

# Mathematical Foundations of Cognitive Radios

Romain Couillet and M erouane Debbah

**Abstract**—Recently, much interest has been directed towards software defined radios and embedded intelligence in telecommunication devices. However, no fundamental basis for cognitive radios has ever been proposed. In this paper, we introduce a fundamental vision of cognitive radios from a physical layer viewpoint. Specifically, our motivation in this work is to embed human-like intelligence in mobile wireless devices, following the three century-old work on Bayesian probability theory, the maximum entropy principle and minimal probability update. This allows us to partially answer such questions as, what are the signal detection capabilities of a wireless device, when facing a situation in which most parameters are missing, how to react and so on. As an introductory example, we will present previous works from the same authors following the cognitive framework, and especially the multi-antenna channel modeling and signal sensing.

**Keywords**—Bayesian inference, cognitive radio, maximum entropy.

## 1. Introduction

In 1948, Claude Shannon introduced a mathematical theory of communications [1], allowing two to three generations of research to design increasingly sophisticated telecommunication tools, whose purpose is to constantly increase the achievable transmission rate over various communication channels. One of the key conclusions of Shannon was to show that a linear increase in the transmission bandwidths is expected to provide linear growth in the channel transmission capacity, while linear transmit power increase only provide sublinear capacity growth.

As a consequence, the last decades of research in telecommunications led to a situation in which the available transmission bandwidth became dramatically scarce and can only be acquired by service providers at extraordinarily high prices. Then, in the end of the nineties, the conclusions of Foschini [2] and Telatar [3] on their work on multiple antenna (MIMO) systems came as a salvation: when increasing the number of embedded antennas in both transmit and receive devices, a potential linear growth (with the number of antennas) in capacity was expected. Since the exploitation of the space dimension can come virtually at a zero cost compared to the exploitation of the frequency dimension, these stunning results rapidly generated lots of research work in the early years of the twenty-first century.

However, practical applications of multiple antenna systems took a long time to be put in place, when it was clearly realized that the exceptional predicted capacity gain could only come at a very strong signal to noise ratio (SNR)

and for low correlated channels; for instance, line of sight components in a transmission almost completely annihilates the gain of multiple antenna systems. However, up to this point in the evolution of wireless devices, the initial result from Shannon was still applicable to the most advanced technologies.

After the MIMO delusion, Joseph Mitola [4] realized that a new virtual dimension could be exploited to increase the achievable transmission rate: making the radio smarter. The basic insight of Mitola was to observe that most allocated bandwidth is not efficiently used in the sense that, most of the time, large pieces of bandwidth are left unoccupied. Enabling the wireless devices to sense the frequency spectrum in a decentralized manner<sup>1</sup> allows for a potentially high increase of *spectral efficiency*, which we define here as the actual averaged transmission rate over the theoretical capacity. These large-scale ideas from Mitola recently motivated a wide range of research with common denominator the introduction of intelligence in wireless devices. For instance, Haykin [5] introduces the concept of interference temperature, which allows to control the level of interference allowed in a network, i.e., if a given user has a rate constraint largely inferior to the effective channel capacity, the excess unused rate could be used by another device, as long as this device does not request more than the available excess rate. This interference temperature brought the new idea of primary and secondary users in a wireless network: primary users are those subscribers who are charged a high price to communicate with high quality of service, while secondary users pay a lower price to communicate over opportunistic excess rates left unused by the primary users, e.g., [6], [7].

However, all these ideas, revolutionary as they may seem, only scratch the surface of a larger entity that is the cognitive radio. Indeed, if the cognitive radio is defined, as was supposedly the prior idea of Mitola or even more certainly the basic view of Haykin<sup>2</sup>, as a radio in which all entities are capable of cognition, then the limitations in the capabilities of these radios is still unknown and not really explored. Concrete works on smart devices date back to Shannon's time as well. Claude Shannon was already interested in ideas such as a robot capable of playing chess [8]; he provided an original viewpoint of the cognitive abilities of future computers back in 1953 [9] and even constructed a mind-reading machine, the circuitry of which is depicted in [10].

<sup>1</sup>So to limit the needs for control signaling.

<sup>2</sup>Remember that the title of his main contribution on cognitive radios [5] refers to "brain-empowered" radios.

In this work, we propose to define a fundamental basis for cognitive radios on a physical layer viewpoint, which enables human-like intelligence in wireless devices. This work comes as a rupture compared to previous telecommunication work, as we will no longer rely on Shannon's work, but will rather extend it. The reasons why we escape from Shannon's framework will be explained and justified in the following sections. The additional mathematical tools needed to extend Shannon's theory of information are the theory of Bayesian probabilities, the maximum entropy principle [11] and the minimal cross-entropy principle [12], [13], among others.

The remainder of this paper unfolds as follows: in Section 2, we present the key philosophical ideas which lead from Shannon's classical information theory to Jaynes' more general probability theory. In Section 3, we provide two examples of direct application of Jaynes' maximum entropy principle to the problems of channel modeling and signal sensing. Then in Section 4, we discuss the present advantages and limitations of cognitive radios, and provide our conclusions in Section 5.

*Notation:* In the following, boldface lower-case symbols represent vectors, capital boldface characters denote matrices ( $\mathbf{I}_N$  is the  $N \times N$  identity matrix).  $X_{ij}$  denotes the  $(i, j)$  entry of  $\mathbf{X}$ . The Hermitian transpose is denoted  $(\cdot)^H$ . The operators  $\text{tr}\mathbf{X}$  and  $|\mathbf{X}|$  represent the trace and determinant, respectively. The symbol  $E[\cdot]$  denotes expectation. The operator  $\text{vec}(\cdot)$  turns a matrix  $\mathbf{X}$  into a vector of the concatenated columns of  $\mathbf{X}$ . Finally, the notation  $P_x(y)$  denotes the probability density function of the variable  $x$  in position  $x = y$ .

## 2. From Shannon to Jaynes

We will first present a simple example to show the inherent limitations of Shannon's theory of information.

### 2.1. Channel Capacity Revisited

Let us consider the simplest communication scheme, modeled as

$$y = x + n, \quad (1)$$

for some transmit signal  $x$ , additive background noise  $n$  and receive signal  $y$ . The Shannon capacity  $C$  of such a system reads

$$C = \sup_{p_x} I(x; y), \quad (2)$$

with  $p_x$  the probability distribution of the variable  $x$  taken in the set of single-variable probability distributions, and  $I$  denotes the mutual information [1]. The Eq. (2) can only be computed if the distribution of  $n$  is known. In practice,  $n$  is often taken as Gaussian, both for simplicity reasons and because this is somehow *often* close to the reality. However, there is no actual way to predict the distribution of the noise before transmitting data, and in reality the expression (2) is impossible to compute. This leads to the conclusion that

all capacity computations are in fact only *approximations* of Eq. (2).

Moreover, it is important to observe that what we call noise is in effect the sum contribution of interfering waves with different properties. If part of this noise can be analyzed by the cognitive device<sup>3</sup>, then the capacity will increase. All these primary observations lead to realize that the channel capacity is largely dependent on the prior information available at the receiver. In particular, two identical receivers, facing the same channel, may have different actual capacities depending on the individual channel state information.

Assuming the noise is known to be Gaussian with zero mean, the receiver is left to estimate the noise variance. In general, only approximative values of the SNR are available. Therefore, the channel capacity might be better seen as a *rate vector*, with entries indexed by every possible values of the SNR and taking different degrees of probability. These degrees of probability differ for each receiver, making the capacity again information-dependent and user-dependent. As a matter of fact, what one would call "real capacity", that would correspond to the capacity if the receiver knows exactly the noise variance, does not carry in itself any actual significance: as recalled by Jaynes [11; p. 634], the channel capacity is not an intrinsic value of the channel but an intrinsic value of the level of knowledge of the system designer<sup>4</sup>.

### 2.2. Limitations of Information Theory

Already in 1963, Leon Brillouin [14] realized the fundamental limitation of Shannon's information theory. In his own words [14], "*The methods of [information] theory can be successfully applied to all technical problems concerning information: coding, telecommunication, mechanical computers, etc. In all of these problems we are actually processing information or transmitting it from one place to another, and the present theory is extremely useful in setting up rules and stating exact limits for what can and cannot be done. But we are in no position to investigate the process of thought, and we cannot, for the moment, introduce into our theory any element involving the human value of the information. This elimination of the human element is a very serious limitation, but this is the price we have so far had to pay for being able to set up this body of scientific knowledge. The restrictions that we have introduced enable us to give a quantitative definition of information and to treat information as a physically measurable quantity. This definition cannot distinguish between information of great importance and a piece of news of no great value for the person who receives it.*"

Within the realm of cognitive devices, this situation in which information carries relevance, which depends on whom receives it, typically arises. Let us go back to

<sup>3</sup>There is no reason why a cognitive device would not be able to infer on what the noise is made of.

<sup>4</sup>The system designer can be seen as a virtual entity sharing the knowledge of both transmitter and receiver.

the channel capacity example above. If the receiver is provided with some additional information concerning the transmission medium, like the typical channel delay spread, the channel Doppler spread, the number of reflections, the presence of buildings in the neighborhood, how much does this affect the channel capacity is an open issue, which cannot be solved within Shannon's framework. And if the receiver experiences a poor decoding rate, what kind of information should it request to the transmitter in order to increase its performance is also an open question, e.g., should the receiver request more pilot symbols at the risk of a huge waste in spectral efficiency, should the receiver request some deterministic information regarding a given parameter of the channel? All these problems do not have deterministic channel-dependent answers but depend on the specific knowledge of the transmitter/receiver pair to which some piece of additional information might or might not be valuable.

To partially answer those questions, we propose in the following to introduce first the notion of *degrees of belief*, which turns every deterministic measurable entity, e.g., the value of the channel capacity, the value of the SNR or the value of the channel fading, into a random variable with an assigned probability distribution: this probability distribution will translate the confidence of the cognitive devices regarding the estimation of the measurable entity in question. Then, we will introduce the notion of *relevance* which enables to estimate the relative importance of information. Finally, we will discuss our general view of the capabilities of a cognitive radio.

### 2.3. The Bayesian Approach

As briefly stated in the previous section, we aim at extending the classical Shannon's information theory to enable cognitive devices with the ability of *plausible reasoning*. That is, a cognitive radio should not rely on empirical (often erroneous) decisions, but rather should be able to express doubt and to reason honestly when provided with limited knowledge. A first step in this approach is to turn empirical decisions into degrees of belief.

#### 2.3.1. Degrees of Belief and the Maximum Entropy Principle

In the Bayesian philosophy, contrary to the orthodox probability philosophy, deterministic parameters of a system, e.g., a weight, a height, the channel delay spread, which a cognitive entity needs to evaluate, must be characterized by the degrees of belief attached to all possible values for this parameter. Therefore this gives a clear meaning, for instance, to the probability that the height of the Eiffel Tower is 50 m. As a consequence, assuming a cognitive telecommunication device is not aware of the intensity  $\sigma^2$  of the background noise, instead of expressing the achievable transmission rate as the scalar  $C = \log(1 + \sigma^2)$ , which is therefore irrelevant to the communication device, it would be more adequate to consider the "vector"

$C(x) = \log(1 + x)$ ,  $x \geq 0$ , attached to a degrees of belief function, i.e., a probability density function,  $p(x)$  for each potential noise variance  $x$ . Two fundamental questions arise at this point: (i) how to use the vector  $C(x)$ ?, and (ii) how to compute  $p(x)$ ?

Answering (i) is a matter of *decision theory*, in the sense that different requirements might come into play to decide on the actual transmission rate to use: if reliability is needed, one will decide to transmit at a rate  $\log(1 + x)$  such that  $\int_0^x p(t)dt$  is less than a given (small) value, while if performance with low reliability is sought for, then  $x$  will take a larger value. This part of the cognitive radio spectrum will not be covered in this contribution.

Question (ii), on the contrary, is the point of interest in the present paper. Given the total amount of prior information at the cognitive device, how to assign degrees of belief in a systematic way? The answer to this question partially appears in the work of Shannon [1] but is better explained and developed by Jaynes [15] thanks to the introduction of the maximum entropy principle (MaxEnt) [16]. The key idea behind MaxEnt is to find a density function  $p$ , which fulfills the constraints imposed by the prior information  $I$  while introducing no additional (unwanted) information. In other words, this density function should maximize the ignorance about unknown parameters of the cognitive device, while satisfying the constraints given in  $I$ . In Jaynes' terms, this density function is maximally non-committal regarding missing information. This function translating ignorance is proven by Jaynes and more accurately later by Shore and Johnson [17] to be the entropy function  $H$ :

$$H(p) = - \int \log(p(t))p(t)dt. \quad (3)$$

When the information contained in  $I$  is of statistical nature, such as first or second order statistics, the function  $p$  which maximizes the entropy while satisfying the constraints in  $I$  is unique and can be computed with Lagrangian multipliers. An example will be given in Subsection 3.1.

#### 2.3.2. Relevance

The problem of relevance of information is a second topic in the establishment of foundations for cognitive radios. If cognitive devices were to act like human beings, they should be able to request additional information when they do not have enough evidence to take decisions. For instance, to obtain a more accurate estimate of the noise variance  $\sigma^2$  in order to have more confidence on the achievable transmission rates, an intelligent device could require the transmitter to stop transmitting so that it can estimate  $\sigma^2$ . But this would be an expensive waste in spectral efficiency, so it could alternatively request deterministic information on a dedicated channel from the transmitter. How accurate this information must be is then another problem. To be able to decide on what *question* to ask to the transmitter, the cognitive device needs to be able to judge the *relevance* of every possible question.

This notion of *questions*, or *inquiries*, is a philosophical topic upon which little literature and very few concrete

results exist. In 1978, Cox [18], who is also at the origin of the immense work from Jaynes on Bayesian probability theory [19], mathematically defined a question as the set of possible answers to this question. Therefore, a question will be relevant if its answers carry valuable information. Assuming the set of questions is seen as an ordered set, with the largest questions being the most relevant (since their answers carry potentially more cogent information), a cognitive device can decide which appropriate request to formulate to the transmitter. The work on relevance and questions is however still in its infancy, but we insist that those are fundamental needs to the cognitive radio field; for instance, interesting contributions are found in the works of Knuth *et al.* [20], who uses lattice theory to create partial orders of finite sets of questions, which is seen as the dual (in the lattice theory terminology) of the set of answers to those questions.

### 2.3.3. What is a Cognitive Radio?

In our viewpoint, a cognitive radio must ideally be able to adapt to its environment, by gathering all cogent information about the propagation channel, the transmitted signal, etc., while never producing undesirable empirical information. This would therefore relieve the telecommunication field from all ad hoc methods, based on empirical decisions concerning unknown parameters. This does not mean that a cognitive device is not prone to making errors; however, these potential errors will never originate from erroneous system assumptions, but rather from lack of information, which would generate *broad* maximum entropy distributions<sup>5</sup>. If more cogent information is provided to a cognitive device, it will integrate it and increase its decision capabilities. In a way, the more signals a cognitive communication device is fed with, the more efficient it is; this would mean, for instance, that cognitive devices age wisely: the older the cognitive device, the more efficient.

Regarding for instance signal sensing, the first steps of which will be detailed in Section 3, we expect a cognitive device to process the received signals as follows:

1. **Initialization:** integrate all cogent information about the communication channel, the properties of the supposedly received informative signal, etc., and compute the degrees of belief associated to all relevant variables.
2. **Update loop:** when the cognitive device is fed with incoming signals, it shall update its degrees of belief regarding all the previous variables and provide the overall probability that the received signal originates from a coherent data source.
3. **Decision:** using some criterion from decision theory, e.g., the evidence for the presence/absence of a co-

<sup>5</sup>When little is known on a given parameter, the maximum entropy distribution attached to this parameter will be broad in the sense that no specific value is preferred to any other, while when more information is available on this parameter, the maximum entropy distribution will be very peaky around the exact value of the parameter.

herent data source is more than a given threshold, the cognitive device declares whether data originating from a coherent source have been received.

This protocol does not necessarily provide the most efficient sensing strategy in specific situations (sometimes it might provide a quick response, sometimes traditional algorithms might provide faster responses), but it provides the most *honest* way to treat the signal sensing problem. It is important to note that no signal detection strategy can be proven superior to any other as long as too much information on the communication environment is missing. If a given algorithm could be proven better than the Bayesian strategy, this would mean this algorithm has an information advantage; honesty would then require that the Bayesian strategy be aware of this additional piece of information. The significant advantage of the Bayesian philosophy and the maximum entropy principle over classical methods is that they do not to take any empirical guess to solve a problem. Therefore, instead of being either luckily very good, or unluckily very bad depending on the accuracy of this “guess”, they perform as best as their prior information allows them to.

Also, a cognitive device ought to be capable of requesting information when it faces a situation where it crucially lacks cogent information; for instance, a cognitive mobile phone in a low network coverage situation, should be able to request information (or even help) to the neighboring cell phones which enjoy better coverage. The interest of this request would be measured by its relevance. Adding the possibilities of formulating inquiries might eventually lead to enabling cognitive devices with the ability of *discussing*, instead of just *transmitting* and *receiving*. Bidirectional communications used to be a point of deep interest when it was realized that Shannon’s theory of communication is in fact precisely a theory of *transmission*, in which past transmitted information is assumed uncorrelated with subsequent transmitted information. In 1973, Marko proposed a generalization of Shannon’s information theory framework to encompass bidirectional communications [21], in the objective to accurately model the social interactions among animals and especially human beings. The lead was then followed by Massey [22] who extends information theory to include feedback in the expression of Shannon’s mutual information.

## 3. Examples of Application

The most elementary requirement of a cognitive radio lies in its sensing capabilities. When a waveform is received at the cognitive device, it must be capable of deciding whether this waveform originates from a coherent source of information or if this waveform is pure background noise. When little is known by the receiver concerning the surrounding environment, this problem is very intricate and has led to lots of different ad hoc techniques. Our purpose in the following is to provide a unique way of deciding on

the presence of a coherent data source given a specific amount of prior information at the receiver. First, we will discuss channel modeling, which is a necessary step to properly handle the Bayesian signal detection method.

### 3.1. MaxEnt Channel Modeling

#### 3.1.1. Introduction

Channel modeling is an entire field of research in telecommunications, which produces every year lots of new contributions. However, this huge amount of previous work on channel models leads to the following paradoxical conclusion: for a given total information gathered by a cognitive device, there exist many different channel models proposed in the literature. In such a situation, which of those channel models is the cognitive device supposed to trust? In reality, the fundamental difference between all those models lies in the additional hypothesis each of them, explicitly or implicitly, carries; some models might implicitly suggest that channels usually have a short delay spread for a given communication technology, or might suggest that it is very likely to have a strong line of sight component, etc.

However, if the receiver is not aware of that implicit information, this very information should honestly *not* be taken into account. What we will provide in the following is a systematic way to model channels, given some cogent information  $I$ , which fulfill the constraints imposed by  $I$  while being non-committal regarding unknown parameters. In brief, we will provide the most elementary models compliant with  $I$ , without introducing unwanted hypothesis.

#### 3.1.2. Gaussian i.i.d. Channels

Surprisingly enough, we will realize that most of the classical channels in the basic literature fall into the maximum entropy channel modeling methodology. This is the case of Gaussian i.i.d. Indeed, let us assume that the information  $I$  known to the cognitive device gathers the following:

- the transmitter is equipped with  $n_T$  transmit antennas,
- the receiver is equipped with  $n_R$  receive antennas,
- the channel carries an energy  $E$ .

The transmission model is

$$\mathbf{y} = \sqrt{\frac{\rho}{n_T}} \mathbf{H} \mathbf{x} + \mathbf{n}, \quad (4)$$

where  $\mathbf{x} \in \mathbb{C}^{n_T}$  is the transmitted symbol vector,  $\mathbf{n} \in \mathbb{C}^{n_R}$  the thermal or interfering noise,  $\rho$  the signal to noise ratio and  $\mathbf{H} \in \mathbb{C}^{n_R \times n_T}$  the channel we want to model.

In mathematical terms, based on the fact that

$$\int d\mathbf{H} \sum_{i=1}^{n_R} \sum_{j=1}^{n_T} |h_{ij}|^2 P_{\mathbf{H}}(\mathbf{H}) = n_T n_R E \quad (\text{finite energy}), \quad (5)$$

$$\int dP_{\mathbf{H}}(\mathbf{H}) = 1 \quad (P_{\mathbf{H}}(\mathbf{H}) \text{ is a probability distribution}), \quad (6)$$

what distribution  $P_{\mathbf{H}}$ <sup>6</sup> 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 consistent with the energy constraint. This statement is mathematically expressed by the maximization of the following expression involving Lagrange multipliers with respect to  $P_{\mathbf{H}}$ :

$$\begin{aligned} L(P_{\mathbf{H}}) = & - \int d\mathbf{H} P_{\mathbf{H}}(\mathbf{H}) \log P_{\mathbf{H}}(\mathbf{H}) \\ & + \gamma \sum_{i=1}^{n_R} \sum_{j=1}^{n_T} [E - \int d\mathbf{H} |h_{ij}|^2 P_{\mathbf{H}}(\mathbf{H})] \\ & + \beta \left[ 1 - \int d\mathbf{H} P_{\mathbf{H}}(\mathbf{H}) \right]. \end{aligned} \quad (7)$$

If we derive  $L(P_{\mathbf{H}})$  with respect to  $P_{\mathbf{H}}$ , we get

$$\frac{dL(P_{\mathbf{H}})}{dP_{\mathbf{H}}} = -1 - \log P_{\mathbf{H}}(\mathbf{H}) - \gamma \sum_{i=1}^{n_R} \sum_{j=1}^{n_T} |h_{ij}|^2 - \beta = 0, \quad (8)$$

which yields

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

with

$$P_{h_{ij}}(x) = e^{-(\gamma |x|^2 + \frac{\beta+1}{n_R n_T})}. \quad (9)$$

One of the most important conclusions of the maximum entropy principle is that, while we have only assumed the knowledge about the variance, this assumption naturally implies 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 solution to the maximization of the entropy is the Gaussian i.i.d. channel. This does not mean that we have supposed independence of the channel fades in the model, nor does it mean that real channels ought to be i.i.d. if those are known to be of energy  $E$ . However, in the generalized  $L(P_{\mathbf{P}})$  expression, there exists no constraint on the dependence of the channel entries and this leads to natural independence as an honest guess on the behavior of the channel entries. 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.

<sup>6</sup>It is important to note that we are concerned with  $P_{\mathbf{H}|I}$ , where  $I$  represents the general background knowledge (here the variance) used to formulate the problem. However, for the sake of readability,  $P_{\mathbf{H}|I}$  will be denoted  $P_{\mathbf{H}}$ .

### 3.1.3. Other Channel Models

In [23], a more complete survey on MaxEnt channel models is proposed. We will gather in the following the main results.

If the information  $I$  at the receiver is the same as previously but the receiver is not aware of the exact value of the channel energy  $E$  but knows that it is contained in the interval  $[0, E_{\max}]$ , then

$$P_{\mathbf{H}}(\mathbf{H}) = \int P_{\mathbf{H},E}(\mathbf{H}, E) dE \quad (10)$$

$$= \int P_{\mathbf{H}|E}(\mathbf{H}) P_E(E) dE. \quad (11)$$

If  $P_E$  is assigned a uniform prior on the set  $[0, E_{\max}]$ , then we obtain<sup>7</sup>:

$$P_{\mathbf{H}}(\mathbf{H}) = \frac{1}{E_{\max} \pi^{n_{\mathbf{R}} n_{\mathbf{T}}}} \int_{\frac{1}{E_{\max}}}^{\infty} u^{n_{\mathbf{R}} n_{\mathbf{T}} - 2} e^{-\sum_{i=1}^{n_{\mathbf{R}}} \sum_{j=1}^{n_{\mathbf{T}}} |h_{ij}|^2 u} du. \quad (12)$$

Note that the distribution is invariant to 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.

If the channel covariance matrix  $\mathbf{Q} = E(\text{vec}(\mathbf{H})\text{vec}(\mathbf{H})^H)$  is known to the receiver, and therefore is part of the side information  $I$ , then, denoting  $\mathbf{Q} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^H$  the spectral decomposition of  $\mathbf{Q}$ , with  $\mathbf{V} = [\mathbf{v}_1, \dots, \mathbf{v}_{n_{\mathbf{R}} n_{\mathbf{T}}}]$  and  $\mathbf{\Lambda} = \text{diag}(\lambda_1, \dots, \lambda_{n_{\mathbf{R}} n_{\mathbf{T}}})$ :

$$P_{\mathbf{H}}(\mathbf{H}) = \frac{1}{\prod_{i=1}^{n_{\mathbf{R}} n_{\mathbf{T}}} \pi \lambda_i} \exp \left\{ - \sum_{i=1}^{n_{\mathbf{R}} n_{\mathbf{T}}} \frac{|\mathbf{v}_i^H \text{vec}(\mathbf{H})|^2}{\lambda_i} \right\}. \quad (13)$$

## 3.2. Signal Detection

Now that channel modeling has been investigated, the multiple antenna signal sensing problem can be completely handled.

### 3.2.1. Channel State Information

In this problem, the cogent information at the receiver is divided into known parameters:

- S-i) the receiver has  $n_{\mathbf{R}}$  antennas;
- S-ii) the receiver samples as many as  $L$  times the input from the RF interface;
- S-iii) the signal sent by the transmitter has a constant unit mean power; it is quite important to note

<sup>7</sup>The assignment of uniform priors on variables defined on a continuous space is a very controversial point of the maximum entropy theory, which is longly discussed in [11]. Another classically used prior, which solves the problem of invariance to variable change is the so-called Jeffreys' uninformative prior [24].

that this hypothesis is very weak and should be made more accurate for communications schemes that are known only to use either QPSK (quadrature phase shift keying), 16-QAM (quadrature amplitude modulation), 64-QAM modulations for instance;

- S-iv) the MIMO channel has a constant mean power.

We similarly define additional information the receiver may be aware of

- V-i) the transmitter possesses (and uses)  $n_{\mathbf{T}}$  antennas;
- V-ii) the noise variance  $\sigma^2$  is known.

### 3.2.2. Signal Model

Given a certain amount of sampled signals, the objective of the signal detection methods is to be able to optimally infer on the following hypothesis:

- $\mathcal{H}_0$ : only background noise is received;
- $\mathcal{H}_1$ : informative data added to background noise is received.

Given hypothesis S-iii), the only information on the transmitted signal (under  $\mathcal{H}_1$ ) is its unit variance. The maximum entropy principle claims that, under this limited state of knowledge, the transmitted data must be modeled as i.i.d. Gaussian [11]. The data vector, at time  $l \in \{1, \dots, L\}$ , is denoted  $\mathbf{s}^{(l)} = (s_1^{(l)}, \dots, s_{n_{\mathbf{T}}}^{(l)})^T \in \mathbb{C}^{n_{\mathbf{T}}}$ . The data vectors are stacked into the receive matrix  $\mathbf{S} = [\mathbf{s}^{(1)}, \dots, \mathbf{s}^{(L)}]$ .

If the noise level  $\sigma^2$  is known, then either under  $\mathcal{H}_0$  or  $\mathcal{H}_1$ , the background noise must be represented, due to the same maximum entropy argument as before, by a *complex standard Gaussian matrix*  $\mathbf{\Theta} \in \mathbb{C}^{n_{\mathbf{R}} \times L}$  (i.e., a matrix with i.i.d. standard complex Gaussian entries  $\theta_{ij}$ ) [25]. Under  $\mathcal{H}_1$ , the channel matrix is denoted  $\mathbf{H} \in \mathbb{C}^{n_{\mathbf{R}} \times n_{\mathbf{T}}}$  with entry  $h_{ij}$  being the link between the  $j$ th transmitting antenna and the  $i$ th receiving antenna. The model for  $\mathbf{H}$  follows the MaxEnt channel modeling rules. In the present situation, only the constant mean power (or equivalently, the energy) of the channel is known. Therefore  $\mathbf{H}$  will be modeled as i.i.d. Gaussian, following the reasoning in the previous section. The received data at sampling time  $l$  are given by the  $n_{\mathbf{T}} \times 1$  vector  $\mathbf{y}^{(l)}$  that we stack, over the  $L$  sampling periods, into the matrix  $\mathbf{Y} = [\mathbf{y}^{(1)}, \dots, \mathbf{y}^{(L)}] \in \mathbb{C}^{n_{\mathbf{R}} \times L}$ .

This leads for  $\mathcal{H}_0$  to the model:

$$\mathbf{Y} = \sigma \mathbf{\Theta}. \quad (14)$$

And for  $\mathcal{H}_1$  to

$$\mathbf{Y} = [\mathbf{H}, \sigma \mathbf{I}_N] \begin{bmatrix} \mathbf{S} \\ \mathbf{\Theta} \end{bmatrix}. \quad (15)$$

We also denote by  $\mathbf{\Sigma}$  the covariance matrix

$$\mathbf{\Sigma} = E[\mathbf{Y}\mathbf{Y}^H] \quad (16)$$

$$= L(\mathbf{H}\mathbf{H}^H + \sigma^2 \mathbf{I}_{n_{\mathbf{R}}}) \quad (17)$$

$$= \mathbf{U}(\mathbf{\Lambda})\mathbf{U}^H, \quad (18)$$

where  $\mathbf{\Lambda} = \text{diag}(v_1 + \sigma^2, \dots, v_{n_R} + \sigma^2)$ , with  $\{v_i, i \in \{1, \dots, n_R\}\}$  the eigenvalues of  $\mathbf{H}\mathbf{H}^H$  and  $\mathbf{U}$  a certain unitary matrix.

Our intention is to make a decision on whether, given the received data matrix  $\mathbf{Y}$ , the probability for  $\mathcal{H}_1$  is greater than the probability for  $\mathcal{H}_0$ . This problem is usually referred to as *hypothesis testing* [11]. The decision criterion is based on the ratio

$$C(\mathbf{Y}) = \frac{P_{\mathcal{H}_1|\mathbf{Y}}(\mathbf{Y})}{P_{\mathcal{H}_0|\mathbf{Y}}(\mathbf{Y})}, \quad (19)$$

which we need to decide is whether lesser or greater than 1.

### 3.2.3. Results and Experiments

At this point in the derivation, computing  $C$  resorts to mere mathematical integration. The details of the calculus are given in [26]. We only provide here the results. First, assume  $\sigma^2$  and  $n_T$  are known, then, denoting  $x_1, \dots, x_{n_R}$  the eigenvalues of  $\mathbf{Y}\mathbf{Y}^H$ :

$$P_{\mathbf{Y}|\mathcal{H}_0}(\mathbf{Y}) = \frac{1}{(\pi\sigma^2)^{n_R L}} e^{-\frac{1}{\sigma^2} \text{tr} \mathbf{Y}\mathbf{Y}^H} \quad (20)$$

$$= \frac{1}{(\pi\sigma^2)^{n_R L}} e^{-\frac{1}{\sigma^2} \sum_{i=1}^{n_R} x_i} \quad (21)$$

and

$$P_{\mathbf{Y}|\mathcal{H}_1}(\mathbf{Y}) = \int_{\mathbf{\Sigma}} P_{\mathbf{Y}|\mathbf{\Sigma}, \mathcal{H}_1}(\mathbf{Y}, \mathbf{\Sigma}) P_{\mathbf{\Sigma}}(\mathbf{\Sigma}) d\mathbf{\Sigma} \quad (22)$$

$$= \int_{\mathcal{U}(n_R) \times (\mathbb{R}^+)^{n_R}} P_{\mathbf{Y}|\mathbf{\Sigma}, \mathcal{H}_1}(\mathbf{Y}, \mathbf{U}\mathbf{\Lambda}\mathbf{U}^H) P_{\mathbf{\Lambda}}(\mathbf{\Lambda}) d\mathbf{U} d\mathbf{\Lambda}, \quad (23)$$

which, after complete derivation, using in particular the Harish-Chandra identity [27], gives

$$P_{\mathbf{Y}|\mathcal{H}_1}(\mathbf{Y}) = \alpha \sum_{\mathbf{a} \subset [1, n_R]} \frac{e^{\frac{\sum_{i=1}^{n_T} x_{a_i}}{\sigma^2}}}{\prod_{\substack{j \neq a_1 \\ j \neq a_i}} (x_{a_i} - x_j)} \times \sum_{\mathbf{b} \in \mathcal{P}(n_T)} (-1)^{\text{sgn}(\mathbf{b})+1} \prod_{l=1}^{n_T} J_{n_R - L - 2 + b_l}(n_T \sigma^2, n_T x_{a_i}) \quad (24)$$

with  $\mathcal{P}(k)$  the ensemble of permutations of  $k$ ,  $\text{sgn}(\mathbf{b})$  the sign of the permutation  $\mathbf{b}$ :

$$J_k(x, y) = \int_x^{+\infty} t^k e^{-t - \frac{y}{t}} dt \quad (25)$$

and

$$\alpha = \frac{(n_R - n_T)! n_T^{(2L - n_T + 1)n_T / 2} e^{n_T^2 \sigma^2 - \frac{\sum_{i=1}^{n_R} x_i}{\sigma^2}}}{n_R! \pi^{n_R L} \sigma^{2(n_R - n_T)(L - n_T)} \prod_{j=1}^{n_T - 1} j!}. \quad (26)$$

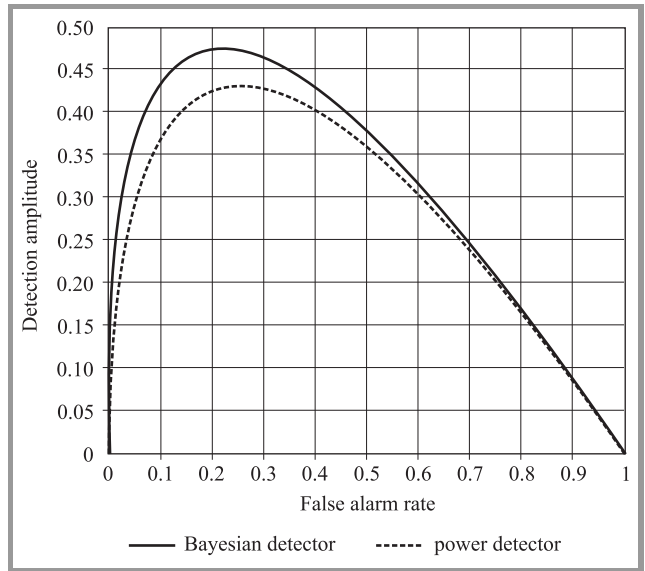


Fig. 1. Detection amplitude comparison in MIMO;  $M = 1, N = 8, L = 20, SNR = -10$  dB.

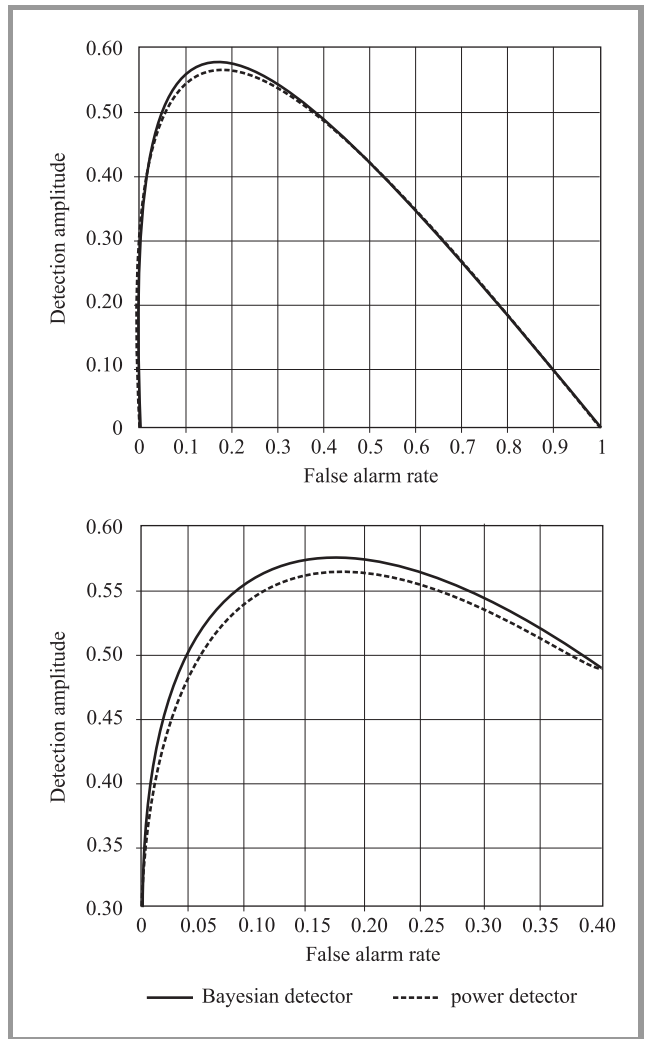


Fig. 2. Detection amplitude comparison in MIMO;  $M = 2, N = 8, L = 10, SNR = -10$  dB.

These expressions are rather complex but show that the Bayesian signal detection, within the state of knowledge  $I$ , only depends on the eigenvalues  $x_1, \dots, x_{n_R}$  of the Gram matrix  $\mathbf{Y}\mathbf{Y}^H$  of the received data  $\mathbf{Y}$ .

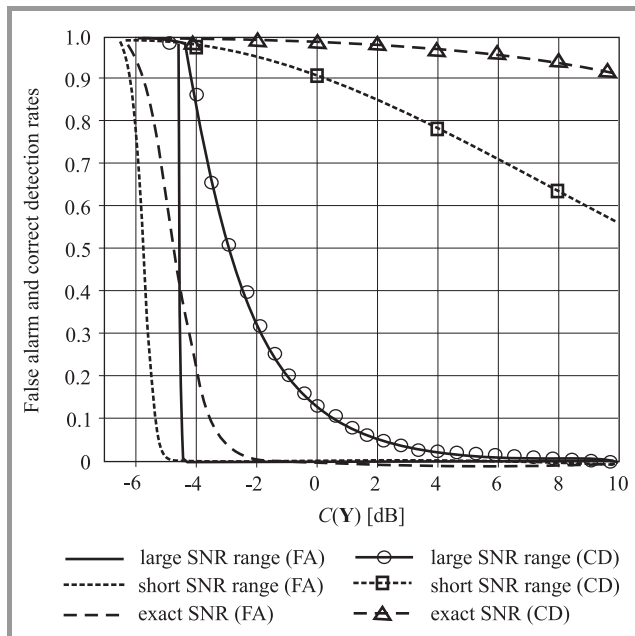
A comparison with the classical power detector, e.g., [28]–[30], which consists in summing all individual powers received on the antenna array is provided in Fig. 1. In the latter,  $n_T = 1$  and the comparison is made between the difference “correct detection rate minus false alarm rate” computed from Monte Carlo simulations for both Bayesian and classical signal detectors.

We observe a slight gain in performance due to the novel Bayesian detector. Especially, for a low false alarm rate (which is often demanded in practice), we observe a large gain in correct detection rate provided by the Bayesian detector. This statement is however only valid for  $n_T = 1$ . When  $n_T$  is larger, then the channel hardening effect reduces the gain of the Bayesian detector. This is shown in Fig. 2 in which  $n_T = 2$ .

Now, if the noise power  $\sigma^2$  is not perfectly known (this is classically the situation since knowledge of the noise power implies prior identification of the background noise), the probability distribution must be updated by marginalizing over  $\sigma^2$ , from the lower bound  $\sigma_-^2$  to the upper bound  $\sigma_+^2$  on  $\sigma^2$ . Therefore:

$$P_{Y|I} = \frac{1}{\sigma_+^2 - \sigma_-^2} \int_{\sigma_-^2}^{\sigma_+^2} P_{Y|\sigma^2, I}(\mathbf{Y}, \sigma^2) d\sigma^2, \quad (27)$$

which is too involved to compute, but can be numerically estimated. An example is provided in Fig. 3 in which



**Fig. 3.** False alarm (FA) and correct detection (CD) rates for exact, short discrete range ( $\{0, 2.5, 5\}$ ) and large discrete range ( $\{-5, -2.5, 0, 2.5, 5\}$ ) SNR;  $M = 1$ ,  $N = 8$ ,  $L = 10$ ,  $SNR = 2.5$  dB.

the intervals  $[\sigma_-^2, \sigma_+^2]$  are taken increasingly large. In the latter, correct detection rate against false alarm rate is de-

picted for different values of  $\sigma_-^2$  and  $\sigma_+^2$ . It is observed that the range of ensured correct detection gets increasingly narrower when  $[\sigma_-^2, \sigma_+^2]$  is large. Note that this situation cannot be compared against classical power detection methods which do not provide solutions when  $\sigma^2$  is not perfectly known.

## 4. Discussion

In addition to these first two studies on maximum entropy considerations for cognitive radios, the authors proposed more practical studies on maximum entropy orthogonal frequency division multiplexing (OFDM) channel estimation [31], maximum entropy carrier frequency offset estimation [32; Chapter 13], minimal update channel estimation [33], etc. From all those studies, we draw the following conclusions.

- Quite often, classical techniques, in particular in the channel estimation field, are rediscovered using MaxEnt. However, it is important to note that, even if the final formulas are the same in the classical and Bayesian MaxEnt approach, the philosophical conclusions are very different. Usually classical methods derive from empirical parameter settings, which could have been chosen differently, while Bayesian approaches give unique deterministic solutions, which stem from honesty in the treatment of prior information.
- The MaxEnt principle allows one to marginalize over all parameters when those are not perfectly known. As a consequence, while classical solutions are found anew, those methods can usually be extended to cope with the lack of information on some key variables. For instance, in the signal sensing proposed in Section 3 and completed in [26], the situations where noise variance and number of transmit antennas are not perfectly known can be easily handled, whereas classical methods stumble on these problems and solve them by empirical (possibly largely erroneous) parametrizations.

On the other hand, MaxEnt calculus and final solutions can turn very rapidly extremely mathematically involved, as exemplified by the final signal sensing formula in Section 3. This is a major problem, and the subject of most criticism towards Bayesian approaches. A missing part in these MaxEnt approach would be a systematic method which, from the general (very involved) solution, would provide approximate solutions. Quite remarkably, Caticha provides a vision of the maximum entropy principle, or more precisely the minimum cross entropy principle, which might help decide on the optimal approximation taken from a set of possible approximations [34]. These considerations might lead to such systematic approximation methods.

Another point of concern in the MaxEnt framework lies in the many integrals that may need to be computed when lit-



tle is known on the surrounding environment. With the increasing capabilities of modern computers, numerical approximations might help to compute those integrals, but these approximations would only be valid if not so many integrals are considered; two reasons explain this fact: first, the complexity increase due to additional integrals is exponential in the number of integrals and second, small errors in inner integrals tend to lead to large errors when integrated many times (this is often referred to as the curse of dimensionality).

As a consequence, while the first MaxEnt results provided by the authors show significant performance increase, many problems remain to be solved for cognitive radios to be fully intelligent, both on fundamental philosophical considerations (many questions raised in the introduction of the present paper are left unanswered) and on practical applications.

## 5. Conclusion

In this paper, we introduced the fundamentals of cognitive radios under a physical layer viewpoint. These fundamentals are based on the extension of Shannon's information theory to the Bayesian probability theory and the maximum entropy principle, which enables the cognitive devices with plausible (human-like) reasoning. Through the first-step studies of maximum entropy channel modeling and signal sensing, we paved the path to the establishment of strong theoretical grounds to the realm of cognitive radios.

## Acknowledgment

This work is partially supported by the European Commission in the framework of the FP7 Network of Excellence in Wireless Communications NEWCOM++.

## References

- [1] C. Shannon, "A mathematical theory of communications", *Bell Syst. Tech. J.*, vol. 27, no. 7, pp. 379–423, 1948.
- [2] G. Foschini and M. Gans, "On limits of wireless communications in a fading environment when using multiple antennas", *Wirel. Pers. Commun.*, vol. 6, no. 3, pp. 311–335, 1998.
- [3] E. Telatar, "Capacity of multi-antenna Gaussian channels", *Eur. Trans. Telecommun.*, vol. 10, no. 6, pp. 585–595, 1999.
- [4] J. Mitola III and G. Q. Maguire Jr, "Cognitive radio: making software radios more personal", *IEEE Pers. Commun.*, vol. 6, no. 4, pp. 13–18, 1999.
- [5] S. Haykin, "Cognitive radio: brain-empowered wireless communications", *IEEE J. Sel. Areas in Commun.*, vol. 23, no. 2, pp. 201–220, 2005.
- [6] A. Ghasemi and E. S. Sousa, "Collaborative spectrum sensing for opportunistic access in fading environments", in *First IEEE Int. Sympo. New Front. Dynam. Spectr. Acc. Netw.*, Baltimore, USA, 2005, pp. 131–136.
- [7] L. S. Cardoso, M. Kobayashi, Ø. Ryan, and M. Debbah, "Vandermonde frequency division multiplexing for cognitive radio", in *IEEE 9th Worksh. Sig. Proces. Adv. Wirel. Commun.*, Recife, Brazil, 2008, pp. 421–425.
- [8] C. Shannon, "Programming a computer for playing chess", *Phil. Mag.*, Ser. 7, vol. 41, no. 314, pp. 256–275, 1950.
- [9] C. Shannon, "Computers and automata", *Proc. IRE*, vol. 41, iss. 10, pp. 1234–1241, 1953.
- [10] C. Shannon, "A mind-reading machine", Bell Laboratories memorandum, 1953 (reprinted in *Claude Elwood Shannon Collected Papers*. New York: IEEE Press, 1993).
- [11] E. T. Jaynes, *Probability Theory: The Logic of Science*. Cambridge: Cambridge University Press, June 2003.
- [12] A. Caticha, "Maximum entropy, fluctuations and priors", in *Maximum Entropy and Bayesian Methods in Science and Engineering*, Ed. A. Mohammad-Djafari, vol. 568, no. 94, 2001, <http://arxiv.org/abs/math-ph/0008017>
- [13] A. Caticha, "Lectures on probability, entropy and statistical physics", 2008, <http://arxiv.org/abs/0808.0012> [physics.data-an]
- [14] L. Brillouin, *Science and Information Theory*. New York: Dover Publ., 1963.
- [15] E. T. Jaynes, "On the rationale of maximum-entropy methods", *Proc. IEEE*, vol. 70, no. 9, pp. 939–952, 1982.
- [16] E. T. Jaynes, "Information theory and statistical mechanics", *Phys. Rev.*, vol. 106, no. 4, pp. 620–630, 1957.
- [17] J. Shore and R. Johnson, "Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy", *IEEE Trans. Inform. Theory*, vol. 26, no. 1, pp. 26–37, 1980.
- [18] R. T. Cox, "Of inference and inquiry, an essay in inductive logic", in *Proceedings 1978 Maximum Entropy Formalism Conference*. Cambridge: MIT Press, 1979, pp. 119–167.
- [19] R. T. Cox, "Probability, frequency and reasonable expectation", *Amer. J. Phys.*, vol. 14, pp. 1–13, 1946.
- [20] K. H. Knuth, A. Caticha, J. L. Center, A. Giffin, and C. C. Rodriguez, "Lattice theory, measures and probability", in *27th Int. Worksh. Bayesian Inference and Maximum Entropy Methods in Science and Engineering, AIP Conf. Proc.*, Melville, USA, 2007, vol. 954, pp. 23–36.
- [21] H. Marko, "The bidirectional communication theory – a generalization of information theory", *IEEE Trans. Commun.*, vol. COM-21, no. 12, pp. 1345–1351, 1973.
- [22] J. Massey, "Causality, feedback and directed information", in *Proc. Int. Symp. Inform. Theory Appl.*, Waikiki, Hawaii, 1990.
- [23] M. Guillaud, M. Debbah, and A. L. Moustakas, "Maximum entropy MIMO wireless channel models", *IEEE Trans. Inform. Theory*, Dec. 2006, <http://arxiv.org/abs/cs.IT/0612101>
- [24] H. Jeffreys, "An invariant form for the prior probability in estimation problems", *Proc. Royal Soc. London, Ser. A, Math. Phys. Sci.*, vol. 186, no. 1007, pp. 453–461, 1946.
- [25] A. M. Tulino and S. Verdú, *Random Matrix Theory and Wireless Communications*. Hanover (USA): Now Publishers, 2004, vol. 1, iss. 1.
- [26] R. Couillet and M. Debbah, "Bayesian inference for multiple antenna cognitive receivers", *IEEE Wirel. Commun. Netwo. Conf.*, Budapest, Hungary, 2009.
- [27] A. B. Balantekin, "Character expansions, Itzykson-Zuber integrals, and the QCD partition function", *Phys. Rev. D*, vol. 62, no. 8, p. 085017, 2000.
- [28] H. Urkowitz, "Energy detection of unknown deterministic signals", *Proc. IEEE*, vol. 55, no. 4, pp. 523–531, 1967.
- [29] V. I. Kostylev, "Energy detection of a signal with random amplitude", in *Proc. IEEE Int. Conf. Commun. ICC'02*, New York, USA, 2002, pp. 1606–1610.
- [30] Z. Quan, S. Cui, A. H. Sayed, and H. V. Poor, "Spatial-spectral joint detection for wideband spectrum sensing in cognitive radio networks", in *Proc. ICASSP Conf.*, Las Vegas, USA, 2008.
- [31] R. Couillet and M. Debbah, "A maximum entropy approach to OFDM channel estimation", in *IEEE Sig. Proces. Adv. Wirel. Commun. Conf.*, Perugia, Italy, 2009.
- [32] L. Song, T. Jiang, and Y. Zhang, *Orthogonal Frequency Division Multiple Access (OFDMA)* to be published by Auerbach Publ., CRC Press, Taylor & Francis Group.
- [33] R. Couillet, A. Ancora and M. Debbah, "Minimal update for OFDM channel estimation" to be submitted.

- [34] C. Y. Tseng and A. Caticha, "Maximum entropy approach to the theory of simple fluids", 2003, Arxiv preprint cond-mat/0310746.



**Romain Couillet** was born in Abbeville, France. He received his Master of Science degree (specialization in mobile telecommunication) from the Eurecom Institute, France in 2007. He received his M.Sc. degree (specialization in communication systems) from the Telecom ParisTech, France, in 2007. In Sept. 2007, he joined NXP,

founded by Philips, which eventually merged into the ST-Ericsson company. In ST-Ericsson, he works as an algorithm development engineer on the Long Term Evolution Advanced (LTE-A) project. In parallel to his position in ST-Ericsson, he is currently pursuing his Ph.D. degree at the Supélec, France. His research interests include wireless communications, multi-user MIMO detection techniques, cognitive radio and random matrix theory. He is the recipient of the Valuetools 2008 best student paper award.

ST-Ericsson

505 route des Lucioles, 06650 Sophia Antipolis, France

e-mail: romain.couillet@supelec.fr

Alcatel-Lucent Chair, Supélec

3 rue Joliot Curie, 91192 Gif sur Yvette, France



**Mérouane Debbah** Mérouane Debbah was born in Madrid, Spain. He entered the Ecole Normale Supérieure de Cachan, France, in 1996, where he received his M.Sc. and Ph.D. degrees, respectively, in 1999 and 2002. From 1999 to 2002, he worked for the Motorola Labs on Wireless Local Area Networks and prospective fourth

generation systems. From 2002 until 2003, he was appointed Senior Researcher at the Vienna Research Center for Telecommunications (FTW), Austria, working on MIMO wireless channel modeling issues. From 2003 until 2007, he joined the Mobile Communications Department of the Institut Eurecom (Sophia Antipolis, France) as an Assistant Professor. He is presently a Professor at the Supélec (Gif-sur-Yvette, France), holder of the Alcatel-Lucent Chair on Flexible Radio. His research interests are in information theory, signal processing and wireless communications. He is the recipient of the 2005 Mario Boella Prize Award, the 2007 General Symposium IEEE GLOBECOM best paper award, the Wi-Opt 2009 best paper award as well as the Valuetools 2007, Valuetools 2008 and CrownCom 2009 best student paper awards. He is a WWRF fellow.

e-mail: merouane.debbah@supelec.fr

Alcatel-Lucent Chair, Supélec

3 rue Joliot Curie, 91192 Gif sur Yvette, France