

# Interference Constraint Active Learning with Uncertain Feedback for Cognitive Radio Networks

Anestis Tsakmalis, *Student Member, IEEE*, Symeon Chatzinotas, *Senior Member, IEEE*,  
and Björn Ottersten, *Fellow, IEEE*

**Abstract**—In this paper, an intelligent probing method for interference constraint learning is proposed to allow a centralized Cognitive Radio Network (CRN) access the frequency band of a Primary User (PU) in an underlay cognitive communication scenario. The main idea is that the CRN probes the PU and subsequently eavesdrops the reverse PU link to acquire the binary ACK/NACK packet. This feedback is implicit channel state information (CSI) of the PU link, indicating whether the probing-induced interference is harmful or not. The intelligence of this sequential probing process lies in the selection of the power levels of the Secondary Users (SUs) which aims to minimize the number of probing attempts, a clearly Active Learning (AL) procedure, and expectantly the overall PU QoS degradation. The enhancement introduced in this work is that we incorporate the probability of each feedback being correct into this intelligent probing mechanism by using a multivariate Bayesian AL method. This technique is inspired by the Probabilistic Bisection Algorithm (PBA) and the deterministic Cutting Plane Methods (CPMs). The optimality of this multivariate Bayesian AL method is proven and its effectiveness is demonstrated through numerical simulations. Computationally cheap CPM adaptations are also presented which outperform existing AL methods.

**Keywords**—Cognitive Radio, Bayesian Active Learning, Probabilistic Bisection Algorithm, Cutting Plane Methods

## I. INTRODUCTION

Current telecommunication networks have reached a critical point of their lifespan. In terms of spectrum usage efficiency, some of these networks do not fully exploit the capacity of their frequency bands as they still rely on old communication technologies. In order to deal with spectrum underutilization without replacing the costly infrastructure of these legacy systems, the research community has proposed the deployment of unlicensed intelligent radio devices which can detect access opportunities in the legacy system bands and take advantage of them for their own service demands. This new kind of radio is called in literature Cognitive Radio (CR) and promotes a flexible spectrum usage by sensing, understanding, adapting and interacting with its surrounding environment [1], [2].

This work was supported by the National Research Fund of Luxembourg under the CORE projects "SeMiGod: SpEctrum Management and Interference mitigation in cOgnitive raDio satellite networks" and "SATSENT: SATellite SEnsor NeTworks for spectrum monitoring" and by the European Union under the HORIZON 2020 research project "SANSa: Shared Access terrestrial-satellite backhaul Network enabled by Smart Antennas".

The authors are with the Interdisciplinary Centre for Security, Reliability and Trust (SnT), University of Luxembourg. Email: {anestis.tsakmalis, symeon.chatzinotas, bjorn.ottersten}@uni.lu.

In the CR framework, the coexistence scenarios of a legacy network and a CRN, also referred to as PU and SU systems respectively, are divided in three categories: underlay, overlay and interweave [3]. In this paper, we focus on an underlay CR scenario, where SUs may transmit in a PU frequency band as long as the induced to the PU receiver interference is under a certain threshold. In general, the underlay approach is related to Power Control (PC) or Beamforming (BF) problems where the CR users must intelligently select their transmit power levels or beamforming vectors to optimize some operation metric and satisfy the PU interference constraint. In all these scenarios, an architecture suggestion for the deployed CRN could be the CR users to be coordinated by a Cognitive Base Station (CBS) using a dedicated control channel [4] which denotes a centralized structure and is more applicable than a decentralized CRN where CR users are partially independent and pass messages among each other.

An essential piece of information of these PC or BF problems regarding the PU interference constraint is accurate Channel State Information (CSI) of the interference channels. However, the legacy system (PU) was not originally designed to exchange any information, hence the two networks are not able to directly communicate. This indicates that no feedback about the induced to the PU interference can be transmitted to the CRN (SUs) in order to infer the interference channel gains. Since no cooperation between the PU and SU systems is expected, the CRN must somehow learn this interference CSI once it is deployed. In the CR context though, a common approach to overcome this issue is the CRs to use the PU reverse link feedback, check how this changes because of the CR operation and thus calculate the SU-to-PU channel gains in a sequential manner. This iterative procedure is clearly a probing scheme which combines carefully selecting the CR transmitting parameters and eavesdropping the PU reverse link feedback. Capturing and exploiting this feedback bridges the gap of PU and SU system segregation and enables learning in the CRN part. In previous work, this was obtained from the binary ACK/NACK packet of the reverse PU link [5], [6] for underlay PC or BF problems.

Another kind of feedback introduced in [7], [8] is the Modulation and Coding Classification (MCC) information. Assuming that the PU link operates under an Adaptive Coding and Modulation (ACM) protocol, whenever the PU link quality deteriorates due to CR induced interference, the PU changes its Modulation and Coding Scheme (MCS) to a more robust one. If the CR is equipped with an MCC module, detecting this MCS transition is feasible and in fact easier than decoding a

PU ACK/NACK packet. The most crucial benefit of exploiting the MCC feedback instead of the ACK/NACK one is that the MCC feedback in contrast with the binary ACK/NACK packet delivers more than 1-bit information and therefore speeds up the learning process as demonstrated in [8]. Nevertheless, in any learning process utilizing either MCC or ACK/NACK feedback there is a crucial issue which rises from classifying the PU signal to an MCS or decoding the PU message to obtain the ACK/NACK feedback. This problem is linked to the low SINR conditions of the sensed PU signal on the CRN side which may occur and make the classification or the decoding unreliable. Therefore, if a learning mechanism which exploits a sensed feedback is to be applied, it should also incorporate the uncertainty of the feedback resulting from realistic SINR conditions.

In this paper, the MCC approach used in our previous work [8] is not adopted and the binary ACK/NACK packet is chosen as the learning facilitator. The main reason for doing so is that the advantages of the MCC feedback were exhibited in [7], [8] and also it was clearly shown that any cognitive scenario considering a binary feedback of the PU link quality can perform even better with the multilevel MCC feedback. Therefore, in this paper we utilize this rudimentary binary feedback, the ACK/NACK packet, in order to focus on developing more sophisticated learning mechanisms.

#### A. Contributions

Herein, an AL probing method suitable for centrally organized CRNs is demonstrated which rapidly estimates the interference channel gains from multiple SU transmitters to a PU receiver. This case study assumes that the PU link is operating under a communications protocol where the receiver sends an ACK/NACK packet to the transmitter to acknowledge positively or negatively the receipt of messages. A common practice in the CR regime which is adopted here as well is the CRN to capture this packet from the PU feedback link and exploit it to learn the SU-to-PU channel gains. In this scenario, obtaining this binary feedback takes place in the CBS using a sensing antenna and a PU feedback packet decoder. This piece of information is utilized to implement a sequential probing technique where the SUs constantly adjust their transmit power levels according to CBS directives and monitor whether the ACK/NACK packet changes state. This intelligent probing design aims to minimize the number of probing attempts so that once the CRN is deployed in the PU system's environment, it may quickly learn the interference channels and then optimize its operation while satisfying the PU interference constraint which depends on the SU-to-PU channel gains.

Nevertheless, when utilizing the binary feedback a practical consideration must be taken into account. Due to low SINR conditions of the sensed PU signal by the CBS sensing antenna, the ACK/NACK packet decoding may be imperfect. Therefore, this feedback uncertainty which is expressed quantitatively by the probability of correct decoding ( $P_{cd}$ ) must also be included in the AL probing mechanism. In this work, a multivariate Bayesian AL method based on the univariate

PBA and the deterministic CPMs is implemented in order to include  $P_{cd}$  within this AL framework.

In summary, this paper delivers specifically the following major contributions:

- The novel construction of a multivariate Bayesian AL method designed for probing the PU and learning fast interference channel gains.
- An optimality proof is provided for the proposed multivariate Bayesian AL method.
- A computationally cheap and analytical CPM adaptation is given as a Bayesian AL technique suitable for high dimensional problems.
- Simulations show convergence rates for our multivariate Bayesian AL method and the cheap CPM adaptation faster than the ones of the method developed in [9] and the Probit Maximum Likelihood (ML) approach of [6] which is used as a benchmark technique.
- Results are given about the PU QoS degradation during all the examined AL methods in order to empirically prove that the faster an AL method is, the more protective it is to the PU link.

#### B. Structure

The remainder of this paper is structured as follows: Section II reviews in detail prior work related to cognitive scenarios using an uncertain PU link feedback. Section III provides the system model and the problem formulation. Section IV presents a set of multivariate Bayesian AL methods for interference channel gain learning and studies the optimality of the proposed technique. Section V elaborates on computational techniques necessary for the implementation of multivariate Bayesian AL. In Section VI, the simulation results which are obtained from the application of the proposed techniques are shown and compared with existing methods. Section VII gives the concluding remarks and future work in this topic. Finally, to improve the readability of the paper, we provide the descriptions of the most frequent abbreviations in Table I.

TABLE I: Definitions of Abbreviations

Abbreviation	Description
PU	Primary User
SU	Secondary User
CBS	Cognitive Base Station
CRN	Cognitive Radio Network
AL	Active Learning
PBA	Probabilistic Bisection Algorithm
CPM	Cutting Plane Method
MCMC	Markov Chain Monte Carlo

## II. RELATED WORK

Previous works in the field of cognitive underlay methods using rudimentary feedback have focused on PC and BF scenarios with different assumptions, protocols, system models and constraints. Assuming no cooperation between the CRN and the PU system, this feedback is acquired by most commonly eavesdropping the PU reverse link channel and decoding the PU ACK/NACK packet or by sensing the

PU signal and applying MCC in order to track the PU MCS change [8]. The general form of these underlay CR scenarios is the optimization of an SU system metric, such as total CRN throughput, worst SU throughput or SU SINR, subject to QoS constraints for PUs, like SINR, data rate or outage probability [3] which the CRN needs to learn. Hence, these study cases involve both learning PU constraints and solving an optimization problem which may be tackled in a centralized manner by a central decision maker or in a decentralized way by each SU individually. Most of the learning techniques are based on a simple iterative scheme of probing the PU system and acquiring the feedback indicating how the PU operation is changed. An additional discrimination of these problems is based on the reliability of the feedback itself which in many cases is questionable and introduces uncertainty into the learning and optimization procedures.

First, we describe the group of these scenarios which incorporate perfect feedback knowledge. A stochastic approximation algorithm is exploited in [10] for distributed BF which exhibits slow convergence rate as it basically is a random exploration technique. In [5], the one-bit null space learning algorithm (OBNSLA) is developed, which essentially is a blind realization of the Cyclic Jacobi Technique, in order to learn the null space of the interference channel matrix in a MIMO underlay cognitive scenario. Finally, a sign algorithm is established by the authors of [11] for transmit BF using 1-bit feedback to coarsely update the antenna weights in an LMS-like manner. It is worth noting that in the latter work, the case of feedback error is discussed but not addressed extensively.

In the uncertain feedback problems studied by the research community, the notable work in [12] has considered the centralized weighted sum-rate maximization topic under average SU power and probabilistic PU interference constraints. In this study, the optimization objective is achieved only after the interference channel gain learning process is terminated, a very common tactic for handling the aforementioned learning and optimization general structure of these problems. In its learning part, the recursive Bayesian estimation is employed by using imperfect CSI feedback which may potentially be as elementary as the binary ACK/NACK packet. Furthermore, significant work has been conducted in this area by tackling the uncertain feedback within a Partially Observable Markov Decision Process framework [13], [14] where uncertainty is introduced with a belief factor related to the reliability of the feedback information. In [13], a binary Spectrum Sensing feedback has been used to enable CRs apply a Reinforcement Learning procedure, the Q-Learning, to regulate the aggregated interference to the PU and in [14] a distributed channel admission solution is formulated based on Dynamic Programming, while in their previous work [15] a SU power control policy is also included using the same formulation but without elaborating on the belief factor enhancement. In [16], a methodical overview of all the Reinforcement Learning applications in CRNs based on the Markov Decision Process framework is provided. Additionally, the authors of [6] proposed a CPM based learning algorithm where probing the PU system targets to both learning interference channel matrices and maximizing the SNR at the SU receiver side in an underlay cognitive BF

scenario. In this work, the feedback error follows a Gaussian model and an ML AL approach is proposed, but without any theoretical convergence guarantees.

In this paper, we rely on the AL rationale of [6] applied in the underlay PC problem by using the ACK/NACK feedback and we focus only on learning the unknown interference channel gains without optimizing any SU system metric. The investigated scenario considers a centralized CRN where SUs are coordinated by a CBS using a dedicated control channel which usually as a structure exhibits faster learning and adaptation rates than the decentralized approach. We propose a multivariate Bayesian AL method for intelligent probing which incorporates the probability of each feedback being correct, prove its convergence optimality and demonstrate its effectiveness compared to other AL methods through numerical simulations.

### III. SYSTEM MODEL AND PROBLEM FORMULATION

Initially, let us consider a PU link and  $N$  SU links existing in the same frequency band as shown in Fig. 1. Furthermore, a Frequency Division Multiple Access (FDMA) method allows SU links to operate in separate sub-bands of the PU frequency band and without interfering with each other, but still aggregate causing interference to the PU system. The structure of the CRN is a centralized one where the SUs are dictated their operational parameters and coordinated by the CBS using a dedicated control channel. The examined scenarios in this study are considering the PU, the sensing and the unknown interference channels to follow the quasi static block fading model which applies for telecommunication links such as the satellite or the backhauling ones, but not for mobile ones where channels change rapidly. Here, we focus on channel power gains  $g$ , which in general are defined as  $g = \|c\|^2$ , where  $c$  is the complex channel gain. From this point, we will refer to channel power gains as channel gains.

As far as the interference to the PU link is concerned, this is caused by the transmitter part of each SU link to the receiver of the PU link. Taking into account that the SU links transmit solely in the PU frequency band, the aggregated interference on the PU side is defined as:

$$I_{PU} = \mathbf{g} \mathbf{p}^T \quad (1)$$

where  $\mathbf{g}$  is the unknown interference channel gain vector  $[g_1, \dots, g_N]$  with  $g_i$  being the  $SU_i$ -to-PU interference channel gain and  $\mathbf{p}$  is the SU power vector  $[p_1, \dots, p_N]$  with  $p_i$  being the  $SU_i$  transmit power. The SU power levels  $[p_1, \dots, p_N]$  are communicated from the CBS to the SUs through the CRN control channel and they define the messaging overhead of this network. In a practical setting, these transmit power commands from the CBS to the SUs demand  $\lceil \log_2(N_{pl}) \rceil$  bits for each SU if we assume that the SU power range is discretized to  $N_{pl}$  power levels. However, in our study, SU power levels are considered to be continuous variables. Additionally, the SINR of the PU is defined as:

$$SINR_{PU} = 10 \log_{10} \left( \frac{g_{PU} p_{PU}}{I_{PU} + N_{PU}} \right) \text{dB} \quad (2)$$

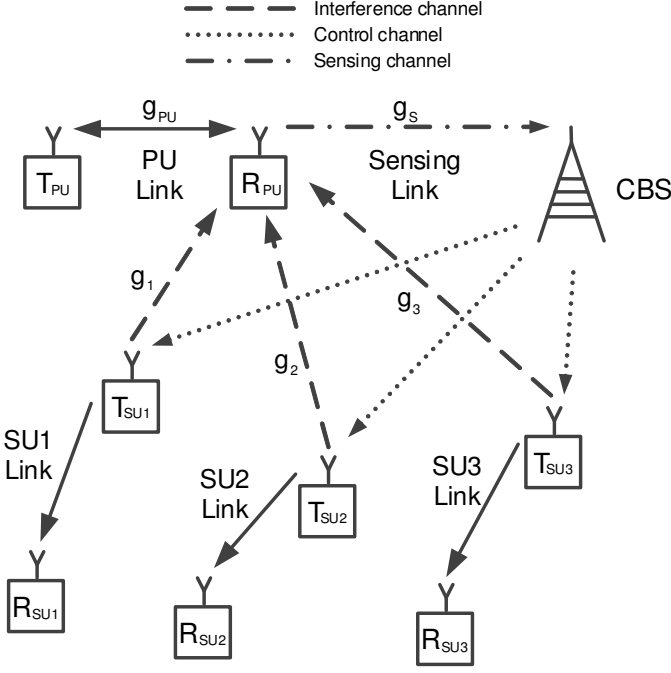


Fig. 1: The PU system and the CRN

where  $g_{PU}$  is the PU link channel gain,  $p_{PU}$  is the PU transmit power and  $N_{PU}$  is the PU receiver noise power.

In this study, we consider that the CBS is equipped with a secondary omnidirectional antenna only for sensing the signal of the PU reverse link and a module for decoding the binary ACK/NACK feedback. From this decoding process, the CRN is able to obtain a feedback observation,  $Z$ , and infer whether the induced interference to the PU,  $I_{PU}$ , is harmful or not for the PU data packet reception by the PU receiver. Assuming that  $N_{PU}$  and the received power remain the same at the PU receiver side, the minimum required  $SINR_{PU}$ ,  $\gamma$ , corresponds to a particular unknown maximum allowed  $I_{PU}$  value,  $I_{th}$ , below which an ACK is sent and over which a NACK is transmitted to the PU transmitter. Subsequently, the observed feedback  $Z$  is defined as:

$$Z = \begin{cases} +1 & \text{if } \mathbf{g} \mathbf{p}^T \leq I_{th} \\ -1 & \text{if } \mathbf{g} \mathbf{p}^T > I_{th} \end{cases}. \quad (3)$$

This piece of information will be exploited in the next section to learn the PU interference constraint determined as:

$$\mathbf{g} \mathbf{p}^T = I_{th}. \quad (4)$$

A necessary simplification of the information gained by (3) is that the  $g_i$  gains normalized to  $I_{th}$  are adequate for defining the interference constraint (4). Therefore, if  $\mathbf{h} = \frac{\mathbf{g}}{I_{th}}$ , the observed feedback can also be written as:

$$Z = \begin{cases} +1 & \text{if } \mathbf{h} \mathbf{p}^T \leq 1 \\ -1 & \text{if } \mathbf{h} \mathbf{p}^T > 1 \end{cases} \quad (5)$$

while the normalized version of (4) is expressed as:

$$\mathbf{h} \mathbf{p}^T = 1. \quad (6)$$

Due to the realistic limitation of low SNR sensing channel, the feedback packet has a probability of being correctly decoded,  $P_{cd}$ . In most Automatic Repeat reQuest (ARQ) and Hybrid ARQ mechanisms [17] which utilize identical mechanisms of request for retransmission, feedback packets are transmitted in blocks as bits which allows us to derive lower bounded analytical expressions or exact numerical values based on Bit Error Rate (BER) curves for the  $P_{cd}$  of each feedback packet depending on the MCS of the ACK/NACK block and the SNR of the sensed PU signal. Thus the feedback observation,  $Z$ , has a probability of being correct and can be expressed as:

$$\Pr[Z \text{ being correct}] = P_{cd} \quad (7)$$

and a complementary probability indicating how likely  $Z$  is incorrect:

$$\Pr[Z \text{ being incorrect}] = 1 - P_{cd}. \quad (8)$$

In section IV, we investigate AL methods which consider this uncertainty information to infer (6) with the least probing attempts possible.

#### A. Multiple PU interference constraint Active Learning

Even though in the rest of this work only one PU link is going to be considered, a multiple PU scenario must also be discussed. In this case, a number of different PU interference constraints can be tackled as in [8]. Here, we assume a PU system with  $M$  users where each PU occupies a separate frequency band. In this subsection, we will show how to decouple this multiple constraint learning problem. To achieve this, the CRN must know how the PUs are assigned within the PU system bandwidth according to the number of the PU channels and their bandwidth. Assuming this piece of knowledge, the CBS may partition the  $N$  SU set to  $M$  subsets and spread them over the PU system bandwidth in an FDMA fashion again as shown in Fig. 2 so that no SU interferes to more than one PU. Each SU subset is defined as  $\{SU_{1,m}, \dots, SU_{N_m,m}\}$  where  $m = 1, \dots, M$  and  $N_m$  is the number of elements of the  $m_{th}$  subset.

This allows the CBS to separate the multiple interference constraint AL to a set of distinct AL sub-problems and thus perform concurrently a learning method for each PU and SU subset. Hence, the original problem can be expressed into the following  $M$  constraint learning sub-problems:

$$\mathbf{g}_m \mathbf{p}_m^T \leq I_{th,m}, \quad m = 1, \dots, M \quad (9)$$

where  $\mathbf{g}_m$  are the interference channel gain vectors  $[g_{1,m}, \dots, g_{N_m,m}]$  with  $g_{i,m}$  being the  $SU_{i,m}$ -to- $PU_m$  interference channel gain,  $\mathbf{p}_m$  are the SU power vectors  $[p_{1,m}, \dots, p_{N_m,m}]$  with  $p_{i,m}$  being the  $SU_{i,m}$  transmit power and  $I_{th,m}$  are the  $PU_m$  interference thresholds.

In order for this approach to work, the CBS must be able to sense each PU channel separately in order to obtain

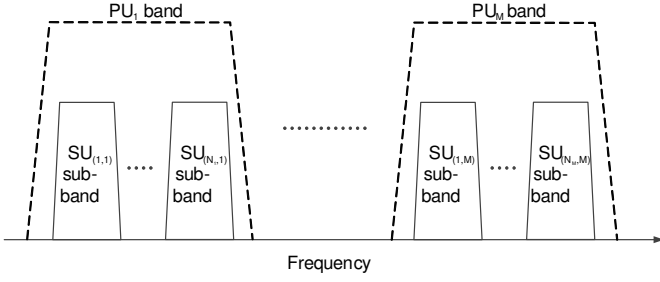


Fig. 2: The multiple PU scenario

$M$  feedback observations,  $\{Z_1, \dots, Z_M\}$ , which deliver the following inequalities:

$$Z_m = \begin{cases} +1 & \text{if } \mathbf{h}_m \mathbf{p}_m^T \leq 1 \\ -1 & \text{if } \mathbf{h}_m \mathbf{p}_m^T > 1 \end{cases}, m = 1, \dots, M. \quad (10)$$

#### IV. BAYESIAN ACTIVE LEARNING METHODS

In this paper, the goal is to design SU probing power vectors,  $\mathbf{p}$ , using uncertain observations of ACK/NACK feedback,  $Z$ , in order to learn as fast as possible the unknown normalized interference channel gain vector,  $\mathbf{h}$ . These unknown parameters define the constraints (4) and (6) which constitute the PU interference constraint in underlay cognitive scenarios and from here on will be referred to as the interference hyperplane. This uncertainty based AL probing idea was first explored as a cognitive BF problem by the authors of [6] who managed to simultaneously learn the null space of an unknown interference channel matrix and maximize the SNR at the SU receiver side. The iterative nature of this proactive probing strategy can also be employed in our scenario as illustrated in Fig. 3, where the CRN probes the PU and subsequently monitors the ACK/NACK feedback sent by the PU receiver in order to infer the interference hyperplane.

In general, sequential uncertain pieces of knowledge are incorporated using a Bayesian approach where recursive Bayesian estimation is the main knowledge extraction tool. To this direction, a multivariate Bayesian Learning method is implemented by using the uncertain observations and their corresponding SU probing power vectors. To describe in detail the recursive Bayesian updating, first we need to define the feedback conditional likelihood in this process using (7) and (8) as the probability of  $Z$  conditioned on the unknown parameter  $\mathbf{h}$ :

$$\Pr[Z|\mathbf{h}] = \begin{cases} P_{cd} & \text{if } Z = +1 \text{ and } \mathbf{h} \mathbf{p}^T \leq 1 \\ 1 - P_{cd} & \text{if } Z = +1 \text{ and } \mathbf{h} \mathbf{p}^T > 1 \\ P_{cd} & \text{if } Z = -1 \text{ and } \mathbf{h} \mathbf{p}^T > 1 \\ 1 - P_{cd} & \text{if } Z = -1 \text{ and } \mathbf{h} \mathbf{p}^T \leq 1 \end{cases}. \quad (11)$$

This expression is actually a robust threshold likelihood metric determined by the uncertainty of the feedback observation,  $Z$ .

Now, let us assume that following  $t$  probing attempts,  $\{\mathbf{p}(0), \dots, \mathbf{p}(t-1)\}$ , the CBS has observed  $t$  pieces of

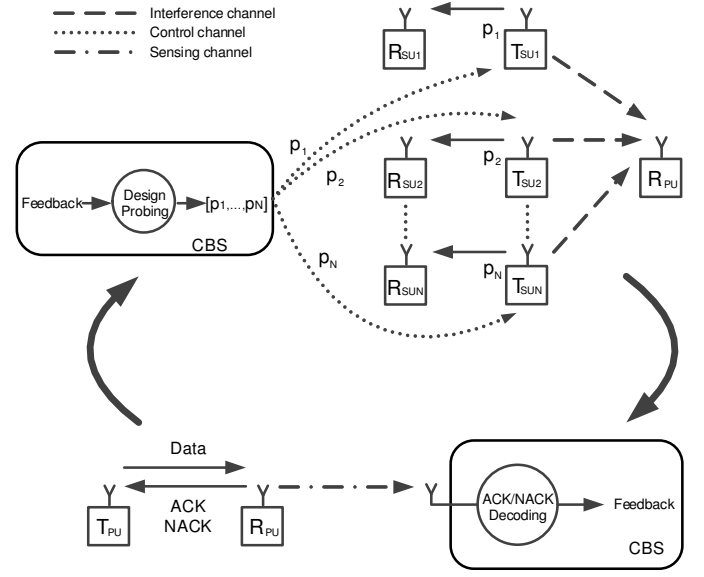


Fig. 3: The Active Learning probing scheme

ACK/NACK feedback,  $Z_{0:(t-1)} = \{Z_0, \dots, Z_{(t-1)}\}$ . After a new probing power vector  $\mathbf{p}(t)$  and a piece of feedback,  $Z_t$ , the  $\mathbf{h}$  posterior probability density function (pdf) according to the recursive form of the Bayes rule is expressed as:

$$f_{t+1}(\mathbf{h}) = \Pr[\mathbf{h}|Z_{0:t}] = \frac{\Pr[Z_t|\mathbf{h}, Z_{0:(t-1)}] \Pr[\mathbf{h}|Z_{0:(t-1)}]}{\Pr[Z_t|Z_{0:(t-1)}]} \quad (12)$$

which indicates the probability of where the true value of  $\mathbf{h}$ ,  $\mathbf{h}^*$ , lies in the  $\mathbf{h}$  space given  $Z_{0:t}$ . In (12), we also show the equivalence of the  $f_{t+1}(\mathbf{h})$  pdf with the condition  $Z_{0:t}$  which represents the knowledge gained until the  $t$  step. Here, a necessary remark about the first term of the numerator in (12) must be made which simplifies (12) and which will also help us later. The observation  $Z_t$  is conditionally independent of the previous observations  $Z_{0:(t-1)}$  given the state  $\mathbf{h} = \mathbf{h}^*$  and therefore  $\Pr[Z_t|\mathbf{h}, Z_{0:(t-1)}]$  can be written as  $\Pr[Z_t|\mathbf{h}]$  which delivers the following form of (12):

$$f_{t+1}(\mathbf{h}) = \frac{\Pr[Z_t|\mathbf{h}] f_t(\mathbf{h})}{\Pr[Z_t|Z_{0:(t-1)}]}. \quad (13)$$

The denominator term is called the marginal likelihood and even though it is difficult to be calculated, it is actually a normalization constant which guarantees that the posterior pdf integrates to 1. Usually, it is computed as the integral of the numerator in (13) which in our case is an  $N$  dimensional integration over the  $\mathbf{h}$  region and computationally intractable. In Section V, we will explain why its computation is not necessary for the application of the considered AL algorithms. A general assumption when applying recursive Bayesian estimation and employed here as well is the prior pdf  $f_0(\mathbf{h})$  to be a uniform non informative pdf [18], which is the maximum entropy pdf for random variables within a bounded domain and

therefore guarantees that no specific value of  $\mathbf{h}$  is favoured in the beginning of this recursive process.

As stated in the beginning of this section, the target of this sequential AL probing is to select SU power vectors which aim to learn  $\mathbf{h}^*$  with the minimum number of probing attempts assuming the coarse likelihood function described in (11). The proposed method in this paper is inspired by a univariate Bayesian AL algorithm, the PBA [19], and the deterministic multivariate AL techniques, the CPMs [20]. In our previous work [8], [21], the effectiveness of both methods has separately been shown in CR learning scenarios with reliable and uncertain PU feedback, but by maintaining their basic form. Here, we suggest a combination in order to formulate a multivariate Bayesian AL technique. In brief, the PBA has proven that assuming a recursive Bayesian updating for estimating a 1-D parameter and an uncertain binary feedback which indicates whether the true value of the 1-D parameter lies right or left of a testing point, the fastest way to learn its value is to always test in the next step the median of the posterior derived from the sequential Bayesian updating. On the other hand, in the deterministic multivariate AL case, the CPMs, which are extensions of the bisection algorithm to higher dimensions, theoretically guarantee that converging strategies to locate a point within a uniform uncertainty region exist and they are implemented by making linear cuts which pass through specific points of the uncertainty region in each step. These points can be the Center of Gravity (CG), the Analytic Center (AC) or the center of the Minimum Volume Ellipsoid (MVE) covering the region. The resulting CPMs are named respectively CGCPM, ACCPM and MVE-CPM.

#### A. The Optimal Cutting Plane in Bayesian Active Learning with Robust Threshold Likelihood functions

In this section, we investigate the optimal design strategy of a SU probing power vector, which represents a hyperplane in the  $\mathbf{h}$  space, that should be chosen in each step of this recursive Bayesian estimation process in order to optimally reduce the expected posterior pdf entropy after  $N_T$  probing power vectors, or cutting hyperplanes, with their corresponding pieces of feedback,  $Z_{0:(N_T-1)}$ . To achieve this, we employ Theorem IV.1, which is a multivariate extension of the PBA [19], and provide its proof, which is also a multivariate adaptation of the PBA optimality proof, in the Appendix to improve the continuity of this manuscript. Additionally, in Section V, a numerical approximation is provided for the optimal cutting plane.

**Theorem IV.1.** *Given a limited number of  $N_T$  probing attempts, pieces of feedback,  $Z$ , with conditional likelihood as in (11) and a Bayesian updating rule for  $f_{t+1}(\mathbf{h})$  as in (13), the probing power vector sequence,  $\{\mathbf{p}(0), \dots, \mathbf{p}(N_T - 1)\}$ , which achieves the minimum expected entropy of  $f_{N_T}(\mathbf{h})$  corresponds to the median regressors of the  $\{f_0(\mathbf{h}), \dots, f_{N_T-1}(\mathbf{h})\}$  pdf's,  $\{\mathbf{p}_{med}(0), \dots, \mathbf{p}_{med}(N_T - 1)\}$ .*

*Proof:* The proof of this theorem can be found in the Appendix. ■

#### B. Suboptimal Cutting Planes in Bayesian Active Learning

In the previous subsection, Theorem IV.1 shows that an AL technique within the Bayesian Learning framework must choose training samples, represented as hyperplanes in the  $\mathbf{h}$  space, which cut as evenly as possible the posterior pdf of each step. This reminds us of the CPMs where instead of designing exact cutting hyperplanes, we rely on the geometric properties of specific points of convex polyhedra for which every hyperplane passing from them cuts the polyhedron in two halfspaces whose volumes have a proven lower bound. As far as pdf's are concerned and not just convex bodies, a fundamental theoretical result for partitioning by hyperplanes in  $N$  dimensions was first given in [22], where it was proven that for any pdf, there exists at least one point for which every hyperplane passing from it divides the pdf in two parts whose probability masses have a proven lower bound of  $1/(N + 1)$ . Nevertheless, this is merely an existence theoretical result which does not define explicitly this point in the support region of a pdf. An additional theorem proven in [23] states that specifically for a log-concave multivariate pdf every hyperplane passing from its mean has the property of cutting the pdf in probability masses of at least  $1/e$ , which is the same theoretical bound for the CGCPM in convex polyhedra [22].

In this subsection, we also consider designing SU probing power vectors which pass through the mean and the mode of each step's posterior pdf, which are equivalents of the CG and the AC respectively. Specifically for the mean case, if  $f_t(\mathbf{h})$  is the posterior pdf attained after the  $(t - 1)$  step and  $\mathbf{h}_{mean}(t)$  is the  $f_t(\mathbf{h})$  posterior pdf mean which is calculated as:

$$\mathbf{h}_{mean}(t) = \frac{\int \mathbf{h} f_t(\mathbf{h}) dV_{\mathbf{h}}}{\int f_t(\mathbf{h}) dV_{\mathbf{h}}} \quad (14)$$

then the probing power vector  $\mathbf{p}(t)$  for which:

$$\mathbf{h}_{mean}(t) \mathbf{p}^T(t) = 1 \quad (15)$$

represents a cutting hyperplane in the  $\mathbf{h}$  space which passes through  $\mathbf{h}_{mean}(t)$ . Alternatively, we examine the performance of an AL method where if the  $f_t(\mathbf{h})$  posterior pdf mode or Maximum A Posteriori (MAP) point,  $\mathbf{h}_{MAP}(t)$ , is defined as:

$$\mathbf{h}_{MAP}(t) = \arg \max_{\mathbf{h}} (f_t(\mathbf{h})) \quad (16)$$

then the  $\mathbf{p}(t)$  for which:

$$\mathbf{h}_{MAP}(t) \mathbf{p}^T(t) = 1 \quad (17)$$

describes a cutting hyperplane which passes through  $\mathbf{h}_{MAP}(t)$ .

Finally, we demonstrate the Shallow-cut deterministic MVE-CPM which is suitable for uncertain pieces of feedback [24]. The standard MVE-CPM is basically an ellipsoidal approximation of the uncertainty region of the true value of  $\mathbf{h}$ ,  $\mathbf{h}^*$ . If  $\mathbf{h}_{EC}(t)$  is the center of this ellipsoid, then a probing power vector  $\mathbf{p}(t)$  for which:

$$\mathbf{h}_{EC}(t) \mathbf{p}^T(t) = 1 \quad (18)$$

defines a cutting plane that passes through  $\mathbf{h}_{EC}(t)$  in the  $\mathbf{h}$  space. With each such cutting hyperplane and its corresponding feedback, the shape of this ellipsoid, represented by its matrix

$\mathbf{P}$ , is updated and specifically it shrinks and its center moves towards  $\mathbf{h}^*$ . The updating equations for  $\mathbf{h}_{EC}(t+1)$  and the ellipsoid matrix  $\mathbf{P}_{t+1}$  are respectively:

$$\mathbf{h}_{EC}(t+1) = \mathbf{h}_{EC}(t) - \frac{1+N\alpha}{N+1} \tilde{\mathbf{p}}(t) \mathbf{P}_t \quad (19)$$

and

$$\mathbf{P}_{t+1} = \frac{N^2(1-\alpha^2)}{N^2-1} \left( \mathbf{P}_t - \frac{2(1+N\alpha)}{(N+1)(1+\alpha)} \mathbf{P}_t \tilde{\mathbf{p}}^\top(t) \tilde{\mathbf{p}}(t) \mathbf{P}_t \right) \quad (20)$$

where  $\tilde{\mathbf{p}}$  is a normalized subgradient term given by:

$$\tilde{\mathbf{p}}(t) = \frac{\mathbf{p}(t)}{\sqrt{\mathbf{p}(t) \mathbf{P}_t \mathbf{p}^\top(t)}} \quad (21)$$

and  $\alpha$  is heuristically determined as:

$$\alpha = -2 \frac{1-P_{cd}}{N}. \quad (22)$$

The design of  $\alpha$  is basically a simple linear function of  $P_{cd}$  which guarantees that in the case of the most uncertain feedback,  $P_{cd} = 0.5$ , the ellipsoid updated based on (19) and (20) remains the same and that when  $P_{cd} = 1$ , (19) and (20) are identical to the updating equations of the neutral-cut MVE-CPM. Lastly, we must mention as practical considerations about the MVE-CPM that the computational complexity in each step of this method is  $\mathcal{O}(N^2)$  and that usually the initial ellipsoid is chosen to represent a circular uncertainty region with an arbitrary center.

Here, we need to point out an important issue in AL which was emphasized in our previous work [8], the necessity of exploration. Reducing the uncertainty for  $\mathbf{h}^*$  must be performed by approaching this exact value uniformly from all directions. This means that the training samples in an AL process, in this case the power probing vectors, must be diversified and this can be accomplished by choosing cutting planes of random direction uniformly. Therefore, we need first to define how to uniformly sample a random direction  $\boldsymbol{\theta}$ , where  $\boldsymbol{\theta}$  is a unit vector. This problem is related to the uniform hypersphere point picking which has been tackled by generating  $N$ , the hypersphere dimensions, random values according to a 1-D Gaussian distribution with zero mean and arbitrary variance,  $\sigma^2$ , and set each one as variable values of a vector  $\boldsymbol{\eta}$ :

$$\eta_i \sim \mathcal{N}(0, \sigma^2), \quad i = 1, \dots, N \quad (23)$$

Then,  $\boldsymbol{\theta}$  is produced by normalizing  $\boldsymbol{\eta}$  to its magnitude,  $\boldsymbol{\theta} = \frac{\boldsymbol{\eta}}{\|\boldsymbol{\eta}\|}$ . The endpoint of the resulting unit vector  $\boldsymbol{\theta}$  is uniformly distributed on the surface of the unit hypersphere.

Now, particularly for our study, in order to produce a power vector which represents a cutting hyperplane of random direction,  $\mathbf{p}(t)$  must be parallel to a randomly generated  $\boldsymbol{\theta}$ ,  $\mathbf{p}(t) = \beta \boldsymbol{\theta}$  where  $\beta \in \mathbb{R}$ , and according to the CPM used, it must satisfy (15), (17) or (18). After some processing,  $\mathbf{p}(t)$  is expressed for the Mean CPM as:

$$\mathbf{p}(t) = \beta \boldsymbol{\theta} = \frac{\boldsymbol{\theta}}{\mathbf{h}_{mean}(t) \boldsymbol{\theta}^\top} \quad (24)$$

for the MAP CPM as:

$$\mathbf{p}(t) = \beta \boldsymbol{\theta} = \frac{\boldsymbol{\theta}}{\mathbf{h}_{MAP}(t) \boldsymbol{\theta}^\top} \quad (25)$$

and for the MVE-CPM as:

$$\mathbf{p}(t) = \beta \boldsymbol{\theta} = \frac{\boldsymbol{\theta}}{\mathbf{h}_{EC}(t) \boldsymbol{\theta}^\top}. \quad (26)$$

Moreover, all the coordinates of  $\mathbf{p}(t)$ , which represent power levels, must be non negative, otherwise a new  $\boldsymbol{\theta}$  has to be generated until a valid power vector is produced. In the next section, besides presenting numerical approximations of  $\mathbf{h}_{mean}(t)$  and  $\mathbf{h}_{MAP}(t)$ , an exploration strategy will be directly introduced in the numerical estimation of  $\mathbf{p}_{med}(t)$ .

## V. NUMERICAL APPROXIMATIONS FOR CUTTING PLANE ESTIMATION

A common problem when dealing with analytically intractable multivariate pdf's, as in our case (13), is how to estimate the mean, the mode and the median hyperplane of these density functions which are key points of the investigated Bayesian AL methods as shown in the previous section. To tackle this issue, the research community has developed sophisticated sampling methods based on Markov chain random walks, the Markov Chain Monte Carlo (MCMC) techniques [25]. Learning about probability models by simulating them and generating random samples from them has proven to be more efficient than theoretical approximations of the desired distributions, but still more computationally expensive especially when the dimension number of the target multivariate pdf grows.

One of the most commonly used sampling algorithms is the *Hit and Run* algorithm which was first thoroughly elaborated in [26]. The simplest form of this sampling mechanism is to start from a point  $\mathbf{x}_0$  in the support region  $S$  of a pdf  $f$ , choose uniformly a random direction  $\boldsymbol{\theta}_0$ , find the linear segment within  $S$  which is defined by the line passing through  $\mathbf{x}_0$  and having direction  $\boldsymbol{\theta}_0$  and compute the conditional density function along this linear segment. Subsequently, perform a 1-D random sampling over the linear segment using the conditional density function in order to find the first point of the random walk,  $\mathbf{x}_1$ , and repeat this process with starting point  $\mathbf{x}_1$  to generate the second one and so on. As far as the practical details of this sampling algorithm are concerned, first, we already defined in Section IV how to uniformly sample a random direction  $\boldsymbol{\theta}$ , a problem which is related to the uniform hypersphere point picking. Additionally, the conditional density function along the linear segment of each random point generating step must be determined. Given a point  $\mathbf{x}$  and a vector  $\boldsymbol{\theta}$ , the parametrized expression of a linear segment defined by them and within  $S$  is described as  $\mathbf{x} + \lambda \boldsymbol{\theta}$ , where  $\lambda \in [\lambda_l, \lambda_u]$ . The conditional density function  $\pi(\lambda)$  based on which the 1-D random sampling is performed can be written as:

$$\pi(\lambda) = \frac{f(\mathbf{x} + \lambda \boldsymbol{\theta})}{\int_{\lambda_l}^{\lambda_u} f(\mathbf{x} + \nu \boldsymbol{\theta}) d\nu}. \quad (27)$$

Once an adequate number of random points is produced by the aforementioned random walk process, all the required characteristics of the  $f$  pdf can be extracted so that the exhibited cutting planes in Section IV can be determined. At first, let us examine how the median hyperplane of a pdf can be estimated which according to our previous analysis is optimal in Bayesian AL with robust threshold likelihood functions. Given a set of points  $\{\mathbf{x}_1, \dots, \mathbf{x}_{N_r}\}$  with real valued weights  $\{f(\mathbf{x}_1), \dots, f(\mathbf{x}_{N_r})\}$  like in our case due to the MCMC technique, a hyperplane expressed as in (34) and which satisfies the condition  $C(\mathbf{w}) = 1/2$ , where  $C(\cdot)$  is defined in the Appendix, is actually called in robust statistics literature a halving hyperplane. The thin difference between a halving and a median hyperplane is that the latter is also a halving hyperplane [27] and moreover it minimizes the sum of the weighted distances of the points from it:

$$\mathbf{w}_{med} = \arg \min_{\mathbf{w}} \left( \sum_{i=1}^{N_r} f(\mathbf{x}_i) \frac{|\mathbf{x}_i \mathbf{w}^\top - 1|}{\|\mathbf{w}\|} \right) \quad (28)$$

where  $N_r$  is the number of the random points. In our work, we focus on the solution of (28) and not just halving hyperplanes, basically because it delivers a unique hyperplane towards a specific direction and not a set of hyperplanes from which we need to select somehow one candidate.

As far as the suboptimal cutting planes are concerned, we are interested in regressors passing from the mean and the mode of a pdf. The mean of  $f$  can be computed as:

$$\mathbf{x}_{mean} = \frac{\sum_{i=1}^{N_r} \mathbf{x}_i f(\mathbf{x}_i)}{\sum_{i=1}^{N_r} f(\mathbf{x}_i)} \quad (29)$$

while the mode or MAP point can be calculated as:

$$\mathbf{x}_{MAP} = \frac{\sum_{i:f(\mathbf{x}_i)=f_{max}} \mathbf{x}_i f(\mathbf{x}_i)}{\sum_{i:f(\mathbf{x}_i)=f_{max}} f(\mathbf{x}_i)}. \quad (30)$$

It is worth noting that the suboptimal MVE-CPM presented previously does not need such numerical techniques to work. The reason is that all its parameters can be analytically calculated without using the computationally expensive MCMC's.

In the linear piecewise optimization problem of (28), the median cutting plane solution is unique for hyperplanes of a specific direction. Nevertheless, this does not mean that it is unique for all directions, a fact which can be taken advantage of to introduce exploration in the final solution. An adaptation which can be made is to uniformly sample a random unit vector,  $\mathbf{w}_{rand}$ , and solve (28) for  $\mathbf{w} = \kappa \mathbf{w}_{rand}$ , where  $\kappa \in \mathbb{R}$ . Consequently, (28) becomes:

$$\kappa_{med} = \arg \min_{\kappa} \left( \sum_{i=1}^{N_r} f(\mathbf{x}_i) \frac{|\kappa \mathbf{x}_i \mathbf{w}_{rand}^\top - 1|}{\|\kappa \mathbf{w}_{rand}\|} \right) \quad (31)$$

and

$$\mathbf{w}_{med} = \kappa_{med} \mathbf{w}_{rand}. \quad (32)$$

All the aforementioned numerical approximations of the mean,  $\mathbf{x}_{mean}$ , the mode,  $\mathbf{x}_{MAP}$ , and the median hyperplane,  $\mathbf{w}_{med}$ , concern an arbitrary pdf  $f(\mathbf{x})$ . Respectively, for our Bayesian AL techniques, these approximations will be used in each time step  $t$  of the AL procedures in order to estimate  $\mathbf{h}_{mean}(t)$ ,  $\mathbf{h}_{MAP}(t)$  and  $\mathbf{p}_{med}(t)$  of  $f_t(\mathbf{h})$ . One last detail about the  $\mathbf{p}_{med}(t)$  estimation is that it must have non negative elements, since they denote power levels. Therefore, if (31) and (32) do not produce a  $\mathbf{w}_{med}$  for which  $\mathbf{w}_{med} \geq \mathbf{0}$  holds, then a new  $\mathbf{w}_{rand}$  must be generated until a valid power vector is delivered. For all these numerical approximations, a critical remark which must be made is that the denominator term or normalization factor of (13) can be omitted either because of fraction reduction in (27), (29), (30) or due to redundancy in (31). Therefore, we can use the unnormalized version of (13) which basically is the product of the collected likelihood functions and the uniform prior pdf instead of the actual  $f$  values.

## VI. RESULTS

In this section, we provide simulation results to compare the performance of the optimal and suboptimal Bayesian AL methods presented in Section IV, the method we previously developed and described in detail in [21] and the ML based mechanism of [6], which is used as a benchmark technique. The mean and mode crossing and median AL methods are based on MCMC sampling in the  $\mathbf{h}$  space and they are numerical, accurate, but also require a great amount of samples. The method in [21] is a computationally cheap AL method which performs consecutive 1-D PBA's in the SU power vector feasible region in order to find the interference hyperplane and the benchmark technique is a Probit ML approach where the binary feedback errors are modelled using an additive Gaussian noise model. The benchmark technique can also be considered as a Probit MAP scheme for estimating the MAP point assuming again that the prior pdf is the uniform one. Specifically, the only difference in the Probit MAP technique is that the likelihood function, instead of the (11) form, is expressed as:

$$\Pr[Z|\mathbf{h}] = \begin{cases} \Phi\left(\frac{1-\mathbf{h}\mathbf{p}^\top}{\sigma_m}\right) & \text{if } Z = +1 \\ \Phi\left(\frac{\mathbf{h}\mathbf{p}^\top-1}{\sigma_m}\right) & \text{if } Z = -1 \end{cases} \quad (33)$$

where  $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{z^2}{2}} dz$  is the standard Gaussian cdf and  $\sigma_m$  is the standard deviation of the likelihood Probit model. Using this model, the MAP estimation becomes a fast and easily solvable convex optimization problem using numerical algorithms, but with less accuracy than the MCMC based MAP calculation.

In subsections VI.B and VI.C, the figures show the channel estimation error depending on the number of time flops where each time flop is the time period necessary for the CBS to decode the ACK/NACK packet and design the SU probing power vector. The interference channel gain vector estimation error metric at each time flop is defined as the normalized root-square error  $\frac{\|\mathbf{h}(t)-\mathbf{h}^*\|}{\|\mathbf{h}^*\|}$  and basically demonstrates the



learning efficiency of each method. The error figure results are obtained as the average of the error metric defined earlier over 100 SU random topologies, which deliver 100 random draws of interference channel gain vectors  $\mathbf{h}^*$ . Subsequently, the figures of subsection VI.D present the aggregated interference caused to the PU during each AL process demonstrated in this paper. As all these AL methods progress in time, it is important to examine the degradation of the PU link quality which can be measured as the induced harmful interference. To this direction, we designed a metric which measures the PU interference exceeding  $I_{th}$  and is defined as  $I_{harm}(t) = H(I_{PU}(t) - I_{th}) * I_{PU}(t)$ , where  $H$  is the Heaviside step function. This parameter of harmful interference is also averaged over the 100 SU random topologies to deliver the corresponding average metric  $I_{harm,av}(t)$ . This metric is a clear indicator of PU protection during the probing process of all the AL schemes. Even though limiting the induced PU interference is not taken into account in any of the presented probing design techniques, useful conclusions can be extracted for PU protection.

In Table II, all the AL methods described in this paper are collected with the corresponding abbreviations for identifying them in the performance figures and with the equations or references required for their implementation or understanding. Later on, the aforementioned techniques are compared and the resulting diagrams are divided in two categories, the MCMC based and the computationally cheap ones. The computationally cheap techniques are the MVE-CPM based AL method, the one proposed in [21] and the benchmark procedure [6]. This second category is expected to have worse learning performance than the first one, but its AL candidates are recommended for learning interference hyperplanes of high dimensions, where MCMC sampling fails.

TABLE II: An overview of the examined AL methods

Method	Figure Abbreviation	Related Equations or References
MCMC based Median CPM	MCMC Median	(31) and (32)
MCMC based Mean CPM	MCMC Mean	(29) and (24)
MCMC based MAP CPM	MCMC MAP	(30) and (25)
MVE CPM	MVE CPM	(19), (20), (21), (22) and (26)
Probit based MAP CPM	Probit MAP	[6]
1-D PBA based AL method	1-D PBA	[21]

An algorithmic description of all the Bayesian AL methods provided in this paper is presented in Algo. 1 where a limited "budget" of  $N_T$  probing attempts is considered.

#### A. Simulation Parameters

As far as the technical parameters of the simulations are concerned, the PU receiver is chosen to normally operate and acknowledge with ACK packets when interference is below  $I_{th} = -97\text{dBm}$ , a limit unknown to the CRN. The examined scenarios consider  $N = 5$  and  $N = 10$  SUs which are dispersed uniformly within a 3km range around the PU receiver. The interference channel gains that are unknown to the CRN are assumed to follow an exponential path loss model

#### Algorithm 1 Bayesian Active Learning for interference hyperplane estimation

---

Assume a uniform  $f_0(\mathbf{h})$  or an initial ellipsoid  $\{\mathbf{P}_0, \mathbf{h}_{EC}(0)\}$   
**for**  $t = 0 : (N_T - 1)$  **do**  
  1: Design  $\mathbf{p}(t)$  using the related equations from Table II and probe the PU  
  2: Acquire uncertain observation  $Z_t$   
  3: Incorporate  $Z_t$  to define  $f_{t+1}(\mathbf{h}|Z_{0:t})$  or update ellipsoid to  $\{\mathbf{P}_{t+1}, \mathbf{h}_{EC}(t+1)\}$   
**end for**  
Choose  $\mathbf{h}^*$  as  $\mathbf{h}_{mean}(N_T)$  or  $\mathbf{h}_{EC}(N_T)$

---

$g_i = \frac{1}{d_i^4}$ , where  $d_i$  is the distance of the  $\text{SU}_i$  from the PU receiver in metres. The remaining scenario parameters are the maximum transmit power,  $p_{max_i}$ , which is set to 23dBm for all SUs, and the probability of the CBS correctly decoding the ACK/NACK packets,  $P_{cd}$ , which is selected to take three values  $\{0.7, 0.8, 0.9\}$ .

Additionally, a practical consideration which must be taken into account is the necessary number of samples for the MCMC based AL methods to be accurate, which in the MCMC literature is not well defined. Using the MCMC convergence diagnostics method of [28], we have concluded that  $N_r = 20000$  is acceptable for median estimation error of 1% in  $N = 5$  dimensions. For the mean and mode estimations, the error is around 0.1% for the same  $N_r$ . In the case of  $N = 10$ , the corresponding errors are 1.6% for the median regressor and 0.1% for the mean and the mode with  $N_r = 150000$  samples, which are similar to the previous ones. In [28], the estimation of a statistical feature of a pdf using MCMCs is monitored as the sampling proceeds. Particularly, after each sample, a series of cumulative sums of residuals concerning the feature of interest is constructed. Subsequently, its smoothness is analysed in order to decide if the MCMC estimation of this feature has converged and thus presume that it is safe to stop sampling. A comparative review of the most popular MCMC convergence diagnostics tools can be found in [29] where the advantages and disadvantages of each method are explained.

#### B. Estimation Performance of MCMC Based AL Methods

Initially, let us see in Fig. 4, 5 and 6 the performance of the first category AL methods for  $N = 5$  SUs. Here, it can be clearly seen that as  $P_{cd}$  is reduced, the three AL techniques require more time flops, meaning probing attempts, to correctly estimate  $\mathbf{h}^*$ . Furthermore, the median based AL method outperforms in speed both the mean and MAP based probing schemes. More specifically, in the case of  $P_{cd} = 0.9$ , Fig. 4, for an estimation error 1% the median method achieves convergence in 105 time flops, whereas the corresponding numbers of time flops for the mean and MAP based techniques are 116 and 158 respectively. This convergence gain is also observed for  $P_{cd} = 0.8$  in Fig. 5 where in 200 time flops the median, mean and MAP based methods have corresponding estimation errors 2.1%, 5.5% and 27% and for  $P_{cd} = 0.7$  in

Fig. 6 where in 500 time flops the respective estimation errors are 1.8%, 2.9% and 23%.

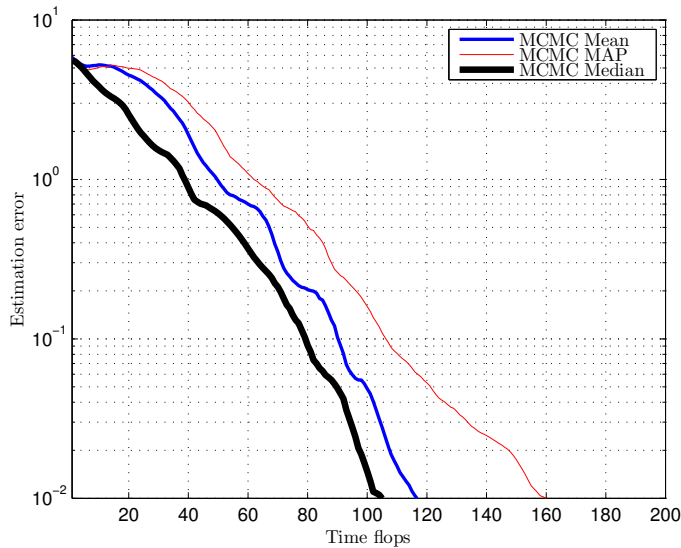


Fig. 4: Interference channel gain vector estimation error progress vs time of the MCMC based AL methods for  $P_{cd} = 0.9$  and  $N = 5$  SUs

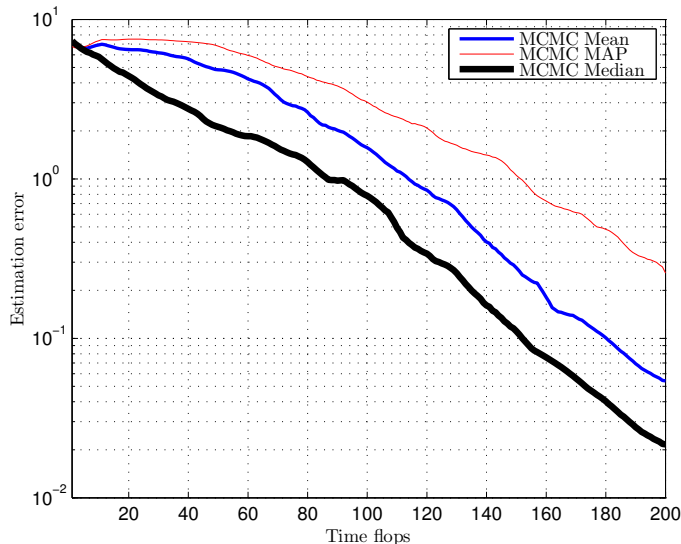


Fig. 5: Interference channel gain vector estimation error progress vs time of the MCMC based AL methods for  $P_{cd} = 0.8$  and  $N = 5$  SUs

Moreover, to clearly show that the median based method is faster than the mean and the MAP based ones, we need to increase the problem dimensions, the number of the SUs. The observed differences among the necessary probing attempts of

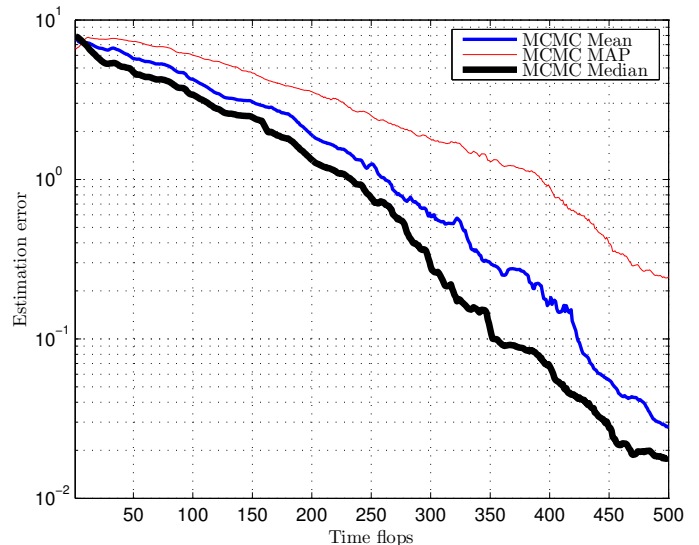


Fig. 6: Interference channel gain vector estimation error progress vs time of the MCMC based AL methods for  $P_{cd} = 0.7$  and  $N = 5$  SUs

the MCMC based methods in order to reach  $\mathbf{h}^*$  within some certain error should be increased as the CRN grows. Otherwise, one could argue that there is a  $N$  limit beyond which the median based technique is not optimal. The next diagram in Fig. 7 shows the estimation performance of the first group AL methods for  $N = 10$  SUs and  $P_{cd} = 0.9$ .

Compared to the diagram in Fig. 4, it is clearly shown that first of all the probing attempts to achieve estimation error of 1% have increased for all methods which is expected to happen, since in the  $N = 10$  case  $\mathbf{h}^*$  has higher dimensions and therefore more coefficients. Second, the convergence gains among the MCMC based methods have also increased which experimentally validates that the higher the problem dimensions, the larger the performance differences among these methods. Specifically, as seen in Fig. 7, the median and mean based AL schemes achieve an estimation error of 1% at 208 and 252 time flops respectively, whereas the MAP based can hardly compete them. This delivers us a convergence gain of 44 time flops between the median and the mean based method for  $N = 10$  SUs, while the convergence gain between the same schemes for  $N = 10$  SUs is 9 time flops as observed in Fig. 4.

However, as mentioned earlier, the high learning speed of the MCMC based probing schemes comes with a penalty. The *Hit and Run* calculation of the median regressors and the mean and MAP points requires the generation of many random samples in the  $\mathbf{h}$  space according to the pdf of each step. The number of these samples grows exponentially with the problem dimensions and can be specified by the aforementioned convergence diagnostics tool [28]. Particularly in our scenarios, as we emphasized in subsection VI.A, for the  $N = 5$  and  $N = 10$  SU cases we used respectively

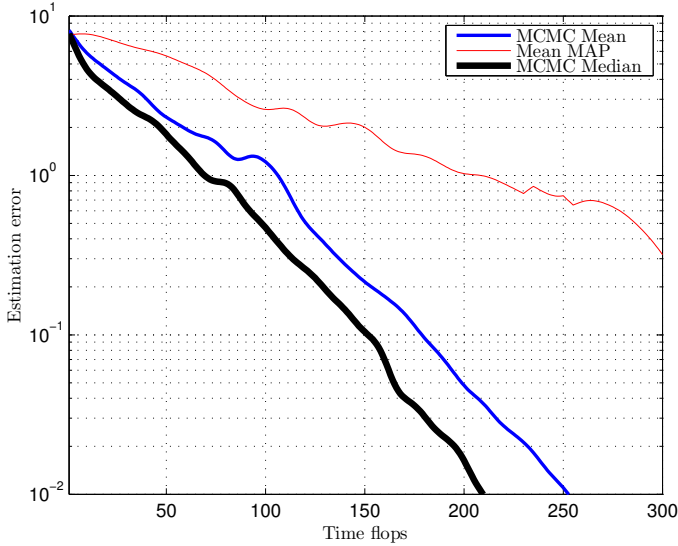


Fig. 7: Interference channel gain vector estimation error progress vs time of the MCMC based AL methods for  $P_{cd} = 0.9$  and  $N = 10$  SUs

$N_r = 20000$  and  $N_r = 150000$  samples. Therefore, in order for the CBS, where all these computations take place, to design the SU probing power vectors an exponentially increasing to  $N$  computational burden is demanded. This means that the larger the CRN a CBS must coordinate, the more computations the CBS needs to perform in order to achieve fast convergence performances.

### C. Estimation Performance of Computationally Cheap AL Methods

To tackle the computational issue due to MCMC usage, we have also tested the performance of the analytical MVE-CPM based AL technique we developed in this paper, of the Probit MAP scheme in [6] and of the Bayesian AL scheme in [21] which is essentially based on Bayesian 1-D grid estimators and thus it is also computationally effective.

At first, as seen in Fig. 8 and 9 for  $P_{cd} = 0.9$  and  $P_{cd} = 0.8$  respectively, the AL mechanisms of this second category have worse convergence rate than the previous ones, which was expected. Furthermore, when comparing the last three techniques, it can easily be observed that the suboptimal MVE-CPM based AL method developed in this paper converges faster than the other two ones. Particularly, for  $P_{cd} = 0.9$  an estimation error of 1% is achieved by the MVE-CPM in 398 time flops, whereas by the scheme of [21] in 452. Similarly, for  $P_{cd} = 0.8$  the MVE-CPM learns  $\mathbf{h}^*$  in 708 time flops, while the [21] scheme in 765 time flops. The benchmark method in both cases does not perform well and exhibits slow learning rate, basically because it uses an additive Gaussian model for the feedback error resulting in approximating the likelihood functions of (11) as Gaussian cdf's, which is a very rough approximation. Here, we must also mention that a reason for

testing the MCMC based MAP method was to prove that choosing the MAP point, or at least an almost exact estimation of it through MCMC sampling, for AL purposes is not optimal. Furthermore, below the barrier of  $P_{cd} = 0.8$ , these techniques do not manage to converge, hence results for  $P_{cd} = 0.7$  have not been included.

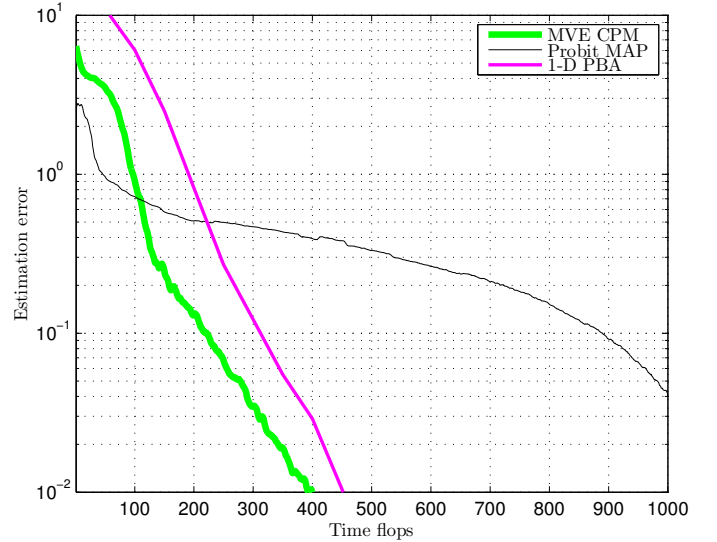


Fig. 8: Interference channel gain vector estimation error progress vs time of the computationally cheap AL methods for  $P_{cd} = 0.9$  and  $N = 5$  SUs

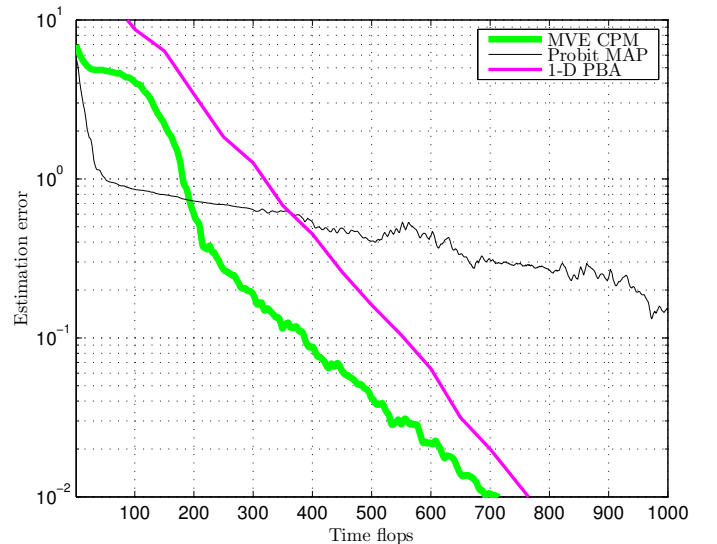


Fig. 9: Interference channel gain vector estimation error progress vs time of the computationally cheap AL methods for  $P_{cd} = 0.8$  and  $N = 5$  SUs

Finally, in order to confirm that the performance ranking

of the computationally cheap AL methods does not change as the problem dimensions grow, the estimation error diagram for  $N = 10$  SUs and  $P_{cd} = 0.9$  is given in Fig. 10. The Probit MAP scheme appears to converge slowly, while the MVE-CPM based AL method and the scheme of [21] achieve an estimation error of 2.3% in 1342 and 2000 time flops respectively. Compared to the convergence gain between the two last techniques for  $N = 5$  SUs as shown in Fig. 8, in this case a greater gain is delivered in favour of the MVE-CPM based AL method. Hence, it can be safely concluded that for this group, the MVE-CPM based scheme is optimal.

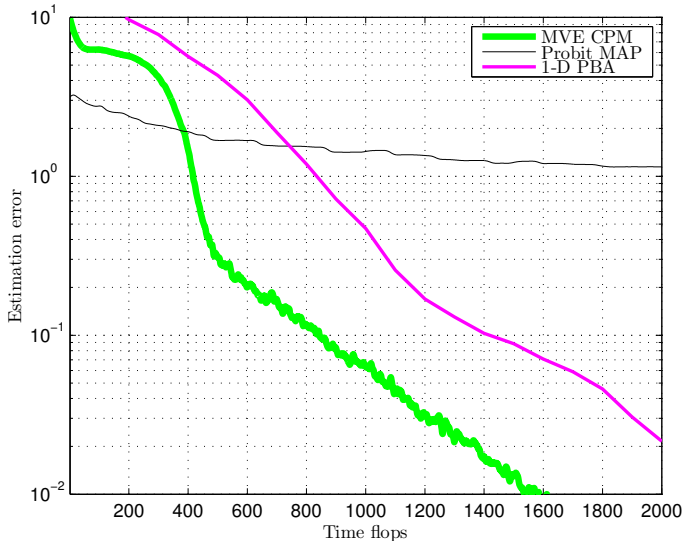


Fig. 10: Interference channel gain vector estimation error progress vs time of the computationally cheap AL methods for  $P_{cd} = 0.9$  and  $N = 10$  SUs

#### D. Interference induced by MCMC Based and Computationally Cheap AL Methods

In this subsection, we provide the  $I_{harm,av}$  diagrams for MCMC based and computationally cheap AL methods in Fig. 11 and Fig. 12 respectively. These diagrams enable us to distinguish which AL methods are optimal in terms of protecting the PU, since an important aspect of all the aforementioned probing schemes is the harmful PU interference caused by each AL process. The results demonstrated here correspond to  $P_{cd} = 0.8$ , firstly because the  $I_{harm,av}$  curves become more distinguishable as the convergence time increases and secondly for the reason that we wish to compare the two method groups, MCMC based and computationally cheap, but below the barrier of  $P_{cd} = 0.8$ , the computationally cheap techniques do not manage to converge. Additionally, the number of SUs is chosen to be  $N = 5$  instead of  $N = 10$ , because specifically the results in Fig. 12 are easier to be discriminated. Hence, choosing  $P_{cd} = 0.8$  and  $N = 5$  SUs is suitable for acquiring readable results and drawing solid conclusions.

Initially, in Fig. 11, we observe that after 200 time flops the mean and median AL schemes induce interference close to the  $I_{th} = -97$ dBm, but still the median based MCMC AL method provides better protection to the PU, since it causes less interference to the PU through time. As far as the computationally cheap AL methods are concerned, in Fig. 12 we first notice that the Probit MAP AL scheme, which is used as a benchmark method, approaches the  $I_{th} = -97$ dBm slower than the MVE-CPM or the 1-D PBA based AL method and also induces more harmful interference to the PU overall. Moreover, the MVE-CPM compared to the 1-D PBA based method of [21] generates less harmful PU interference, without any high interference spikes and with smoother convergence to the  $I_{th} = -97$ dBm. Conclusively, it can be derived from Fig. 11 and Fig. 12 that the MVE-CPM is worse than all MCMC based AL schemes, but still comparable in terms of induced harmful interference, and that in general the faster the estimation performance of an AL is, the less interference it causes to the PU. Additionally, the more informative an AL process is, the closer to the PU interference threshold it probes. This indicates that in an AL setting, even though the only probing design metric is the information gain, a probing power vector which delivers more information is also more cautious towards the PU link operation.

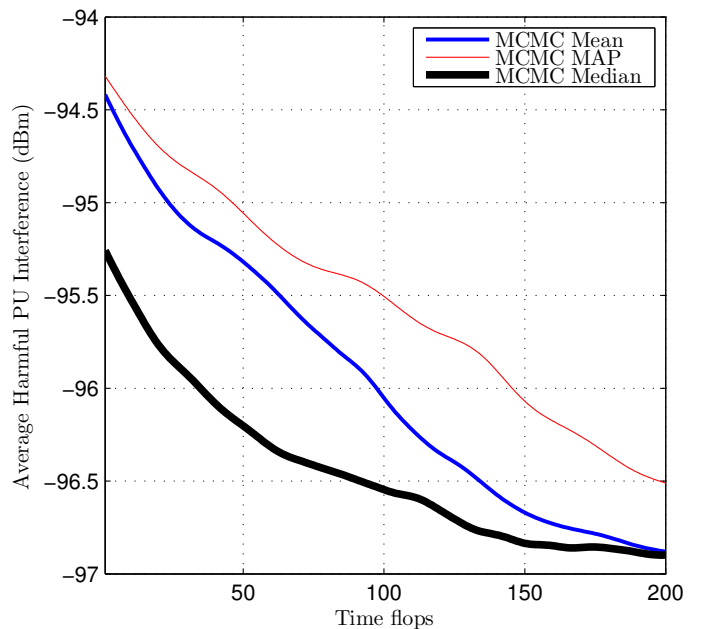


Fig. 11:  $I_{harm,av}$  progress vs time of the MCMC based AL methods for  $P_{cd} = 0.8$  and  $N = 5$  SUs

## VII. CONCLUSIONS AND FUTURE WORK

In this paper, we proposed probing methods which can be used by a centralized CRN for PU interference constraint fast learning using uncertain ACK/NACK PU feedback. The

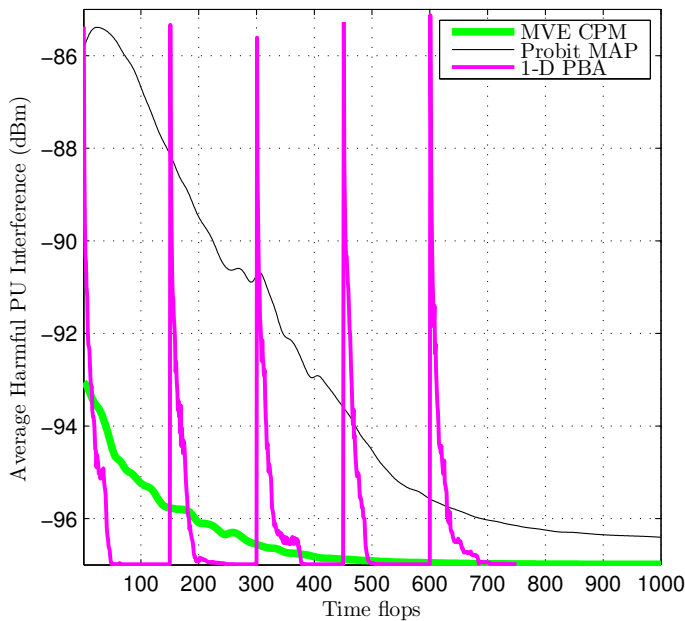


Fig. 12:  $I_{harm,av}$  progress vs time of the computationally cheap AL methods for  $P_{cd} = 0.8$  and  $N = 5$  SUs

proposed techniques were inspired by the deterministic multivariate CPMs and the univariate PBA. The first method we suggest, whose optimality is also proven, is a median based Bayesian AL design of the SU probing power vectors using MCMC sampling and the second one is an MVE-CPM adaptation that is less accurate, but computationally affordable and suitable for large CRN's. The superior performance of these methods compared to existing ones in the AL field [6], [21] was demonstrated through numerical simulations in static channel scenarios for interference channel gain learning. Additionally, results were given for the induced PU interference, which prove that the median based Bayesian AL method and the MVE-CPM adaptation are more protective to the PU among the MCMC based and the computationally cheap AL techniques respectively.

As part of our future work, the uncertainty driven methods demonstrated in this paper will also be tested for fading interference channel gain learning based on the forgetting factor approach of [6]. In addition, the same Bayesian AL techniques will be applied to cognitive beamforming and cognitive relay network scenarios to confirm their proactive learning efficiency. Within the AL framework, we also plan to develop variations of our current methods which will be suitable for a decentralized CRN structure with a message passing mechanism between the SUs. Such AL methods are closely related to decentralized learning schemes and could tackle issues like computational complexity and scalability. Moreover, we intend to develop more accurate analytical approximations of the exhibited feedback likelihood functions and thus improve the convergence rate of computationally cheap AL probing schemes. A natural extension of the work

presented in this paper is handling the acquired probabilistic knowledge of the interference channel gains in Stochastic Programming or Robust Optimization problems for defining the optimal average or worst case CRN operation. Finally, we are interested in deriving probing designs which not only maximize information gain, but also incorporate other metrics such as CRN throughput, CRN energy consumption and PU induced interference. All these metrics could be incorporated into an optimization problem, but just one of them as an objective, while the rest as constraints, so that we could still guarantee some lower or upper bounds for the latter. Another alternative could be to formulate a Multi-objective optimization problem to study the trade-off's in satisfying these different objectives.

#### APPENDIX PROOF OF THEOREM IV.1

First of all, let us define the multivariate cumulative distribution function (cdf) in a more "natural" than the usual way. Assuming a multivariate pdf  $f$  in  $S \subseteq \mathbb{R}^n$  and a vector  $\mathbf{x} = [x_1, \dots, x_N]$ , usually its cdf  $F$  is defined as  $F(\mathbf{x}) = \Pr[X_1 \leq x_1, \dots, X_N \leq x_N]$  which is the joint probability of its components  $X_1, \dots, X_N$ , that are scalar valued random variables, being less or equal than the values  $x_1, \dots, x_N$  respectively. Nevertheless, this definition is not geometrically smooth and commonly used just because it is easy to be computed in case of independent  $\mathbf{x}$  components. Here, we describe it more strictly and not just by using a "box limit"-like definition. Assuming a hyperplane in  $\mathbb{R}^n$ :

$$\mathbf{x} \mathbf{w}^\top = 1 \quad (34)$$

we alternatively determine the cdf  $C$  of a multivariate pdf  $f$  as:

$$C(\mathbf{w}) = \Pr[\mathbf{x} \mathbf{w}^\top \leq 1] = \int_{\mathbf{x} \mathbf{w}^\top \leq 1} f(\mathbf{x}) dV_{\mathbf{x}}. \quad (35)$$

Specifically for our case study, the posterior cdf after the  $(t-1)$  step,  $C_t(\mathbf{p})$ , is expressed as:

$$C_t(\mathbf{p}) = \Pr[\mathbf{h} \mathbf{p}^\top \leq 1 | Z_{0:(t-1)}] = \int_{\mathbf{h} \mathbf{p}^\top \leq 1} f_t(\mathbf{h}) dV_{\mathbf{h}} \quad (36)$$

and the support region of  $f_t(\mathbf{h})$  is limited to the positive orthant of the  $\mathbf{h}$  space,  $\mathbb{R}_+^N$ , because the interference channel gains can only have non negative values.

Further on, we elaborate on the marginal likelihood of (13). In the event of  $Z_t = +1$ , the conditional probability  $\Pr[Z_t | Z_{0:(t-1)}]$  can also be written according to the Bayes sum rule as:

$$\begin{aligned} \Pr[Z_t = +1 | Z_{0:(t-1)}] = & \\ \Pr[\mathbf{h} \mathbf{p}^\top(t) \leq 1 | Z_{0:(t-1)}] \Pr[Z_t = +1 | \mathbf{h} \mathbf{p}^\top(t) \leq 1, Z_{0:(t-1)}] + & \\ \Pr[\mathbf{h} \mathbf{p}^\top(t) > 1 | Z_{0:(t-1)}] \Pr[Z_t = +1 | \mathbf{h} \mathbf{p}^\top(t) > 1, Z_{0:(t-1)}]. & \end{aligned} \quad (37)$$

By exploiting the conditional independence of  $Z_t$  to the previous observations  $Z_{0:(t-1)}$  given  $\mathbf{p}(t)$  and the equivalence

of the conditions  $Z_{0:(t-1)}$  and  $f_t(\mathbf{h})$  which were underlined earlier,  $\Pr[Z_t = +1|Z_{0:(t-1)}]$  can also be expressed as:

$$\begin{aligned} & \Pr[Z_t = +1|Z_{0:(t-1)}] = \\ & \Pr[\mathbf{h} \mathbf{p}^\top(t) \leq 1|f_t(\mathbf{h})] \Pr[Z_t = +1|\mathbf{h} \mathbf{p}^\top(t) \leq 1] + \\ & \Pr[\mathbf{h} \mathbf{p}^\top(t) > 1|f_t(\mathbf{h})] \Pr[Z_t = +1|\mathbf{h} \mathbf{p}^\top(t) > 1] = \\ & C_t(\mathbf{p}(t)) P_{cd} + (1 - C_t(\mathbf{p}(t))) (1 - P_{cd}) \end{aligned} \quad (38)$$

which in the spirit of [19] is a function of  $\mathbf{p}(t)$ ,  $\gamma_t(\mathbf{p}(t))$ , and can be written as:

$$\gamma_t(\mathbf{p}(t)) = C_t(\mathbf{p}(t)) P_{cd} + (1 - C_t(\mathbf{p}(t))) (1 - P_{cd}). \quad (39)$$

A similar expression can also be derived for the  $Z_t = -1$  event:

$$\Pr[Z_t = -1|f_t(\mathbf{h})] = 1 - \gamma_t(\mathbf{p}(t)). \quad (40)$$

After all this preliminary work, we continue with our original goal of this proof, which is to find the optimal probing rule that maps  $\{f_0, \dots, f_{N_T-1}\}$  to  $\{\mathbf{p}(0), \dots, \mathbf{p}(N_T-1)\}$  in order to achieve the maximum average entropy reduction from the  $f_0(\mathbf{h})$  to the  $f_{N_T}(\mathbf{h})$  pdf. Based on the Dynamic Programming (DP) proof of the univariate case [19], our goal expressed in a formal manner is to seek the optimal probing design policy  $\pi_{0:(N_T-1)}^* = \{\mathbf{p}(0) = \mu^*(f_0), \dots, \mathbf{p}(N_T-1) = \mu^*(f_{N_T-1})\}$  which maximizes the expected information gain over all possible feedback sequences derived by this policy  $\pi^*$ :

$$\pi_{0:(N_T-1)}^* = \arg \max J_{\pi_{0:(N_T-1)}}. \quad (41)$$

The expected information gain of an arbitrary policy  $\pi$ ,  $J_{\pi_{0:(N_T-1)}}$ , is expressed as the conditional expectation:

$$J_{\pi_{0:(N_T-1)}} = E_\pi[\mathcal{H}(f_0) - \mathcal{H}(f_{N_T})|\mathbf{p}(N_T-1), f_{N_T-1}] \quad (42)$$

where  $\mathcal{H}$  is the entropy operator of a pdf. To create a multistage version of our objective in the DP spirit [30], we rewrite (42) as:

$$\begin{aligned} J_{\pi_{0:(N_T-1)}} &= E_\pi[\mathcal{H}(f_0) - \mathcal{H}(f_{N_T})|\mathbf{p}(N_T-1), f_{N_T-1}] = \\ & E_\pi[\mathcal{H}(f_0) - \mathcal{H}(f_1) + \dots \\ & + \mathcal{H}(f_{N_T-1}) - \mathcal{H}(f_{N_T})|\mathbf{p}(N_T-1), f_{N_T-1}] = \\ & E_\pi \left[ \sum_{k=0}^{N_T-1} (\mathcal{H}(f_k) - \mathcal{H}(f_{k+1})|\mathbf{p}(k), f_k) \right] \end{aligned} \quad (43)$$

where we added and subtracted all the entropy terms of the intermediate pdf's to form an additive gain over time.

To solve (41), DP is using a backward induction logic, where if we define  $J_{\pi_{k:(N_T-1)}}$  as:

$$\begin{aligned} J_{\pi_{k:(N_T-1)}} &= E_\pi \left[ \sum_k^{N_T-1} (\mathcal{H}(f_k) - \mathcal{H}(f_{k+1})|\mathbf{p}(k), f_k) \right] = \\ & E_\pi [\mathcal{H}(f_k) - \mathcal{H}(f_{k+1})|\mathbf{p}(k), f_k] + J_{\pi_{(k+1):(N_T-1)}} \end{aligned} \quad (44)$$

then we first need to solve:

$$\begin{aligned} & \max_\mu J_{\pi_{N_T-1}} = \\ & \max_\mu E_\pi [\mathcal{H}(f_{N_T-1}) - \mathcal{H}(f_{N_T})|\mathbf{p}(N_T-1), f_{N_T-1}] \end{aligned} \quad (45)$$

and secondly we must solve:

$$\begin{aligned} & \max_\mu J_{\pi_{k:(N_T-1)}} = \\ & \max_\mu E_\pi [\mathcal{H}(f_k) - \mathcal{H}(f_{k+1})|\mathbf{p}(k), f_k] + J_{\pi_{(k+1):(N_T-1)}}^* \end{aligned} \quad (46)$$

for an arbitrary  $k \in \{0, \dots, (N_T-2)\}$ . If the two resulting rules, meaning the functions which assign a probing power vector, our decision, to a posterior pdf, our state, are identical, then by induction we may say that this is the optimal design policy  $\mu^*$  and that it satisfies (41).

Now, let us start from solving (45). Using the updating equation (13), we evaluate  $E_\pi [\mathcal{H}(f_{N_T-1}) - \mathcal{H}(f_{N_T})|\mathbf{p}(N_T-1), f_{N_T-1}]$  over the two possible events  $Z_{N_T-1} = +1$  and  $Z_{N_T-1} = -1$ :

$$\begin{aligned} & E_\pi [\mathcal{H}(f_{N_T-1}) - \mathcal{H}(f_{N_T})|\mathbf{p}(N_T-1), f_{N_T-1}] = \\ & E_\pi [E_{\mathbf{h}} [-\log(f_{N_T-1})] - E_{\mathbf{h}} [-\log(f_{N_T})]] - \\ & E_\pi [E_{\mathbf{h}} [-\log(\Pr[Z_{N_T-1}|\mathbf{h}])] |\mathbf{p}(N_T-1), f_{N_T-1}] + \\ & E_\pi [E_{\mathbf{h}} [-\log(\Pr[Z_{N_T-1}|Z_{0:(N_T-2)}])] |\mathbf{p}(N_T-1), f_{N_T-1}]. \end{aligned} \quad (47)$$

The last two remaining terms can be further processed. With the help of (11) for  $\Pr[Z_{N_T-1}|\mathbf{h}]$ , the third term can be analyzed as:

$$\begin{aligned} & E_\pi [E_{\mathbf{h}} [-\log(\Pr[Z_{N_T-1}|\mathbf{h}])] |\mathbf{p}(N_T-1), f_{N_T-1}] = \\ & E_\pi [E_{\mathbf{h}} [-\log(\Pr[Z_{N_T-1}|\mathbf{h}])] |\mathbf{p}(N_T-1)] = \\ & -P_{cd} \log(P_{cd}) - (1 - P_{cd}) \log(1 - P_{cd}) \end{aligned} \quad (48)$$

since  $Z_{N_T-1}$  does not depend on neither  $f_{N_T-1}$  nor  $\mathbf{p}(N_T-1)$ , the ‘‘query’’. This result tells us that this term does not depend on the design of  $\mathbf{p}(N_T-1)$  and therefore it does not participate in the maximization of  $J_{\pi_{(N_T-1)}}$ . Additionally, by using (38) and (40) which again lead us to omit  $E_{\mathbf{h}}$ , the fourth term becomes:

$$\begin{aligned} & E_\pi [E_{\mathbf{h}} [-\log(\Pr[Z_{N_T-1}|Z_{0:(N_T-2)}])] |\mathbf{p}(N_T-1), f_{N_T-1}] = \\ & E_\pi [E_{\mathbf{h}} [-\log(\Pr[Z_{N_T-1}|f_{N_T-1}])] |\mathbf{p}(N_T-1), f_{N_T-1}] = \\ & E_\pi [-\log(\Pr[Z_{N_T-1}|f_{N_T-1}])|\mathbf{p}(N_T-1)] = \\ & -\gamma_{N_T-1}(\mathbf{p}(N_T-1)) \log(\gamma_{N_T-1}(\mathbf{p}(N_T-1))) - \\ & (1 - \gamma_{N_T-1}(\mathbf{p}(N_T-1))) \log((1 - \gamma_{N_T-1}(\mathbf{p}(N_T-1)))). \end{aligned} \quad (49)$$

After elaborating on  $J_{\pi_{(N_T-1)}}$ , we reached the conclusion that the probing rule that maximizes  $J_{\pi_{(N_T-1)}}$  is achieved by maximizing the quantity derived from (49), which occurs for  $\gamma_{N_T-1}(\mathbf{p}(N_T-1)) = 1/2$  and consequently for  $C_{N_T-1}(\mathbf{p}(N_T-1)) = 1/2$ . The same result is derived from solving (46) with a similar analysis which delivers that the optimal  $\mathbf{p}(k)$  design rule for maximizing  $J_{\pi_{k:(N_T-1)}}$  is  $C_k(\mathbf{p}(k)) = 1/2$ . Thus, the overall design policy that solves (41) is  $\mathbf{p}(k) = \mu^*(f_k) = C_k^{-1}(1/2) = \mathbf{p}_{med}(k)$ . This result indicates that in order to reach as fast as possible our learning solution, the probing power vectors should always be chosen as median regressors (bisectors) of the current posterior pdf estimate,  $\{\mathbf{p}(0), \dots, \mathbf{p}(N_T-1)\} = \{\mathbf{p}_{med}(0), \dots, \mathbf{p}_{med}(N_T-1)\}$ .

## REFERENCES

- [1] J. Mitola, "Cognitive radio an integrated agent architecture for software defined radio," Ph.D. dissertation, KTH Royal Institute of Technology Stockholm, Stockholm, Sweden, 2000.
- [2] S. Haykin, "Cognitive Radio: Brain-Empowered Wireless Communications," *IEEE J. Sel. Areas Commun.*, vol. 23, no. 2, pp. 201–220, Feb. 2005.
- [3] A. Goldsmith, S. Jafar, I. Maric, and S. Srinivasa, "Breaking Spectrum Gridlock With Cognitive Radios: An Information Theoretic Perspective," *Proc. IEEE*, vol. 97, no. 5, pp. 894–914, May 2009.
- [4] I. F. Akyildiz, W.-Y. Lee, M. C. Vuran, and S. Mohanty, "NeXt generation/dynamic spectrum access/cognitive radio wireless networks: A survey," *Computer Networks Journal (Elsevier)*, vol. 50, pp. 2127–2159, Sept. 2006.
- [5] Y. Noam and A. J. Goldsmith, "The One-Bit Null Space Learning Algorithm and Its Convergence," *IEEE Trans. Signal Process.*, vol. 61, no. 24, pp. 6135–6149, Dec. 2013.
- [6] B. Gopalakrishnan and N. D. Sidiropoulos, "Cognitive Transmit Beamforming from Binary CSIT," *IEEE Trans. Wireless Commun.*, vol. 14, no. 2, pp. 895–906, Feb. 2014.
- [7] A. Tsakmalis, S. Chatzinotas, and B. Ottersten, "Modulation and Coding Classification for Adaptive Power Control in 5G Cognitive Communications," in *Proc. IEEE 14th Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, Jun. 2014, pp. 234–238.
- [8] A. Tsakmalis, S. Chatzinotas, and B. Ottersten, "Centralized Power Control in Cognitive Radio Networks Using Modulation and Coding Classification Feedback," *IEEE Trans. Cognitive Commun. and Networking*, vol. 2, no. 3, pp. 223–237, Sept. 2016.
- [9] A. Tsakmalis, S. Chatzinotas, and B. Ottersten, "Power Control in Cognitive Radio Networks Using Cooperative Modulation and Coding Classification," in *Proc. 10th Int. Conf. on Cognitive Radio Oriented Wireless Netw. (CROWNCOM)*, Apr. 2015.
- [10] R. Mudumbai, J. Hespanha, U. Madhow, and G. Barriac, "Distributed Transmit Beamforming Using Feedback Control," *IEEE Trans. Inf. Theory*, vol. 56, no. 1, pp. 411–426, Jan. 2010.
- [11] B. C. Banister and J. R. Zeidler, "A Simple Gradient Sign Algorithm for Transmit Antenna Weight Adaptation With Feedback," *IEEE Trans. Signal Process.*, vol. 51, no. 5, pp. 1156–1171, May 2003.
- [12] A. G. Marques, L. M. Lopez-Ramos, G. B. Giannakis, and J. Ramos, "Resource Allocation for Interweave and Underlay CRs Under Probability-of-Interference Constraints," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 10, pp. 1922–1933, Nov. 2012.
- [13] A. Galindo-Serrano and L. Giupponi, "Distributed Q-Learning for Aggregated Interference Control in Cognitive Radio Networks," *IEEE Trans. Veh. Technol.*, vol. 59, no. 4, pp. 1823–1834, May 2010.
- [14] F. E. Lapicciarella, Z. Ding, and X. Liu, "Improved spectrum access control of cognitive radios based on primary ARQ signals," *IET Commun.*, vol. 6, no. 8, pp. 900–908, May 2012.
- [15] F. E. Lapicciarella, S. Huang, X. Liu, and Z. Ding, "Feedback-based access and power control for distributed multiuser cognitive networks," *Information Theory and Application (ITA) Workshop*, pp. 85–89, Feb. 2009.
- [16] W. Wang, A. Kwasinski, D. Niyato, and Z. Han, "A Survey on Applications of Model-Free Strategy Learning in Cognitive Wireless Networks," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 3, pp. 1717–1757, 3rd Quart. 2016.
- [17] S. B. Wicker, *Error Control Systems for Digital Communication and Storage*. Prentice-Hall, 1995.
- [18] K. P. Murphy, *Machine Learning: A Probabilistic Perspective*. Cambridge, MA, USA: MIT Press, 2012.
- [19] R. Waeber, P. I. Frazier, and S. G. Henderson, "Bisection Search with Noisy Responses," *SIAM J. Control Optim.*, vol. 51, no. 3, pp. 2261–2279, May 2013.
- [20] S. Boyd and L. Vandenberghe, "Localization and Cutting-Plane Methods," EE364b Lecture notes, Stanford University, Apr. 2008.
- [21] A. Tsakmalis, S. Chatzinotas, and B. Ottersten, "Active Interference Constraint Learning with Uncertain Feedback for Cognitive Radio Networks," in *Proc. IEEE International Conf. on Commun. (ICC)*, May 2016.
- [22] B. Grunbaum, "Partitions of mass-distributions and of convex bodies by hyperplanes," *Pacific Journal of Mathematics*, vol. 10, no. 4, pp. 1256–1261, Dec. 1960.
- [23] A. Caplin and B. Nalebuff, "Aggregation and Social Choice: A Mean Voter Theorem," *Econometrica*, vol. 59, no. 1, pp. 1–23, Jan. 1991.
- [24] S. Boyd and C. Barratt, *Linear Controller Design: Limits of Performance*. Prentice-Hall, 1991.
- [25] W. R. Gilks, S. Richardson, and D. J. Spiegelhalter, *Markov Chain Monte Carlo in Practice*. Boca Raton, FL: Chapman & Hall/CRC, 1996.
- [26] C. J. P. B elisle, H. E. Romeijn, and R. L. Smith, "Hit-and-run algorithms for generating multivariate distributions," *Math. Oper. Res.*, vol. 18, no. 2, pp. 255–266, May 1993.
- [27] H. Martini and A. Sch obel, "Median hyperplanes in normed spaces - a survey," *Discrete Applied Mathematics (Elsevier)*, vol. 89, pp. 181–195, Dec. 1998.
- [28] B. Yu and P. Mykland, "Looking at Markov Samplers through Cusum Path Plots: a single diagnostic idea," *Statist. Comput.*, vol. 8, no. 3, pp. 275–286, Aug. 1998.
- [29] M. Cowles and B. Carlin, "Markov Chain Monte Carlo Convergence Diagnostics: A Comparative Review," *Journal of the American Statistical Association*, vol. 91, no. 434, pp. 883–904, Jun. 1996.
- [30] D. P. Bertsekas, *Dynamic Programming and Optimal Control*. Belmont, MA: Athena Scientific, 1995.



**Anestis Tsakmalis** received the M.Eng. degree in telecommunications from Aristotle University of Thessaloniki, Thessaloniki, Greece, in 2009, the M.Sc. degree in telecommunications from University of Thessaly, Volos, Greece, in 2011 and he is currently pursuing his Ph.D. degree in electrical engineering from the Interdisciplinary Center for Security and Trust, (SnT), University of Luxembourg. His research interests include machine learning and statistics for wireless communications with focus on interference management and cognitive radios.



**Symeon Chatzinotas** is currently the Deputy Head of the SIGCOM Research Group, Interdisciplinary Centre for Security, Reliability, and Trust, University of Luxembourg, Luxembourg. He received the M.Eng. degree in telecommunications from Aristotle University of Thessaloniki, Thessaloniki, Greece, and the M.Sc. and Ph.D. degrees in electronic engineering from the University of Surrey, Surrey, U.K., in 2003, 2006, and 2009, respectively. In the past, he has worked on numerous R&D projects for the Institute of Informatics Telecommunications, National

Center for Scientific Research Demokritos, Institute of Telematics and Informatics, Center of Research and Technology Hellas, and Mobile Communications Research Group, Center of Communication Systems Research, University of Surrey, Surrey, U.K. Since 2004, he has authored more than 200 technical papers in refereed international journals, conferences and scientific books. He is the corecipient of the 2014 Distinguished Contributions to Satellite Communications Award, and Satellite and Space Communications Technical Committee, IEEE Communications Society, and CROWNCOM 2015 Best Paper Award. His research interests include multiuser information theory, cooperative/cognitive communications and wireless networks optimization.



**Björn Ottersten** was born in Stockholm, Sweden, in 1961. He received the M.S. degree in electrical engineering and applied physics from Linköping University, Linköping, Sweden, in 1986, and the Ph.D. degree in electrical engineering from Stanford University, Stanford, CA, USA, in 1989. He has held research positions at the Department of Electrical Engineering, Linköping University, the Information Systems Laboratory, Stanford University, the Katholieke Universiteit Leuven, Leuven, Belgium, and the University of Luxembourg, Luxembourg.

From 1996 to 1997, he was the Director of Research at ArrayComm Inc, a start-up in San Jose, CA, based on his patented technology. In 1991, he was appointed a Professor of Signal Processing with the Royal Institute of Technology (KTH), Stockholm, Sweden. From 1992 to 2004, he was the Head of the Department for Signals, Sensors, and Systems, KTH, and from 2004 to 2008, he was the Dean of the School of Electrical Engineering, KTH. Currently, he is the Director for the Interdisciplinary Centre for Security, Reliability and Trust, University of Luxembourg. As Digital Champion of Luxembourg, he acts as an Adviser to the European Commission. His research interests include security and trust, reliable wireless communications, and statistical signal processing. He is a Fellow of the EURASIP and a Member of the IEEE Signal Processing Society Board of Governors. He has served as an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING and on the Editorial Board of IEEE Signal Processing Magazine. He is currently Editor-in-Chief of EURASIP Signal Processing Journal and a Member of the Editorial Boards of EURASIP Journal of Applied Signal Processing and Foundations and Trends in Signal Processing. He has coauthored journal papers that received the IEEE Signal Processing Society Best Paper Award in 1993, 2001, 2006, and 2013 and three IEEE conference papers receiving Best Paper Awards. He was the recipient of the IEEE Signal Processing Society Technical Achievement Award in 2011. He was the first recipient of the European Research Council Advanced Research Grant.