) + v_{nb}(t) + v_{pa}(t) + v_{ps}(t) + v_{imp}(t)|_{t=nT_s}, \tag{6.2}$$

where $T_s$ is the sampling rate and

- $v_{bkgr}(t)$ is the colored background noise with a relatively low *power spectral density* (PSD), which is caused by the sum of numerous noise sources of low power. It covers the communication frequency range. Contrary to the white noise, which is a random noise having a continuous and uniform spectral density that is substantially independent of the frequency over the specified frequency range, the colored background noise shows strong dependency on the considered frequency. In the frequency range above 1 MHz, background noise is below -110dBm/Hz,even dropping to -130dBm/Hz, and far below the noise level of -80dBm/Hz or -90dBm/Hz under 500kHz. As a result, its PSD can be approximated by a exponential function. It is caused, for example, by common household appliances like computers, dimmers, or hair dryers, which can cause disturbances in the frequency range of up to 30 MHz. For residential and industrial environment, the PSD is given by [9]

$$PSD(f) = -35 + 35 \exp\left(-\frac{f}{3.6}\right) \tag{6.3}$$

  and

$$PSD(f) = -35 + 40 \exp\left(-\frac{f}{8.6}\right), \tag{6.4}$$

  respectively, where $f$ is in MHz.
- $v_{nb}(t)$ is a narrowband noise caused by the ingress of radio broadcasting stations usually in the frequency range of 1-22 MHz. It is mostly amplitude modulated sinusoidal signals. Their amplitude generally varies along the daytime, becoming higher by night when the reflection properties of the atmosphere become stronger. It is mathematically modeled by

$$v_{nb}(t) = \sum_{i=1}^{N} A_i(t) \cos\left(\omega_i t + \theta_i(t)\right) \tag{6.5}$$

  where $N$ is the number of broadcasting stations and $A_i(t)$, $\omega_i$ and $\theta_i(t)$ are random variables.
- $v_{pa}(t)$ is a periodical impulsive noise asynchronous to the fundamental component of power system, which is mostly caused by switched-mode power supplies. It usually has a repetition rate between 50 and 200 kHz, and which results in the spectrum with discrete lines with frequency spacing according to the repetition rate. It is produced by extended use of switching power supplies found in various household appliances today

- $v_{ps}(t)$ is a periodic impulsive noise synchronous to the fundamental component of power system with a repetition rate of 50 or 100 Hz and synchronous with the power frequency. Such impulses have a short duration, in the order of microseconds, and have a power spectral density that decreases with the frequency. This type of noise is generally caused by power supply operating synchronously with the power frequency, such as the power converters connected to the main supply, rectifiers within DC power supplies and appliances such as thyristor- or triac-based light dimmers. it can be expressed by [40]

$$v_{ps}(t) = \sum_{i=1}^{N} A_i(t) \sin\left(\omega_i(t - t_{arr,i}) + \theta_i(t)\right) e^{-\frac{t-t_{arr,i}}{\tau_i}} \prod\left(\frac{t - t_{arr,i}}{\tau_i}\right) \tag{6.6}$$

where $\omega_i$ is the angular frequency of the sinusoid, and $\theta_i$ the phase, of the *i*th damped sinusoid. $\prod(t)$ is defined as a square pulse of duration $t_{w,s}$s, with constant amplitude in the interval $0 < t \leq 1$ and zero elsewhere. $t_{arr,s}$ is the periodic arrival time, and $A_i$ denotes the amplitude of the *i*th sinusoid.We assume $A_i \sim \mathcal{N}(0, G_i\sigma_i^2)$, where $G_i$ represents the increase over the variance of Gaussian background noise $\sigma_i^2$, and $G_i$ can range from 20 to 30 dB. The gain $G_i$ of sinusoids at higher pseudo-frequencies is selected to match the typical low-frequency content observed in impulsive noise measurements, usually below 1 MHz. The term $\tau_i$ denotes the damping factor.
- $v_{imp}(t)$ is an asynchronous impulsive noise which is the hardest one. These impulses have durations of some microseconds up to a few milliseconds with an arbitrary interarrival time. Their PSD can reach values of more than 50 dB above the level of the background noise, making them the principal cause of error occurrences in the digital communication over PLC networks. This type of noise contains a broadband portion significantly exceeding the background noise, and a narrowband portion appearing only in certain frequency ranges. It is mainly caused by switching transients, which occur all over a power supply network at irregular intervals. It can be denoted by [39]

$$v_{imp}(t) = \sum_{i=1}^{M} A_i \prod_i \left(\frac{t - t_{arr,i}}{t_{w,i}}\right) \tag{6.7}$$

where $\prod_i(t)$ denotes the asynchronous impulse noise pulse waveform and $A_i$, $t_{w,i}$ and $t_{arr,i}$ refer to three random variables: amplitude, impulse width, and interarrival time (the time between the arrival of two impulses), respectively. For a majority of impulses we find amplitudes around 1 V, impulse widths in the range of 100 $\mu$s, and interarrival times of 100 ms. Fortunately, even in heavily disturbed environments such as industrial zones, the average disturbance ratio is well below 1 percent, meaning that 99 percent of the time is absolutely free of asynchronous impulsive noise.

For illustration purpose, Fig. 6.5 portrays a sample of background and asynchronous impulsive noises in LV distribution power lines [41].

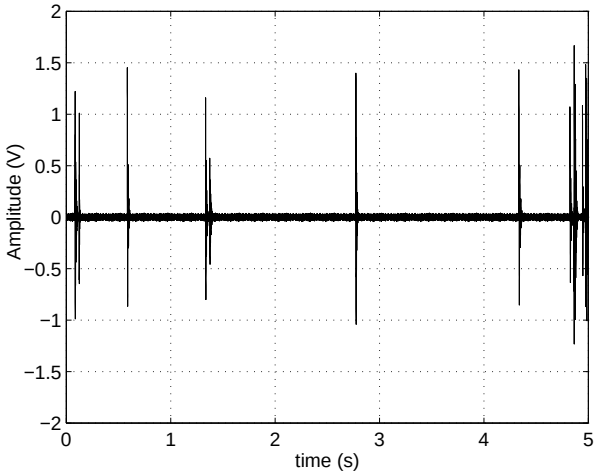



**Fig. 6.5.** Background noise and impulse noise in LV distribution power lines.

## 6.5 PL Channel Equalization with Fuzzy Techniques

The PL channels introduce strong ISI. Also, the deep frequency notches present in the channel transfer function prevent the use of linear equalizers, as the noise enhancement they cause is a serious drawback on a very noisy channel. As a result, a *decision feedback equalizers* (DFEs) could be a very promising solution due to its optimal results [42]. Against the use of DFEs is the fact that PLC channels have long fast-varying impulse responses and strong noise bursts, concentrated in both time and frequency. Then, it is plausible to envision that noise bursts cause catastrophic error propagation, and that fast time variations may not allow channel parameter tracking when adaptation cannot be fast enough. In this regard, the use of other kinds of nonlinear equalization techniques capable of coping with strong noise bursts as well as long and fast-varying impulse responses demand investigations.

Computational intelligence-based techniques have been widely applied for communication channel equalization that present these aforementioned issues [43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53]. Among them, fuzzy systems have been applied to the following problems [54, 55, 47, 56, 57, 58, 59, 60, 61, 62, 44, 63] *i*) equalization of time-varying and -invariant channels, *ii*) equalization of channel corrupted by nonGaussian noise, *iii*) *co-channel interference* (CCI), *iv*) development of one and two dimensional equalizers, and *v*) adaptive fuzzy equalizers. Recently, the use of FSs in a turbo equalization scheme [64], different strategies for parameters updating [54], type-2 FSs [50] and blind equalization [65] are being investigated.

A discrete-time version of linear, causal and time invariant model of baseband PL channels for BoPLC and NaPLC applications is given by

$$y(n) = \tilde{y}(n) + v(n) = \sum_{i=0}^{L_h-1} h(i)x(n-i) + v(n), \tag{6.8}$$

where $\tilde{y}(n)$ denotes the PL channel output free of noise, $v(n)$ is the additive noise, the sequence $\{x(n)\}$ is constituted by transmitted symbols previously obtained from a OFDM symbol, from a *pulse amplitude modulation* (PAM), a *quadrature amplitude modulation* (QAM), or any other symbol constellation, and $\{h(n)\}_{n=0}^{L_h-1}$ is the bandlimited, dispersive, and linear *finite impulse response* (FIR) PL channel model.

By considering a *binary phase shift keying* BPSK constellation for a single-carrier or a *code division multiple access* (CDMA) system, the transmitted sequence $\{x(n)\}$ is taken from $\{+1, -1\}$ of symbols and it is assumed to be an equiprobable and independent sequence with $E\{x(n-k)x(n-l)\} = \sigma_x^2\delta(k-l)$ and $E\{x(n) = 0$. $v(n)$ is the additive impulsive noise and $\widetilde{y}(n)$ denotes the noise-free channel output. The channel outputs observed by the linear equalizer $\mathbf{w}(n) = [w(n)\cdots w(n-L_w+1)]^T$ can be written as vector $\mathbf{y}(n) = [y(n)\cdots y(n-L_w+1)]^T$. The vector of the transmitted symbols that influence the equalizer decision is expressed by $\mathbf{x}(n) = [y(n)\cdots y(n-L_h-L_w+1)]^T$. As a result, there are $n_s = 2^{L_h+L_w}$ possible combinations of the channel input sequence. And, $n_s$ different values of the noise-free channel output vector $\widetilde{\mathbf{y}}(n) = [\widetilde{y}(n)\cdots\widetilde{y}(n-L_w+1)]^T$ are possible. Each of these noise-free channel output vector is called channel output state vector $\widetilde{\mathbf{y}}_j(n),\ j = 1, \cdot n_s$ given by

$$\widetilde{\mathbf{y}}(n) = \mathbf{H}\mathbf{x}_j(n), \tag{6.9}$$

where $\mathbf{x}_j(n) == [x(n)\cdots x(n-L_h+L_w+1)]^T$ denotes the *j*th input vector and $\mathbf{H}$ is a matrix channel impulse response given in the form of

$$\mathbf{H} = \begin{bmatrix} h_0 & h_1 & \cdots & h_{L_h-1} & \cdots & 0 \\ 0 & h_0 & \cdots & h_{L_h-1} & \cdots & 0 \\ \vdots & \vdots & & & & \vdots \\ 0 & 0 & h_0 & \cdots & h_{L_h-2} & h_{L_h-1} \end{bmatrix}, \tag{6.10}$$

The equalizer output $\hat{x}(n-d)$ is delayed by $d$ samples form of the transmitted sequence.

The PL channels can be equalized by using two categories of adaptive equalization techniques, namely [66, 67]: sequence estimation and symbol decision. The optimal solution for sequence estimation is achieved by using *maximum-likelihood sequence estimation* (MLSE) [68]. The MLSE is implemented by using Viterbi algorithm [69], which determines the estimated transmitted sequence $\{x(n)\}_{n=0}^{\infty}$ when the cost function defined by

$$J_{cost} = \sum\nolimits_{n=1}^{\infty}\left[y(n) - \sum\nolimits_{i=0}^{L_h-1} h(i)\hat{x}(n-i)\right]^2, \tag{6.11}$$

is minimized. Although this algorithm demands the highest computational cost, it provides the lowest error rate when the channel is known. On the other hand, the optimal solution for symbol decision equalization is obtained from the

Bayes' probability theory [70, 71]. If the additive noise is modeled by a random process with normal distribution as $\mathcal{N}(0, \sigma_v^2)$, then the *normalized optimal bayesian equalizer* (NOBE) is defined by

$$f_b(\mathbf{y}(n)) = \frac{\left\{ \sum_{\mathbf{y}_i \in \mathcal{Y}_d^+} \exp\left(\frac{-\|\mathbf{y}(n)-\tilde{\mathbf{y}}_i\|^2}{2\sigma_v^2}\right) - \sum_{\mathbf{y}_j \in \mathcal{Y}_d^-} \exp\left(\frac{-\|\mathbf{y}(n)-\tilde{\mathbf{y}}_j\|^2}{2\sigma_v^2}\right) \right\}}{\sum_{\mathbf{y}_k \in \mathcal{Y}_d} \exp\left(\frac{-\|\mathbf{y}(n)-\tilde{\mathbf{y}}_k\|^2}{2\sigma_v^2}\right)}, \tag{6.12}$$

where $\mathcal{Y}_d = \mathcal{Y}_d^- \cup \mathcal{Y}_d^-$ denote the state space composed of $L_w$-length vectors $\mathbf{y}_i$, $\mathcal{Y}_d^+ = \{\mathbf{y_i}(n),\ i = 1, \cdots, 2^{L_h-1} | x(n-d) = +1\}$ and $\mathcal{Y}_d^- = \{\mathbf{y_j}(n),\ j = 1, \cdots, 2^{L_h-1} | x(n-d) = -1\}$. The separation between $\mathcal{Y}_d^+$ and $\mathcal{Y}_d^-$ is given by

$$f_b(\mathbf{y}(n)) = 0. \tag{6.13}$$

Now, let Gaussian functions be the membership functions that model the uncertainty of the input data and fuzzy rules and $\star$ be a t-norm named product of a *nonsingleton* type-1 FS [72]. Also, assume that defuzzification contributions associated with $\mathcal{Y}_d^+ = \{\mathbf{y_i}(n),\ i = 1, \cdots, 2^{L_h-1} | x(n-d) = +1\}$ and $\mathcal{Y}_d^- = \{\mathbf{y_j}(n),\ j = 1, \cdots, 2^{L_h-1} | x(n-d) = -1\}$ be equal to $1/2^{l_h-1}$ and $-1/2^{l_h-1}$, respectively. Then, the *normalized and optimal nonsigleton fuzzy equalizer* (NONFE) is given by

$$f_{ns1}(\mathbf{y}(n)) = \frac{\left\{ \sum_{\mathbf{y}_i \in \mathcal{Y}_d^+} \exp\left(\frac{-\|\mathbf{y}(n)-\tilde{\mathbf{y}}_i\|^2}{2(\sigma_r^2+\sigma_v^2)}\right) - \sum_{\mathbf{y}_j \in \mathcal{Y}_d^-} \exp\left(\frac{-\|\mathbf{y}(n)-\tilde{\mathbf{y}}_j\|^2}{2(\sigma_r^2+\sigma_v^2)}\right) \right\}}{\sum_{\mathbf{y}_k \in \mathcal{Y}_d} \exp\left(\frac{-\|\mathbf{y}(n)-\tilde{\mathbf{y}}_k\|^2}{2(\sigma_r^2+\sigma_v^2)}\right)}, \tag{6.14}$$

where $\sigma_r^2$ is the variance from the membership functions modelling the input uncertainties and antecedent rules. The separation surface obtained by NONFE is given by

$$f_{ns1}(\mathbf{y}(n)) = 0. \tag{6.15}$$

One can note that

$$\frac{f_b(\mathbf{y}(n))}{f_{ns1}(\mathbf{y}(n))} = 1 \tag{6.16}$$

if $\sigma_r^2 = 0$.

Following [18, 41], let Gaussian functions be the membership functions that model the uncertainty of the input data and fuzzy rules and $\star$ be a t-norm named product,

$$\overline{\mu}_{Y_i}(y_i(n)) = \exp\left[-\frac{1}{2}\left(\frac{y_i(n) - \overline{m}_{y_i}(n)}{\overline{\sigma}_{y_i}(n)}\right)^2\right] \tag{6.17}$$

and

$$\underline{\mu}_{Y_i}(y_i(n)) = \exp\left[-\frac{1}{2}\left(\frac{y_i(n) - \underline{m}_{y_i}(n)}{\underline{\sigma}_{y_i}(n)}\right)^2\right] \tag{6.18}$$

denote the upper and lower "worst-case" membership functions associated with the non-stationary behavior of the additive noise at the inputs of the interconnected type-1 fuzzy algorithm,

$$\overline{\mu}_{F_i^l}(y_i(n)) = \exp\left[-\frac{1}{2}\left(\frac{y_i(n) - \overline{m}_{F_i^l}(n)}{\overline{\sigma}_{F_i^l}(n)}\right)^2\right] \tag{6.19}$$

and

$$\underline{\mu}_{F_i^l}(y_i(n)) = \exp\left[-\frac{1}{2}\left(\frac{y_i(n) - \underline{m}_{F_i^l}(n)}{\underline{\sigma}_{F_i^l}(n)}\right)^2\right] \tag{6.20}$$

refer to the upper and lower "worst-case" membership functions that define the uncertainties of the rules' antecedents. Now, considering that

$$\overline{\mu}_{Q_i^k}(\overline{y}_{i,\max}^k(n)) = \sup_{y_i(n)\in Y_i} \overline{\mu}_{Y_i}\left(y_i(n)\right) \star \overline{\mu}_{F_i^k}\left(y_i(n)\right) \tag{6.21}$$

and

$$\underline{\mu}_{Q_i^k}(\underline{y}_{i,\max}^k(n)) = \sup_{y_i(n)\in Y_i} \underline{\mu}_{Y_i}\left(y_i(n)\right) \star \underline{\mu}_{F_i^k}\left(y_i(n)\right), \tag{6.22}$$

where $y_i(n)$ is the *i*th element of the input vector $\mathbf{y}(n) = [y_0(n)\cdots y_{L-1}(n)]^T$. Equations (6.21)-(6.22) attain their maximums under the following conditions [73]:

$$\overline{y}_{i,max}^k(n) = \frac{\overline{\sigma}_{y_i}^2(n)\overline{m}_{F_i^k}(n) + \overline{\sigma}_{F_i^k}^2(n)\overline{m}_{y_i}(n)}{\overline{\sigma}_{y_i}^2(n) + \overline{\sigma}_{F_i^k}^2(n)} \tag{6.23}$$

and

$$\underline{y}_{i,max}^k(n) = \frac{\underline{\sigma}_{y_i}^2(n)\underline{m}_{F_i^k}(n) + \underline{\sigma}_{F_i^k}^2(n)\underline{m}_{y_i}(n)}{\underline{\sigma}_{y_i}^2(n) + \underline{\sigma}_{F_i^k}^2(n)}. \tag{6.24}$$

Assuming that $m_{y_i(n)} = y_i(n)$ [73], then the nonsingleton interconnected type-1 fuzzy algorithm output is expressed by

$$f_{ns2}(\mathbf{y}(n)) = f_{ns2,\{1\}}(\mathbf{y}(n)) - f_{ns2,\{-1\}}(\mathbf{y}(n)) \tag{6.25}$$

where

$$f_{ns2,\{1\}}(\mathbf{y}(n)) = \frac{\underline{f}_{ns2,\{1\}}(\mathbf{y}(n)) + \overline{f}_{ns2,\{1\}}(\mathbf{y}(n))}{\underline{f}_{ns2}(\mathbf{y}(n)) + \overline{f}_{ns2}(\mathbf{y}(n))}, \tag{6.26}$$

$$f_{ns2,\{-1\}}(\mathbf{y}(n)) = \frac{\underline{f}_{ns2,\{-1\}}(\mathbf{y}(n)) + \overline{f}_{ns2,\{-1\}}(\mathbf{y}(n))}{\underline{f}_{ns2}(\mathbf{y}(n)) + \overline{f}_{ns2}(\mathbf{y}(n))}, \tag{6.27}$$

$$\underline{f}_{ns2,\{1\}}(\mathbf{y}(n)) + \overline{f}_{ns2,\{1\}}(\mathbf{y}(n)) = \sum_{\underline{\tilde{\mathbf{y}}}_i, \overline{\tilde{\mathbf{y}}}_i \in \mathcal{Y}_d^+} \left[\exp\left(\frac{-\left\|\mathbf{y}(n)-\underline{\tilde{\mathbf{y}}}_i\right\|^2}{2\left(\underline{\sigma}_r^2+\underline{\sigma}_v^2\right)}\right) + \exp\left(\frac{-\left\|\mathbf{y}(n)-\overline{\tilde{\mathbf{y}}}_i\right\|^2}{2(\bar{\sigma}_r^2+\bar{\sigma}_v^2)}\right)\right], \tag{6.28}$$

$$\underline{f}_{ns2,\{-1\}}(\mathbf{y}(n)) + \overline{f}_{ns2,\{-1\}}(\mathbf{y}(n)) = \sum_{\underline{\tilde{\mathbf{y}}}_j, \overline{\tilde{\mathbf{y}}}_j \in \mathcal{Y}_d^-} \left[\exp\left(\frac{-\left\|\mathbf{y}(n)-\underline{\tilde{\mathbf{y}}}_j\right\|^2}{2\left(\underline{\sigma}_r^2+\underline{\sigma}_v^2\right)}\right) + \exp\left(\frac{-\left\|\mathbf{y}(n)-\overline{\tilde{\mathbf{y}}}_j\right\|^2}{2(\bar{\sigma}_r^2+\bar{\sigma}_v^2)}\right)\right], \tag{6.29}$$

and

$$\underline{f}_{ns2}(\mathbf{y}(n)) + \overline{f}_{ns2}(\mathbf{y}(n)) = \sum_{\underline{\tilde{\mathbf{y}}}_k, \overline{\tilde{\mathbf{y}}}_k \in \mathcal{Y}_d} \left[\exp\left(\frac{-\left\|\mathbf{y}(n)-\underline{\tilde{\mathbf{y}}}_k\right\|^2}{2\left(\underline{\sigma}_r^2+\underline{\sigma}_v^2\right)}\right) + \exp\left(\frac{-\left\|\mathbf{y}(n)-\overline{\tilde{\mathbf{y}}}_k\right\|^2}{2(\bar{\sigma}_r^2+\bar{\sigma}_v^2)}\right)\right]. \tag{6.30}$$

Note that $\underline{\tilde{\mathbf{y}}}_k$ and $\overline{\tilde{\mathbf{y}}}_k$ refer to the uncertainties associated with the center of the *k*th vector state at the output channel. The separation between $\mathcal{Y}_d^+$ and $\mathcal{Y}_d^-$ is given by

$$f_{ns2}(\mathbf{y}(n)) = 0. \tag{6.31}$$

Now, supposing that there are not uncertainties in the membership functions, then $\underline{\sigma}_r = \overline{\sigma}_r = \sigma_r$, $\underline{\sigma}_v = \overline{\sigma}_v = \sigma_v$, $\underline{\tilde{\mathbf{y}}}_k = \overline{\tilde{\mathbf{y}}}_k = \tilde{\mathbf{y}}_k$. And, as a result

$$f_{ns2}(\mathbf{y}(n)) = f_{ns1}(\mathbf{y}(n)), \tag{6.32}$$

Furthermore, assuming that $\sigma_r^2 = 0$, then

$$f_{ns2}(\mathbf{y}(n)) = f_{ns1}(\mathbf{y}(n)) = f_{bs}(\mathbf{y}(n)). \tag{6.33}$$

Equation (6.33) means that under some assumptions, the type-1 interconnected fuzzy algorithm reduces to NOBE. The main advantages of the interconnected type-1 fuzzy system over type-1 and type-2 FLS are [41]: *i*) it deals with rule uncertainties, in a "worse case" sense, but in a way that is enough to capture considerable level of rules uncertainties; *ii*) the heavy machinery of type-2 FLS is not applicable. In fact, only the machinery of type-1 fuzzy is needed; *iii*) it presents the same number of parameters of type-2 FLS and twice the number of parameters of type-1 FLS and demands twice the computational effort of the type-1 FLS algorithm; and, finally, *iv*) updating rules of the parameters of the

algorithm can be obtained similarly to those of a type-1 FLS. As a result, better equalization of PL channel corrupted by high-power impulsive can be performed [18]. One has to note that the interconnected type-1 fuzzy algorithm can also be easily derived by using the decision-feedback scheme to reduce the number of channel states required for the decision purposes [74]. Simulation results presented in [18] show that equalization techniques based on interconnected type-1 fuzzy system can surpass the performance of traditional channel equalization approach applied to PL channels.

For illustration purpose, Fig. 6.6 shows the convergent behavior of adaptive and *decision feedback* (DF) version of the interconnected type-1 fuzzy equalizer introduced in [18] when applied to PL channel equalization. In this plot, fuzzy2-S-DFE and fuzzy2-NS-DFE denotes the singleton and nonsingleton adaptive interconnected type-1 fuzzy equalizers. Fuzzy-S-DFE and Fuzzy-NS-DFE denote the singleton and nonsingleton adaptive type-1 fuzzy equalizers proposed in [54]. From this plot, one can note that the interconnected type-1 fuzzy equalizers improves the convergence ratio of standard type-1 fuzzy equalizers.

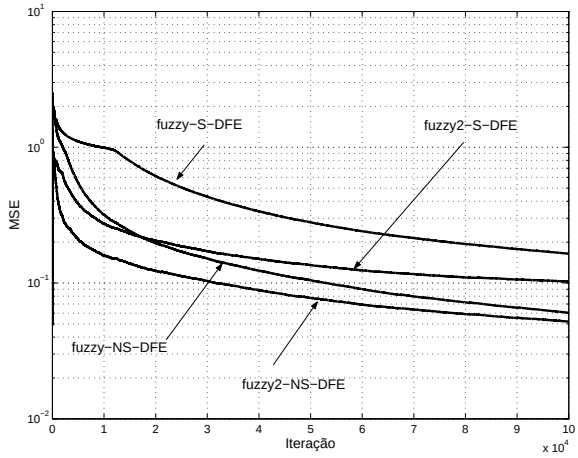



**Fig. 6.6.** Convergence of type-1 and interconnected type-1 fuzzy equalizers.

## 6.6 Mitigation of Impulsive Noise

The impulsive noise encourages investigation on communication techniques that can effectively cope with it. In fact, PLC system designs can be regarded as a highly challenging task as the engineer has to deal with very limited resources in a hostile environment, which in no way has been or could be prepared for communication purposes. For enhancing the data rate is not possible to extend bandwidth or assign new frequency ranges. Then, only more sophisticated modulation schemes with improved spectral efficiency or adaptation strategies such as impulsive noise cancellation can push technology forward. In this regard, several approaches have been proposed so far to cope with the impulsive noises in PLC as well as in other communication systems. These approaches follow two basic strategies:

*i*) the employment of powerful channel coding, interleaving, and modulation techniques [75, 76, 77, 78, 9, 79, 80, 81]. In fact, it is recognized that a great deal of research effort has been made towards the use of channel coding together with interleaving technique to mitigate impulsive noises in multicarrier, single-carrier, spread-spectrum-based systems for PLC applications. Improvements in terms of bit-rate as well as immunities to impulse noise attained by using turbo-coding, space-time codes, *Reed Solomon* (RS) codes, among others, have been reported since the last years [82, 83, 84, 85, 86, 42]. Recently, the use of the *low density parity check* (LDPC) codes isolated or combined with other codes has shown that a single-carrier system can be more efficient than a multicarrier one for PLC systems designed for dealing with impulse noise scenario [82]. Although the use of channel coding techniques shows improvement, two reasons can limit their full deployment: *a*) high power impulsive noises demand very long lengths interleaving procedure. As a result, end-to-end delay and, consequently, QoS for real-time applications could no longer be guaranteed, and *b*) the decoding errors generated by the channel decoding process degrade the performance of such techniques because of their intrinsic error propagation property.

Based on the premise that channel coding techniques are developed to guarantee reliable transmission through communication channels, it can be pointed out that improvements can be attained if a nonlinear technique is applied to reduce the impulsive noises at the channel output independently of the communication system applied (multi-

carrier, single-carrier and spread-spectrum). The use of these nonlinear techniques can significantly reduce not only the interleaving length, but also the decoding errors if well designed and implemented.

*ii*) the use of nonlinear techniques for impulsive noise cancellation at the channel output [87, 88, 89, 90, 91]. In this approach, a noise cancellation technique is added at the output of PL channel. A standard communication system as well as one with noise cancellation technique are illustrated in Fig. 6.7. In Fig. 6.7 (b), the block noise cancel implement the nonlinear technique applied to reduce the noise power at the input of the signal processing, demodulation and decoding techniques implemented in the PHY (physical layer) block, which is responsible for digital communication and signal processing algorithm running in the physical layer.

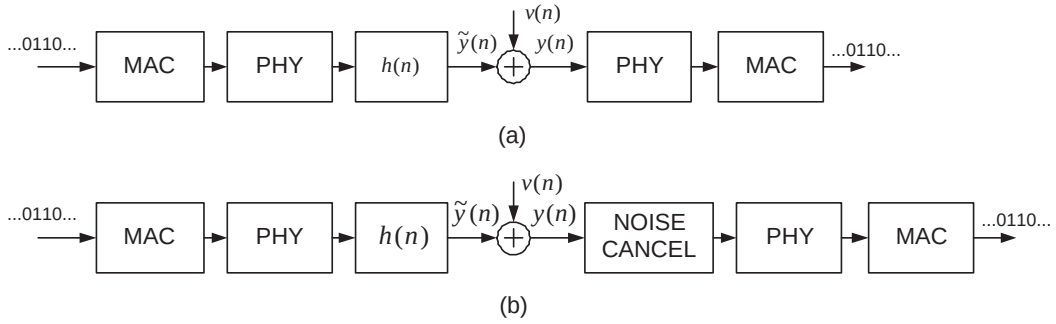



**Fig. 6.7.** a) A standard digital communication system and b) a digital communication system with noise cancellation technique.

The use of nonlinear techniques instead of linear ones to reduce the presence of impulsive noises are motivated by the following reasons [88, 92]: *i*) the impulsive noises can be seen as a nonlinear component at the channel output; *ii*) the linear techniques are unable to change features of impulsive noises such as regularity [92] and, finally, *iii*) the linear techniques are not appropriate for reducing the presence of non-stationary and broadband impulse noise.

The effect of impulsive noises in digital communication systems has been recently analyzed and the performance of new nonlinear techniques were analyzed in [41]. It can be stated that the standard techniques are two nonlinear functions given by [93]

$$\hat{y}(n) = \begin{cases} y(n), & \text{if } |y(n)| < \mathcal{A}_{\tilde{y}} \\ \mathcal{A}_{\tilde{y}} \exp\{j\angle y(n)\}, & \text{otherwise} \end{cases} \tag{6.34}$$

and

$$\hat{y}(n) = \begin{cases} y(n), & \text{if } |y(n)| < \mathcal{A}_{\tilde{y}} \\ 0, & \text{otherwise} \end{cases}, \tag{6.35}$$

where $y(n)$ is *n*th output sample of a PL channel corrupted by the presence of impulsive noises, $\mathcal{A}_{\tilde{y}} = \max_n |\tilde{y}(n)|$, $\angle y(n)$ is the phase of $y(n)$.

According to [41], a very interesting technique showing good results when applied for impulsive noises mitigation in digital communication systems is the one derived from *median filtering* (MF) approach [91, 90]. The *weighted median filtering* (WMF) technique can be expressed by

$$\hat{y}(n+K+1) = \begin{cases} y(n+K+1), & \text{if } \Psi(\mathbf{y}_{2K+1}(n)) \text{ is true} \\ \text{MEDIAN}\left(\mathbf{y}_{2K+1}(n), \mathbf{W}_{2K+1}^{MF}\right) & \\ \text{otherwise} & \end{cases}, \tag{6.36}$$

where $\mathbf{W}_{2K+1}^{MF} = \mathcal{W}_1, \mathcal{W}_2, ..., \mathcal{W}_{2K+1}$ are $2K+1$ weight samples, $\mathbf{y}_{2K+1}(n)$ is a $(2K+1)$-length vector whose elements are given by 6.8, $\Psi(\mathbf{y}_{2K+1}(n))$ is a condition that can or cannot be satisfied, $K$ is the maximum length of impulsive noises that can be canceled, and $\text{MEDIAN}\left(\mathbf{y}_{2K+1}(n), \mathbf{W}_{2K+1}^{MF}\right)$ is the weighted median filter. Note that if $\Psi(\mathbf{y}_{2K+1}(n))$ is not considered and $\mathcal{W}_1 = \mathcal{W}_2 = \cdots = \mathcal{W}_{2K+1} = 1$, then the MF-based technique turns into a traditional MF. On the other hand, if

$$\begin{aligned} &\Psi(\mathbf{y}_{2K+1}(n)) = \\ &|y(n+K+1) - \text{MEDIAN}\left(\mathbf{y}_{2K+1}(n), \mathbf{W}_{2K+1}^{MF}\right)| < \rho, \end{aligned} \tag{6.37}$$

where $\rho$ is a threshold, then a conditioned (C) WMF is accomplished. In addition, if $\mathcal{W}_1 = \mathcal{W}_2 = \cdots = \mathcal{W}_{2K+1} = 1$ in (6.37), the *conditioned* MF (CMF) technique proposed in [91] is obtained. In fact, numerical simulations have revealed that only conditioned MF versions are able to show some interesting results in PL channels. This affirmation

makes sense because the CMF is only an effective solution for impulsive noises mitigation during the occurrence periods of such noises. Note that (6.34)-(6.35) can also provide this kind of improvement.

Another promising class of nonlinear technique capable of reducing the presence of impulse noise is the one emerged from soft computing or computational intelligence paradigms such as *neural networks* (NN's), *wavelet networks* (WN's) and *fuzzy systems* (FS's). These techniques are gaining widespread acceptance in a large variety of fields, from engineering to commercial, from forecasting to artificial intelligence. The reason for such an increasing interest resides in their intrinsic generality, flexibility and good performance in many applications where other methods either tend to fail, or become cumbersome (in particular impulse noise cancellation).

Among NNs, WNs and FSs, one has to note that FSs (in particular type-2 FSs) are being successfully applied to several problems in which the degree of uncertainties is high and, consequently, traditional techniques have failed in coping with such uncertainties in the following situation [73]:

- Measurement noise is nonstationary, but the nature of the nonstationarity cannot be expressed mathematically ahead of time, e.g. time-series forecasting under variable *signal-to-noise ratio* (SNR) measurements, communication channel under impulse noise presence;
- A data-generating mechanism is time-varying, but the nature of the time variations cannot be expressed mathematically ahead of time, e.g. equalization and co-channel interference reduction for nonlinear and time-varying digital communication channels;
- Features are described by statistical attributes that are nonstationary, but the nature of the nonstationarity cannot be expressed mathematically ahead of time, e.g. rule-based classification of video traffic.

As shown in [41], the use of computational intelligence based techniques to reduce the hardness of impulsive noise at the channel output is a powerful technique. However, one has to note that the use of such technique demands a training sequence at the Rx side for its training and, consequently, considerable computational burden.

**Impulsive noise cancellation in DMT and OFDM systems**

The following definition and nomenclature are considered for both *orthogonal frequency division multiplexing* (OFDM) and *discrete multitone transceiver* (DMT) systems:

*i*) $\mathbf{h} = [h_0 \cdots h_{L_h-1}]$ refers to the impulse response vector of the PL channel, $\mathbf{w} = [w_0 \cdots w_{L_w-1}]$ is the vector representation of the time domain equalizer responsible for shortening the channel impulse response. $\mathbf{H} = [H_0 \cdots H_{N-1}]$ and $\mathbf{W} = [W_0 \cdots W_{N-1}]$ are vectors constituted by the DFT samples expressed by

$$H_k = \sum_{n=0}^{N-1} h_n e^{-j(2\pi/N)nk},\ k = 0, ..., N-1, \tag{6.38}$$

and

$$W_k = \sum_{n=0}^{N-1} w_n e^{-j(2\pi/N)nk},\ k = 0, ..., N-1, \tag{6.39}$$

respectively;

*ii*) $\mathbf{c} = \mathbf{h} * \mathbf{w} = [c_0 \cdots c_{L_c-1}]$ denotes the vector representation of the shortened impulse response, whose length and effective length are $L_h + L_w - 1$ and $L_b$, respectively. Since the filter for shortening impulse response, $\mathbf{w}$, is not ideal, then $\mathbf{c} = \mathbf{c}^{ISI} + \mathbf{c}^{ISI_{res}}$, where $\mathbf{c}^{ISI_{res}} = [c_0 \cdots c_{L-1}\ \mathbf{0}_{L_b}\ c_{L+L_b-1} \cdots c_{L_c-1}]$ is responsible for ISI component that results from the fact that $\mathbf{w}$ is not ideal and $\mathbf{c}^{ISI} = [\mathbf{0}_L\ c_L \cdots c_{L+L_b-1}\ \mathbf{0}_{L_c-L-L_b}]$ produces ISI if the length of the *cyclic prefix* (CP) $L_{cp} \geq L_b - 1$; $\mathbf{0}_{L_O}$ denotes a vector composed of $L_O$ zeros. The *k*th samples of $C_k^{ISI}$ and $C_k^{ISI_{res}}$ of the vector $\mathbf{C} = [C_0 \cdots C_{N-1}] = [C_0^{ISI} + C_0^{ISI_{res}} \cdots C_{N-1}^{ISI} + C_{N-1}^{ISI_{res}}]$ can be approximated by

$$C_k^{ISI} \cong \sum_{n=0}^{N-1} c_n^{ISI} e^{-j(2\pi/N)nk},\ k = 0, \cdots, N-1, \tag{6.40}$$

e

$$C_k^{ISI_{res}} \cong \sum_{n=0}^{N-1} c_n^{ISI_{res}} e^{-j(2\pi/N)nk},\ k = 0, \cdots, N-1; \tag{6.41}$$

*iii*) the frequency representation of the additive noise in the $k$th sub-channel is given by

$$V_k = V_{bkgr,\,k} + V_{nb,\,k} + V_{pa,\,k} + V_{ps,\,k} + V_{imp,\,k}\,, \tag{6.42}$$

where $k = 0, \cdots, N-1$,

$$V_{bkgr,\,k} = \sum_{n=0}^{N-1} v_{bkgr}(n) e^{-j(2\pi/N)nk}, \tag{6.43}$$

$$V_{nb,\,k} = \sum_{n=0}^{N-1} v_{nb}(n) e^{-j(2\pi/N)nk}, \tag{6.44}$$

$$V_{pa,\,k} = \sum_{n=0}^{N-1} v_{pa}(n) e^{-j(2\pi/N)nk}, \tag{6.45}$$

$$V_{ps,\,k} = \sum_{n=0}^{N-1} v_{ps}(n) e^{-j(2\pi/N)nk}, \tag{6.46}$$

and

$$V_{imp,\,k} = \sum_{n=0}^{N-1} v_{imp}(n) e^{-j(2\pi/N)nk}; \tag{6.47}$$

*iv*) the PSD of $x(n)$, $v(n)$, $v_{bkgr}(n)$, $v_{nb}(n)$, $v_{pa}(n)$, $v_{ps}(n)$, and $v_{imp}(n)$ in the $k$th sub-channel are denoted by $\mathcal{S}_k$, $\mathcal{S}_{v,k}$, $\mathcal{S}_{bkgr,k}$, $\mathcal{S}_{nb,k}$, $\mathcal{S}_{pa,k}$, $\mathcal{S}_{ps,k}$, and $\mathcal{S}_{imp,k}$, respectively.

Assuming that PL channel is time-invariant during a symbol transmission and its discrete time version is expressed by (6.8), then the SNR for multicarrier (in particular OFDM and DMT) are given by

$$\mathrm{SNR_k} = \begin{cases} \frac{S_k \left|C_k^{ISI}\right|^2}{S_{v,\,k}|W_k|^2 + S_k|C_k^{ISI_{res}}|^2}, & \text{if DMT} \\ \frac{S_k|H_k|^2}{S_{v,\,k}}, & \text{if OFDM} \end{cases} \tag{6.48}$$

where $N$ is the number of sub-channels and $k = 0, \cdots, N-1$. $H_k$ is the $k$th coefficient of the channel in the frequency domain. The terms $S_k|C_k^{ISI_{res}}|^2,\ k = 0, ..., N-1$, appear due to the fact that the *time domain equalizer* (TEQ) [94] is unable to ideally shorten the channel impulse response. $\left|C_k^{ISI}\right|^2,\ k = 0, ..., N-1$, are frequency attenuations of the transmitted signals produced by the coefficients of the $\mathbf{c}(n)$ that lie inside the effective impulse response whose length is $L_b$. $S_{v,\,k}\,|W_k|^2,\ k = 0, ..., N-1$, are the filtered noise generated by filtering the additive noise with the TEQ, and $W_k,\ k = 0, ..., N-1$, are frequency domain weights of the TEQ. As, all terms at the right side of (6.48) are positive, a nonlinear technique for impulsive noise cancellation can be applied. If this technique does not affect other components, then the SNR in the $k$th sub-channel can be expressed by

$$\mathrm{SNR_k} = \begin{cases} \frac{S_k \left|C_k^{ISI}\right|^2}{S_{bkgr,\,k}|W_k|^2 + S_k|C_k^{ISI_{res}}|^2}, & \text{if DMT} \\ \frac{S_k|H_k|^2}{S_{bkgr,\,k}}, & \text{if OFDM} \end{cases}. \tag{6.49}$$

As a result, the following inequalities are valid:

$$\frac{S_k \left|C_k^{ISI}\right|^2}{S_{bkgr,\,k}\,|W_k|^2 + S_k|C_k^{ISI_{res}}|^2} \geq \frac{S_k \left|C_k^{ISI}\right|^2}{S_{v,\,k}\,|W_k|^2 + S_k|C_k^{ISI_{res}}|^2} \tag{6.50}$$

and

$$\frac{S_k\,|H_k|^2}{S_{bkgr,\,k}} \geq \frac{S_k\,|H_k|^2}{S_{v,\,k}}. \tag{6.51}$$

By applying the same idea to single-carrier and *code division multiple access* (CDMA) systems, equivalent SNR improvement can be achieved.

## 6.7 Channel Coding

the use of suitable error control strategies must be applied under the following two restrictions in order to ensure reliable communication in hostile environments such as power lines and to reduce the bit error rate [9]: *i*) PLC networks shall operate with a signal power that has to be below a limit defined by the regulatory bodies and *ii*) the signal level has to keep data transmission over PLC medium possible. That means there should be a certain SNR level in the network making communications possible. Applications of *forward error correction* (FEC) and interleaving mechanisms [75, 80, 77, 78, 79] are commonly found in digital communication systems to cope with the disturbances in PL channels such as impulsive noises as well as strong ISI. FEC can improve the net throughput in noisy networks such as PLC networks so that the requirement to repeat data packets is reduced. In a noisy PLC environment, particularly in the presence of line synchronous impairments causing burst errors, modems with FEC can offer more reliable data communication than non-FEC modems.

In case of errors with FEC and interleaving, the damaged data have to be retransmitted by an *automatic repeat request* (ARQ) mechanism. The application of ARQ can reduce the error probability to a very low value and it is only

limited by the remaining error probability of *cyclic redundancy check* (CRC) code used for error recognition, or error tolerance specified by a particular application [9].

According to [95], the interleaving of modulated symbols at the Tx side, which are transmitted over a communication channel, followed by deinterleaving at the receiver disperse burst of channel-symbol errors or corrupted symbols over a number of codewords or constraint lengths, thereby facilitating the removal of the errors by the decoding. Ideally, the interleaving and deinterleaving ensure that the decoder encounters statistically independent symbol decisions or metrics, as it would if the channel was memoryless. Interleaving of channel symbols is useful when error bursts are caused by fast fading, interference, or even decision directed equalization.

To deal with the high-power impulsive noise bursts as well as strong ISI in NaPLC and BoPLC systems implemented with several modulation schemes, the use of Reed-Solomon codes, convolutional codes, turbo codes, bit-interleaved coded modulation, trellis codes, space-time codes, and low density parity check LDPC codes [86, 96, 97, 98, 99, 83, 42, 100, 9, 82, 85, 84, 101, 82, 102, 103, 104] have been investigated the last few years.

From recent results in this field, one can note that a great research effort on the use of turbo coding and LDPC coding techniques to improve the performance of digital communication system for PLC networks.

## 6.8 Multi-Carrier, Spread Spectrum, and Single-Carrier

The selection of a modulation scheme for PLC system has to take into account three major issues: *i*) the presence of noise and impulse disturbances causing a relatively low SNR, *ii*) the time-varying frequency-selective nature of the channel, and *iii*) regulatory constraints with regard to electromagnetic compatibility that limit the transmitted power. A choice for a robust solution against PLC impairments as well as an adaptive one capable of overcoming the PLC hardness should be selected. Regarding the former choice, one has to note that it refers to using standard digital communication strategies to develop a reliable and robust digital communication system capable of delivering QoS for the target applications. The latter choice is something totally new, because an adaptive digital communication system has to be capable of selecting appropriate modulation, access, coding techniques and providing dynamically spectrum allocation to achieve the system requirements. For doing so, cognitive radios, which are adaptive and extremely programmable, learning users' preferences automatically adjusted to changes in the operating environment, have to be applied to future and powerful PLC modems [105, 106, 107, 108].

In fact, PLC systems have to manage point-to-multipoint multiuser communications, share spectrum with other well-established telecommunication system and deal with adverse and time-varying PL channel. Then, depending on the target application, each modulation technique has certain advantages. For instance, for a lower-cost and low-data-rate power line system, FSK seems to be a good solution. For higher data rates up to 1Mb/s CDMA offers the advantage of using the its processing gain to meet radiation allowance limits [11]. Additionally, for low-data-rate communication or NaPLC applications we can modulate digital signals onto the power lines by using modulation techniques such as BPSK, *phase shift keying* (PSK), *frequency shift keying* (FSK), *minimum shift keying* (MSK), and *Gaussian minimum shift keying* (GMSK) [109, 76]. However, for broadband applications, advanced techniques such as M-ary PSK (MPSK), M-ary *quadrature amplitude modulation* (MQAM), and M-ary FSK (MFSK) have to be applied [110, 9]. Additionally, to cope with the impairments of PL channels, PLC systems require robust and efficient modulation techniques such as *spread-spectrum* (SS) and *multicarrier* (MC) schemes [111, 112, 113]. Among them, OFDM [114], DMT [86, 41, 88, 115] and CDMA have been analyzed for PLC applications. And, by performing a fair comparison between DS-CDMA and OFDM systems for BoPLC downstream, one can note that OFDM achieve remarkable performance and high flexibility in resources management, while CDMA guarantees good performance and satisfactory allocation policies with simple and cheap receivers [116, 117]. Recently, various combinations of both schemes (MC-SS), like *multicarrier* CDMA (MC-CDMA), have been proposed so far for BoPLC [100, 118]. The advantage offered by MC-SS schemes refer to the fact that they achieve very good performances in the case of multiuser communications in difficult environments and, therefore, represent a potential solution for PLC system.

### 6.8.1 Single Carrier System

Single-carrier modulation is an attractive proposition from the complexity point of view [8, 119]. Despite that, single-carrier modulation cannot offer more than 1 bit/s/Hz for PLC applications [120]. This is due to the fact that PL channels introduce strong *intersymbol interference* (ISI), then powerful detection and equalization techniques are demanded. Additionally, deep frequency notches present in the PL channels prevent the use of linear equalizers, as the noise enhancement they cause is a serious drawback on a very noisy channel, so only poor performance can be usually achieved. In the access domain where typical delay spreads are around 10 $\mu$s dramatic ISI would occur already for data rates far below 100 kb/s [120]. Thus, the application of expensive channel equalizers cannot be avoided. So the

advantage of simplicity in single-carrier modulation is lost. However, in [82] it is demonstrated that under the hardness of PL channels corrupted by strong noise bursts, a single carrier system using a PAM technique along with a powerful coding scheme based on LDPC coding is capable of surpassing the performance of DMT and OFDM system for PLC applications. In fact, according to [121], a DMT or OFDM system is severely degraded by the hardness of PL channel as well as the high-power impulsive noises.

### 6.8.2 Spread-Spectrum System

Spread spectrum is a type of modulation that spreads data to be transmitted across the entire available frequency band, exceeding the minimum bandwidth required to send the information. SS systems come in several varieties: direct sequence, frequency hopping, time hopping, chirp, and hybrid methods. The *direct sequence* (DS) spread spectrum techniques have the ability to realize a multiple access structure in a simple way by choosing suitable spreading sequences, that is, code-division multiple access, and hence is widely used in practical communication systems [100, 111]. In PL environments, SS system would show improved noise immunity over narrowband systems. They possess numerous advantages such as robustness against non-Gaussian noise, interception difficulty, alleviation of multipath interference using coherent RAKE reception, and narrowband interference suppression [122]. Additionally, it demands low power spectrum density for data transmission.

Past contributions advocated that the effect of spread spectrum modulation is considered limited for PL channels, because the useful bandwidth in these channels is under 30 MHz. However, recent contributions point out that the useful bandwidth ranges from 1 up to 100 MHz [6, 5]. Then, one can take full advantage of the interference suppression of SS techniques, because a large bandwidth is available in PL channel and, consequently, the maximum data-rate of this transmission bandwidth can be enough for the use of a multiple access technique called CDMA, which allows several users, possibly with different rate demands, to access the PLC channel simultaneously.

From recent results, it is expected a great research effort on the analysis and deployment of the CDMA for PLC systems operating in the frequency range from 1 MHz to 100 MHz.

### 6.8.3 OFDM and DMT System

Recently OFDM and DMT have been proposed as very good candidates for transmission due to their merits in simplifying channel estimation, and high bandwidth efficiency and flexibility in high bit rates when applied to frequency-selective channels such as PL ones. Following Shannon's optimum transmission suggestion, as pointed out in [123], the OFDM or the DMT achieves the highest performance in channels with frequency-selective fading and severe ISI.

The OFDM is, in essence, a kind of parallel digital modulation technique, which can overcome the multipath effect by prolonging the symbol period of the transmitted signal. In the OFDM, a stream of high rate serial data is divided into a number of streams of lower-rate parallel data by orthogonal subcarriers so that the symbol period of every subchannel is lengthened. By doing so, then subband is so narrow that the associated subchannel has a flat frequency response. As a result, a frequency-selective channel becomes equivalent to a set of multiple flat-fading suhchannels and ISI is decreased. The *interchannel interferences* (ICI) of the subcarriers can be eliminated theoretically by the orthogonality between the subcarriers, and equalization of the data with coherent detection simply amounts to a normalization by a complex scalar, while incoherent detection does not require any further equalization. More important, each symbol can be detected separately, an approach that can be extremely successful in preventing errors caused by strong channel attenuation or noise in specific subhands. Hence, OFDM is robust against narrowband interference and high noise levels [42, 124, 125]

The total bandwidth demanded by OFDM or DMT is divided into N parallel subchannels, and bits are assigned to subchannels in direct proportion to the subchannel SNRs. The scheme that assigns the energy and hits to the different subchannels is called a loading algorithm. In spite of its robustness against the frequency selectivity, which is seen as an advantage of OFDM, any time-varying character of the channel is known to pose limits to the system performance because it results in the lost of orthogonality. The occurrence of ICI appears because the signal components of a subcarrier interfere with those of the neighboring subcarriers.

For PLC applications, coded-OFDM and coded-DMT systems have been not only investigated by scientific communities, but also applied to commercial BoPLC modems [86, 126, 127, 128, 21, 42, 124, 120, 12, 129, 2, 15]. Further development to increase the coded-OFDM or coded-DMT performance is on the way to overcome the weakness of OFDM- and DMT-based system, (see [130, 41].

### 6.8.4 MC-SS Systems

The basic MC-CDMA signal is generated by a serial concatenation of classical DSCDMA and OFDM. Each chip of the direct sequence spread data symbol is mapped onto a different sub-carrier. Thus, with MC-CDMA the chips

of a spread data symbol are transmitted in parallel on different sub-carriers, in contrast to a serial transmission with DS-CDMA.

The MC-DS-CDMA signal is generated by serial-to-parallel converting data symbols into $N_c$ sub-streams and applying DS-CDMA on each individual sub-stream. Thus, with MC-DS-CDMA, each data symbol is spread in bandwidth within its sub-channel, but in contrast to MC-CDMA or DS-CDMA not over the whole transmission bandwidth for Nc¿1. An MC-DS-CDMA system with one sub-carrier is identical to a single-carrier DS-CDMA system. MC-DS-CDMA systems can be distinguished in systems where the sub-channels are narrowband and the fading per sub-channel appears flat and in systems with broadband sub-channels where the fading is frequency-selective per sub-channel. The fading over the whole transmission bandwidth can be frequency-selective in both cases.

The use of MC-CDMA systems was analyzed in [100].The inherent spectral diversity of the MC-CDMA system significantly improves the PLC system performance. Also, the MC-CDMA system is also more robust to the influence of the impulse noise for the same reasons. Furthermore, for the MC-CDMA system, the infected tones can be easily zeroed, resulting in almost no loss in the PLC system performance in the presence of impulse noise.

The performance of DS-MC-based system was discussed in [118]. In this contribution, the authors show that adaptive DS-MC system with multiple access is suitable to backward and forward links of PLC networks.

Revising the literature about the use of combined SS and MC for PLC networks, few contributions are noted. In fact, few researchers have put their attention on the advantages offered by the combination of SS and MC for BoPLC and NaPLC applications.

## 6.9 Multiple Access

Multiple access is the ability of many users to communicate with each other while sharing a common transmission medium. PL multiple-access communications are facilitated if the transmitted signals are orthogonal or separable in some sense. The choice for a multiple access scheme depends on the applied transmission system within the physical layer and its features.

Signals may be separated in time (time-division multiple access or TDMA), frequency (frequency-division multiple access or FDMA), or code (code-division multiple access or CDMA).

In a TDMA scheme, the time axis is divided into “time slots” that represent the accessible portions of transmission resources provided by the multiple access scheme. The benefit of using TDMA is that it provides an upper bound on access delay; thus, QoS is guaranteed. However, the difficulty in generating a synchronized clock signal in power line networks between devices remains a problem [2].

FDMA scheme can be implemented in different transmission systems, such as spread spectrum and OFDM-based transmission systems, which are considered suitable for BoPLC systems. Recently, the use of FDMA was investigated in [118].

CDMA is implemented by using spread spectrum modulation while transmitting signals from multiple users in the same frequency band at the same time. Information theory indicates that in an isolated environment, CDMA systems achieve the same spectral efficiency as TDMA or FDMA systems only if optimal multiuser detection is used. However, due to the low bit-rate achieved with CDMA, its use is very limited. However, one has to note that recent conclusions about frequency range between 1 and 100 MHz for viable operation of PLC systems open the possibility for the use of CDMA, wideband CDMA, and multi-user detection techniques for multiple access in PLC systems [131, 100].

## 6.10 Impact on Video, Image, and Audio Transmission

Multimedia content transmissions over PLs can provide low-cost and a simple solution for those applications that need to transmit audio/video data. This can be accomplished if requirements such as bandwidth, latency, and jitter to the different nodes and traffic profiles coexisting in a PLC network are guaranteed by effectively combining resource allocation algorithms, prioritization of outgoing traffic in different categories (best effort, low priority, high priority service), node classification, and intelligent buffer management policies. For example, a real time video surveillance system with both visual and motion detectors which transmits video over the power line would not require the expensive investment needed to install video cables and telephone connection that offers security and good speech quality. [132, 133, 16]

Although the power line network is an extremely harsh medium environment for multimedia content transmission, rapid development of technology is showing that this can be overcome in the near future. Since the existing PL networks is already the most pervasive network in the world. PL networks reach into corners and rooms where most

electronic devices are already plugged into power outlets and would not need additional networks. Then, PLC networks can serve as the backbone network for multimedia content transmissions [134].

Despite that, video transmissions over power lines is not an easy task because of the characteristics of the PL channels which typically exhibits high error probability and relatively low transmission rates. Besides, the low-delay requirements of multimedia signals can make unpractical the usual mechanism of acknowledgement and packet retransmission. The nature of these target applications brings with itself a new, more stringent set of requirements: higher data rates to provide support for multiple HDTV streams, QoS metrics must be met in terms of latency, jitter, and very low frame error rates (for video and voice applications), and since many target applications require embedding of the communications circuitry in consumer electronics equipment, the add-in cost must be as low as possible.

The development of powerful signal processing techniques capable of diminishing the impairments of PL channels to video, audio and image transmissions will provide in the near future reliable and robust PLC networks capable of delivering a wide range of multimedia content to residential, commercial and industrial facilities. In fact, few contributions have addressed the issues that impair multimedia content transmissions in PLC networks.

## 6.11 Concluding Remarks

This contribution discussed the power lines for narrowband and broadband data communications. Several key issues, problems, and demands of NaPLC and BoPLC applications for investigations were highlighted. Special emphasis was given to PL channel modeling, additive noise modeling, channel equalization with powerful fuzzy techniques, impulsive noise cancellation, channel coding, modulation and multiple access schemes, and some issues related to multimedia content transmission through PLC networks.

Some important questions that deserve attention by those involved or interested in the PLC are: *i*) the use of cognitive ratio concepts for PLC modems; *ii*) investigation of impulse noise cancellation and denoising techniques together with interleaving and channel coding ones; *iii*) development of BoPLC and NaPLC systems for MV lines; *iv*) analysis of combined SS and MC systems, and *v*) investigation of schemes to provide reliable multimedia content transmissions.

## References

1. T. Zahariadis and K. Pramataris, “Broadband consumer electronics networking and automation,” *Int. J. of Commun. Syst.*, no. 17, pp. 27–53, 2004.
2. Y.-J. Lin, H. A. Latchman, R. E. Newman, and S. Katar, “A comparative performance study of wireless and power line networks,” *IEEE Comm. Magazine*, vol. 41, no. 4, pp. 54–63, Apr. 2003.
3. W. Sanderson, “Broadband communications over a rural power distribution circuit,” in *Proc. IEEE Southeastcon*, 2000, pp. 497–504.
4. P. J. Langfeld, “The Capacity of typical power line reference channels and strategies for system design,” in *Proc. International Symposium on Power-line Communication and its Applications*, 2001, pp. 271–278.
5. E. Liu, Y. Gao, G. Samdani, O. Mukhtar, and T. Korhonen, “Broadband powerline channel and capacity analysis,” in *Proc. International Symposium on Power-line Communication and its Applications*, 2005, pp. 7–11.
6. P. Amirshahi and M. Kavehrad, “High-frequency characteristics of overhead multiconductor power lines for broadband communications,” *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 7, pp. 1292–1303, July 2006.
7. M. Kuhn, S. Berger, I. Hammerström, and A. Wittneben, “Power line enhanced cooperative wireless communications,” *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 7, pp. 1401–1410, July 2006.
8. K. Dostert, *Power Line Communications*, ser. NY. Prentice Hall, 2001.
9. H. Hrasnica, A. Haidine, and R. Lehnert, *Broadband Powerline Communications: Network Design*, ser. NY. John Wiley & Sons, 2005.
10. M. Gebhardt, F. Weinmann, and K. Dostert, “Physical and regulatory constraints for communication over the power supply grid,” *IEEE Communications Magazine*, vol. 41, no. 5, pp. 84–90, May 2003.
11. N. Pavlidou, A. J. H. Vinck, J. Yazdani, and B. Honary, “Power line communications: State of art and future trends,” *IEEE Comm. Magazine*, vol. 41, no. 4, pp. 34–40, Apr. 2003.
12. “Homeplug powerline alliance,” www.homeplug.org.
13. M. Y. Chung, M.-H. Jung, T.-J. Lee, and Y. Lee, “Performance analysis of HomePlug 1.0 MAC with CSMA/CA,” *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 7, pp. 1411–1420, July 2006.
14. “Open plc european research alliance (opera),” www.ist-opera.org.
15. “Ieee p1901 draft standard for broadband over power line networks: Medium access control and physical layer specifications,” http://grouper.ieee.org/groups/1901/.
16. Y.-J. Lin, H. A. Latchman, M. Lee, and S. Katar, “A power line communication network infrastructure for the smart home,” *IEEE Wireless Communications*, vol. 9, no. 6, pp. 104–111, Dec. 2002.

17. T. Esmailian, “Multi mega bit per second data transmission over in-building power lines,” Ph.D. Dissertation, University of Toronto (Canada), 2003.
18. M. V. Ribeiro, “Signal processing techniques for powerline communications and power quality applications,” Ph.D. Dissertation, University of Campinas (Brazil), 2005.
19. V. Gungor and F. Lambert, “A survey on communication networks for electric system automation,” *Computer Networks*, no. 50, p. 877-897, 2006.
20. P. Sutterlin, “A power line communication tutorial: Challenges and technologies,” in *Proc. IEEE International Conference on Power Line Communications and Its Applications*, 1998.
21. “Mitsubishi,” http://global.mitsubishielectric.com/bu/plc.
22. P. R. S. Pimentel, A. Baldissin, L. S. Cesar, R. R. Framil, and A. Pascalicch, “Revolution in the distribution (use of the technology power line communication in the transmission of data, voice and images),” in *Proc. Latin America Transmission and Distribution Conference and Exposition*, Jan. 2004, pp. 314–320.
23. “Luz para todos (light for everyone),” http://www.mme.gov.br/, 2004.
24. A. Cunha, P. Biggs, and H. Quintella, “The challenges of integrating BPL into existing telecom business models: market visions from Brazilian research,” in *Proc. International Symposium on Power Line Communications and Its Applications*, 2005, pp. 395–399.
25. “Fórum aptel de plc/brasil,” http://aptel.com.br/.
26. M. Zimmermann and K. Dostert, “A Multipath model for the powerline channel,” *IEEE Trans. on Communications*, vol. 50, no. 4, pp. 553–559, Apr. 2002.
27. D. Cooper and T. Jeans, “Narrowband, low data rate communications on the low-voltage mains in the CENELEC frequencies I. Noise and attenuation,” *IEEE Trans. on Power Delivery*, vol. 17, no. 3, pp. 718–723, July 2002.
28. ——, “Narrowband, low data rate communications on the low-voltage mains in the CENELEC frequencies. II. Multiplicative signal fading and efficient modulation schemes,” *IEEE Trans. on Power Delivery*, vol. 17, no. 3, pp. 724–729, July 2002.
29. S. Galli and T. Banwell, “A novel approach to the modeling of the indoor power line channel-Part II: transfer function and its properties,” *IEEE Trans. on Power Delivery*, vol. 20, no. 3, pp. 1869–1878, July 2005.
30. T. Banwell and S. Galli, “A novel approach to the modeling of the indoor power line channel part I: circuit analysis and companion model,” *IEEE Trans. on Power Delivery*, vol. 20, no. 2, Part 1, pp. 655–663, Apr. 2005.
31. S. Galli and T. C. Banwell, “A deterministic frequency-domain model for the indoor power line transfer function,” *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 7, pp. 1304–1316, July 2006.
32. P. S. Henry, “Interference characteristics of broadband power line communication systems using aerial medium voltage wires,” *IEEE Communications Magazine*, vol. 43, no. 4, pp. 92–98, Apr. 2005.
33. N. Suljanovic, A. Mujcic, M. Zajc, and J. F. Tasic, “Computation of high-frequency and time characteristics of corona noise on HV power line,” *IEEE Trans. on Power Delivery*, vol. 20, no. 1, pp. 71–79, Jan. 2005.
34. G. Jee, C. Edison, R. D. Rao, and Y. Cern, “Demonstration of the technical viability of PLC systems on medium- and low-voltages lines in the United States,” *IEEE Comm. Magazine*, vol. 41, no. 5, pp. 108–112, May 2003.
35. P. A. A. F. Wouters, P. C. J. M. van der Wielen, J. Veen, P. Wagenaars, and E. F. Steennis, “Effect of cable load impedance on coupling schemes for MV power line communication,” *IEEE Trans. on Power Delivery*, vol. 20, no. 2, Part 1, pp. 638–645, Apr. 2005.
36. F. Corripio, J. Arrabal, L. D. D. Rio, and J. Munoz, “Analysis of the cyclic short-term variation of indoor power line channels,” *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 7, pp. 1327–1338, July 2006.
37. S. Barmada, A. Musolino, and M. Raugi, “Innovative model for time-varying power line communication channel response evaluation,” *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 7, pp. 1317–1326, July 2006.
38. M. Katayama, T. Yamazato, and H. Okada, “A mathematical model of noise in narrowband power line communication systems,” *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 7, pp. 1267–1276, July 2006.
39. M. Zimmermann and K. Dostert, “Analysis and modeling of impulse noise in broad-band powerline communications,” *IEEE Trans. on Electromagnetic Compatibility*, vol. 44, no. 1, pp. 249–258, May 2002.
40. V. Degardin, M. Lienard, A. Zeddam, F. Gauthier, and P. Degauque, “Classification and characterization of impulsive noise on indoor power lines used for data communications,” *Electronics Letters*, vol. 48, no. 4, p. 913-918, Nov. 2002.
41. M. V. Ribeiro, C. A. Duque, and J. M. T. Romano, “An interconnected type-1 fuzzy algorithm for impulsive noise cancellation in multicarrier-based power line communication systems,” *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 7, pp. 1364–1376, July 2006.
42. E. Biglieri, “Coding and modulation for a horrible channel,” *IEEE Comm. Magazine*, vol. 41, no. 5, pp. 92–98, May 2003.
43. C. Luo, S. Cheng, L. Xiong, J. Nguimbis, Y. Zhang, and J. Ma, “A nonlinear equalization method based on multilayer perceptron for OFDM power-line communication,” *IEEE Trans. on Power Delivery*, vol. 20, no. 4, pp. 2437–2442, Oct. 2005.
44. M. V. Ribeiro, “On Fuzzy-DFE-LMS and Fuzzy-DFE-RLS algorithms to equalize power line channels,” in *IEEE International Symposium on Industrial Electronics*, 2003, pp. 1001–1006.
45. S. Chen, G. J. Gibson, C. F. N. Cowan, and P. M. Grant, “Adaptive equalization of finite nonlinear channels using multilayer perceptrons,” *Signal Processing*, vol. 20, pp. 107–119, 1990.
46. J. M. Mendel, “Unicertainty, fuzzy logic, and signal processing,” *Signal Processing*, vol. 80, pp. 913–933, 2000.
47. S. K. Patra and B. Mulgrew, “Efficient architecture for bayesian equalization using fuzzy filters,” *IEEE Trans. on Circuits and Systems - II, Analog and Digital Signal Processing*, vol. 45, no. 7, pp. 812–820, 1998.
48. M. Ibnkahla, “Applications of neural networks to digital communications - a survey,” *Signal Processing*, vol. 80, no. 7, pp. 1185–1215, 2000.

49. L. M. Reyneri, "Unification of neural and wavelet networks and fuzzy systems," *IEEE Trans. on Neural Networks*, vol. 10, no. 4, pp. 801–814, Julho 1999.
50. Q. Liang and J. M. Mendel, "Overcoming time-varying co-channel interference using type-2 Fuzzy adaptive filter," *IEEE Trans. on Circuits and Systems*, vol. 47, no. 12, pp. 1419–1428, Dec. 2000.
51. A. P. Engelbrecht, *Computational intelligence : An introduction*. John Wiley & Sons, 2002.
52. J. C. Patra, R. N. Pal, B. N. Chatterji, and G. Panda, "Nonlinear channel equalization for QAM signal constellation using artificial neural networks," *IEEE Transactions on Systems, Man, and Cybernetics, Part B*, vol. 29, no. 2, pp. 262–271, 1999.
53. B. Mulgrew, "Applying radial basis functions," *IEEE Signal Processing Magazine*, vol. 13, pp. 50–65, Mar. 1996.
54. M. V. Ribeiro, "Learning rate updating methods applied to adaptive fuzzy equalizers for broadband power line communications," *EURASIP Journal on Applied Signal Processing, Special Issue on Nonlinear and Image Processing*, vol. 2004, no. 16, pp. 2592–2599, 2004.
55. L. X. Wang and J. M. Mendel, "Fuzzy adaptive filters, with application to nonlinear channel equalization," *IEEE Trans. on Fuzzy Systems*, vol. 1, pp. 161–170, 1993.
56. S. K. Patra and B. Mulgrew, "Fuzzy implementation of a Bayesian equaliser in the presence of intersymbol and co-channel interference," *IEE Proc.-Commun.*, vol. 145, no. 5, pp. 323–330, 1998.
57. K. Y. Lee, "Complex RLS fuzzy adaptive filter and its application to channel equalisation," *Electronics Letters*, vol. 30, no. 19, pp. 1572–1574, 1994.
58. K. Lee, "Fuzzy adaptive decision feedback equaliser," *Electronics Letters*, vol. 30, no. 10, pp. 749–751, May 1994.
59. P. Sarwal and M. Srinath, "A fuzzy logic system for channel equalization," *IEEE Trans. on Fuzzy Systems*, vol. 3, no. 2, pp. 246–249, May 1995.
60. K. Y. Lee, "Complex fuzzy adaptive filter with LMS algorithm," *IEEE Trans. on Signal Processing*, vol. 44, no. 2, pp. 424–427, Feb. 1996.
61. Q. Liang and J. M. Mendel, "Equalization of nonlinear time-varying channels using type-2 fuzzy adaptive filters," *IEEE Trans. on Fuzzy Systems*, vol. 8, no. 5, pp. 551–563, Oct. 2000.
62. S. K. Patra and B. Mulgrew, "Fuzzy techniques for adaptive nonlinear equalization," *Signal Processing*, vol. 80, no. 6, pp. 985–1000, 2000.
63. C. Dualibe, P. Jespers, and M. Verleysen, "Embedded fuzzy control for automatic channel equalization after digital transmissions," in *IEEE International Symposium on Circuits and Systems*, vol. 2, 2001, pp. 173–176.
64. M. B. Loiola, M. V. Ribeiro, and J. M. T. Romano, "A turbo equalizer using fuzzy filters," in *Proc. IEEE workshop on Machine Learning for Signal Processing*, 2004, pp. 695–704.
65. R. Ferrari, C. M. Panazio, R. R. F. Attux, C. C. Cavalcante, L. N. de Castro, F. J. V. Zuben, and J. M. T. Romano, "Unsupervised channel equalization using fuzzy prediction-error filters," in *IEEE Neural Networks for Signal Processing*, 2003, pp. 1680–1686.
66. S. Haykin, *Adaptive filter theory*. NJ: Englewood Cliffs, Prentice Hall, 1996.
67. S. Qureshi, "Adaptive equalization," *Proc. of the IEEE*, vol. 73, pp. 1349–1387, Sep. 1985.
68. G. D. Forney, "Maximum likelihood sequence estimator of digital sequences in the presence of intersymbol interference," *IEEE Trans. on Information Theory*, vol. IT-18, pp. 363–378, May 1972.
69. G. D. F. Jr., "The Viterbi algorithms," *Proc. of the IEEE*, vol. 61, pp. 268–278, Março 1973.
70. R. O. Duda and P. E. Hart, *Pattern Classification and Scene Analysis*. John Wiley and Sons, 1973.
71. H. L. V. Trees, *Detection, estimation and modulation theory, Parts I and III*. NY: Springer-Verlag, 1968/1971.
72. G. C. Mouzouris and J. M. Mendel, "Nonsingleton fuzzy logic systems: Theory and design," *IEEE Trans. on Fuzzy Systems*, vol. 5, no. 1, pp. 56–71, Feb. 1997.
73. J. M. Mendel, *Uncertain Rule-Based Fuzzy Logic Systems. Introduction and New Directions*. NJ: Prentice Hall, 2001.
74. S. Chen, B. Mulgrew, and S. McLaughlin, "Adaptive bayesian equalizer with decision feedback," *IEEE Trans. on Signal Processing*, vol. 41, no. 9, pp. 2918–2927, Sep. 1993.
75. S. G. Wilson, *Digital Modulation and Coding*. Prentice Hall, 1995.
76. J. G. Proakis, *Digital Communications*, 4th ed. McGraw-Hill, 2000.
77. B. Vucetic and J. Yuan, *Space-Time Coding*, ser. NJ. John Wiley & Sons, 2003.
78. C. B. Schlegel and L. C. Pérez, *Trellis and Turbo Coding*, ser. NJ. IEEE Press, 2004.
79. L. Hanzo, C. H. Wong, and M. S. Yee, *Adaptive Wireless Transceivers: Turbo-Coded, Turbo-Equalized and Space-Time Coded TDMA, CDMA, and OFDM systems*, 1st ed. John Wiley & Sons, 2002.
80. T. K. Moon, *Error Correction Coding Mathematical Methods and Algorithms*. John Wiley & Sons, 2005.
81. F. Abdelkefi, P. Duhamel, and F. Alberge, "Impulsive noise cancellation in multicarrier transmission," *IEEE Trans. on Communications*, vol. 53, no. 1, pp. 94–106, Jan. 2005.
82. R. Hormis, I. Berenguer, and X. Wang, "A simple baseband transmission scheme for power line channels," *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 7, pp. 1351–1363, July 2006.
83. L. Lampe, R. Schober, and S. Yiu, "Distributed space-time coding for multihop transmission in power line communication networks," *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 7, pp. 1389–1400, July 2006.
84. J. Haring and J. H. Vinck, "Iterative decoding of codes over complex numbers for impulsive noise channels," *IEEE Trans. on Information Theory*, vol. 49, no. 5, pp. 1251–1260, May 2003.
85. ——, "Coding and signal space diversity for a class of fading and impulsive noise channels," *IEEE Trans. on Information Theory*, vol. 50, no. 5, pp. 887–895, May 2004.

86. S. Baig and N. D. Gohar, "A discrete multitone transceiver at the heart of the PHY layer of an in-home power line communication local-area network," *IEEE Comm. Magazine*, vol. 41, no. 4, pp. 48–53, Apr. 2003.
87. S. V. Zhidkov, "Impulsive noise suppression in OFDM based communication systems," *IEEE Trans. on Consumer Electronics*, vol. 49, no. 4, pp. 944–948, Nov. 2003.
88. M. V. Ribeiro, R. R. Lopes, C. A. Duque, and J. M. T. Romano, "Impulse noise mitigation for improved bit-rate in PLC-DMT," *IEEE Trans. on Power Delivery*, vol. 21, no. 1, pp. 94–101, Jan. 2005.
89. O. R. H. T. C. Chuah, B. S. Sharif, "Robust adaptive spread-spectrum receiver with neural-net preprocessing in non-Gaussian noise," *IEEE Trans on Neural Networks*, vol. 12, no. 3, pp. 546–558, Mai. 2001.
90. V. Milosevic and B. Ristic, "Effect of impulse noise rejection with median filter on binary digital receiver performance," *Electronics Letters*, vol. 6, no. 12, pp. 392–394, Mar. 1989.
91. T. Kasparis, "Frequency independent sinusoidal suppression using median filters," in *Proc. 1991 IEEE Acoustics, Speech, and Signal Processing International Conference (ICASSP)*, vol. 3, 1991, pp. 1969–1972.
92. X.-P. Zhang, "Thresholding neural network for adaptive noise reduction," *IEEE Trans. on Neural Networks*, vol. 12, no. 3, pp. 567–584, May 2001.
93. T. C. Giles, "On the design of HF radio modems," Ph.D. Thesis, University of Adelaide (Australia), 1995.
94. G. Arslan, B. L. Evans, and S. Kiaei, "Equalization for discrete multitone transceivers to maximize bit rate," *IEEE Trans. on Signal Processing*, vol. 49, no. 12, pp. 3123–3135, Dec. 2001.
95. D. Torrieri, *Principles of Spread-spectrum Communications SystemsDON TORRIERI*. Springer Science, 2005.
96. M. H. L. Chan and D. F. R. W. Donaldson, "Performance enhancement using forward error correction on power line communication channels," *IEEE Trans. on Power Delivery*, vol. 9, no. 2, pp. 645–653, Apr. 1994.
97. T. Faber, T. Scholand, and P. Jung, "Turbo decoding in impulsive noise environments," *IEEE Electronics Letters*, vol. 39, no. 14, pp. 1069–1071, July 2003.
98. H. C. Ferreira, A. J. H. Vinck, T. G. Swart, and I. de Beer, "Permutation trellis codes," *IEEE Trans. on Communications*, vol. 53, no. 11, pp. 1782–1789, Nov. 2005.
99. C. J. Colbourn, T. Kløve, and A. C. H. Ling, "Permutation arrays for powerline communication and mutually orthogonal latin squares," *IEEE Trans. on Information Theory*, vol. 50, no. 6, pp. 1289–1291, June 2004.
100. H. Day and H. V. Poor, "Advanced signal processing for power line communications," *IEEE Communications Magazine*, pp. 100–107, May 2003.
101. J. O. Onunga and R. W. Donaldson, "A simple packet retransmission strategy for throughput and delay enhancement on power line communication channels," *IEEE Trans. on Power Delivery*, vol. 8, no. 3, pp. 818–826, July 1993.
102. T. Wada, "A study on performance of LDPC codes on power line communications ," in *Proc. IEEE International Conference on Communications*, vol. 1, June 2004, pp. 109–113.
103. ——, "A study on the performance of turbo coding for noise environments in power lines," in *Proc. IEEE International Conference on Communications*, vol. 5, May 2003, pp. 3071–3075.
104. Q. H. Spencer, "Short-block LDPC codes for a low-frequency power-line communications system," in *Proc. International Symposium on Power Line Communications and Its Applications*, Apr. 2005, pp. 95–99.
105. J. M. III and G. Q. M. Jr., "Cognitive radio: making software radios more personal," *IEEE Personal Communications*, vol. 6, no. 4, pp. 13–18, Aug. 1999.
106. H. Zhang, X. Zhou, K. Y. Yazdandoost, and I. Chlamtac, "Multiple signal waveforms adaptation in cognitive ultra-wideband radio evolution," *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 4, Part 1, pp. 878–884, Apr. 2006.
107. S. haykin, "Cognitive radio: brain-empowered wireless communications," *IEEE Journal on Selected Areas in Communications*, vol. 23, no. 2, pp. 201–220, Feb. 2005.
108. N. Devroye, P. Mitran, and V. Tarokh, "Achievable rates in cognitive radio channels," *IEEE Trans. on Information Theory*, vol. 52, no. 5, pp. 1813–1827, May 2006.
109. E. A. Lee and D. G. Messerschmitt, *Digital Communication*, 2nd ed. Kluwer Academic Publishers, 1994.
110. L. Hanzo, W. Webb, and T. Keller, *Single- and Multi-Carrier Quadrature Amplitude Modulation: Principles and Applications for Personal Communications, WLANS and Broadcasting*, 1st ed. John Wiley & Sons, 2000.
111. K. S. Zigangirov, *Theory of Code Division Multiple Access Communication*, ser. on Digital & Mobile Communication. IEEE Press and John Wiley & Sons, 2004.
112. K. Fazel and S. Kaiser, *Multi-Carrier and Spread Spectrum Systems*. John Wiley & Sons, 2003.
113. L. Hanzo, M. Münster, B. J. Choi, and T. Keller, *OFDM and MC-CDMA for Broadband Multi-User Communications, WLANs and Broadcasting*, 1st ed. John Wiley & Sons, 2003.
114. A. R. S. Bahai, B. R. Satzberg, and M. Ergen, *Multi-Carrier Digital Communications: Theory and Applications of OFDM*, 2nd ed. Springer-Verlag Telos, 2004.
115. T. Starr, J. M. Cioffi, and P. J. Silverman, *Understanding Digital Subscriber Line Technology*. NJ: Prentice Hall, 1998.
116. E. D. Re, R. Fantacci, S. Morosi, and R. Serravalle, "Comparison of CDMA and OFDM techniques for downstream power-line communications on low voltage grids," *IEEE Trans on Power Delivery*, vol. 18, no. 4, pp. 1104–1109, Oct. 2003.
117. R. Fantacci and S. Morosi, "Multicarrier spread spectrum techniques for downstream power-line communications on low voltage grids," *Int. J. of Commun. Syst.*, no. 16, pp. 401–416, 2003.
118. M. Crussiere and J.-Y. B. J.-F. Helard, "Adaptive spread-spectrum multicarrier multiple-access over wirelines," *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 7, pp. 1377–1388, July 2006.
119. I. H. Cavdar, "Performance analysis of FSK power line communications systems over the time-varying channels: measurements and modeling," *IEEE Trans. on Power Delivery*, vol. 19, no. 1, pp. 111–117, Jan. 2004.

120. M. Gotz, M. Rapp, and K. Dostert, “Power line channel characteristics and their effect on communication system design,” *IEEE Communications Magazine*, vol. 42, no. 4, pp. 78–86, Apr. 2004.
121. Y. H. Ma, P. L. So, and E. Gunawan, “Performance analysis of OFDM systems for broadband power line communications under impulsive noise and multipath effects,” *IEEE Trans. on Power Delivery*, vol. 20, no. 2, Part 1, pp. 674–682, Apr. 2005.
122. K. S. Surendran and H. Leung, “An analog spread-spectrum interface for power-line data communication in home networking,” *IEEE Trans. on Power Delivery*, vol. 20, no. 1, pp. 80–89, Jan. 2005.
123. J. A. C. Bingham, “Multicarrier modulation for data transmission: An idea whose time has come,” *IEEE Comm. Magazine*, vol. 28, no. 5, pp. 5–14, May 1990.
124. Y. Zhang and S. Cheng, “A novel multicarrier signal transmission system over multipath channel of low-voltage power line,” *IEEE Trans. on Power Delivery*, vol. 19, no. 4, pp. 1668–1672, Oct. 2004.
125. H. Meng, Y. L. Guan, and S. Chen, “Modeling and analysis of noise effects on broadband power-line communications,” *IEEE Trans. on Power Delivery*, vol. 20, no. 2, Part 1, pp. 630–637, Apr. 2005.
126. Ascom, “Nothing is as powerful as an idea for which the time is ripe: welcome to the second discovery of electricity,” www.ascom.com, Junho 2001.
127. “Ds2,” www.ds2.es, 2004.
128. “Intellon,” www.intellon.com/.
129. M. K. Lee, R. E. Newman, H. A. Latchman, S. Katar, and L. Yonge, “HomePlug 1.0 powerline communication LAN - protocol description and peformance results,” *Int. J. of Commun. Syst.*, no. 16, pp. 447–473, 2003.
130. Y. Zhang and S. Cheng, “A novel multicarrier signal transmission system over multipath channel of low-voltage power line,” *IEEE Trans. on Power Delivery*, vol. 19, no. 4, pp. 1668–1672, Oct. 2004.
131. M. Kuhn, S. Berger, I. Hammerstrom, and A. Wittneben, “Power line enhanced cooperative wireless communications,” *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 7, pp. 1401–1410, July 2006.
132. K. H. Afkhamie, S. Katar, L. Yonge, and R. Newman, “An overview of the upcoming HomePlug AV standard,” in *Proc. International Symposium on Power Line Communications and Its Applications*, Apr. 2005, pp. 400–404.
133. R. Bernardini, M. Durigon, R. Rinaldo, A. Tonello, and A. Vitali, “Robust transmission of multimedia data over power-lines,” in *Proc. International Symposium on Power Line Communications and Its Applications*, Apr. 2005, pp. 295–299.
134. G. Markarian and H. Xufeng, “Distribution of digital TV signals over home power line networks,” in *Proc. International Symposium on Power Line Communications and Its Applications*, Apr. 2005, pp. 409–413.

# 7

# New Trends in Biometrics

Bernadette Dorizzi[1]

GET/INT, 9 rue Charles Fourier, 91011 Evry, France
Bernadette.Dorizzi@int-evry.fr

## 7.1 Introduction

The proliferation of information access terminals and the growing use of applications (such as e-commerce, e-voting and e-banking) involving the transfer of personal data make it essential to provide reliable systems that are user-friendly and generally acceptable. With conventional identity verification systems for access control, such as passports and identity cards, passwords and secret codes can easily be falsified. Biometrics seem to provide a way of overcoming some of these systems' drawbacks, by basing verification on aspects that are specific to each individual.

For a long time the use of biometrics has remained limited to policing applications, but in view of its potential advantages, this technology is now being considered for many other tasks. Commercial applications have thus been developed, most often based on fingerprints or iris prints; these are currently considered to be the most reliable methods, but have the disadvantage of being intrusive, and are often disliked by users. This partly explains why their use remains limited at present to professional applications (e.g. airport personnel access control), and why, up to now, they have hardly ever been used for the general public, even though commercial products of this type are available.

Clearly, users are more familiar with methods based on face, voice or handwritten signature recognition, but the level of performance of such applications is not yet high enough for their large-scale use to be a realistic proposition. In view of this, combining several methods would seem to be a promising way forward, which remains to be validated. Several American studies forecast a skyrocketing of the biometrics market, mainly as a result of the development of electronic data transfer, particularly on the Internet. At the European level, there are few studies available at present, and one of the EU's current concerns is to rapidly obtain reliable forward-looking studies.

What has changed recently is the ability to digitize, store and retrieve biometric patterns and have them processed by computers. Large scale deployments can thus be envisaged for example border control, voter ID cards, national ID cards, driver's license, welfare disbursement etc. In these types of applications, biometrics must be considered only as one element of a whole system that involves the use of sensors to acquire the biometric sample, the transmission of data from the sensor to a computer where matching will be performed after access to a huge database of stored templates. It means that biometrics should not be evaluated alone but it is this system that must be designed and evaluated in its entirety.

## 7.2 Overview of technical challenges: An already established technology still under development

### 7.2.1 Architecture of a biometric system

Generally speaking, there are two phases in a biometric system (see Fig. 7.1): a learning phase (enrolment) and a recognition phase (verification). In all cases, the item considered (e.g. finger print or voice) is recorded using a sensor and digital data are then available (a table of pixels, a digital signal, etc.). In most cases the data themselves are not used directly; instead the relevant characteristics are first extracted from the data to form a **template**. This has two advantages: the volume of data to be stored is reduced, and greater anonymity is achieved in data storage (because it is not possible to recover the original signal by referring to these characteristics).

The role of the **learning** module is to create a model of a given person by reference to one or more recordings of the item considered. Most of the models used are statistical models, which make it possible to allow for a certain variability in individual data.

The **recognition** module enables a decision to be taken. In identification mode, the system compares the measured signal with the various models contained in the data base and selects the model corresponding most closely to the signal. In verification mode, the system will compare the measured signal with just one of the data base models and then authorise the person or reject him. Identification may be a very difficult task if the data base contains thousands of individuals. Access time problems then become crucial.

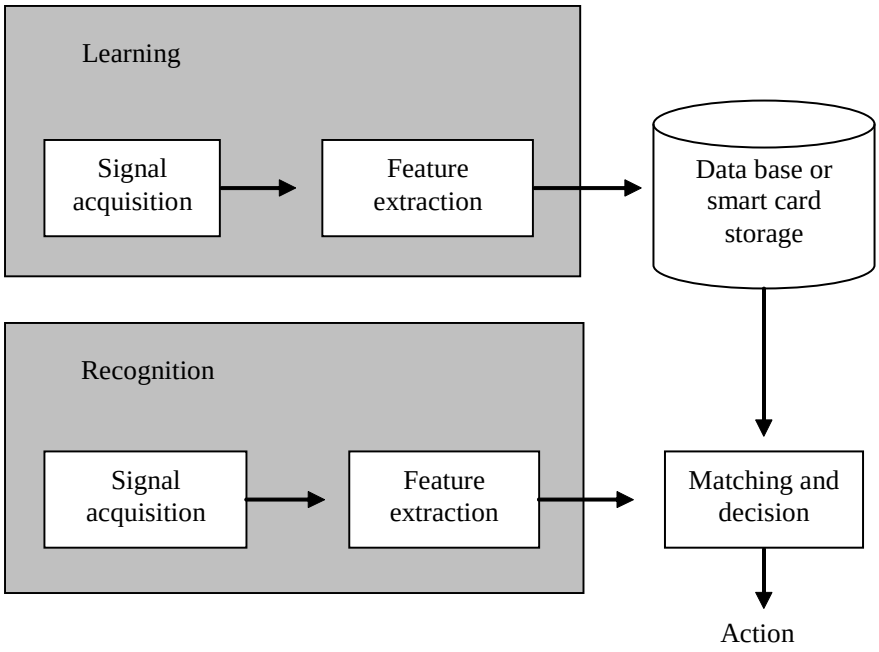



**Fig. 7.1.** The various modules of a biometric system

### 7.2.2 What are the different types of errors a biometric system can make?

Biometric systems are often evaluated solely on the basis of recognition system performance. But it is important to note that other factors are involved in the deployment of a biometric system. One factor is the quality and ruggedness of the sensors used. Clearly the quality of the sensors used will affect the performances of the associated recognition algorithms. What should be evaluated is therefore the sensor/algorithm combination, but this is difficult because often the same sensors are not used in both the enrolment and test phases. In practice therefore the evaluation is made on the basis of the recognition algorithm's resistance to the use of various types of sensor (interoperability problem). Another key factor in determining the acceptability of a biometric solution is the quality of the associated communication interface. In addition to ease of use, acquisition speed and processing speed are key factors, which are in many cases not evaluated in practice.

In the case of a verification system, two error rates are evaluated which vary in opposite directions: the **false rejection rate** (FRR) [rejection of a legitimate user called "the client"] and the **false acceptance rate** (FAR) [acceptance of an impostor]. In Figure 7.2 are drawn the distributions of clients and impostors according to the response of the system which in general is a real number (likelihood) (see [1]). The decision of acceptance or rejection of a person is thus taken by comparing the answer of the system to a threshold (called the decision threshold). The values of FAR and FRR are thus dependent on this threshold which can be chosen so as to reduce the global error of the system.

The decision threshold must be adjusted according to the desired characteristics for the application considered. High security applications require a low FAR which has the effect of increasing the FRR, while Low security applications are less demanding in terms of FAR (see Figure 7.3). EER denotes Equal Error Rate (FAR = FRR). This threshold must be calculated afresh for each application, to adapt it to the specific population concerned. This is done in general using a small database recorded for this purpose.

### 7.2.3 Different problems with different scopes and challenges

We refer to [2] to identify three different ways of using a biometric system:

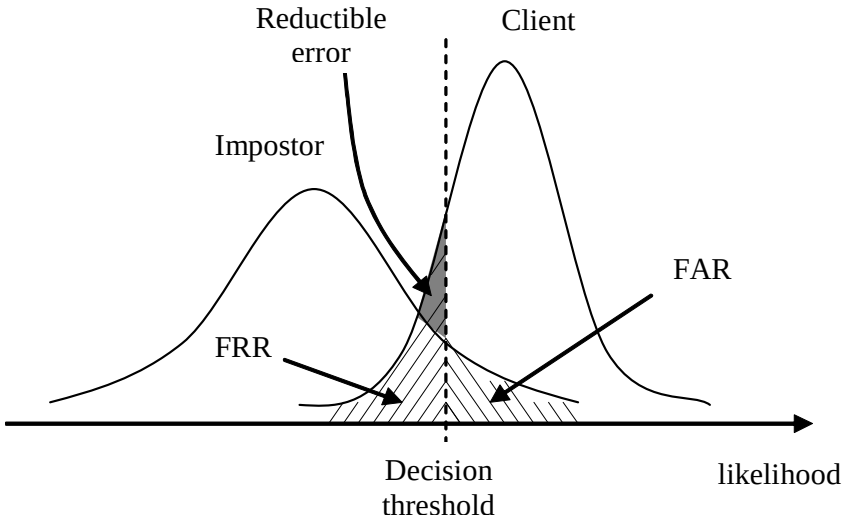



**Fig. 7.2.** False rejection rate and false acceptance rate of a biometric verification system

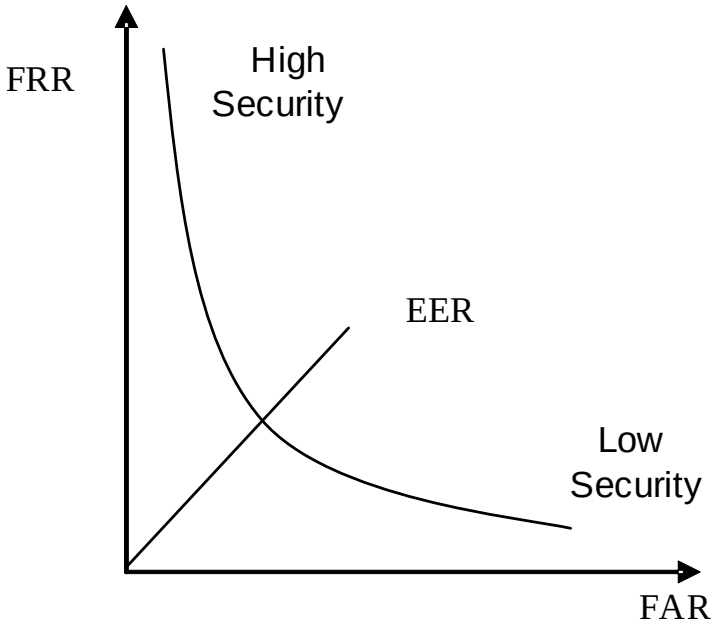



**Fig. 7.3.** ROC (Receiver Operating Characteristic) curve

(i) **Positive Identification ("Is this person truly known to the system?")**. Biometrics can verify with high certainty the authenticity of a claimed enrolment based on the input biometric sample. For example, a person claims that he is known as Mr. X within the authentication system and offers his fingerprint; the system then either accepts or rejects the claim based on a comparison performed between the offered pattern and the enrolled pattern associated with the claimed identity. Commercial applications such as computer network logon, electronic data security, ATMs, credit card purchases, physical access control, cellular phones, PDAs, medical records management, and distance learning are sample authentication applications. Authentication applications are typically cost sensitive with a strong incentive to be user friendly.

(ii) **Large Scale Identification ("Is this person in the database?")**. Given an input biometric sample, a large-scale identification determines if the pattern is associated with any of a large number (e.g., millions) of enrolled identities. Typical large-scale identification applications include welfare-disbursement, national ID cards, border control, voter ID cards, driver's license, criminal investigation, corpse identification, parenthood determination, missing children identification, etc. These large-scale identification applications require a large sustainable throughput

with as little human supervision as possible. In this case, the data base must include data that can characterize each of the people in the base, and the system must then search for the person that best corresponds to what it observes.

(iii) **Screening ("Is this a wanted person?")**. Screening applications covertly and unobtrusively determine whether a person belongs to a watch-list of identities. Examples of screening applications could include airport security, security at public events, and other surveillance applications. The screening watch list consists of a moderate (e.g., a few hundred) number of identities. By their very nature, these screening applications do not have a well-defined "user" enrolment phase, can expect only minimal control over their subjects and imaging conditions and require large sustainable throughput with as little human supervision as possible.

Neither large scale identification nor screening can be accomplished without biometrics (e.g., by using token-based or knowledge-based identification).

| Service | Device | Storage | Database size | Accuracy |
|---|---|---|---|---|
| 1:N Large Scale Identification | PC or smart card | Local or central database | Millions of people | Low FAR |
| 1:1 Positive Identification/Verification | PC or hard disks | Smart card or local database | No database needed | Low FRR |
| Screening | PC or hard disk | Local or central databases | Few hundred of people | Low FAR |

**Table 7.1.** Summary of some characteristics associated to the above mentioned 3 types of applications.

Table 7.1 summarizes different characteristics associated to the three situations depicted above. Note in particular that Large Scale Identification involves the storage of the data on a central database, contrary to verification applications with which the information concerning the person can be recorded, for example on a smart card held by the user, which ensures a higher degree of confidentiality, but offer the disadvantage of potential theft or loss.

## 7.3 Description of different biometric modalities

Even if some modalities like iris or fingerprint can be considered as "sufficiently efficient", it is interesting to also envisage other inputs as the choice of one modality is linked to acceptability or usage purposes. In this report, we describe in more details 4 modalities: iris, fingerprint, DNA and face, even if there are other possible choices such as voice, signature or gaiting.

### 7.3.1 Iris recognition

**How does iris recognition work?**

The iris (see Fig. 7.4) is an overt body that is available for remote (non invasive) assessment. Unlike other modalities, face for example, the variability in appearance of any iris might be well enough constrained to make an automated system possible based on currently available machine vision technologies [3].

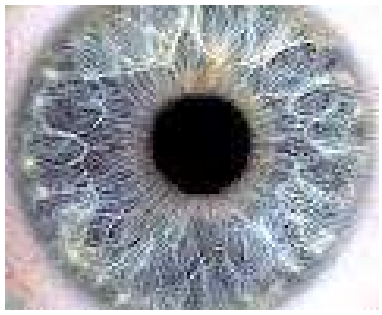


**Fig. 7.4.** Iris image.

J.Daugman is a pioneer in the iris recognition area. He published his first results in 1993 [4], relying on the use of Gabor wavelets in order to process the image at several resolution levels. An iris code composed of binary vectors is

computed in this way and a statistical matcher (XOR, logical exclusive OR operator) analyses basically the average Hamming distance between two codes (bit to bit test agreement). This works has been patented by a US Patent (No. 4,641,349 entitled “Iris Recognition System”) held by Iridian Technologies, Inc (e.g. Sansar and IrisScan).

Another approach, in the framework of iris verification, introduced by Wildes [3], consists of measuring the correlation between two images using different small windows of several levels of resolution and Linear discrimination analysis to make the decision. Also, other methods for iris verification have been proposed, in particular relying on ICA : Independent Component Analysis (a lot of research is now conducted in Asia on this modality).

**Summary**

The iris code obtained in the corresponding encoding process is the most precise print of all existing biometric techniques, at the expense of rather constrained acquisition conditions (the camera must be infra-red, the eyes must be at a very precise distance from the camera). These elements provide a very good quality of the initial image which is necessary to ensure such a high level of performance. On the other hand they may generate a long time during the enrolment phase and the necessity of personal assistance [2]. This method also requires a relatively expensive acquisition system and necessarily involves the scanning of the eye, which can initially prove offputting to users. The resulting reliability means it can be successfully used both for identification and authentication, an advantage which few other techniques can offer.

### 7.3.2 Fingerprint recognition

**State of the art**

Most fingerprint processing approches use specific features called minutiae as a description of the fingerprint. These minutiae are composed of big details like starting lines, splitting lines and line fragments and smaller ones like ridges ending, incipient ridges, bifurcation, pores, delta and line-shapes (see Fig. 7.5).

Automatic minutiae detection is an extremely critical process, especially in low-quality fingerprint images, where noise and contrast deficiency can originate pixel configurations similar to minutiae or hide real minutiae.

Several approaches to automatic minutiae extraction have been proposed [5]: Although rather different from each other, most of these methods transform fingerprint images into binary images. The images obtained are submitted to a postprocessing which allows the detection of ridge-lines. The different fingerprint authentication systems can be classified by their level of accuracy. The greater the accuracy needed, the more complex and naturally the more expensive the system is. The classification of a system is based on the number of features (minutiae) that it can deal with. The high-tech systems are able to exploit up to 80 points and also to distinguish between a real fingerprint and a forged one (synthetic fingerprint). The most widely used in general use employs some 20 particular points. Since the fingerprint is never captured in the same position, the verification algorithm must perform rotation and translation of the captured fingerprint in order to adjust the fingerprint minutiae with the template minutiae.

The final stage in the matching procedure is to compare the sampled template with a set of enrolled templates (identification), or a single enrolled template (authentication). It is highly improbable that the sample is bit-wise identical to the template. This is due to approximations in the scanning procedure, misalignment of the images and errors or approximations introduced in the process of extracting the minutiae. Accordingly, a matching algorithm is required to test various orientations of the image and the degree of correspondence of the minutiae, and it assigns a numerical score to the match. Different methods exist for processing fingerprints:

- The direct optical correlation is practically not used, because it is not very efficient for large databases.
- The general shape of the fingerprint is generally used to pre-process the images, and reduce the search in large databases. This uses the general directions of the lines of the fingerprint, and the presence of the core and the delta. Several categories have been defined in the Henry system: whorl, right loop, left loop, arch, and tented arch.
- Most methods use minutiae, the specific points like ridge endings, bifurcations, etc. Only the position and direction of these features are stored in the signature for further comparison. - Some methods count the number of ridges between particular points, generally the minutiae, instead of the distances computed from the position.
- Other Pattern matching algorithms use the general shape of the ridges. The fingerprint is divided into small sectors, and the ridge direction, phase and pitch are extracted and stored.
- Very often, algorithms use a combination of all these techniques.

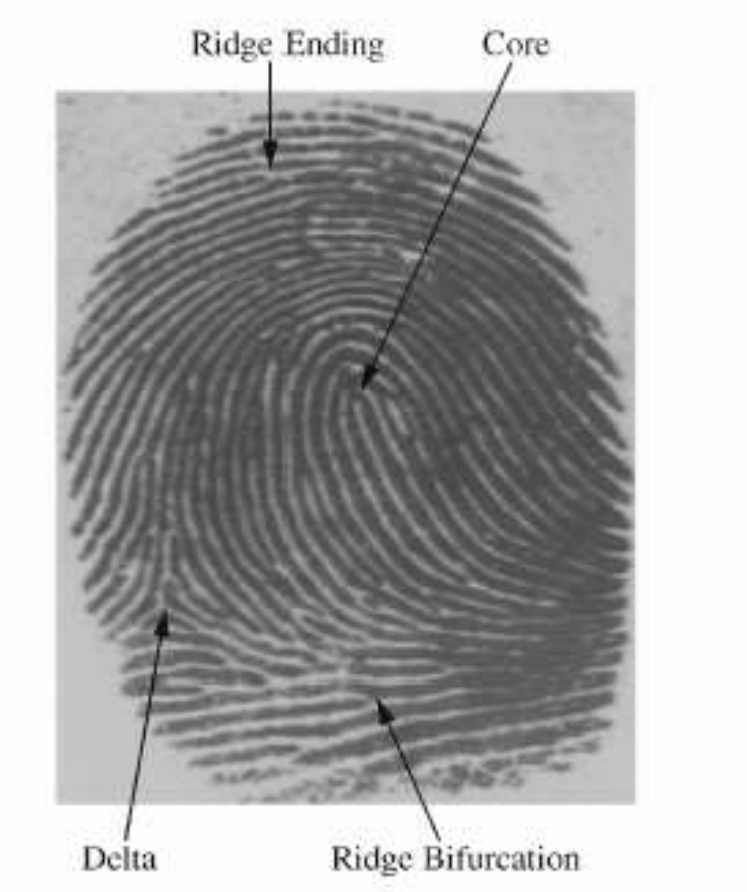



**Fig. 7.5.** Minutiae of a fingerprint.

**Summary**

Fingerprint is, up to now, the modality that allows the best compromise between price, acceptability and accuracy [5] and a lot of systems based on this modality are already operational. However, the latest evaluation results [6] show that the performance of such a system deeply relies on the quality of the acquired images, in particular during the enrolment phase. Moreover it seems that a not so negligible part of the population cannot be enrolled through fingerprints (manual workers, persons with too wet or too dry hands etc.); the percentage is estimated at up to 1 or 2 % but it seems that this number can be decreased with the use of two or more fingers and adequate specific enrolment processes for persons who present problems. Another point is the existence of a great number of different sensors associated with various technologies which make the interoperability problem more difficult to solve because it is the coupling of sensor and algorithms that is optimized by the designer of the biometric system and dissociating them may lead to a noticable decrease in performance. Fingerprint is in general rather well accepted even if it has some forensic connotations and it allows both identification and verification.

### 7.3.3 Face Recognition

**State of the art**

In more than twenty years of research, several methods have been tested with the aim of recognizing people from the image of their face. Some of them are dealing with local features [7] like eyes, noise and mouth, while others consider the appearance of a face (Eigen face [8] and Fisher face [9] methods) but no method up to now performs sufficiently well, which means that a lot of research is still going on.

By 1993 several algorithms were claimed to have accurate performance in minimally constrained environments. To better understand the potential of these algorithms, DARPA and the Army Research Laboratory established the FERET program with the goals of both evaluating their performance and encouraging advances in the technology [10]. The FERET database testing employs faces with variable positions, scales, and lighting in a manner consistent

with mug hots or driver's license photography. On databases of fewer than 200 people and images taken under similar conditions, all algorithms produce nearly perfect performance. Interestingly, even simple correlation matching can sometimes achieve similar accuracy for databases of only 200 people [11]. This is strong evidence that any new algorithm should be tested with databases of at least 200 individuals, and should achieve a performance over 95% on mug shot-like images before it can be considered potentially competitive.

The latest face verification competition FAT 2004, held in conjunction with the 17th International Conference on Pattern Recognition has shown that the best performance is obtained with a method based on independent feature analysis rather than those working with either Eigen or Fisher face methods. It also showed that by degrading the recording conditions (leading to images of lower quality), the results dropped from 1.39 to more than 13% of EER, for a database of only 52 persons. Aging also significantly degraded the performances [12].

**Summary**

Face is considered at this moment as a relatively non accurate modality due to the presence of a lot of variability (from 1.39% to more than 13% EER). Some are due to the different changes that can occur to the person over time, like aging, wearing bears or not, glasses, hair etc. while others are related to environmental conditions (illumination, textured background, poses, facial expressions). Therefore the performance highly varies depending on the recording conditions and the context of application (static images or video, with uniform background or not, with constant lighting conditions or not).

Face recognition is not efficient enough at this moment to deal with Large Scale Identification but it can be useful in the context of verification or limited access control with constrained acquisition conditions. It means that, during enrolment, the person must face the camera at a fixed distance and that the background is uniform. This will ease the verification process while remaining acceptable for the user. In the video context, no system can be considered as sufficiently developed [10, 11] but there are promising new efforts using 3-D modeling in order to cope with the problem of pose [13, 14]. Of course this may mean the use of sophisticated 3-D scanners in place of standard medium-cost cameras, therefore increasing the cost of the global system which otherwise remain practicable.

However, due to the fact that this modality is well accepted and natural for the users, and that it has been introduced as a standard in travel documents by the ICAO, a lot of research is being conducted to improve the accuracy of the systems. A big increase in performance can be expected in the next 5 years but this modality can never be expected to be as accurate as fingerprint or iris due to its intrinsic variability and behavioral character.

Nevertheless for comfort applications (like access control to car, home or personalization of environment) which imposes limited FAR constraints, using the face is still very interesting as it can be transparent but in this case an association with other modalities has to be considered in order to reduce the error rates or to do a preselection of the database.

### 7.3.4 DNA in Forensic Authentication

**State of the Art**

**DNA Sequencing:** DNA sequencing consists in the ordering the bases (A, T, G or C) of the DNA or of a fragment of the DNA. This procedure is quite error-prone, depending on the quality of data.
In 1970, the dot matrix or diagram method was proposed by A.J.Gibbs and G.A. McIntyre to analyze the similarity of the nucleotide or protein sequences (but not the whole DNA sequence)[15].
At the same time, Needleman and Wunsch used a dynamic programming algorithm to compute the similarity between 2 DNA fragments. The disadvantage of this method is that it is time consuming, therefore it is impossible to compare two sequences of 300 proteins (1088 comparisons, Waterman 1989) [15].
For this reason, the comparison of DNA segments is not used for the forensic applications, but DNA sequencing is useful to store the DNA in a computer for further research.

**DNA fingerprinting:** the main type of forensic DNA testing consists of DNA fingerprint: DNA fingerprinting is based on Restriction Fragment Length Polymorphism (RFLP) or Polymerase Chain Reaction (PCR).
Obtaining DNA fingerprinting is the result of a complicated laboratory procedure [16], which consists of taking DNA from a cell, cutting it into many DNA segments by using appropriate enzymes, separating DNA segments on a gel by using electrophoresis, attaching a colored probe (a small piece of DNA) to the gel and a pattern is produced. This procedure makes a single probe's DNA fingerprint.
The final DNA fingerprint is built by using several probes (5-10 or more) simultaneously. Figure 7.6 shows the first DNA fingerprint constructed by Professor Jeffreys using 11 probes.

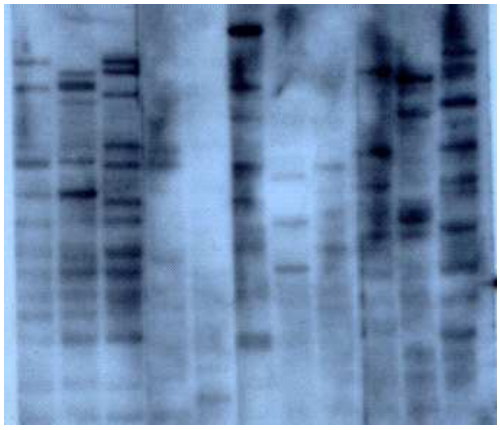


**Fig. 7.6.** The first DNA fingerprint, produced by P. Jeffreys, September 1984.

As a conservative estimate in [17], when using one probe to make the DNA fingerprint, the chance that two people have the same fingerprint is 25% (in reality, this probability is less than 0.05). But, if we use 14 probes, the chance that two people have the same final DNA fingerprint is about (0.25)14 or 1 per 268 million.
The set of probes, which is thus the DNA fingerprint, is visualised as a 1 D "bar code" and the DNA fingerprint matching is just a numerical comparison which allows large-scale identification.
While effective at producing highly discriminatory patterns, this technology is slow, cumbersome and manual - it couldn't be automated.

**Summary**

Except for identical twins, each person's DNA is unique. It can thus be considered as a "perfect" modality for identity verification. However, the human DNA length is too big to allow the examination of the whole DNA sequence; in general identification techniques look at some specific areas of DNA, which are known to vary widely between people. The accuracy of this technique is thus very high allowing thus both identification and verification. Enrolment can be done from any cell that contains a structure called the nucleus. This includes blood, semen, saliva or hair samples. Acquiring these data may thus be felt as intrusive by people. At this moment, DNA analysis is performed in specialized laboratories and is cost and time-consuming (roughly 4 or 5 hours for the whole procedure). Moreover there is a complete lack of standardization which means that we are very far from being able to produce interoperable systems.

DNA usage is mainly forensic. Anyhow, future introduction, if possible, of a rapid, cheap DNA biometric authentication technique will face acceptability problems as, unlike iris or fingerprint which just correspond to performing some internal measure of the man, DNA is something that's intrinsic to, and very detailed about, the person. Anyhow, when using DNA for authentication, only a very small part of the genetic information is coded and this information is unsufficient to go back to the initial genetic heritage of the person.

So, it seems that it will be a long time before DNA printing can become a real time routine for biometric authentication. However, a Canadian Laboratory recently announced a proprietary DNA extraction process which would need only simple equipment and need a time of only 15 minutes.

According to James F. Crow [18], the technical improvements in the future will be of two types : more automation and more accuracy in the existing processes. The author even foresees that DNA analysis could be made in real time. The other improvement concerns the building of new systems requiring very small amounts of material to provide an identification. Future DNA forensics are likely to involve plants and animals, they will not be confined to human DNA as each of us carries a unique constellation of viruses, bacteria and other parasites.

## 7.4 Evaluation of biometric systems

First, one can try to compare the error rates of different systems in each modality, using estimations of FAR (False Acceptance Rate) and FRR (False Rejection Rate), but these results are only indicative as only some systems in a restricted number of environments per application have been tested. In fact, the performance of the systems is highly dependent on the test conditions and very often, the systems are evaluated in laboratory conditions with a small database and relatively good quality data. Moreover fair evaluation should include forgeries (natural or simulated) in

the database and this is very rarely the case. Fingerprints and face are the subject of yearly independent international evaluation [6, 10, 19] which now aims at testing more operational situations while no open evaluation on iris is being conducted.

One must also notice that new systems show up continuously and that the performance is thus going to improve continuously. This is in particular true for face authentication, the performance of which is still unsufficient for a real applications. When trying to compare the results of different systems one has to note that the test results do not use similar test methodology or datasets of similar scale. It would be fairer to provide the global ROC curves, instead of FMR or FNMR.The technologies may not be directly comparable in the extent of specific applications.

As an illustration, we report in Table 7.2 what we consider as the most accurate table available at this moment in the literature [2]. FNMR denotes rejection rates also quoted as FRR in the literature. FMR are also called FAR (False Acceptance Rates) in the literature. n/a denotes data non-availability.

| **Biometric** | **Face** | **Finger** | **Iris** |
|---|---|---|---|
| FTE % failure to enroll | n/a | 4 | 7 |
| FNMR % rejection rates | 4 | 2.5 | 6 |
| FMR1 % verification match error rate | 10 | $< 0.01$ | $< 0.001$ |
| FMR2 % large-scale identification for database sizes of 1 million | 40 | 0.1 | n/a |
| FMR3 % screening match error rates for database sizes of 500 | 12 | $< 1$ | n/a |

**Table 7.2.** Typical biometric accuracy performance numbers reported in large third party tests.

The face recognition results are based on FRVT 2002 [20] and its extrapolation using Eyematic data. The fingerprint authentication errors are from [21] and assume use of right index fingers with no failures-to-enroll.

Both fingerprint screening and identification assume the use of 2 fingers. Fingerprint identification performance reflects the state of the art AFIS (Automatic Fingerprint Identification System) performance based on 2 fingers against a 6 million person database with no failures-to-enroll [21]. Iris FTE is from [22] and fingerprint FTE is from [23]. Iris error rates are from ([24, p.121]). These numbers are based on what the authors [2] believe to be order of magnitude estimates of the performance of the state of the art systems.

Concerning iris, Table 2 gives no results on Large Scale Identification and Screening because there is unfortunately, at the date of writing, no public data base which enables the free evaluation of the algorithms that are commercially available. The only independent known evaluation comes from [25] which indicates an FMR of 0% with a database of roughly 2,000 people. Only Iridian [26] computed Large Scale Identification with 0% FMR and a Database of roughly 10,000 persons.

More generally, in the evaluation of operational biometric systems, other criteria than performance have to be taken into account, such as robustness, acceptability, facility of data acquisition, ergonomy of the interface, enrolment and identification time. For example one has to take into account while choosing a practical fingerprint system, the robustness of the sensor to impacts, wrong or clumsy manipulation, dirtiness [5]. Likewise, the high constraints imposed for the acquisition of irises may significantly increase the global enrolment or verification time capable of producing compression problems in some applications. Moreover, wearing contact lenses or glasses may produce errors.

Note, that for any modality a relatively large part of the population is unable to enroll and this has to be taken into account when facing a specific application. Alternative processes have always to be envisaged.

### 7.4.1 Resistance of the system to forgeries

A major characteristic of biometric recognition methods, against knowledge- and possession based methods is that it is more difficult to reproduce or steal our biometric data than to try to guess a password or to steal an object. Nevertheless, fraudulently reproducing biometric data is possible, but is very dependent on the modality, application and resources being considered and availability of the data to be reproduced.

Different points should be considered in order to answer the question: Is it possible to fool a biometric system? : the technical possibility to reproduce them artificially; the availability of the biometric data (with or without the cooperation of the person); the possibility to design biometric sensors that are resistant to this kind of imposture.

Let us consider the case of fake fingerprints: it is possible to reproduce artificially a fake fingerprint. Producing a gummy clone of an available real finger is technically possible. [27] and [28] have shown that it is possible to fool commercially available fingerprint scanners. While reproducing the ridge and valley structure is relatively easy, simulating all the physiological properties of the finger becomes more complicated, without being impossible. Considering the finger vitality detection problem, fingerprint sensors could be designed to detect physiological characteristics, such as the pulse, the blood pressure, the sweating process. But, based on the knowledge of which of these characteristics is checked in the sensor it is always possible to create a fake finger with these characteristics. [28] argue that many scanners can be fooled by heating up, flashing, humidifying. But as pointed out in [5] a fake finger that will fool all the vitality detections barriers implemented in a fingerprint sensor, can still be built given enough resources.

As far as the availability of the biometric data, it is not easy to have a good three dimensional image of the finger, while it is possible to use latent fingerprints left on different surfaces and objects by the person. However, reconstruction of a fake finger built from latent fingerprints remains quite complicated, and is less reliable. Concerning iris, [29], it is also possible to introduce liveness tests into the cameras, but J. Daugman himself quotes that "Some low-cost, consumer-oriented, handheld iris cameras do not incorporate any effective liveness tests, as Professor Matsumoto demonstrated". In fact, living tissue differs in many respects from artificial objects made of plastic, or printed images on paper. In the case of the eye, there are aspects of physiology, and of the optics of a working eye, that can be exploited. There are also behavioural tests of liveness. Some of these tests rely only on software, but some require special hardware as well, to detect by means of physics that this is living tissue and not fake.

It is important to notice that, in general, the performance of the systems in verification mode (in terms of FAR) are not published taking into account deliberate forgeries. The impostors are simply the other peole in the database but not those who deliberately try to imitate the real trait of the person (except for signature verification where the database contains imitations). So there is almost no way to predict the behavior of a biometric system in the presence of deliberate impostors.

### 7.4.2 Multimodality

The use of several modalities can be considered in order to:

- **Improve the efficiency of the global system**
  A single modality biometric system can be subject to a certain number of defaults leading to an expected or unexpected high level of errors.
  Some errors can be due to *some noise* associated with the sensed data. It may be introduced in such data in many different ways: by sensors (for example, a dirty fingerprint), by ambient conditions (for instance, poor illumination for capturing someone's face), or by the user (the voice of someone having a cold is a noisy input to the system). As a consequence, when comparing the sensed data to a client's references (the stored templates of such a client), the biometric input may be incorrectly matched, and the user may be falsely rejected.
  A high level of errors can also be related to *Intra-class variability*. Biometric data may be variable from one acquisition to another (depending for instance on the emotional state of the person). This intra-class variability may be stronger for some individuals, especially when talking about behavioral biometrics (like signature or voice or gaiting). Therefore, the biometric data acquired during authentication may be very different from the data acquired during enrolment. This affects, of course, the matching process and may lead the system to failure.
  A biometric trait is expected to be *differential across clients*, i.e. it has to vary significantly from one person to another. Some modalities do indeed permit the identification of a person (fingerprints, iris), because of their high discrimination capability while others are less effective .
  *Impostor attacks /liveness.* A biometric system may be attacked with forged data, or genuine data of a dead person may be presented to the sensor. Generally, behavioral biometrics (signature, voice) is more susceptible to attacks since it is easier for forgers, but physical biometrics is also the object of such attacks (there is extensive literature on how to make fake fingers) [28].
  Using several **different modalities together** should help to deal with the points mentioned above, mostly when using complementary biometrics such as behavioral and physical, discriminative or not etc. [30]. Indeed, multi-modality has a clear impact on performance: research works have shown that multimodal systems enhance authentication systems'performance significantly, relative to unimodal systems. Such systems have by construction a higher discrimination capability and are more difficult to attack by impostors. Indeed combining fingerprint with hand shape, or face recognition may circumvent the usage of fake fingerprints, as faces and hands are more difficult to imitate than fingers. This is also the case for voice and lip movements which are naturally correlated.

- **Provide an alternative**
  Considering two (or more) modalities does not mean using them at the same time. Indeed if we build a biometric system relying on both fingerprint and face and if a person cannot enroll its fingerprint, because of the bad quality of his finger, then it will be possible to use only his face image for verification. Non availability of a biometric trait can also be temporary. Imagine a system functioning with iris and fingerprints. If one person during a short period has a problem with his eye, so that it is impossible to perform the iris scan, the fingerprint system can be used instead. The same thing occurs with people which would refuse to use a specific modality (for religious or health purposes for instance). So the multimodality of the system allows a flexibility by providing an alternative to the identification process.
  When the modalities are considered at the same time, the fusion of the information provided by the different modalities can be done at several different levels [31]: At the decision level : in this case each system takes a decision and only at the end, the two (eventually contradictory decisions) are combined in general by a simple AND or OR operation, leading to a lot of errors or rejection. In this framework, the performance of a "bad" system can degrade those of the bi-modal system. More interesting is the case of fusion of scores. In this case, the combination concerns the scores produced by the system before producing its final decision. In this framework, the performance is always increased provided that the fusion scheme is adequately chosen [32, 33, 34]. In particular this scheme allows the reduction of the importance of a less accurate modality to the benefit of a more accurate one (case of face and fingerprint for example). In certain cases, the two modalities that are combined may be very correlated. This is the case for example, of lip movement and voice recorded together when a person is speaking. In these cases, it is possible and interesting to fuse the information at an even earlier stage, namely just after feature extraction and to build a unique system taking as input a combination of these features [35].

### 7.4.3 Application Issues

One has to distinguish between "Mass Identification" applications (border control, National ID cards, Visas etc.) which demand a great level of security (very low FAR) and domestic, personal applications (ambient intelligence, personal accesses to computers) for which the constraints are low FRR and friendly interfaces.

Mass identification involves:

- Storage of the data on a central Database
- High accuracy level
- User constraints for high quality enrolment

The size of the population may be a problem, when considering access times to database, fluidity of the entire process.

Interoperability is another issue : if we want a border control system to be used in several Schengen area entry points, either we have to use the same system all over Europe, or we need the different systems to be interoperable (which means the ability of software and hardware on multiple machines from multiple vendors to communicate). Interoperability between different systems is achieved by using common standards and specifications. At this moment, the standardization of the data formats (for iris, face, fingerprint) is in rapid progress in the ISO- SC37 commission.

As far as only raw data are considered, these formats will indeed assure interoperability but some propositions are also made concerning the standardization of templates (like minutiaes for fingerprints or face images with explicit position of the eyes). Storing templates instead of raw data presents the advantage of compressing the information and insuring more security in case of attack. However, it will not allow complete interoperability. Moreover, as soon as functional interoperability is wanted (interchange of algorithms from different vendors), there is a need for some software interchange formats. The BioAPI specification language has been introduced for this purpose, but its use burdens the whole identification system. It seems that such constraints are essentially suitable for verification systems (1:1) while in the context of Large Scale Identification systems, they increase the processing time which can be prejudicial for an operational system. Very few tests have been conducted so far concerning real interoperability issues which thus remain a fundamental question. Let us just quote here from [36] that with the cooperation of organizations representing seafarers and shipowners, the International Labour Office ILO has just completed a six-week test involving 126 volunteer seafarers on the M.V. Crystal Harmony, a vessel operated by Crystal Cruises. The seafarers included men and women from 30 countries and covered a wide distribution of ages and a diverse set of seafaring job categories.

The testing exercise involved seven biometric products submitted by various manufacturers. The ILO has found that two of them met the requirement of global interoperability.

In the second type of applications, the focus is on transparency and comfort for the user. So non-intrusive biometrics may be used such as video recording. In this case one can recover from a sequence of images, different types of correlated information such as gaiting [37], voice in correlation with the face images. As none of these modalities is

effective enough by itself they cannot be used alone. However, the redundancy that the joint use of all this information will provide, will be an important tool to insure a final good reliability in the identification of the people.

Biometrics can therefore be seen as a way to simplify everyday life, reducing the number of keys, passwords and pin codes that we have to memorize.

### 7.4.4 Biometrics as a way to increase privacy and anonymity and security

Biometrics presents users with the ability to protect and secure their privacy despite the ubiquitous nature of the Internet and almost all forms of commerce, but due to the fact that in general, biometric data are stored in Large Data Bases, one can also consider that privacy and security issues are not satisfied by the actual implementations of biometrics. One interesting issue to this problem is to consider Biometric Encryption which is defined in [38] by Dr. George Tomko as the use of the unique pattern in a person's fingerprint as one's private encryption or coding key. Fingerprint is only an example but iris or other stable biometrics can be envisaged. This way, the fingerprint of one person can be used to code the PIN allowing access to a bank machine. The coded PIN has no connection whatsoever with the finger pattern. What is stored in the database is only the coded PIN. The finger pattern only acts as the coding key of that PIN, any PIN. The fingerprint pattern, encrypted or otherwise, is not stored anywhere during the process. It stays on the finger, where it belongs. But there is another benefit to all of this, and that is, true privacy: the operation of successfully decoding the PIN confirms my eligibility for a service without having to reveal any personal identifiers. Since I'm the only one that can decode the PIN, it is not based on trust. It is "absolute" privacy.

There is also an indirect benefit to privacy. We can continue to have a multitude of PINs and passwords, thereby achieving "safety through numbers" versus one single identification with which to link everything.

The development of biometric encryption techniques is not an easy task because the image of the finger pattern itself is "plastic" and does not remain stable. Each time that you place the fingerprint on a finger scanner, from a holistic view, it may appear to be the same, but there are actually small differences in the pattern due to dryness, moisture and elasticity conditions of the skin. In addition, cuts and scratches can change the pattern. It is somewhat like "fuzzy" decryption. That is the challenge that developers of this new technology face. Furthermore, the system must be built so that these PINs or keys can only be decoded with a "live" finger pattern. It must be immune from a latent print lifted off an object, or a 3-D silicone replica of someone's finger pattern. Some [39] solutions have been already proposed and some patents [40] applied for, but this still remains a research topic as the fact that biometric patterns are never exactly the same while doing different acquisition, renders the production of a private key, that needs to be similar at each stage, very difficult.

### 7.4.5 Biometric data storage

As soon as only verification is requested, the biometric data can be stored on smart cards kept by the user, which provides him with the insurance that the data cannot be used without his own authorization, contrary to what happens with a centralized database. Biometric verification/identification can also be realized through remote access, in this case there is a need for a transmission of the biometric image or template through a network. This means having a highly secure connection. Watermarking may also be used in this case to insure that the transmitted data have not been corrupted.

Of course smart cards can be lost or stolen. For this reason, the data that it contains must be encrypted but if the information has been retrieved by a robber, it is necessary to be able to revoke it and to produce another template which could be used for further identification. Revocation is easy when dealing with pin codes or passwords but not with biometric traits as we cannot change our iris or our fingerprint. Cancellable biometrics is a new research field and some preliminary propositions have been made.

One method of cancellable face biometrics is described in [41]. Considering a face image of a person, it is possible to generate new images obtained after a filtering of the original image. The coefficients of the filter are randomly generated thanks to a PIN code. Changing the PIN code means changing the filter and therefore changing the new face image generated. It has been demonstrated that for face recognition and if the matching algorithm relies on correlations this process does not affect the result of recognition. More research is needed to confirm these results on other face recognition methods.

The use of such filtering is not straightforward for fingerprint or iris recognition, because it affects the quality of the images and the accuracy of the minutiae detection (fingerprint) or texture analysis (iris). For iris, one solution is to extract from the 2048 bit length code a smaller length and to use only this information in the matching process.

Much work has been made in the framework of encryption and watermarking in which the concern is how we can protect the stored data. Even if the aims are different, cancellable biometrics can adapt some of watermarking or

encryption method. The two most known techniques in watermarking or encryption are those developed by Rathan Connell and Bolle [42] and Jain et al [43]. The first one, which is more interesting for cancellable biometrics, relies on the use of special deformations on an image (face, fingerprint) like grid morphing or random transformation like block permutation. This watermarking technique can be easily adapted and turned into a cancellable biometric technique by introducing a PIN code to generate the grid morphing coefficients or the block permutation parameters.

The second watermarking technique consists of embedding face information into the fingerprint one. In the recognition process, the system reconstructs the face and the fingerprint image from the embedded fingerprint image.

It seems, at this stage of research, that cancellable biometrics can't work without introducing a PIN code. We can use the right iris instead of the left iris, or one of our ten fingerprints in order to introduce some cancellation in our biometrics data, provided that enrollment has been realized with a large set of data, but it is not a completely satisfactory solution.

From my point of view, there are still a lot of unsolved problems in the first type of applications (interoperability, storage, accuracy). At this moment, different experiments are being undertaken by several entities (Africa, Asia, US, Europe, EU member-states, etc..) without any coordination. Let's quote for example, the Nigerian National ID card where fingerprints have been stored on 2D bar codes. 50 millions people were enrolled in 4 months, two years ago. A project of a National ID card is now in progress in the Arab Emirates. The fingerprints will be stored on a smart card and the check will be on the card itself (match on card).

Malaysia is developing a multi-usage biometric smart card while Large Scale Experiments of a system relying on iris identification for border control are in progress. It will concern 2 million of air passengers. No smart card is envisaged, the checking will be through identification in a database (1:n). Australia is developing an access checking application for the flying staff relying on face verification from the chip included in the passport (which means, with the ICAO standard, the storage of the face on a contactless chip inserted in the passport). Interoperability is requested [ref SAGEM, private communication]. France will introduce the principle of INES (Identité Nationale Electronique Sécurisée) National Electronic Secured Identity, which will be certified by the government and necessary to obtain either the CNIe (National electronic ID card) and/or the passport. This "identity" will associate alphanumeric data from the civilian state and physical data (photo, fingerprint, signature).

In the second type of application, some products are already available (see for example [44] for a list of products and companies) but there is a lack of certification procedures for the market to develop widely. Use of videos (face, gaiting) in the home environment is still an open issue and a potentially good topic for research which should lead to interesting results in the near future. Note that the corresponding results will be also useful in a wider field than purely biometrics, namely videosurveillance and tracking.

## 7.5 Conclusion

The introduction of biometric elements in a global identity authentication system is often seen as a way to increase security. There are still challenges considering biometic systems for large scale identity verification. Large scale experiments have been conducted in the past in non-european countries (Africa, Asia) and some are now being conducted in Europe but it seems that this is done without any coordination and that the results are not widely spread and directly reusable.

There is still a lack of independent evaluations and testings. NIST (National Institute of Standards) in the US and CBAT (Centre for Biometric Authentication & Testing)[1] in China are working in this direction. The BioSecure Network of Excellence[2], aims at becoming such an European Center for evaluation and testing of biometrics algorithms and systems. Biometrics can also be seen as a way to increase privacy and anonymity when considering personal security needs. It would be nice to increase the education of the future users and operators in order to demystify this topic which is still considered to some extent as science-fiction, as well as to continue to develop technical research in relation to the actual needs of the citizens and the states.

## References

1. B. Dorizzi, P. Lamadeleine, C. Guerrier, and J. L. L. Jardins, "Biométrie : Techniques et usages," *Revue des sciences et techniques de l'ingenieur*, Avril 2004.
2. A. K. J. et al, "Biometrics: A grand Challenge," in *Proceedings of International Conference on Pattern Recognition*, Cambridge, UK, August 2004.

[1] `http://www.sinobiometrics.com/pdf/xuchenghua.pdf`
[2] `http://www.biosecure.inf`

3. R. Wildes, "Iris recognition: an emerging biometric technology," *Proceedings of the IEEE*, vol. 85, no. 9, pp. 1348–1363, September 1997.
4. J. Daugman, "High confidence recognition of persons by rapid video analysis of iris texture," in *European Convention on Security and Detection*, May 1995, pp. 244–251.
5. D. Maltoni, D. Maio, A. K. Jain, and S. Prabhakar, *Handbook of Fingerprint Recognition*. Springer, 2003.
6. *Figerprint Vendor Technology Evaluation*, http://fpvte.nist.gov/, 2003.
7. K. Okada, J. Steffens, T. Maurer, H. Hong, E. Elagin, H. Neven, and C. von der Malsburg., *The Bochum/USC Face Recognition System And How it Fared in the FERET Phase III test*. Springer-Verlag, 1998, pp. 186–205.
8. M. A. Turk and A. P. Pentland, "Face recognition using eigenfaces," in *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, June 1991, pp. 586–591.
9. J. R. Beveridge, K. She, B. Draper, and G. H. Givens, "A nonparametric statistical comparison of principal component and linear discriminant subspaces for face recognition," in *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, December 2001, pp. 535–542.
10. NIST, *FERET Database*, http://www.itl.nist.gov/iad/humanid/feret/, 2001.
11. P. J. Phillips, H. J. Moon, S. A. Rizvi, and P. J. Rauss, "The feret evaluation methodology for face-recognition algorithms," *IEEE Trans. on Pattern Analysis and Machine Learning*, vol. 22, no. 10, pp. 1090–1104, October 2000.
12. *International Conference on Pattern Recogniton*, Cambridge, UK, August 2004.
13. C. Xu, Y. Wang, T. Tan, and L. Quan, "A new attempt to face recognition using 3d eigenfaces," in *6th Asian Conference on Computer Vision (ACCV)*, ser. 884-889, vol. 2, 2004.
14. Chang, Bowyer, and Flynn, "Face recognition using 2d and 3d facial data," in *IEEE International Workshop on Analysis and Modeling of Faces and Gestures*, October 2003, pp. 187–194.
15. D. W. Mount, *Bioinformatics, Sequence and Genome Analysis*. Cold Spring Harbor Laboratory Press, 2004.
16. http://www.alumni.ca/~fren4j0/dna_fingerprinting.htm.
17. http://users.rcn.com/jkimball.ma.ultranet/BiologyPages/R/RFLPs.html.
18. J. F. Crow, "Dna forensics : Past, present and future," in *Genetic Identity Conference Proceedings, Tenth International Symposium on Human Identification,*, 1999.
19. D. M. Blackburn, M. Bone, and P. J. Phillips, *Facial Recognition Vendor Test 2002*, DOD, DARPA and NIJ, http://www.dodcounterdrug.com/facialrecognition/frvt2002/frvt2002.htm, 2002.
20. P. J. Phillips, P. Grother, R. J. Micheals, D. M. Blackburn, E. Tabassi, and J. M. Bone, *FRVT2002: Evaluation Report*, http://www.frvt.org/DLs/FRVT_2002_Evaluation_Report.pdf, March 2003.
21. A. K. Jain, C. Wilson, M. Garris, and C. Watson, *Matching Performance for the US-VISIT IDENT System Using Flat Fingerprints*, NISTIR 7110, http://www.itl.nist.gov/iad/893.03/pact/ir_7110.pdf, May 2004.
22. B. News, *Long lashes thwart ID scan trial*, news.bbc.co.uk/2/hi/uk_news/politics/3693375.stm, 7 May 2004.
23. N. report to the United States Congress, *Summary of NIST Standards for Biometric Accuracy, Tamper Resistance, and Interoperability*, ftp://sequoyah.nist.gov/pub/nist_internal_reports/NISTAPP_Nov02.pdf, November 2002.
24. R. M. Bolle, J. H. Connell, S. Pankanti, N. K. Ratha, and A. W. Senior, *Guide to Biometrics*. Springer, 2003.
25. National Physical Laboratory (UK), "Biometric product testing final report," CESG contract X92A/4009309, http://www.cesg.gov.uk/technology/biometrics, Tech. Rep., For comparisons against 8 competing biometrics, see Figure 6, page 12 of the NPL report 2001.
26. J. Daugman, "The importance of being random: Statistical principles of iris recognition," *Pattern Recognition*, vol. 36, no. 2, pp. 279–291, 2003.
27. T. Putte and J. Keuning, "Biometrical fingerprint recognition: Don't get your fingers burned," in *Proc. Working Conf. on Smart Card Research and Advanced Applications (4th); Proc. TC8/WG8.8*, 2000, pp. 289–303.
28. T. Matsumoto, H. Matsumoto, K. Yamada, and S. Hoshino, "Impact of artificial "gummy" fingers on fingerprint systems," in *Proc. of SPIE*, vol. 4677, February 2002, pp. 275–289.
29. http://www.findbiometrics.com/Pages/feature_daugman_camuk.htm.
30. A. K. Jain and A. Ross, "Multibiometric systems," *Communications of the ACM*, vol. 47, no. 1, 2004.
31. J. Kittler, M. Hatef, R. Duin, and J. Matas, "On combining classifiers," *IEEE Trans. On Pattern Analysis and Machine Intelligence*, vol. 20, no. 3, pp. 226–239, March 1998.
32. S. Garcia-Salicetti, C. Beumier, G. Chollet, B. Dorizzi, J. L.-L. Jardins, J. Lunter, Y. Ni, and D. Petrovska-Delacretaz, "Biomet: a multimodal person authentication database including face, voice, fingerprint, hand and signature modalities," in *Proc. of 4th International Conference on Audio and Video-Based Biometric Person Authentication*, Guildford, UK, July 2003, pp. 845–853.
33. B. L. Van, R. Blouet, S. Renouard, S. Garcia-Salicetti, B. Dorizzi, and G. Chollet, "Signature with text-dependent and text-independent speech for robust identity verification," in *Workshop on Multimodal User Authentication*, Santa Barbara, USA, 2003, pp. 13–18.
34. C. Sanderson, S. Bengio, H. Bourlard, J. Mariéthoz, R. Collobert, M. F. BenZeghiba, F. Cardinaux, and S. Marcel, "Speech & face based biometric authentication at idiap," IDIAP Research Report 03-13, Tech. Rep., February 2003.
35. C. C. Brown, X. Zhang, R. M. Mersereau, and M. Clements, "Automatic speech reading with application to speaker verification," in *Proc. of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2002.
36. http://www.ilo.org/public/english/bureau/inf/pr/2004/53.htm.
37. *Baseline Algorithm and Performance for Gait Based Human ID Challenge Problem*, http://gaitchallenge.org.

38. G. Tomko, "Biometrics as a privacy-enhancing technology: Friend or foe of privacy?" in *Proc. of Data Protection Authorities Workshop*. Hotel Reyes Catolicos, Santiago de Compostela, Spain: http://www.dss.state.ct.us/digital/tomko.htm, September 15th 1998.
39. U. Uludag, S. Pankanti, S. Prabhakar, and A. K. Jain, "Biometric cryptosystems: Issues and challenges," *Proceedings of the IEEE, Special Issue on Enabling Security Technologies for Digital Rights Management*, vol. 92, no. 6, June 2004.
40. C. Soutar, D. Roberge, A. Stoianov, R. Gilroy, and B. V. K. V. Kumar, *Biometric Encryption*. McGraw-Hill, 1999, ch. 22.
41. M. Savvides, B. V. K. V. Kumar, and P. K. Khosla, "Cancelable biometric filters for face recognition," in *Proc. of International Conference of Pattern Recognition (ICPR)*, Cambridge, UK, 2004.
42. Ratha, Connell, and Bolle, "Enhancing security and privacy in biometrics-based authentication systems," *IBM Systems Journal*, vol. 40, no. 3, pp. 614–634, 2001.
43. Jain, Uludag, and Hsu, "Hiding a face in a fingerprint image," in *Proc. of International Conference of Pattern Recognition (ICPR)*, August 2002.
44. http://www.alibaba.com/.
45. *ANSI Homeland Security Standards Panel: Biometric Workshop Report*, www.itl.nist.gov/iad/highlights/2001/fingerprint.html, www.ncits.org/tc_home/m1htm/docs/m1040266.pdf, www.incits.org/tc_home/m1htm/docs/m1030014.pdf (see part 2 Scope), April 2004.
46. vismod.media.mit.edu/tech-reports/TR-516.pdf.
47. http://www.psc.edu/general/software/packages/genbank/genbank.html.
48. http://www.biosci.ohio-state.edu/~genomes/mthermo/.
49. http://europa.eu.int/comm/research/success/fr/soc/0356f.html.

# 8

# Modeling of Complex Photonic Structures Devices and Applications

Vitaly F. Rodríguez-Esquerre[1], Hugo E. Hernández-Figueroa[2] and M. Koshiba[3]

[1] Federal Center of Technological Education of Bahia CEFET-BA (DTEE), Salvador, Bahia, Brazil 40300-016
vitaly@cefetba.br
[2] Department of Microwave and Optics, University of Campinas, Campinas, São Paulo, Brazil 13083-970
hugo@dmo.fee.unicamp.br
[3] Information Communication Systems, Hokkaido University, Sapporo, Japan 060-0814 koshiba@ist.hokudai.ac.jp

## 8.1 Introduction

Photonic crystals (PCs) are periodical structures of dielectric materials in one, two or three dimensions which allow handling the light properties [1], [2], [3]. They are called crystals for the periodicity of the structure, similar to crystals, and photonic because they interact with the light. The development in this area happened because of the analogy between the photonic and electronic crystalline structures introduced by Yablonovitch [4], [5], [6]. These structures have been studied for more than 50 years in one dimension, as an example of it we have filters and mirrors.

Photonic crystals (PCs) have attracted the attention of researchers because of their fascinating ability to suppress, enhance, or otherwise control the emission of light in a selected frequency range by judicious choice of the constant lattice, filling factor, and refractive indexes obtaining a complete photonic band gap (PBG), where light cannot propagate through the crystal in any direction. Photonic crystal structures have a number of potential applications: resonant cavities, lasers, waveguides, low-loss waveguide bends, junctions, couplers, optical switching, wavelength multiplexing and demultiplexing and many more [1], [2], [3], [7], [8], [9].

In order to design new photonic crystal based devices, several approaches have been used, such as the Plane Wave expansion (PWE) [1], [3], finite differences [2], the scattering matrix method [10], the cell method [11] and the finite element method [7], [8], [9], [12], [13], which can overcome the limitations of the methods above referred (convergence, starcaising, etc).

In this tutorial, the basic theory of the PC structures and the application of the finite element method in time and frequency domains using perfectly matched layers and isoparametric curvilinear elements [14] for the analysis of finite size photonic structures and devices such as cavities, filters and waveguides will be presented. The time domain approach includes current sources, the full band scheme, and the slowly varying envelope approximation, consequently, bigger time steps can be used independent of the size of the elements [8], [15]. The transmission characteristics of the structures under analysis can be determined and for the case of resonant cavities, the resonant frequency, quality factor, effective modal area, and field distribution for each mode can be obtained in a single simulation [13]. A computational tool developed by the author, for the determination of band gaps of arbitrarily shaped geometries in 2D periodical lattices, will be provided and its use will be explained [12] Additionally, a brief introduction to future works in this area and several applications will be also presented.

## 8.2 Modal Analysis of 2D Photonic Crystal Structures

To study the properties of photonic band gap materials it is important to find if some bands of frequencies exist (photonic band gap) where the propagation of the electromagnetic waves is prohibited in all the directions [1], [2]. In order to find these frequencies a lot of methods have been proposed up to now. Here, we use the Finite Elements Method (FEM) which works with inhomogeneous media having curved shapes as well and a generalized eigenvalues problem with large sparse (quasi band diagonal) matrices has to be solved [14], [16]. The 2D photonic crystals are structures which have a transverse periodicity (*yz*-plane) and a longitudinal uniformity ($x$-axis). In order to study their properties it is enough to deal with a unit cell, the atomic part of the periodic structure (Fig. 8.1), with appropriate periodic boundary conditions [1], [2], [12].

We start from the second order wave equation in 2D given by,

$$\frac{\partial}{\partial y}\left(p\frac{\partial \Phi}{\partial y}\right)+\frac{\partial}{\partial z}\left(p\frac{\partial \Phi}{\partial z}\right)=-q\frac{\omega_0^2}{c^2}\Phi \tag{8.1}$$

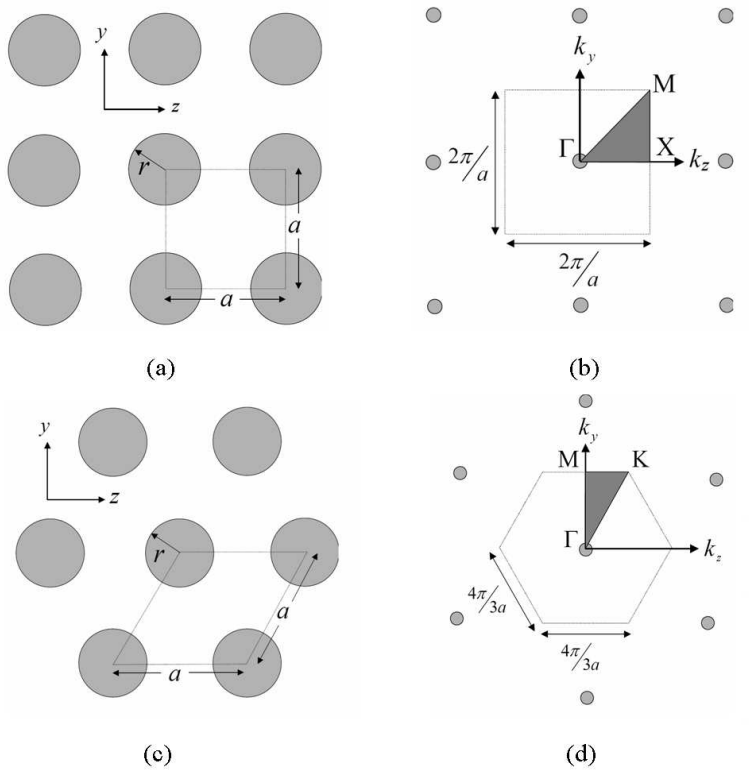



**Fig. 8.1.** Periodic structures showing the unitary cell and the first Brillouin region for (a, b) square and (c, d) triangular lattices respectively [12].

where $c$ is the speed of ligth in free space, $\Phi = E_x$, $p$ = 1, $q = n^2$ for TE modes, $\Phi = H_x$, $p = 1/n^2$, $q$ =1 for TM. The fields $E_x$ and $H_x$ can be separated in their envelope and their fast component as,

$$\begin{aligned} H_z &= h_z e^{-jk_z z} e^{-jk_y y} \\ E_z &= e_z e^{-jk_z z} e^{-jk_y y} \end{aligned} \tag{8.2}$$

By substituting (8.2) into (8.1) and applying the well known Galerkin method [14], [16], and discretizing the computational domain using 6-node isoparametric second-order triangular elements [14] (Fig. 8.2) we obtain an eigenvalue problem given by,

$$[A]\{\phi\} = \left(\frac{\omega_0}{c}\right)^2 [B]\{\phi\} \tag{8.3}$$

For TE modes, $[A]$ and $[B]$ are given by [12],

$$\begin{aligned} [A] = \sum_e \Bigg[ &\iint_e \tfrac{1}{n^2} \tfrac{\partial\{N\}}{\partial y} \tfrac{\partial\{N\}^T}{\partial y} dydz + \iint_e \tfrac{1}{n^2} \tfrac{\partial\{N\}}{\partial z} \tfrac{\partial\{N\}^T}{\partial z} dydz + \iint_e \tfrac{1}{n^2} \tfrac{\partial\{N\}}{\partial z} \tfrac{\partial\{N\}^T}{\partial z} dydz \\ &+ \iint_e jk_z \tfrac{1}{n^2} \left( \tfrac{\partial\{N\}}{\partial z} \{N\}^T - \{N\} \tfrac{\partial\{N\}}{\partial z}^T \right) dydz + \iint_e jk_y \tfrac{1}{n^2} \left( \tfrac{\partial\{N\}}{\partial y} \{N\}^T - \{N\} \tfrac{\partial\{N\}}{\partial y}^T \right) dydz \\ &+ \iint_e \tfrac{1}{n^2} \left(k_z^2 + k_y^2\right) \{N\}\{N\}^T dydz \Bigg] \\ [B] = \sum_e &\left[ \iint_e \{N\}\{N\}^T dydz \right] \end{aligned} \tag{8.4}$$

And for TM modes, $[A]$ and $[B]$ are given by,

$$\begin{aligned} [A] = \sum_e \Bigg[ &\iint_e \tfrac{\partial\{N\}}{\partial y} \tfrac{\partial\{N\}^T}{\partial y} dydz + \iint_e \tfrac{\partial\{N\}}{\partial z} \tfrac{\partial\{N\}^T}{\partial z} dydz + \iint_e \tfrac{1}{n^2} \tfrac{\partial\{N\}}{\partial z} \tfrac{\partial\{N\}^T}{\partial z} dydz \\ &+ \iint_e 2jk_z \left( \tfrac{\partial\{N\}}{\partial z} \{N\}^T \right) dydz + \iint_e 2jk_y \left( \tfrac{\partial\{N\}}{\partial y} \{N\}^T \right) dydz + \iint_e \left(k_z^2 + k_y^2\right) \{N\}\{N\}^T dydz \Bigg] \\ [B] = \sum_e &\left[ \iint_e n^2 \{N\}\{N\}^T dydz \right] \end{aligned} \tag{8.5}$$

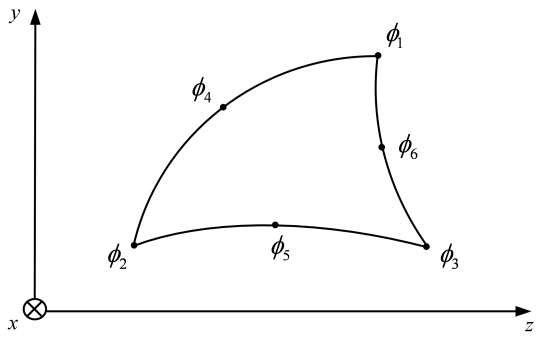



**Fig. 8.2.** 6-node isoparametric second-order triangular element [14].

The matrices $[A]$ and $[B]$ are sparse and symmetric. The matrix $[B]$ is positive and real, $\{\phi\}$ is the field $h_z$ or $e_z$ and $(\omega/c)$ is the eigenvalue.

In order to deal with only the unitary cell, the periodical boundary conditions are imposed as shown in Fig. 8.3.

The dispersion curves computed by using the present algorithm for periodical structures in square and triangular lattices of dielectric rods in air are shown in Fig. 8.4.

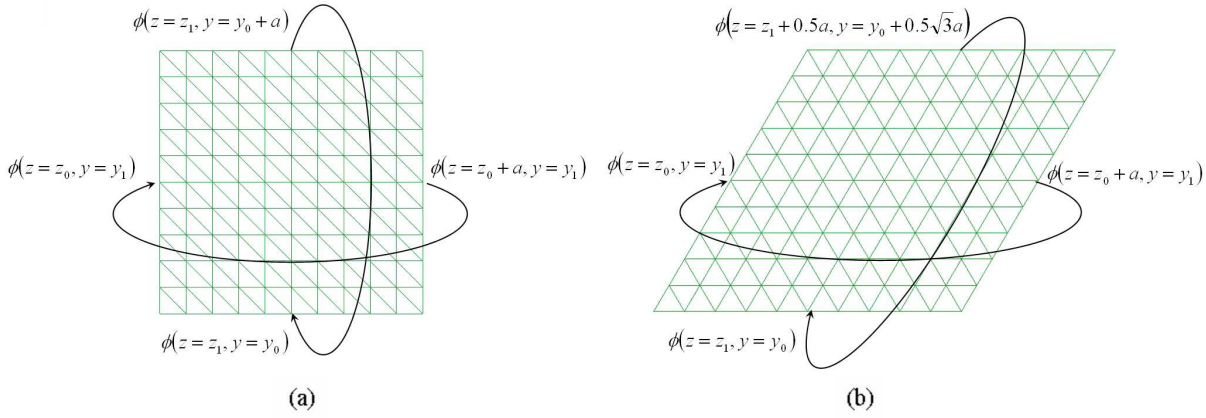



**Fig. 8.3.** Periodical boundary conditions for the unitary cell correspondent to (a) square and (b) triangular lattice, respectively.

## 8.3 Modal Analysis of 2D Photonic Crystal Waveguides

2D Photonic crystal waveguides can be obtained by introducing a defect along the structure [1], [2], [12], it can be the complete removing, the change of the size or the refractive index value of some rows or columns. Waveguides formed by removing a complete line of dielectric rods in square and triangular lattices as well as their respective unitary cell are shown in Fig 8.5

To obtain the dispersion characteristics of waveguides, the same eigenvalue problem given in (8.3) should be solved. In order to obtain the dispersion curves of the waveguides shown in Fig. 5, the eigenvalue problem should be solved along the $\Gamma$M and $\Gamma$K directions for the square and triangular lattice, respectively. The periodical boundary conditions are imposed for the fields at the right and at the left boundaries.

The guided mode for a waveguide obtained by removing a line of rod in a square lattice with parameters $n$=3.4, $a$=0.58 $\mu$m and $r$=0.18$a$ is shown in Fig. 8.6. The structure with no defect has a band gap between the normalized frequencies [0.30 - 0.46]. When a line of dielectric rods is removed, a guided mode appears inside the band gap.

The developed algorithm for modal analysis of structures and waveguides based on the above formulation is currently being used for the design and fabrication of photonic crystal structures and waveguides with circular and elliptical holes [17], [18], [19], [20].

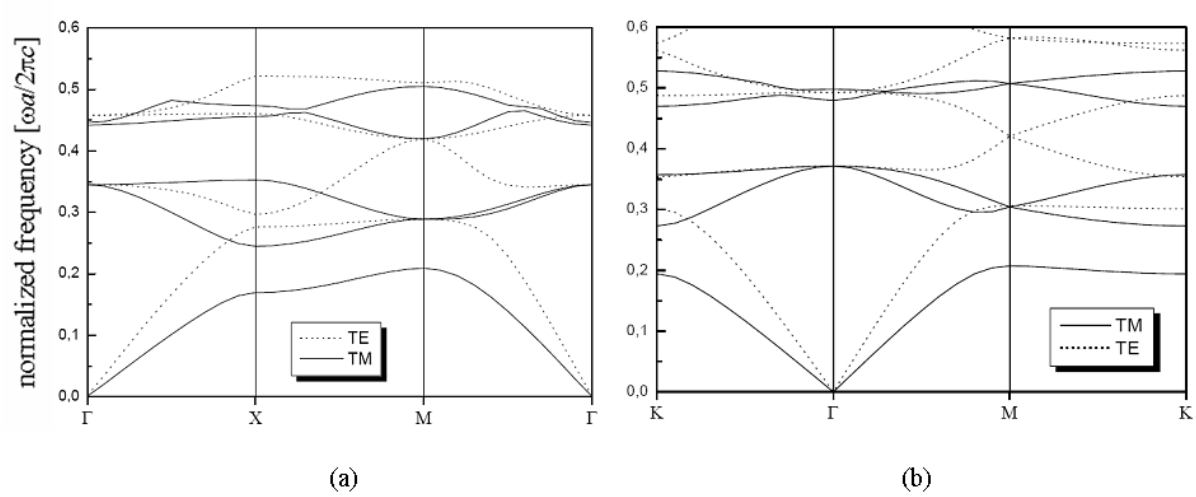



**Fig. 8.4.** Dispersion curves for (a) square lattice of dielectric rod with $n$=3.6 in air and $r/a$=0.35 (b) triangular lattice of dielectric rods with $n$=3.6 and $r/a$=0.33.

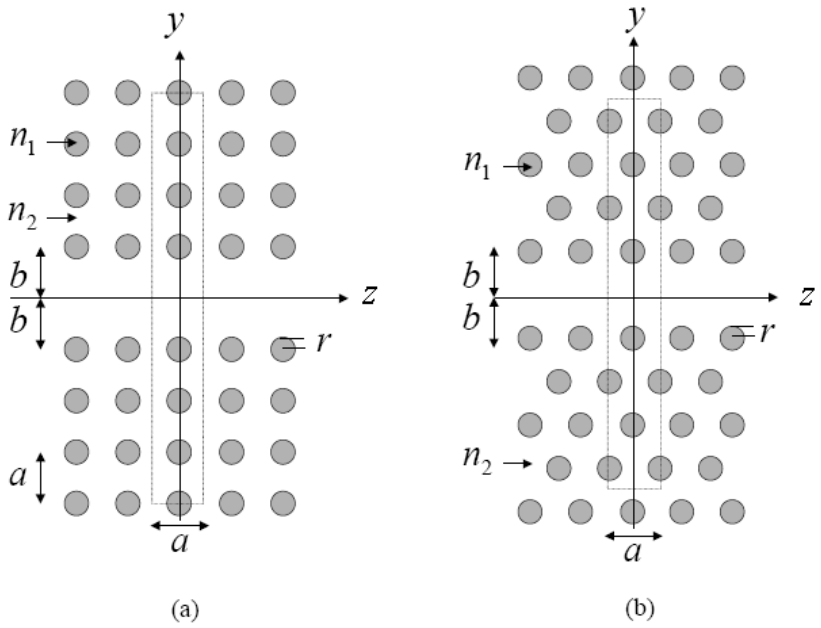



**Fig. 8.5.** Waveguides formed by removing a complete line of dielectric rods in (a) square and (b) triangular lattices, respectively.

## 8.4 Analysis of 2D Photonic Crystal Cavities: Time and Frequency Domains

Resonant cavities are formed by introducing point defects in the periodic lattice. These structures exhibit localized modes in the band gap region with a very narrow spectra and high quality ($Q$) factor in a small area, becoming a good candidate for laser fabrication [2].

Numerical simulation of defects in photonic crystals is essential for the study of the localized modes and it is divided in frequency and time domain methods. In this tutorial, the finite element method (FEM) is applied in both, the time and frequency domains [7], to further investigate the properties of PC resonant cavities.

We consider a finite size photonic crystal cavity on a 2D spatial domain on the $y-z$ plane with the axis of the rods/holes parallel to the $x$ axis as shown in Fig. 8.7 The cavities are formed by introducing point defects, in this case by removing or filling the central rod/hole in a periodic array. The cavity is surrounded by PMLs to simulate open boundaries and the variation in the $x$ direction is neglected ($\partial/\partial x$ = 0). As we can see from the geometry, the cavity has two folded symmetry and only a quarter of the cavity needs to be computed.

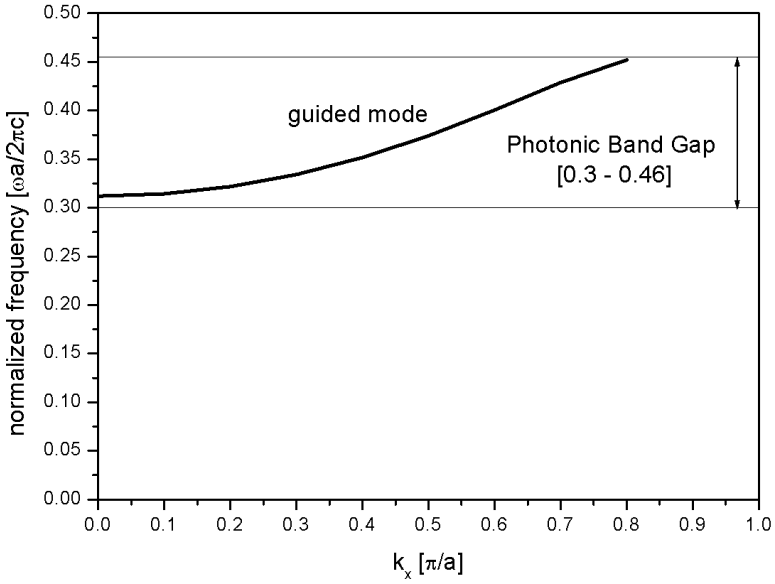



**Fig. 8.6.** Guided mode for a photonic crystal waveguide with parameters $n$=3.4, $a$=0.58 $\mu$m and $r$=0.18$a$.

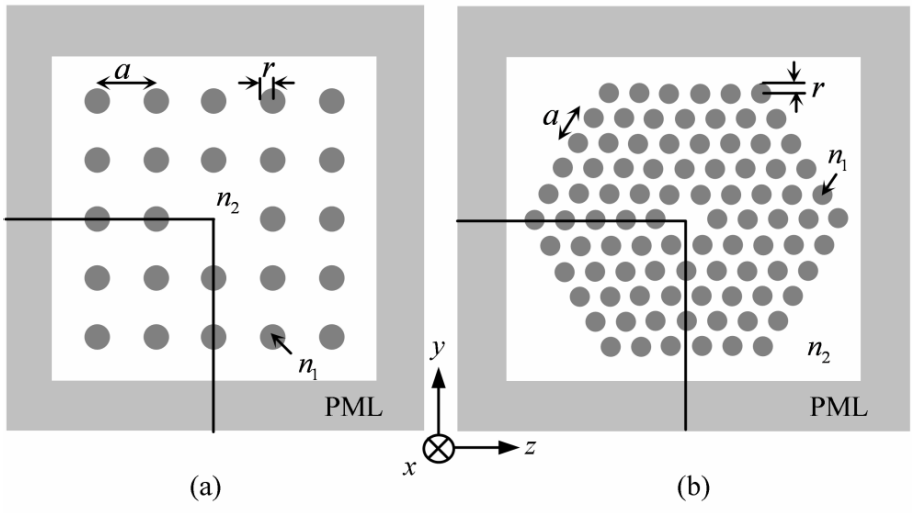



**Fig. 8.7.** Finite size 2D photonic crystal cavities with the rods/holes perpendicular to the $y-z$ plane. The boundaries are surrounded by PMLs to simulate open boundaries (a) square and (b) hexagonal cavities.

### 8.4.1 Time Domain Analysis

The transverse electric (TE) and transverse magnetic (TM) fields are compactly described by the Helmhothz equation considering an external current density $J$, as follows

$$-s\frac{q}{c^2}\frac{\partial \Phi}{\partial t^2}+s_y\frac{\partial}{\partial y}\left(p\frac{s_y}{s}\frac{\partial \Phi}{\partial y}\right)+s_z\frac{\partial}{\partial z}\left(p\frac{s_z}{s}\frac{\partial \Phi}{\partial z}\right)=\Theta\left(J\right)\bullet\hat{a}_x \tag{8.6}$$

where $c$ is the speed of ligth in free space, $J$ is the density of current (external excitation), $\Phi = E_x$, $p$ = 1, $q = n^2$, $\Theta$ = -$\mu_0\partial/\partial t$ for TE modes, $\Phi = H_x$, $p$ = 1/$n^2$, $q$ =1, $\Theta = -\nabla \quad \times(1/n^2)$ for TM modes, $n$ is the refractive index, $\mu_0$ is the free space permeability, and $\hat{a}_x$ is the unit vector in the $x$ direction.

Here, $s$, $s_y$ and $s_z$ are parameters related to the absorbing boundary conditions of PML type, and the parameter $s$ is given by

$$s = \begin{cases} 1 - j\frac{3c}{2\omega_0 nd}\left(\frac{\rho}{d}\right)^2 \ln\frac{1}{R} \text{ in PML regions} \\ 1 \text{ in other regions} \end{cases} \tag{8.7}$$

where $j = \sqrt{-1}$, $\omega_0$ is the angular frequency, $d$ is the thickness of the PML layer, $\rho$ is the distance from the beginning of PML, and $R$ is the theoretical reflection coefficient. The other parameters, $s_y$ and $s_z$, take the values described in Table 8.1.

**Table 8.1.** Values of $s_y$ and $s_z$

| $s_y$ | $s_z$ | PML's location |
|---|---|---|
| 1 | $s$ | Perpendicular to $y$ axis |
| $s$ | 1 | Perpendicular to $z$ axis |
| 1 | 1 | on the corners |

Since the wave is considered to be centered at the frequency $\omega_0$, the field is given by $\Phi = \phi(y,z,t)\exp(j\omega_0 t)$, where $\phi(y,z,t)$ represents the wave's slow-varying envelope and the current density $J$ is given in the same way, i.e. $J = V(y,z,t)\exp(j\omega_0 t)$ with $V(y,z,t)$ being the slowly varying current density. Substituting this expression into (8.6), dividing the spatial domain into curvilinear triangular elements, and applying the conventional Galerkin/FEM procedure, we obtain the following equation for the slowly varying envelope,

$$-\frac{1}{c^2}[M]\frac{\partial^2\{\phi\}}{\partial t^2} - 2j\frac{\omega_0}{c^2}[M]\frac{\partial\{\phi\}}{\partial t} + \left([K] + \frac{\omega_0^2}{c^2}[M]\right)\{\phi\} = \{f\} \tag{8.8}$$

where the matrices $[M]$ and $[K]$ are given by

$$[M] = \sum_e \iint_e sq\{N\}\{N\}^T dydz \tag{8.9}$$

$$[K] = \sum_e \iint_e \left[-p\frac{s_y^2}{s}\frac{\partial\{N\}}{\partial y}\frac{\partial\{N\}^T}{\partial y} - p\frac{s_z^2}{s}\frac{\partial\{N\}}{\partial z}\frac{\partial\{N\}^T}{\partial z}\right] dydz \tag{8.10}$$

where $\{N\}$ is the shape function vector, $T$ denotes a transpose, and $\Sigma_e$ extends over al different elements and $\{f\}$ represents the external excitation and is given by

$$\{f\} = \begin{cases} -\int \mu_0\{N\}\left(j\omega_0\{V_x\} + \frac{\partial\{V_x\}}{\partial t}\right) dydz \text{ for TE modes} \\ -\int \{N\}\left(\nabla\times\frac{1}{n^2}\{V\}\right)_x dydz \text{ for TM modes} \end{cases} \tag{8.11}$$

Here, the vector $\{V\}$ is non zero only at the positions corresponding to the nodal points, where it is applied.

We use isoparametric curvilinear 6-node elements for the spatial discretization [14], see Fig. 8.2 The discretization in the time domain is based on Newmark-Beta formulation [21] and following [22] we obtain

$$\begin{aligned} \frac{d^2\{\phi\}}{dt^2} &= \frac{1}{\Delta t^2}\left[\{\phi\}_{i+1} - 2\{\phi\}_i + \{\phi\}_{i-1}\right] \\ \frac{d\{\phi\}}{dt} &= \frac{1}{2\Delta t}\left[\{\phi\}_{i+1} - \{\phi\}_{i-1}\right] \\ \{\phi\} &= \left[\beta\{\phi\}_{i+1} + (1-2\beta)\{\phi\}_i + \beta\{\phi\}_{i-1}\right] \end{aligned} \tag{8.12}$$

where $\Delta t$ is the time step, the subscripts $i-1$, $i$ and $i+1$ denote the $(i-1)$-th, $i$-th and the $(i+1)$-th time steps, respectively, and $\beta$ ($0 \le \beta \le 1.0$) controls the stability of the method. The marching relation is given as

$$\begin{aligned} \left[-\frac{1}{c^2\Delta t^2}[M] - j\frac{\omega_0}{c^2\Delta t}[M] + \beta\left([K] + \frac{\omega_0^2}{c^2}[M]\right)\right]\{\phi\}_{i+1} = \left[-\frac{2}{c^2\Delta t}[M] - (1-2\beta)\left([K] + \frac{\omega_0^2}{c^2}[M]\right)\right]\{\phi\}_i + \\ \left[\frac{1}{c^2\Delta t^2}[M] - j\frac{\omega_0}{c^2\Delta t}[M] - \beta\left([K] + \frac{\omega_0^2}{c^2}[M]\right)\right]\{\phi\}_{i-1} + \{f\}_i \end{aligned} \tag{8.13}$$

We solved (8.13) by LU decomposition at the first time step and by forward and backward substitutions at each time step to obtain the subsequent field. The initial conditions are $\{\phi\}_1$=$\{\phi\}_0$=0.

The $Q$ factor is an important parameter of the cavity and tells us the number of oscillations for which the energy decays to $e^{-2\pi}$ of its initial value. From energy time variation, we can obtain the $Q$ factor as

$$Q = 2\pi \frac{|U_t|^2}{|U_t|^2 - |U_{t+T}|^2} \tag{8.14}$$

where $U_t$ is the energy at an arbitrary time position, $U_{t+T}$ is the energy after one cycle, and the cycle $T$ corresponds to the resonant frequency. A more general way to compute $Q$ is

$$Q = \frac{\omega_0(t_1 - t_0)}{2\ln(\phi_1/\phi_0)} \tag{8.15}$$

where $\phi_0$ and $\phi_1$ are the amplitudes of the electromagnetic fields at the arbitrary times $t_0$ and $t_1$, respectively. Here the time step used is at least 5 times larger than that one necessary if the fast variation is taken into account. It is important when cavities with higher $Q$ values are analyzed.

### 8.4.2 Frequency Domain Analysis

The frequency domain scalar equation governing the transverse TE and TM modes, over a 2D spatial domain $y - z$ including PMLs, free of charges ($J$=0), is obtained by replacing the operator $\partial/\partial t$ with the factor $j\omega_0$ in (8.1) as follows

$$sq\frac{\omega_0^2}{c^2}\phi + s_y\frac{\partial}{\partial y}\left(p\frac{s_y}{s}\frac{\partial\phi}{\partial y}\right) + s_z\frac{\partial}{\partial z}\left(p\frac{s_z}{s}\frac{\partial\phi}{\partial z}\right) = 0 \tag{8.16}$$

The parameters in (8.16) are the same given in (8.6). Applying the Galerkin method to (8.16), we obtain

$$[A]\{\phi\} = \left(\frac{\omega_0}{c}\right)^2 [B]\{\phi\} \tag{8.17}$$

where $\{\phi\}$ is the vector field, and the matrices $[A]$ and $[B]$ are given by

$$[A] = \sum_e \iint_e \left[p\frac{s_y^2}{s}\frac{\partial\{N\}}{\partial y}\frac{\partial\{N\}^T}{\partial y} + p\frac{s_z^2}{s}\frac{\partial\{N\}}{\partial z}\frac{\partial\{N\}^T}{\partial z}\right] dydz \tag{8.18}$$

$$[B] = \sum_e \iint_e sq\{N\}\{N\}^T dydz \tag{8.19}$$

The resulting sparse complex eigenvalue system is efficiently solved by the subspace iteration method [23]. We obtain then the field distribution and its complex resonant frequency $\omega_0$. The $Q$-factor of the mode associated to the complex frequency $\omega_0$ is given by

$$Q = \frac{Re\,[\omega_0]}{2Im\,[\omega_0]} \tag{8.20}$$

where Re and Im stand for the real and imaginary parts, respectively. The modal area $A$ is computed from the field distribution as in [24],

$$A = \frac{\int n^2|\phi|^2 dydz}{\max(n^2|\phi|^2)} \tag{8.21}$$

where $\phi$ is the $E_x$ or $H_x$ field.

We analyzed several cavity configurations [7] and the results for a hexagonal cavity are presented here. This cavity is formed by removing the central rod in 4-ring and 5-ring hexagonal lattices of dielectric rods with refractive index $n_1$ = 3.0, $r$ = 0.378$a$ in air ($n_2$ = 1.0) [2]. See Fig. 8.7(a). We chose the lattice constant $a$= 0.7254 $\mu$m to obtain the resonance frequency at $\lambda_c$ = 1.55 $\mu$m.

Strategies to compute higher order modes will also be given and explained.

## 8.5 Power splitters for waveguides composed by ultralow refractive index metallic nanostructures

The power splitting properties of metallic photonic crystal waveguides, where the mechanism of propagation is the total external reflection, are investigated [25]. These waveguides are composed by an air core surrounded by a cladding

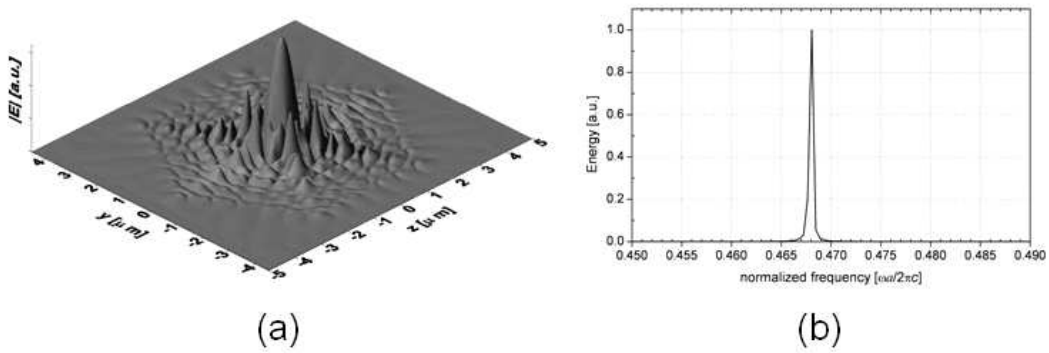



**Fig. 8.8.** (a) Electric field distribution of the resonant mode for the 5-ring cavity after 1,024 fs, and (b) Energy spectra in the cavity, showing the resonance at the normalized frequency $\omega a/2\pi c$ = 0.468 [2].

formed by a periodic arrangement of metallic wires on square or triangular lattices [26], [27]. Power splitters can be achieved by modifying the geometry at the splitting region by introducing a reflecting structure in order to increase the transmission coefficient; some splitters based on these structures are analyzed by an efficient frequency domain finite element approach. The proposed concept may open the possibility to design more sophisticated devices based on these waveguides and splitters. It has been shown that waveguides based on an air core and a cladding composed by a periodic structure of metallic wires, can effectively guide the light by total external reflection in the visible and near-infrared spectral ranges (450-700 nm) with low losses. This guidance mechanism can be explained by the fact that a composite metamaterial of metallic nanowires arranged into square or triangular lattices can be treated as an homogeneous material for an incident electromagnetic field with a wavelength much larger than the unit-cell dimensions. In addition, they can be engineered to show an equivalent refractive index lower than the air. Another peculiar characteristic is the fact that in these kinds of waveguides it is not necessary the presence of photonic band gap regions in order to obtain confinement of the light. The only constraint is that the equivalent refractive index $n_{eq}$ has to be lower than 1.0.

We analyzed splitters based on square and triangular lattices of metallic nanowires, see Fig. 8.9, by and efficient frequency domain finite element method [28]. The metamaterial is composed by 10 layers of cylindrical metallic wires with constant lattice $a$=200 nm, and radius $r$ =0.15$a$ and $r$ =0.1$a$ for the square and triangular lattices, respectively. In this way, the confinement losses in the cladding are minimized. We analyze such structures at single wavelengths and the operating wavelength for each material as well as its correspondent complex refractive index are: silver: 500nm, 0.13 – $j$ 2.90; copper 652.5nm, 0.214 – $j$ 3.67; and gold 652.6, 0.166 – $j$ 3.15. With these parameters we can ensure that the equivalent homogeneous refractive index will be lower than 1.0 with low losses and the mechanism of guidance will be the total external reflection.

Several splitters configurations have been analyzed and we found that the enhancement of the transmission coefficient T is obtained by increasing the reflectivity at the splitting region. It can be done by including a solid metallic structure (Fig. 8.9a) or by placing additional metallic rods marked in dark in Fig. 8.9b and Fig. 8.9c.

In all the simulations the input or illumination is a monochromatic wave with Gaussian profile in the transversal section, with a beam-waist of 400 nm and centered at the entrance of the line defect (waveguide) and only the TE polarization has been considered (Electric field parallel to the wires). The normalized transmission coefficient is obtained by dividing the power at the output of the splitter by the power at the output of the straight waveguide. In both cases the propagation distance is considered to be the same. In this way, we can obtain a transmission coefficient independent of the waveguide losses and mode mismatch. In the absence of all loss mechanisms the maximum value for the normalized transmission coefficient is 0.5 for each branch of the splitter.

First, we analyzed the case of the solid metallic structure (Fig. 8.9a) on square lattice. This splitter is formed by removing the row of columns at $y$=0 and the column of rows at $x$=7.4 $\mu$m. Then, we placed a solid structure with vertices at (7.37, 0), (7.60, $\pm$0.3), and (7.63, $\pm$0.3). When the points at coordinates (7.37, 0) and (7.60, $\pm$0.3) are joined by using a straight line, the computed normalized transmission coefficient for the structure made of silver is 26% and if these point are joined by using an arc of circle, it will increase to 36%. The curved surface acts as a parabola, focusing the light coming from the straight waveguide to the branches. The arc or circle used pass by the following three points: (7.37, 0), (7.50, $\pm$0.046), and (7.60, $\pm$0.3).

Another way to increase the transmission is by placing additional metallic rods with the same radius in the splitting region (Fig. 8.9b). We observed an increment of the transmission to 37% (silver).

Following the same principle, we proceed to design and analyze a splitter based on a waveguide on triangular lattice of metallic rods (Fig. 8.9c). The waveguide was obtained by removing three rows in the $\Gamma$K direction and splitters for this particular waveguide have been analyzed, the transmission coefficient has been calculated to be 0.24 for silver rods. In this case, the resulting geometry for the splitter on triangular lattices seems to be more complicated than the one corresponding to square lattices and also its performance is low. Therefore, in the triangular case we can split the light in two branches forming $60^o$ respect to the input waveguide while in square lattices the light can be split perpendicularly. The electric field distributions for the silver splitters on square and triangular lattices are shown in Fig. 8. The computed values for the normalized transmission coefficients corresponding to splitters in the three configurations and for different materials are shown in Table 8.2.

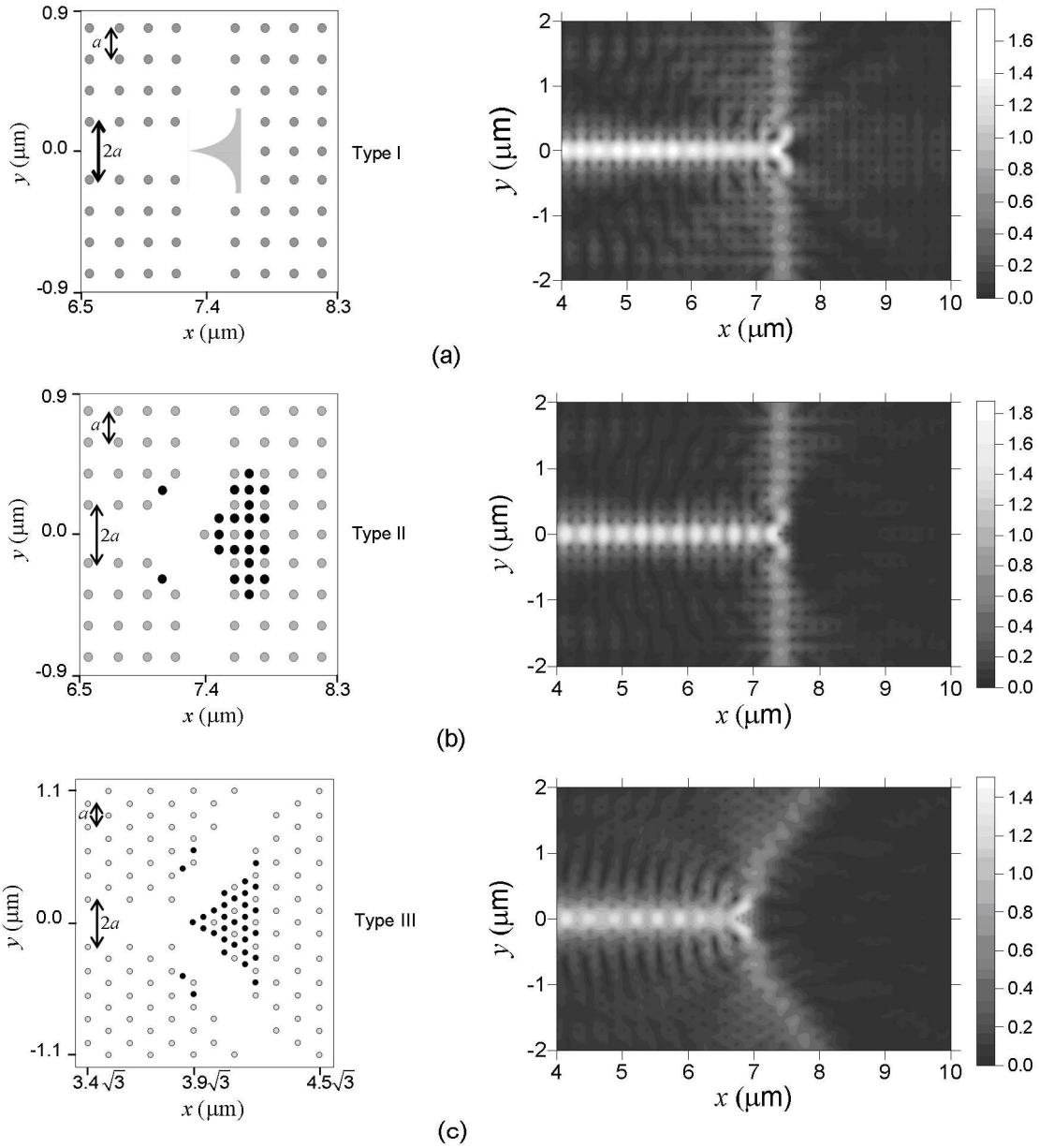



**Fig. 8.9.** Power splitter in waveguide composed by metallic silver rods and their correspondent electric field distributions [25]

The air core waveguides have some advantages over dielectric ones in the visible wavelength interval, such as, the flexibility to engineer the cladding in order to obtain arbitrarily low values of equivalent refractive index and the minimization of the losses by choosing adequately the geometrical parameters. It is expected also that air core waveguide migth be less lossy than dieletric waveguides. In contrast, the metamaterial waveguide designs presented

here do not have the same confinement characteristics than dielectric waveguides but can propagate visible light for several wavelengths.

**Table 8.2.** Transmission coefficients for the splitters analyzed (each branch).

| | TYPE I | TYPE II | TYPE III |
|---|---|---|---|
| silver | 0.35 | 0.37 | 0.24 |
| copper | 0.40 | 0.43 | 0.24 |
| gold | 0.33 | 0.37 | 0.21 |

Air core waveguides are also less lossy than waveguides based on metallic wires where the propagation mechanism is the photonic band gap effect [29] and the existence of photonic band gap is not necessary for their operation. And finally, the metallic structure is mechanically more robust, can operate at higher temperatures and metal components can be used for wiring of future optoelectronic integrated circuits.

## 8.6 Frequency Dependent Envelope Finite Element Time Domain Analysis of Dispersive Materials and Metallic Structures Operating at Optical Frequencies

To accurately perform wide-band electromagnetic simulations, the effects of the medium dispersion should be incorporated. Several approaches have been proposed for the analysis of dispersive media by using the frequency dependent finite difference time domain (FD$^2$TD) method [30], [31] and a few approaches by using the finite element time domain method have been proposed until now [32], [33].

A formulation taking into account the total field and the recursive evaluation of the convolution is presented in [32], its counterpart using the auxiliary differential equation (ADE) is presented in [33], where the formulation for the slowly wave varying envelope and the use of the mutual difference method or ADE to evaluate the convolution integral has been proposed [33]. The formulation in [32] has the limitation of the use of small time steps because the fast component of the electromagnetic field is computed. In [33], although bigger time steps can be used, it increases the number of unknowns and two simultaneous differential equations have to be solved.

In the method here presented [34] called slowly varying envelope approximation frequency dependent finite element time domain by using the recursive convolution (SVEA-FD-FETD-RC), an accurate way to evaluate the recursive convolution is introduced, consequently, larger time steps can be used without sacrificing calculation accuracy neither increasing the number of unknowns.

The 2D scalar wave equation in frequency domain for the electric field can be written as

$$\omega^2 \mu_0 \varepsilon_0 \varepsilon_r(\omega) E_x + \frac{\partial^2 E_x}{\partial y^2} + \frac{\partial^2 E_x}{\partial z^2} = j\mu_0 \omega J_x \tag{8.22}$$

where $\mu_0$ is the permeability of free space, $\varepsilon_0$ is the permittivity of free space, $E_x$ is the $x$ component of the electric field, $\omega$ is the angular frequency, and $J_x$ is the $x$ component of the current density.

In dispersive media like plasma and Debye materials, the relative permittivity is given as

$$\varepsilon_r(\omega) = \begin{cases} 1 + \frac{\omega_p^2}{j\omega(j\omega+\nu_c)} & \text{for plasma materials} \\ \varepsilon_\infty + \frac{\varepsilon_s - \varepsilon_\infty}{1+j\tau_0\omega} & \text{for Debye materials} \end{cases} \tag{8.23}$$

where $\omega_p$ and $\nu_c$ are the plasma and the damping frequencies, respectively, $\varepsilon_s$ and $\varepsilon_\infty$ are the relative permittivity in low and high frequencies, respectively, and $\tau_0$ is the relaxation time.

Substituting (8.23) into (8.22) and applying the Fourier transform, the time domain equations are obtained

$$-\frac{\partial^2 E_x}{\partial y^2} - \frac{\partial^2 E_x}{\partial z^2} + \frac{1}{c^2}\frac{\partial^2 E_x}{\partial t^2} + \frac{1}{c^2}\omega_p^2 (1 - \nu_c \varphi\otimes) E_x = -\mu_0 \frac{\partial J_x}{\partial t} \tag{8.24}$$

for plasma materials, and

$$-\frac{\varepsilon_\infty}{c^2}\frac{\partial^2 E_x}{\partial t^2} - \frac{\varepsilon_s - \varepsilon_\infty}{c^2}\frac{1}{\tau_0}\left(\frac{\partial}{\partial t} - \frac{1}{\tau_0} + \frac{1}{\tau_0^2}\varphi\otimes\right) E_x + \frac{\partial^2 E_x}{\partial y^2} + \frac{\partial^2 E_x}{\partial z^2} = \mu_0 \frac{\partial J_x}{\partial t} \tag{8.25}$$

for Debye materials, where $\varphi$ = exp(-*bt*)$u(t)$, $u(t)$ is the unit step function, $\otimes$ denotes convolution, and $b$ is equal to $\nu_c$ and $1/\tau_0$ for the plasma and Debye materials, respectively.

Defining $\Psi = \varphi \otimes \quad E_x$, under the consideration that the modulated signal varies slower than the carrier, we can separate the signal in the slowly varying envelope and the fast component in the following way: $E_x = U$ exp(*j*$\omega_0 t$), $\Psi$ = $\psi$ exp(*j*$\omega_0 t$) and $J_x = V$ exp(*j*$\omega_0 t$), where $U$, $\psi$ and $V$ are the slowly varying envelope signals and $\omega_0$ is the central angular frequency.

With this, the time derivatives of $E_x$ and $J_x$ can be written as function of their envelopes $U$ and $V$, through the relations,

$$\begin{array}{l} \frac{\partial^2}{\partial t^2} = \frac{\partial^2}{\partial t^2} e^{j\omega_0 t} + 2j\omega_0 \frac{\partial}{\partial t} e^{j\omega_0 t} - \omega_0^2 e^{j\omega_0 t} \\ \frac{\partial}{\partial t} = \frac{\partial}{\partial t} e^{j\omega_0 t} + j\omega_0 e^{j\omega_0 t} \end{array}, \tag{8.26}$$

here $\omega_0$ is the central angular frequency. Substituting these expressions into (8.24) and (8.25), we obtain

$$\frac{1}{c^2}\frac{\partial^2 U}{\partial t^2} + \frac{2j\omega_0}{c^2}\frac{\partial U}{\partial t} - \frac{\partial^2 U}{\partial y^2} - \frac{\partial^2 U}{\partial z^2} + \frac{1}{c^2}\left(\omega_p^2 - \omega_0^2\right) U - \frac{\omega_p^2}{c^2}\nu_c \psi = -\mu_0 \left(\frac{\partial V}{\partial t} + j\omega V\right) \tag{8.27}$$

for plasma materials, and

$$\begin{array}{r} -\frac{\varepsilon_\infty}{c^2}\frac{\partial^2 U}{\partial t^2} - \left(2j\omega_0 \frac{\varepsilon_\infty}{c^2} + \frac{\varepsilon_s - \varepsilon_\infty}{c^2 \tau_0}\right)\frac{\partial U}{\partial t} + \left(\frac{\varepsilon_\infty \omega_0^2}{c^2} - j\omega_0 \frac{\varepsilon_s - \varepsilon_\infty}{c^2 \tau_0} + \frac{\varepsilon_s - \varepsilon_\infty}{c^2 \tau_0^2}\right) U + \frac{\partial^2 U}{\partial y^2} + \frac{\partial^2 U}{\partial z^2} - \frac{\varepsilon_s - \varepsilon_\infty}{c^2 \tau_0^3}\psi \\ = \mu_0 \left(\frac{\partial V}{\partial t} + j\omega V\right) \end{array} \tag{8.28}$$

for Debye materials, applying the conventional Galerkin procedure, and using second-order triangular elements for the spatial discretization, the following matricial differential equation is obtained,

$$[A]\frac{\partial^2 \{U\}}{\partial t^2} + [B]\frac{\partial \{U\}}{\partial t} + [C]\{U\} + [D]\{\psi\} = \{f\} \tag{8.29}$$

Here, matrices $[A]$, $[B]$, $[C]$ and $[D]$, and, the vector $\{f\}$, are the discrete counterparts of the operators and source related to (8.27) and (8.28) under the SVEA, respectively. They can be easily identified from (8.27) or (8.28). Although the resulting convolution for the SVEA includes a fast varying component into the integral, it can be computed accurately, the slow component is linearly approximated in the interval ($t$, $t + \Delta t$), and the fast component is analytically solved. Consequently, the following recursive relation is obtained,

$$\psi_{i+1} = e^{-(b+j\omega_0)\Delta t}\psi_i + \frac{1}{2j\omega_0}\left[U_{i+1} + e^{-b\Delta t}U_i\right]\left(1 - e^{-j\omega_0 \Delta t}\right) \tag{8.30}$$

where $\Delta t$ is the time step, and the subscripts $i$ and $i$+1 denote the $i$-th and the ($i$+1)-th time steps, respectively. This expression let us evaluate the recursive convolution using larger time steps.

The discretization of the envelope $U$ in the time domain is based on Newmark-Beta formulation [21],

$$\begin{array}{l} \frac{d^2\{U\}}{dt^2} = \frac{1}{\Delta t^2}\left[\{U\}_{i+1} - 2\{U\}_i + \{U\}_{i-1}\right] \\ \frac{d\{U\}}{dt} = \frac{1}{2\Delta t}\left[\{U\}_{i+1} - \{U\}_{i-1}\right] \\ \{U\} = \left[\beta\{U\}_{i+1} + (1-2\beta)\{U\}_i + \beta\{U\}_{i-1}\right] \end{array} \tag{8.31}$$

the subscripts$i - -1$, $i$ and $i$+1 denote the ($i - -1$)-th, $i$-th and the ($i$+1)-th time steps, respectively. The resulting linear system has been solved by LU decomposition at the first time step and by forward and backward substitutions at each time step to obtain the subsequent field. The initial conditions are $\{U\}_2$=$\{U\}_1$=$\{\psi\}_1$=0. The size of the time step is just limited by the bandwidth of the simulation and the number of unknowns remains unaltered. In order to avoid reflections, PMLs have been used [35].

Metal-dielectric composite nanostructures formed by periodical lattices of thin metallic wires embedded in a dielectric substrate have attracted the attention of researchers. These devices are good candidates for applications in optical frequencies due to the possibility to obtain equivalent homogeneous media with effective refractive index lower than one or even negative values [sch03, sch04] by a judicious choice of the radius of the wires and the constant lattice. The metal-dielectric structure analyzed is composed by an infinite in extension in the $y$ direction periodical triangular lattice of thin metallic wires of silver with radius $r$ = 15 nm in free space with constant lattice $a$ = 200 nm as shown in Fig. 9. The frequency dependent relative permittivity of silver at optical frequencies corresponds to a plasma like behavior and can be obtained by using $\omega_p = 2\pi \quad \times \nu_p = 2\pi \quad \times$ 2074.8 THz and $\nu_c$ = 105.2 THz in the plasma

expression given in (8.23), these values are different from [26]. However, they fit in an excellent way the experimental data corresponding to the permittivity of the silver given in [26].

Due to the curvilinear nature of the metallic wires, isoparametric triangular elements were implemented to model accurately the geometry and numerical integration based on 7-points formulas [36] was applied to assemble the elemental matrices.

The transmitted, reflected and absorbed normalized powers of a plane wave for normal incidence to a metal-dielectric composite nanostructure with 5 layers in the propagation direction have been computed by normalizing the Fourier transform of the time variation of the electric field at the monitoring points 1 (reflection) and 2 (transmission) with the Fourier transform of the incident electric field. In Fig. 8.10 are shown the results obtained by using the proposed method, which are in a good agreement whit those obtained in frequency domain [28]. The plane wave is generated by a current density with a Gaussian variation in the time as in (8.32), where $T_0$ = 8 fs and $t_0$ = 50 fs were defined to have a sufficiently wide bandwidth, the time step size used was 0.4 fs and the central wavelength used was $\lambda_0$ = 1.0 $\mu$m.

$$J_x = \exp\left[-4\left(t - t_0\right)^2 / T_0^2\right] \tag{8.32}$$

We analyzed the photonic structure, discretizing only one unitary cell in the $y$ direction due to the periodicity and PMLs were used in the $z$ direction to end the computational domain.

Our results were found to be in good agreement with the analytical ones for the plasma and debye material slabs or with those obtained by frequency domain simulations for the nanostructures. The time step used in SVEA is at least four times larger than in the total field simulation to attain the same convergence, when the total field is taken into account, the $\Delta t$ necessary for the convergence has to be at least $T$/30, where $T$ is the period corresponding to the central frequency of the simulation.

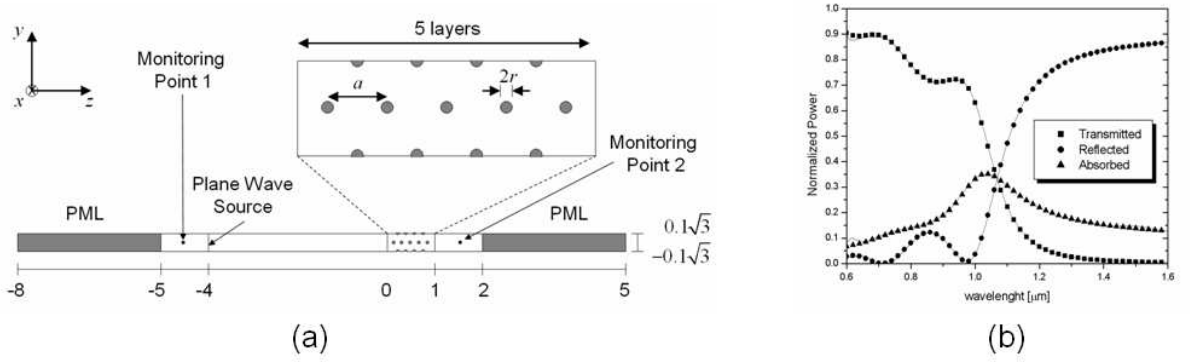



**Fig. 8.10.** (a) Computational domain used for the simulation of the metallic-dielectric composite nanostructure formed by a triangular array of cylindrical wires of silver in air, with radius $r$=0.015 $\mu$m, constant lattice $a$=0.200 $\mu$m, and 5-layers in the propagating direction (units in $\mu$m) (b) normalized transmitted, reflected, and absorbed powers of a plane wave at normal incidence at the dispersive metallic-dielectric composite structure. Results obtained by the proposed time domain method (continuous curves) and frequency domain results (squares, circles and triangles)[28].

## 8.7 Implicit and Explicit Frequency-Dependent Finite-Element Time-Domain FETD Methods for Metallic Nanostructures Analysis

An explicit frequency-dependent algorithm based on the FETD method was newly presented [37] and its performance was compared with an implicit one [38]. In this model, the frequency dependence of the permittivity is represented by using a plasma model, $\varepsilon_r(\omega) = 1 + \omega_p^2 / (j\omega(j\omega + \nu_c))$, where $\omega_p$ and $\nu_c$ are the plasma and the damping frequencies, respectively. In this way the dispersion and losses are taken into account. The explicit algorithm is obtained following the formulation of the implicit method [38], and using the central difference scheme for the time discretization in (8.29) instead of the Newmark method, the conventional lumping process [39] can be applied in order to transform the final linear system of equations in a simple multiplication process. This reduced considerably the computational effort and resources.

In order to assess both the implicit and explicit algorithms, the transmission properties of a metallic PC in air on square array have been computed for normal incidence of a plane wave to it.

The metallic structure analyzed is a periodical square lattice of metallic rods of silver with radius $r$ = 100 nm in free space as shown in Fig. 8.11, where the lattice constant$a$ = 600 nm and the central column of rods has been removed. The structure is composed of 5 layers of rods in the $z$ direction and infinite number of rods in the $y$ direction. The frequency dependent relative permittivity of silver is obtained by using $\omega_p = 2\pi \times \nu_p = 2\pi \times 2074$ THz and $\nu_c$ = 105 THz in the plasma model. These values fit in an excellent way the experimental data corresponding to the permittivity of the silver.

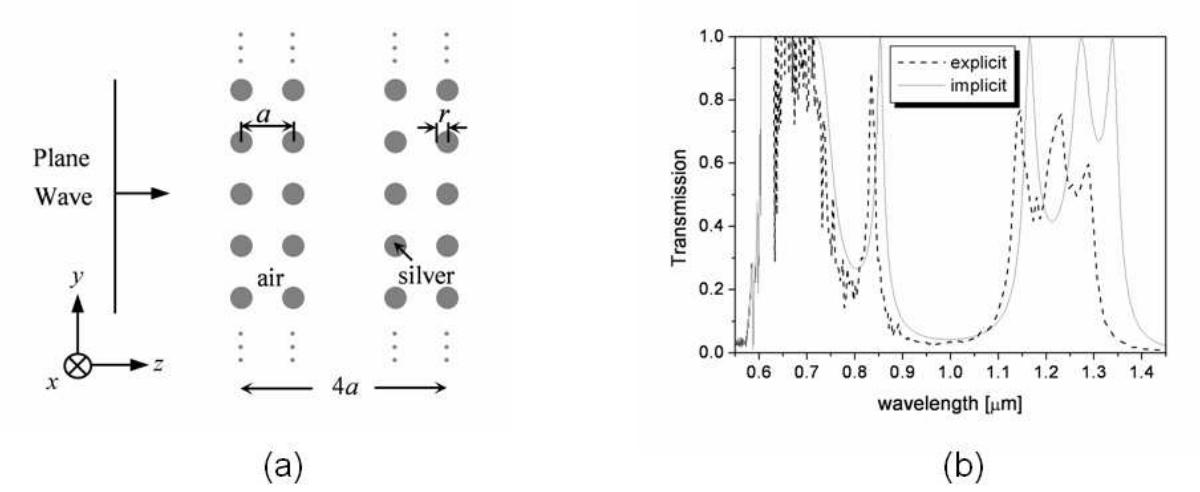



**Fig. 8.11.** (a) Geometry of the problem (b) transmission characteristics [37]

## 8.8 Numerical Platform for Band Gap Computation

A code developed by the authors for the calculation of dispersion curves and band gaps in periodical structures will be freely distributed and its use will be also explained.

The software can be obtained from http://www.cefetba.br/professores/vitaly/pcmodeling, any doubt please contact the author at vitaly@cefetba.br. The zip file contains the project files needed for the computation of PBGs in square and triangular lattices. Please choose adequately the project you want to use. The computation of the PBG is explained briefly and can be obtained by following the next steps.

First, the geometry of the unit cell to be analyzed should be drawn by using the commercial mesh generator Geometry and Data® (http://gid.cimne.upc.es). This software can be used in its educational version with the constrain of meshing only some hundred elements. When the new geometry is already created, the refractive indexes should be assigned, then the periodical boundary conditions have to be assigned in the following way: P1 (top), P2 (right), P3 (bottom) and P4 (left). Finally, the mesh is generated and the data that represents the discretized geometry (nodes numeration, element numeration, refractive indexes of the elements and number of nodes with periodical boundary conditions) should be exported.

The next step is to use Matlab® and run the command PBG which call an m-file where the tranversal propagation constants are generated and it also calls the executable file modopbg.exe. The executable file was obtained by compiling the file modopbg.for, written in Fortran language, where the finite element algorithm has been implemented. It assembles the global matrices $[A]$ and $[B]$, given in (8.4) and (8.5), and saves them as data files. These files are loaded by the m-file and the eigenvalues system is then solved. More details can be found in the readme.doc file.

## 8.9 Applications, Future Work and Trends

Besides the applications here described, several other applications of photonic crystal devices will be discussed such as

- Couplers
- Optical Filters

- Bends in Waveguides
- WDM devices
- Switches.

Regarding future works in this fascinant area, firstly, we are considering the extension of the algorithms here presented for the analysis of 3D (three dimensional) structures in order to obtain results from more realistic simulations. These algorithms will be able to model structures with complex geometries and the refractive index will be dependent of polarization, light intensity, temperature and electric field intensity and for this purpose efficient solvers for the resulting linear system of equations are required [40]. Secondly, the problem of synthesis of optical devices with a determined transmission characteristic will be also analyzed. We will consider for this purpose the use of optimization techniques such as genetic algorithms [41]. Finally, the analysis of photonic crystal fibers and devices is also an interesting topic of research. [42], [43], [44].

## 8.10 Conclusions

In this tutorial, the basic theory about photonic crystals structures and several approaches by using the finite element method in both, time and frequency domains for the analysis of photonic devices based on periodical structures have been presented. It has been shown that FEM is suitable for this kind of analysis and our results are in good agreement with those previously published. Also, several applications of devices based on photonic crystals structures have been discussed.

**ACKNOWLEDGMENTS**

This work has been supported by FAPESP (98/16202-0, CePOF, and 05/51339-1), $21^{st}$ COE Program of Japan, CNPq process 301209/94-4 and FAPESB.

## References

1. J. D. Joannopoulos, R. D. Meade, and J. N. Winn, *Photonic Crystals: Molding the Flow of Light*. Princenton, September 1995.
2. K. Sakoda, *Optical Properties of Photonic Crystals*. Springer-Verlag New York, April 2001.
3. S. G. Johnson, *Photonic crystals: from theory to practice*. Springer, January 2002.
4. E. Yablanovitch, "Inhibited spontaneus emission in solid state physics and electronics," *Physical Review Letters*, vol. 58, no. 20, pp. 2059–2062, May 1987.
5. E. Yablonovitch and T. J. Gmitter, "Photonic band structure: The face-centered-cubic case," *Physical Review Letters*, vol. 63, no. 18, pp. 1950–1953, October 1989.
6. E. Yablonovitch, "Photonic band-gap structures," *J. Opt. Soc. Am. B*, vol. 10, no. 2, pp. 283–295, February 1993.
7. V. F. Rodríguez-Esquerre, M. Koshiba, and H. E. Hernández-Figueroa, "Finite element analysis of photonic crystal resonant cavities: Time and frequency domains," *IEEE/OSA J. Lightwave and Technology*, vol. 23, no. 3, pp. 1514–1521, March 2005.
8. M. Koshiba, Y. Tsuji, and M. Hikari, "Time-domain beam propagation method and its application to photonic crystal circuits," *IEEE/OSA J. Lightwave and Technology*, vol. 18, no. 1, pp. 102–110, January 2000.
9. M. Koshiba, "Wavelength division multiplexing and demultiplexing with photonic crystal waveguide couplers," *IEEE/OSA J. Lightwave and Technology*, vol. 19, no. 12, pp. 1970–1975, December 2000.
10. J. Yonekura, M. Ikeda, and T. Baba, "Analysis of finite 2-d photonic crystals of columns and lightwave devices using the scattering matrix method," *IEEE/OSA Journal of Lightwave and Technolgy*, vol. 17, no. 8, pp. 1500–1508, August 1999.
11. M. Marrone, V. F. Rodríguez-Esquerre, and H. E. Hernández-Figueroa, "Novel numerical method for the analysis of 2d photonic crystals: the cell method," *Optics Express*, vol. 10, no. 22, pp. 1299–1304, November 2005. [Online]. Available: http://www.opticsexpress.org/abstract.cfm?URI=OPEX-10-22-1299
12. V. F. Rodríguez-Esquerre, "Photonic structures modeling by 2d finite elements in time domain," Ph.D. dissertation, FEEC Unicamp, April 2003.
13. V. F. Rodríguez-Esquerre, M. Koshiba, and H. E. Hernández-Figueroa, "Finite-element time-domain analysis of 2-d photonic crystal resonant cavities," *IEEE Photonics Technology Letters*, vol. 16, no. 3, pp. 816–818, March 2004.
14. M. Koshiba, *Optical waveguide theory by the finite element method*. Kluwer Academic Publishers, London, 1992.
15. V. F. Rodríguez-Esquerre and H. E. Hernández-Figueroa, "Novel time-domain step-by-step scheme for integrated optical applications," *IEEE Photonics Technology Letters*, vol. 13, no. 4, pp. 311–313, April 2001.
16. J. Jin, *The finite element method in electromagnetics*. John Wiley and Sons, Inc, second edition, 2002.
17. F. Quinonez, J. W. Menezes, L. Cescato, V. F. Rodríguez-Esquerre, H. E. Hernández-Figueroa, and R. D. Mansano, "Band gap of hexagonal 2d photonic crystals with elliptical holes recorded by interference lithography," *Optics Express*, vol. 14, no. 11, pp. 4873–4879, May 2006.

18. J. W. Menezes, F. Quinonez, L. Cescato, and V. F. Rodríguez-Esquerre, "Design and fabrication of 2d photonic crystal waveguides," in *Proceedings of IV Meeting of SBPMAT Sociedade Brasileira de Pesquisas em Materiais*, Recife, Brazil, October 2005, p. MOPSE 519.
19. F. Quinonez, J. W. Menezes, V. F. Rodríguez-Esquerre, L. Cescato, and R. D. Mansano, "Fabrication of 2d triangular photonic crystals by holographic litography," in *Proceedings of IV Meeting of SBPMAT Sociedade Brasileira de Pesquisas em Materiais*, Recife, Brazil, October 2005, p. MOPSE 512.
20. F. Quinonez, E. L. Rigon, L. Cescato, V. F. Rodríguez-Esquerre, and H. E. Hernández-Figueroa, "Design and fabrication of 2d photonic crystals by holographic lithography," in *Proceedings of SPIE 5th Iberoamerican Meeting on Optics and 8th Latin American Meeting on Optics, Lasers, and Their Applications*, vol. 5622, Porlamar-Isla de Margarita, Venezuela, October 2004, pp. 896–900.
21. N. M. Newmark, "A method of computation for structural dynamics," *Journal of the Engineering Mechanics Division, ASCE*, vol. 85, no. EM3, pp. 67–94, July 1959.
22. A. Frasson and H. E. Hernández-Figueroa, "Envelope full-wave 3d finite element time domain method," *Microwave and Optical Technology Letters*, vol. 35, no. 5, pp. 351–354, December 2002.
23. B. N. Parlet, *The symmetric eigenvalue problem*. Prentice-Hall, 1990.
24. M. R. Watts, S. G. Johnson, H. A. Haus, and J. D. Joannopoulos, "Electromagnetic cavity with arbitrary q and small modal volume without a complete photonic bandgap," *Optics Letters*, vol. 27, no. 2, pp. 1785–1787, October 2002.
25. V. F. Rodríguez-Esquerre, M. Koshiba, H. E. Hernández-Figueroa, and C. E. Rubio-Mercedes, "Power splitters for waveguides composed by ultralow refractive index metallic nanostructures," *Applied Physisc Letters*, vol. 87, p. 091101, 2005.
26. B. T. Schwartz and R. Piestun, "Total external reflection from metamaterials with ultralow refractive index," *Journal of Optical Society of America B*, vol. 20, no. 12, pp. 2448–2453, December 2003.
27. ——, "Waveguiding in air by total external reflection from ultralow index metamaterials," *Applied Physics Letters*, vol. 85, no. 1, pp. 1–3, July 2004.
28. C. E. Rubio-Mercedes and H. E. Hernández-Figueroa, "Padé boundary conditions for the finite-element solution of arbitrary planar junctions," *IEEE/OSA J. Lightwave and Technology*, vol. 22, no. 2, pp. 669–676, February 2004.
29. V. Poborchii, T. Taya, T. Kanayama, and A. Moroz, "Silver-coated silicon pillar photonic crystals: Enhancement of a photonic band gap," *Applied Physics Letters*, vol. 82, no. 4, pp. 508–510, January 2003.
30. R. J. Luebbers, F. Hunsberger, K. S. Kunz, R. B. Standler, and M. Schneider, "A frequency-dependent finite-difference time-domain formulation for dispersive materials," *IEEE Transactions on Electromagnetic Compatibility*, vol. 32, no. 8, pp. 222–227, August 1990.
31. D. M. Sullivan, "A frequency-dependent fdtd method for biological applications," *IEEE Transactions on Microwave Theory and Techniques*, vol. 40, no. 10, pp. 532–539, October 1992.
32. D. Jiao and J. Jin, "Time-domain finite-element modeling of dispersive media," *IEEE Microwave Wireless Compon Lett*, vol. 11, no. 5, pp. 220–222, May 2001.
33. W. Yao, Y. Wang, and T. Itoh, "2-d envelope finite element (evfe) technique-implementation of pml and dispersive medium," in *Proceedings of IEEE MTT-S International Microwave Symposium Digest*, vol. 3, Seattle, Washington, June 2002, pp. 2061–2064.
34. V. F. Rodríguez-Esquerre, M. Koshiba, and H. E. Hernández-Figueroa, "Frequency-dependent envelope finite-element time-domain analysis of dispersive materials," *Microwave and Optical Technology Letters*, vol. 44, no. 1, pp. 13–16, January 2005.
35. J. P. Berenger, "A perfectly matched layer for the absorption of electromagnetic waves," *Journal of Computational Physics*, vol. 114, no. 2, pp. 185–200, October 1994.
36. G. R. Cowper, "Gaussian quadrature formulas for triangles," *International Journal of Numerical Methods in Engineering*, vol. 7, pp. 405–408, 1973.
37. V. F. Rodríguez-Esquerre, M. Koshiba, and H. E. Hernández-Figueroa, "Comparative assessment between implicit and explicit frequency-dependent finite-element time-domain methods for metallic nanostructures analysis," in *Proceedings of IPR2005 Integrated Photonic Research Topical Meeting on CDROM*, San Diego, California, April 2005, p. ITuD3.
38. ——, "Frequency-dependent finite-element time-domain analysis of metal-dielectric composite nanostructures," in *Proceedings of IPR2004 Integrated Photonic Research Topical Meeting on CDROM*, San Francisco, California, July 2004, p. IFE2.
39. O. C. Zienkiewicz and R. L. Taylor, *The Finite Element Method*. Oxford, U.K.: Butterworth-Heinemann, 2000.
40. K. Cláudio, V. F. Rodríguez-Esquerre, J. P. da Silva, M. F. G. Hernández, and H. E. Hernández-Figueroa, "Iterative-solvers for numerical analysis of photonic devices," vol. 48, no. 6, pp. 1182–1186, June 2006.
41. D. Correia, V. F. Rodríguez-Esquerre, and H. E. Hernández-Figueroa, "Genetic-algorithm and finite-element approach to the synthesis of dispersion-flattened fiber," *Microwave and Optical Technology Letters*, vol. 31, no. 4, pp. 245–248, November 2001.
42. J. Knight, "Photonic crystal fibers," *Nature*, vol. 424, pp. 847–851, August 2003.
43. K. Saitoh and M. Koshiba, "Numerical modeling of photonic crystal fibers," *IEEE/OSA Journal of Lightwave Technology*, vol. 23, no. 11, pp. 3580–3590, November 2005.
44. P. S. J. Russell, "Photonic crystal fibers," *Science*, vol. 299, pp. 358–362, January 2003.