Mathematical Foundations of Cognitive Radios

Romain Couillet, Merouane Debbah

Mathematical Foundations of Cognitive Radios

Romain Couillet
ST-Ericsson, Supélec
505 Route des Lucioles
06560 Sophia Antipolis, France
Email: romain.couillet@supelec.fr

Mérouane Debbah
Alcatel-Lucent Chair, Supélec
Plateau de Moulon, 3 rue Joliot-Curie
91192 Gif sur Yvette, France
Email: merouane.debbah@supelec.fr

***Abstract*—Recently, much interest has been directed towards software defined radios and embedded intelligence in telecommunication devices. However, no theoretical framework for cognitive radios has ever been proposed. In this paper, we introduce an information theoretic point of view on cognitive radios. 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?" or "when facing a situation in which most parameters are missing, how to react?". As an introduction, we will present two examples from the same authors using the cognitive framework namely multi-antenna channel modelling and signal sensing.**

I. Introduction

In 1948, Claude Shannon introduced a mathematical theory of communications [1], allowing two to three generations of researchers to design increasingly sophisticated telecommunication systems, whose purpose was to constantly increase the achievable transmission rate over various communication channels. One of the outcomes of Shannon's work showed that a linear increase in the transmission bandwidths provides a linear growth in the channel transmission capacity, while a linear transmit power increase provide logarithmic 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 could only be acquired by service providers at an extraordinarily high price. In the end of the nineties, the work of Foschini [2] and Telatar [3] 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 can be 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 provided a new research field in the early years of the twenty-first century. However, practical applications of multiple antenna systems is still questionnable, as the predicted multiplexing gain can appear at 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. Still, with these new dimensions, the classical Shannon Theory was still the landmark theory to analyze these systems.

Within 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 dedicated 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[1] allows for a potentially high increase of *spectral efficiency*, which we define here as the actual transmission rate averaged over the theoretical capacity. The ideas of Mitola recently motivated a wide range of research with a common denominator: the introduction of intelligence in wireless devices. For instance, Haykin [5] introduced 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 a 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. [29], [30].

However, all these ideas, revolutionary as they may seem, only scratch the surface of a larger entity known as 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[2], 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. Note the idea of smart devices date back to Shannon's work as well. He was already interested in ideas such as a robot capable of playing chess [6]; providing already in 1953 an original viewpoint of the cognitive abilities of future computers and even constructing a mind-reading machine, the circuitry of which is depicted in [8].

In this work, we propose to introduce an information theoretic viewpoint for cognitive radios, to enable human-

[1]In order limit useless overhead signaling.

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

like intelligence in wireless devices. This work does not rely on Shannon's work but tries to extend it through the use of additional mathematical such as the theory of Bayesian probabilities, the maximum entropy principle [9] and the minimal cross-entropy principle [13], [14], among others.

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

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 $(\cdot)^{\mathsf{H}}$. The operators $\mathrm{tr}\mathbf{X}$ and $|\mathbf{X}|$ represent the trace and determinant, respectively. The symbol $\mathrm{E}[\cdot]$ denotes expectation. The operator $\mathrm{vec}(\cdot)$ 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 $x = y$.

II. From Shannon to Jaynes

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

A. Channel Capacity Revisited

Let us consider the simplest communication scheme, modelled as

$$y = x + n \tag{1}$$

for some transmit signal x, additive background noise n and receive signal y. The Shannon capacity C of such a system can be computed as:

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

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 equality (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 equation (2).

Moreover, it is important to observe that what we call noise is in fact, in addition to the thermal noise, the sum contribution of interfering waves with different properties. If part of this noise can be analyzed by the cognitive device[3], 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' is capacity if the receiver knows exactly the noise variance which does not carry in itself any actual significance: as recalled by Jaynes [9] (see pp. 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"[4].

B. Limitations of Information Theory

Already in 1963, Leon Brillouin [22] realized the fundamental limitation of Shannon's information theory. In his own words, '*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.*" - Leon Brillouin, 1963.

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 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 this knowledge affects 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

[3]There is no reason why a cognitive device would not be able to infer on what the noise is made of.

[4]The system designer can be seen as a virtual entity sharing the knowledge of both transmitter and receiver.

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.

C. 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.

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. As a consequence, assuming a cognitive telecommunication device is not aware of the signal to noise radio related to 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 signal to noise ratio 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 [12] thanks to the introduction of the maximum entropy principle (MaxEnt) [11]. 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 [23] to be the entropy function H,

$$H(p) = -\int \log(p(t))p(t)dt \tag{3}$$

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 Section III-A.

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 σ^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 σ^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 [26], who also made important contributions on Bayesian probability theory [10], 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 [27], 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.

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[5]. 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 wiselly: the older the cognitive device, the more efficient.

Regarding for instance signal sensing, the first steps of which will be detailed in Section III, 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 coherent 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. Indeed, through the maximum entropy framework, one only uses the available information and does not take the risk to introduce unwanted assumptions, such as empirical values or imprecise estimates of unknown parameters. 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 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 [24], in the objective to accurately model the social interactions among animals and especially human beings. The lead was then followed by Massey [25] who extended information theory to include feedback in the expression of Shannon's mutual information.

III. 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 many 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 modelling, which is a necessary step to properly handle the Bayesian signal detection method.

A. MaxEnt Channel Modelling

1) Introduction: Channel modelling 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

[5]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.

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 specific 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.

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 modelling methodology. This is the case of Gaussian i.i.d. channels. Indeed, let us assume that the information I known to the cognitive device gathers the following,

1) the transmitter is equipped with n_{T} transmit antennas
2) the receiver is equipped with n_{R} receive antennas
3) the channel carries an energy E.

The transmission model is

$$\mathbf{y} = \sqrt{\frac{\rho}{n_{\mathsf{T}}}}\mathbf{H}\mathbf{x} + \mathbf{n} \tag{4}$$

where $\mathbf{x} \in \mathbb{C}^{n_{\mathsf{T}}}$ is the transmitted symbol vector, $\mathbf{n} \in \mathbb{C}^{n_{\mathsf{R}}}$ the thermal or interfering noise, ρ the signal to noise ratio (SNR) and $\mathbf{H} \in \mathbb{C}^{n_{\mathsf{T}} \times n_{\mathsf{R}}}$ the channel we want to model.

In mathematical terms, based on the fact that

$$\int d\mathbf{H} \sum_{i=1}^{n_{\mathsf{R}}} \sum_{j=1}^{n_{\mathsf{T}}} |h_{ij}|^2 P_{\mathbf{H}}(\mathbf{H}) = n_{\mathsf{T}} n_{\mathsf{R}} E \qquad \text{(Finite energy)} \tag{5}$$

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

what distribution $P_{\mathbf{H}}$[6] should the cognitive device assign to the channel? The cognitive device would like to derive the most general model complying with those constraints, in other words the one which maximizes 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 $P_{\mathbf{H}}$

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

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

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

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

which yields

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

with

$$P_{h_{ij}}(x) = e^{-(\gamma |x|^2 + \frac{\beta+1}{n_{\mathsf{R}} n_{\mathsf{T}}})} \tag{9}$$

One of the most important conclusions of the maximum entropy principle is that, while we have only assumed the knowledge about the energy, 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 the cognitive device has 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.

3) Other Channel Models: In [16], 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

$$\begin{aligned} P_{\mathbf{H}}(\mathbf{H}) &= \int P_{\mathbf{H},E}(\mathbf{H}, E) dE && (10) \\ &= \int P_{\mathbf{H}|E}(\mathbf{H}) P_E(E) dE && (11) \end{aligned}$$

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

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

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.

[7]The assignment of uniform priors on variables defined on a continuous space is a very controversial point of the maximum entropy theory, which is lenghtly discussed in [9]. Another classically used prior, which solves the problem of invariance to variable change is the so-called Jeffreys' uninformative prior [15].

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

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

If $\mathbf{Q}$ is unknown, which would often be the case, one needs to integrate out the *nuisance* parameter $\mathbf{Q}$. As a result of this integration (the detailed proof is provided in [16]), we show that the distribution of $\mathbf{H}$ only depends on the norm $x = \mathrm{vec}(\mathbf{H})^{\mathsf{H}}\mathrm{vec}(\mathbf{H})$, and that

$$P_x(x) = (-1)^{-N^2-1} \sum_{n=1}^{N} f_n(x) \frac{(-\gamma)^{N+n-1}}{[(n-1)!]^2(N-n)!} \quad (14)$$

where

$$f_i(x) = 2\left(\frac{x}{\gamma}\right)^{(i+N-2)/2} K_{i+N-2}(2(\gamma x)^{1/2}) \quad (15)$$

with $-\gamma = E/N^2$, K_j the Bessel-K function of order j.

B. Signal Detection

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

1) Channel State Information: Let us first state the information available at the cognitive receiver:

- S-i) the receiver has n_{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 that this hypothesis is very weak and should be made more accurate for communications schemes that are known only to use either QPSK, 16-QAM, 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_{T} antennas.
- V-ii) the noise variance σ^2 is known.

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 modelled as i.i.d. Gaussian [9]. The data vector, at time $l \in \{1, \ldots, L\}$, is denoted $\mathbf{s}^{(l)} = (s_1^{(l)}, \ldots, s_{n_{\mathsf{T}}}^{(l)})^{\mathsf{T}} \in \mathbb{C}^{n_{\mathsf{T}}}$. The data vectors are stacked into the receive matrix $\mathbf{S} = [\mathbf{s}^{(1)}, \ldots, \mathbf{s}^{(L)}]$.

If the noise level σ^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* $\boldsymbol{\Theta} \in \mathbb{C}^{n_{\mathsf{R}} \times L}$ (i.e. a matrix with i.i.d. standard complex Gaussian entries θ_{ij}) [28]. Under $\mathcal{H}_1$, the channel matrix is denoted $\mathbf{H} \in \mathbb{C}^{n_{\mathsf{R}} \times n_{\mathsf{T}}}$ 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 modelling rules. In the present situation, only the constant mean power (or equivalently, the energy) of the channel is known. Therefore $\mathbf{H}$ will be modelled as i.i.d. Gaussian, following the reasoning in the previous section. The received data at sampling time l are given by the $n_{\mathsf{T}} \times 1$ vector $\mathbf{y}^{(l)}$ that we stack, over the L sampling periods, into the matrix $\mathbf{Y} = [\mathbf{y}^{(1)}, \ldots, \mathbf{y}^{(L)}] \in \mathbb{C}^{n_{\mathsf{R}} \times L}$.

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

$$\mathbf{Y} = \sigma\boldsymbol{\Theta} \quad (16)$$

And for $\mathcal{H}_1$ to

$$\mathbf{Y} = [\mathbf{H}\ ,\ \sigma\mathbf{I}_N] \begin{bmatrix} \mathbf{S} \\ \boldsymbol{\Theta} \end{bmatrix} \quad (17)$$

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

$$\begin{aligned} \boldsymbol{\Sigma} &= \mathrm{E}[\mathbf{Y}\mathbf{Y}^{\mathsf{H}}] && (18) \\ &= L\left(\mathbf{H}\mathbf{H}^{\mathsf{H}} + \sigma^2\mathbf{I}_{n_{\mathsf{R}}}\right) && (19) \\ &= \mathbf{U}\left(L\boldsymbol{\Lambda}\right)\mathbf{U}^{\mathsf{H}} && (20) \end{aligned}$$

where $\boldsymbol{\Lambda} = \mathrm{diag}\left(\nu_{\mathbf{1}} + \sigma^{\mathbf{2}}, \ldots, \nu_{\mathbf{n_{\mathsf{R}}}} + \sigma^{\mathbf{2}}\right)$, with $\{\nu_i, i \in \{1, \ldots, n_{\mathsf{R}}\}\}$ the eigenvalues of $\mathbf{H}\mathbf{H}^{\mathsf{H}}$ 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* [9]. The decision criterion is based on the ratio

$$C(\mathbf{Y}) = \frac{P_{\mathcal{H}_1|\mathbf{Y}}(\mathbf{Y})}{P_{\mathcal{H}_0|\mathbf{Y}}(\mathbf{Y})} \quad (21)$$

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

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 [17]. We only provide here the results. First, assume σ^2 and n_{T} are known, then, denoting $x_1, \ldots, x_{n_{\mathsf{R}}}$ the eigenvalues of $\mathbf{Y}\mathbf{Y}^{\mathsf{H}}$,

$$\begin{aligned} P_{\mathbf{Y}|\mathcal{H}_0}(\mathbf{Y}) &= \frac{1}{(\pi\sigma^2)^{n_{\mathsf{R}}L}} e^{-\frac{1}{\sigma^2}\mathrm{tr}\mathbf{Y}\mathbf{Y}^{\mathsf{H}}} && (22) \\ &= \frac{1}{(\pi\sigma^2)^{n_{\mathsf{R}}L}} e^{-\frac{1}{\sigma^2}\sum_{i=1}^{n_{\mathsf{R}}} x_i} && (23) \end{aligned}$$

and

$$P_{\mathbf{Y}|\mathcal{H}_1}(\mathbf{Y})$$
$$= \int_{\mathbf{\Sigma}} P_{\mathbf{Y}|\mathbf{\Sigma}\mathcal{H}_1}(\mathbf{Y}, \mathbf{\Sigma}) P_{\mathbf{\Sigma}}(\mathbf{\Sigma}) d\mathbf{\Sigma} \tag{24}$$
$$= \int_{\mathcal{U}(n_{\mathsf{R}})\times(\mathbb{R}^+)^{n_{\mathsf{R}}}} P_{\mathbf{Y}|\mathbf{\Sigma},\mathcal{H}_1}(\mathbf{Y}, \mathbf{U}\mathbf{\Lambda}\mathbf{U}^{\mathsf{H}}) P_{\mathbf{\Lambda}}(\mathbf{\Lambda}) d\mathbf{U} d\mathbf{\Lambda} \tag{25}$$

which, after complete derivation, using in particular the Harish-Chandra identity [18], gives

$$P_{\mathbf{Y}|\mathcal{H}_1}(\mathbf{Y}) = \alpha \sum_{\mathbf{a}\subset[1,n_{\mathsf{R}}]} \frac{e^{\frac{\sum_{i=1}^{n_{\mathsf{T}}} x_{a_i}}{\sigma^2}}}{\prod_{a_i} \prod_{\substack{j\neq a_1 \\ \cdots \\ j\neq a_i}} (x_{a_i} - x_j)}$$
$$\times \sum_{\mathbf{b}\in\mathcal{P}(n_{\mathsf{T}})} (-1)^{\mathrm{sgn}(\mathbf{b})+1} \prod_{l=1}^{n_{\mathsf{T}}} J_{n_{\mathsf{R}}-L-2+b_l}(n_{\mathsf{T}}\sigma^2, n_{\mathsf{T}} x_{a_l}) \tag{26}$$

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

$$J_k(x,y) = \int_x^{+\infty} t^k e^{-t-\frac{y}{t}} dt \tag{27}$$

and

$$\alpha = \frac{(n_{\mathsf{R}} - n_{\mathsf{T}})! n_{\mathsf{T}}^{(2L-n_{\mathsf{T}}+1)n_{\mathsf{T}}/2} e^{n_{\mathsf{T}}{}^2\sigma^2 - \frac{\sum_{i=1}^{n_{\mathsf{R}}} x_i}{\sigma^2}}}{n_{\mathsf{R}}! \pi^{n_{\mathsf{R}} L} \sigma^{2(n_{\mathsf{R}}-n_{\mathsf{T}})(L-n_{\mathsf{T}})} \prod_{j=1}^{n_{\mathsf{T}}-1} j!} \tag{28}$$

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

A comparison with the classical power detector, e.g. [19], [20], [21], which consists in summing all individual powers received on the antenna array is provided in Figure 1. In the latter, $n_{\mathsf{T}} = 1$ 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 $n_{\mathsf{T}} = 1$. When n_{T} is larger, then the channel hardening effect reduces the gain of the Bayesian detector. This is shown in Figure 2 in which $n_{\mathsf{T}} = 2$.

Now, if the noise power σ^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 σ^2, from the lower bound σ_-^2 to the upper bound σ_+^2 on σ^2. Therefore,

$$P_{\mathbf{Y}|I} = \frac{1}{\sigma_+^2 - \sigma_-^2} \int_{\sigma_-^2}^{\sigma_+^2} P_{\mathbf{Y}|\sigma^2,I}(\mathbf{Y}, \sigma^2) d\sigma^2 \tag{29}$$

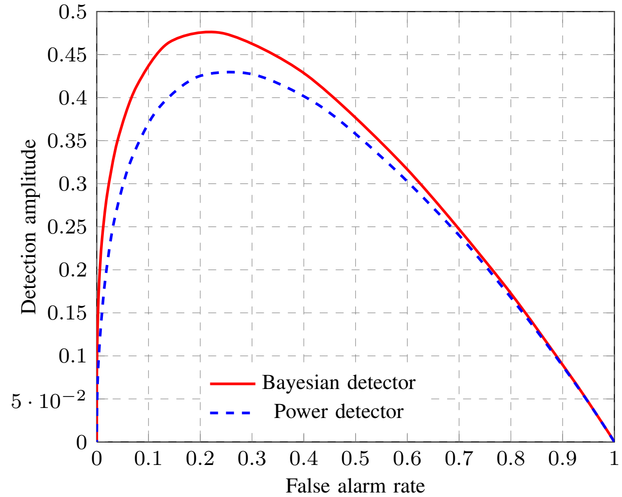

Fig. 1. Detection amplitude comparison in SIMO - $M = 1$, $N = 8$, $L = 20$, $\mathrm{SNR} = -10$ dB

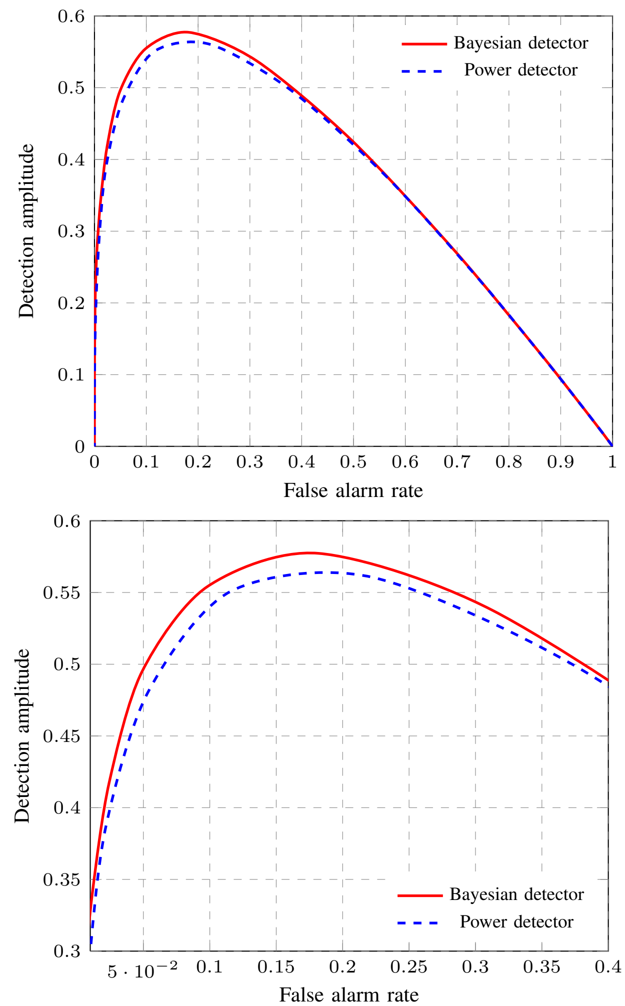

Fig. 2. Detection amplitude comparison in MIMO - $M = 2$, $N = 8$, $L = 10$, $SNR = -10$dB

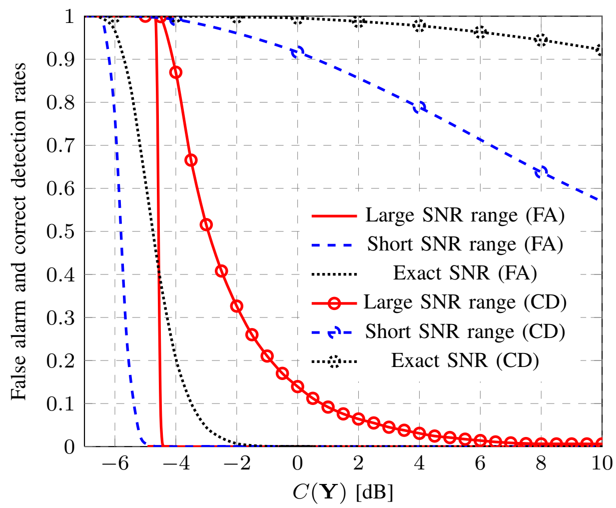

Fig. 3. False alarm (FA) and correct detection (CD) rates for exact, short discrete range ($\{0, 2.5, 5\}$) and large discrete range ($\{-5, -2.5, 0, 2.5, 5\}$) SNR - $M = 1$, $N = 8$, $L = 10$, SNR $= 2.5$ dB

which is too involved to compute, but can be numerically estimated. An example is provided in Figure 3 in which the intervals $[\sigma_-^2, \sigma_+^2]$ are taken increasingly large. In the latter, correct detection rate against false alarm rate is depicted for different values of σ_-^2 and σ_+^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 σ^2 is not perfectly known.

IV. Discussion

In addition to these first two studies on maximum entropy considerations for cognitive radios, the authors proposed more practical studies on maximum entropy OFDM channel estimation [32], maximum entropy carrier frequency offset estimation [33], minimal update channel estimation [34] 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 new, 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 III and completed in [17], 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 III. 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 [31]. 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 little 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).

Also, as exemplified in the previous sections, MaxEnt can treat problems where prior information is of statistical nature, e.g. knowledge of certain moments. However, when it comes to deterministic knowledge, e.g. the position of two buildings facing the signal transmitter/receiver, the determination of the maximum entropy distribution under this type of constraints is difficult to formulate mathematically; for the time being, no systematic method, in the same trend as the Lagrangian multipliers, is known to determine these MaxEnt distributions.

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.

V. Conclusion

In this paper, we introduced a theoretical framework for cognitive radios. These fundamentals notions are based on the extension of Shannon's information theory to the Bayesian probability theory and the maximum entropy principle, which enable the cognitive devices with plausible (human-like) reasoning. These preliminary results (maximum entropy channel modelling and signal detection) are the first steps towards a general mathematical framework for cognitive radios [35].

VI. 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 System Technical Journal, 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,” Wireless Personal Communications, vol. 6, no. 3, pp. 311-335, 1998.

[3] E. Telatar, “Capacity of multi-antenna Gaussian channels,” European transactions on telecommunications, vol. 10, no. 6, pp. 585-595, 1999.

[4] Mitola III, J., and Maguire Jr, GQ, “Cognitive radio: making software radios more personal”, Personal Communications, IEEE [see also IEEE Wireless Communications] 6(4), 13-18, 1999.

[5] S. Haykin, “Cognitive Radio: Brain-Empowered Wireless Communications”, IEEE Journal on Selected areas in Comm. , vol 23., no. 2, Feb 2005.

[6] C. Shannon, “Programming a Computer for Playing Chess,” Philosophical Magazine, Series 7, Vol. 41 (No. 314, March 1950), pp. 256-275

[7] C. Shannon, “Computers and Automata,” C. Shannon, Proceedings of the IRE, Vol. 41, Issue 10, Oct. 1953 Pages:1234-1241

[8] C. Shannon, “A Mind-Reading Machine,” Bell Laboratories memorandum, 1953. Reprinted in the Collected Papers of Claude Elwood Shannon, IEEE Press.

[9] E. T. Jaynes, “Probability Theory: The Logic of Science”, Cambridge University Press, June 2003.

[10] R. T. Cox, “Probability, frequency and reasonable expectation”, American journal of physics, vol. 14 (1946), pp. 1-13.

[11] E. T. Jaynes, “Information theory and statistical mechanics,” Physical review, vol. 106, no. 4, pp. 620-630, 1957.

[12] E. T. Jaynes, “On the rationale of maximum-entropy methods”, Proc. IEEE, vol. 70, no. 9, pp. 939-952, 1982.

[13] A. Caticha, “Maximum Entropy, fluctuations and priors”, A.C., in Maximum Entropy and Bayesian Methods in Science and Engineering, ed. by A. Mohammad-Djafari, A.I.P. Vol. 568, 94 (2001); arXiv.org/abs/math-ph/0008017.

[14] A. Caticha, “Lectures on Probability, Entropy and Statistical Physics”, arXiv:0808.0012v1 [physics.data-an], 2008.

[15] H. Jeffreys, “An Invariant Form for the Prior Probability in Estimation Problems”, Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences 186 (1007): 453-461, 1946.

[16] M. Guillaud, M. Debbah, A. L. Moustakas, “Maximum Entropy MIMO Wireless Channel Models”, Submitted to IEEE Trans. Information Theory Dec. 2006, http://arxiv.org/abs/cs.IT/0612101

[17] R. Couillet and M. Debbah, “Bayesian Inference for Multiple Antenna Cognitive Receivers,” IEEE Wireless Communication and Networking Conference, Budapest, April 2009.

[18] A. B. Balantekin, “Character expansions, Itzykson-Zuber integrals, and the QCD partition function”, *Phys. Rav. D,* vol. 62, no. 8, Oct. 2000.

[19] H. Urkowitz, “Energy detection of unknown deterministic signals,” Proc. of the IEEE, vol. 55, no. 4, pp. 523-531, Apr. 1967

[20] V. I. Kostylev, “Energy detection of a signal with Random Amplitude”, Proc IEEE Int. Conf. on Communications (ICC’02). New York City, pp. 1606-1610, May 2002.

[21] Z. Quan, S. Cui, A. H. Sayed, and H. V. Poor, “Spatial-spectral joint detection for wideband spectrum sensing in cognitive radio networks,” Proc. ICASSP, Las Vegas, April 2008.

[22] L. Brillouin, “Science and information theory,” Dover Publications, 1963.

[23] J. Shore and R. Johnson, “Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy,” IEEE Trans. on Information Theory, vol. 26, no. 1, pp. 26-37, 1980.

[24] H. Marko, “The Bidirectional Communication Theory - A Generalization of Information Theory,” IEEE Trans. on Communications, vol. COM-21, no. 12, 1973.

[25] J. Massey, “Causality, Feedback and Directed Information,” Proceedings of International Symposium on Information Theory and its Applications, Waikiki, Hawaii, 1990.

[26] R. T. Cox, “Of inference and inquiry, an essay in inductive logic,” The maximum entropy formalism, pp. 119-167, 1978.

[27] K.H. Knuth, A. Caticha, J.L. Center, A. Giffin, C.C. Rodriguez, “Lattice theory, measures and probability,” Bayesian Inference and Maximum Entropy Methods in Science and Engineering, Saratoga Springs, NY, USA, 2007, AIP Conference Proceedings 954, American Institute of Physics, Melville NY, 23-36.

[28] A. M. Tulino, S. Verdú, “Random Matrix Theory and Wireless Communications”, Now Publishers, vol. 1, Issue 1, 2004.

[29] A. Ghasemi and E. S. Sousa, “Collaborative spectrum sensing for opportunistic access in fading environments,” First IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks, pp. 131-136, 2005.

[30] L. S. Cardoso, M. Kobayashi, Ø. Ryan and M. Debbah, “Vandermonde frequency division multiplexing for cognitive radio,” IEEE 9th Workshop on Signal Processing Advances in Wireless Communications, pp. 421-425, 2008.

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

[32] R. Couillet and M. Debbah, “A maximum entropy approach to OFDM channel estimation”, *submitted to IEEE Signal Processing Advances in Wireless Communications Conference, Perugia, 2009*.

[33] R. Couillet and M. Debbah, “Information theoretic approach to synchronization: the OFDM carrier frequency offset example”, *submitted to IEEE Signal Processing Advances in Wireless Communications Conference, Perugia, 2009*.

[34] R. Couillet, A. Ancora and M. Debbah, “Minimal Update for OFDM Channel Estimation,” *to be submitted*

[35] R. Couillet and M. Debbah, “A Mathematical Theory of Cognitive Radios,” *in preparation*