Hybrid UCB-HMM: A Machine Learning Strategy for Cognitive Radio in HF Band

Laura Melián-Gutiérrez, Navikkumar Modi, Christophe Moy, *Member, IEEE*, Faouzi Bader, *Senior Member, IEEE*, Iván Pérez-Álvarez, *Member, IEEE*, and Santiago Zazo, *Member, IEEE*

Abstract—**Multiple users transmit in the HF band with worldwide coverage but collide with other HF users. New techniques based on cognitive radio principles are discussed to reduce the inefficient use of this band. In this paper, we show the feasibility of the Upper Confidence Bound (UCB) algorithm, based on reinforcement learning, for an opportunistic access to the HF band. The exploration vs. exploitation dilemma is evaluated in single-channel and multi-channel UCB algorithms in order to obtain their best performance in the HF environment. Furthermore, we propose a new hybrid system, which combines two types of machine learning techniques based on reinforcement learning and learning with Hidden Markov Models. This system can be understood as a metacognitive engine that automatically adapts its data transmission strategy according to HF environment's behaviour to efficiently use spectrum holes. The proposed hybrid UCB-HMM system increases the duration of data transmission's slots when conditions are favourable, and is also able to reduce the required signalling transmissions between transmitter and receiver to inform which channels have been selected for data transmission. This reduction can be as high as 61% with respect to the signalling required by multi-channel UCB.**

Index Terms—**Cognitive Radio, HF, Opportunistic Spectrum Access, Upper Confidence Bound, Hidden Markov Model.**

I. INTRODUCTION

THE HF band offers worldwide coverage without relying on satellite links, and includes the radio frequency band from 3 to 30 MHz. Beyond-line-of-sight (BLOS) links with remote sites can be established by using the ionosphere as a passive reflector but with variable propagation characteristics. This variability is dependent on the use of a natural medium formed by several ionised layers that change with solar radiation, time of day, season and sunspot cycle [1]. Nevertheless, it is the multiple collisions between HF users, including licensed

ones, not the propagation characteristics of the HF band, which induce an inefficient use of the spectrum resources in terms of satisfactory access to the HF band. All users around the world access this band of only 27 MHz bandwidth where trans-horizon links of thousand of kilometres can be established. Legacy users run the risk of colliding with other legacy users because HF frequency bands are usually allocated per country. These collisions need to be reduced, and a dynamic mechanism would not only achieve this goal but also increase the efficient use of the HF band in terms of satisfactory spectrum access without interfering with other active users in the band.

Although the Automatic Link Establishment (ALE) protocol [2] for HF radios has been presented as a primitive form of cognitive radio (CR) in [3], it could not be applied in its current form because CR requires a more dynamic mechanism. ALE is mainly used in military HF stations but is also present in some civilian stations. This protocol follows a listen-before-transmit strategy to sense the channels not in use for transmission and then ranks them according to their propagation characteristics. One drawback to this mechanism is that it can take several seconds to correctly establish a link in a single channel. Furthermore, it does not monitor users' activity in the channels in the recent past. Due to these limitations, we proposed in [4] and [5] to add some adaptability and cognition into the exploitation of the HF band to reduce its inefficient use.

Classical HF modems with single-carrier modulation such as [6] and [7] and current HF standards for link management as [8] depend on long interleavers of several seconds and cannot support a dynamic spectrum access as required by CR [3]. New modems and strategies are needed in order to successfully apply cognitive capabilities to HF systems. The HFDVL (HF Data+Voice Link) system, developed and evaluated with real HF links by the authors in [9], [10], and [11], brings a new communication paradigm in the HF band. This system is based on OFDM-CDM (Orthogonal Frequency Division Multiplexing - Code Division Multiplexing), which does not require inter-leavers, and has a point-to-point transmission delay of 125 ms with a net data rate of 2460 bps. Thus, it is an appropriate platform to implement cognitive based strategies such as the proposal of this paper.

In order to fulfil CR requirements, machine learning (ML) techniques can be used to predict primary users' activity in the surrounding environment, and to decide which channels are the most suitable to transmit. Two models for activity prediction in the HF band were developed with Hidden Markov Models (HMM) in [4], and with Neural Networks in [12]. Both models were trained with real measurements to predict the

activity of different HF users: amateur and non-amateur users in [4], and broadcasting users in [12]. These works differ from previous ML proposals where HF propagation characteristics have been widely modelled [13], [14].

The use of wideband HF transceivers has been specified in [15], instead of single-channel transceivers of 3 kHz bandwidth. An advantage of these transceivers is that they allow multi-band techniques with relatively low hardware complexity to be applied. In this CR context, the application of dynamic spectrum access [16] to the HF band could be carried out to allow access to unoccupied channels at a particular time and area. Different studies have presented and analysed the challenges and opportunities of applying CR to HF communications [17], [18] and [19], in which the authors proposed new specifications based on CR principles for future evolution of the ALE protocol to avoid collisions between HF users. However, to the best of the authors' knowledge, no proposals on dynamic spectrum access have been undertaken for the HF environment.

The HF activity prediction model in [4] was designed to represent all possible activity patterns present in the HF band. However, its straight application in a multi-band scenario for opportunistic spectrum access (OSA) requires as much HMM based prediction models as the number of channels to sense and transmit. Each prediction model learns from a channel and predicts the activity in that particular channel for the next minute. Therefore, it allows long-term data transmissions while the amount of transmissions needed for link management is reduced.

Furthermore, a strategy to select the best channels for data transmission is also required. This last limitation can be overcome if a reinforcement learning (RL) approach is jointly used with HMM based prediction models. In this paper, we propose a new hybrid scheme where two learning approaches are used to learn from the environment and to transmit in a short-term or a long-term basis according to the environment's conditions. A short-term learning approach based on Upper Confidence Bound (UCB) and a long-term learning approach based on HMM are developed and analysed. The proposed approach will ensure that, whenever the activity in the band decreases, longer data transmissions can be established whereas the amount of transmissions for link management will be reduced. This hybrid UCB-HMM proposal can be considered as a metacognitive radio engine [20], where the HMM based prediction model and UCB algorithm are cognitive engines working in parallel and each one will be selected according to the environment's behaviour.

The goal of this work is twofold; to show the feasibility of RL in a real environment, such as the HF band, and to propose a new hybrid system which combines two types of ML techniques, reinforcement learning (RL) and learning with HMMs. The main goals of this new hybrid system are: to adapt data transmission's slots to the environment's behaviour, to reduce the amount of channel signalling for link management, and to reduce the complexity of multiple HMMs working in parallel. The HF band is selected in this work due to its variability and high activity load, where multiple transmission patterns are present. Moreover, the availability of the HF spectrum activity database HFSA_IDeTIC_F1_V01 [4], which contains wideband measurements of the HF band including several types

of transmissions, allows us to show the feasibility of RL in this band and to validate the proposed hybrid scheme in a real environment.

The first part of this paper deals with the feasibility of Upper Confidence Bound algorithm in the HF environment. A description of the UCB algorithm is presented in Section II whereas its validation is detailed in Section III, including the analysis and validation of the exploration vs. exploitation dilemma for single and multiple channel UCB. The second part of the paper starts at Section IV, where the new proposed hybrid UCB-HMM system is described and analysed in a multi-band scenario. Finally, the main analyses and conclusions of the proposed work are drawn in Section V.

II. UPPER CONFIDENCE BOUND ALGORITHM

The UCB algorithm is one of the RL algorithms that was introduced in [21] and [22] as an approach for solving Multi-Armed Bandit (MAB) problems [23]. The MAB problem is analogous to the traditional slot machine except that it has more than one lever. In [24], the UCB algorithm was proposed for the CR environment in order to provide secondary users with opportunistic spectrum access (OSA) through a set of channels following the MAB paradigm. In this paradigm, the spectrum is divided into a set of N channels, each one with the same bandwidth and modulation scheme, and each represents one lever of the MAB. At each time slot, the RL algorithm plays one or multiple levers and obtains a reward from each one.

The UCB algorithm allows decision making in an OSA cognitive radio context to maximise the transmission opportunities of secondary users. Since it is based on RL, it learns from previously observed rewards starting from scratch, e.g. without any a priori knowledge of the activity within the set of channels. The UCB algorithm continuously learns (in the exploration phase) and predicts the next available channel to transmit on within the set of channels (in the exploitation phase) during the entire process. One of the advantages of the UCB algorithm is that the exploitation phase starts at the beginning of the process even when the knowledge is not mature enough [25]. Reciprocally, the exploration is maintained along the overall process, even during the exploitation phase. Thus, both exploration and exploitation are continuously superposed for improved management of the exploration versus exploitation dilemma.

At each time slot t, the algorithm updates UCB indices named as $B_{t,k,T_k(t)}$, where $T_k(t)$ is the number of times channel k has been selected by the algorithm in previous slots, and returns the channel index k of the maximum UCB index (see Algorithm 1). From the different UCB indices existing in the literature, UCB$_1$ index [22] is computed in this paper and evaluated with one (UCB$_1$) or multiple channel (UCB$_1$-M, known as UCB$_1$ - multiple plays) selection of the most available channels within the set. Other algorithms based on the computation of UCB indices could also have been considered but the use of UCB$_1$ produces no loss of generality. The UCB$_1$ index is defined as follows:

$$B_{t,k,T_k(t)} = \overline{X}_{k,T_k(t)} + A_{t,k,T_k(t)}, \qquad (1)$$

Algorithm 1. UCB$_1$ algorithm

Inputs: $N, \alpha, \{a_0, r_0, a_1, r_1, \ldots, a_{t-1}, r_{t-1}\}$
Outputs: a_t
1: **loop**
2:    **if** $t \leq N$ **then**
3:       $a_t = t + 1$
4:    **else**
5:       $T_k(t) \leftarrow \sum_{m=0}^{t-1} \mathbf{1}_{\{a_m=k\}}, \forall k$
6:       $\overline{X}_{k,T_k(t)} \leftarrow \frac{\sum_{m=0}^{t-1} r_m \mathbf{1}_{\{a_m=k\}}}{T_k(t)}, \forall k$
7:       $A_{t,k,T_k(t)} \leftarrow \sqrt{\frac{\alpha \ln(t)}{T_k(t)}}, \forall k$
8:       $B_{t,k,T_k(t)} \leftarrow \overline{X}_{k,T_k(t)} + A_{t,k,T_k(t)}, \forall k$
9:       $a_t = \arg\max_k \left(B_{t,k,T_k(t)} \right)$
10:   **end if**
11:   **return** a_t
12: **end loop**

where $\overline{X}_{k,T_k(t)}$ and $A_{t,k,T_k(t)}$ are two terms that represent the exploitation and the exploration contributions, respectively. $\overline{X}_{k,T_k(t)}$ is the empirical mean and it is defined as

$$\overline{X}_{k,T_k(t)} = \frac{\sum_{m=0}^{t-1} r_m \mathbf{1}_{\{a_m=k\}}}{T_k(t)}, \qquad (2)$$

where r_m and a_m are the achieved reward and selected channel, respectively, at the m-th time slot, and the indicator function is as follows

$$\mathbf{1}_{\{a_m=k\}} = \begin{cases} 1 & \text{if } a_m = k \\ 0 & \text{if } a_m \neq k \end{cases} \qquad (3)$$

The empirical mean is also proportional to the accumulated throughput by the UCB$_1$ algorithm when channel k has been selected. $A_{t,k,T_k(t)}$ is the UCB$_1$ bias [22] defined as follows

$$A_{t,k,T_k(t)} = \sqrt{\frac{\alpha \ln(t)}{T_k(t)}}, \qquad (4)$$

where α is the exploitation-exploration factor of the algorithm. If α increases, UCB$_1$ bias $A_{t,k,T_k(t)}$ dominates and UCB$_1$ algorithm explores new channels. Otherwise, if α decreases, UCB$_1$ bias also does and $\overline{X}_{k,T_k(t)}$ dominates the UCB$_1$ index $B_{t,k,T_k(t)}$ forcing the algorithm to mostly exploit previous channels [24].

III. UCB$_1$ VALIDATION IN THE HF ENVIRONMENT

Before going into the details of the combination of both UCB$_1$ and HMM approaches in Section IV, we discuss hereafter the feasibility of RL in the HF environment, particularly the performance of the Upper Confidence Bound (UCB$_1$) and multi-channel Upper Confidence Bound (UCB$_1$-M) algorithms previously described in Section II.

A. HF Database Parameters

In this paper, we use the same HF spectrum activity database HFSA_IDeTIC_F1_V01 used in [4] to train and validate the HMM based prediction model in the HF environment. This

TABLE I
MAIN PARAMETERS OF THE HF DATABASE IN [4]

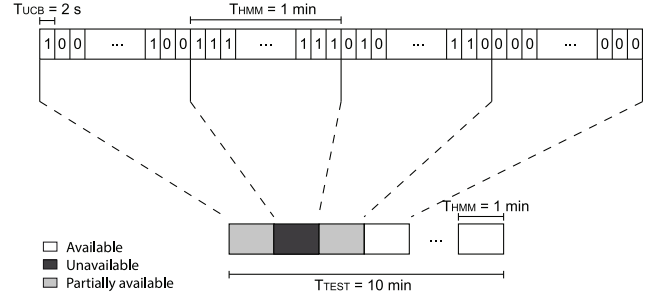| HF database Parameters	
Number of measurements	63
Frequency location	13.875 MHz - 14.475 MHz
Bandwidth	600 kHz (200 channels of 3 kHz each)
Time duration (T_{TEST})	10 min
Segmentation time for HMM (T_{HMM})	1 min
Sample time slot (T_{UCB})	2 s

Fig. 1. Data segmentation and classification in the HF database [4].

HF database consists of 63 wideband measurements of the 14 MHz band where amateur and non-amateur HF stations are operating during a testing time (T_{TEST}) of 10 minutes. Measurements were undertaken during weekdays and weekends, when amateur contests were scheduled. The database includes real measurements with different degrees of spectrum activity in the HF band. The main parameters of this HF database [4] are described in Table I.

According to the used energy detector, these measurements are divided in samples with '0' and '1' values that indicate whether the channel is free or occupied, respectively. The sample time slot is 2 seconds and is referred to as T_{UCB} in this paper since UCB$_1$ and UCB$_1$-M algorithms will be executed every T_{UCB}. Furthermore, due to the structure defined in the HMM based prediction model, each measured channel over a slot of 1 minute (referred hereafter in this paper as T_{HMM}) is classified based on the following categories: available, unavailable, and partially available channels [4] as indicated in Fig. 1.

From the HF database in [4], we have selected those measurements with the worst conditions in terms of availability to validate the performance of UCB$_1$ and UCB$_1$-M algorithms. Fig. 2 shows an example of these measurements acquired during several weekends which corresponds to the High Activity scenario classified in the HF database in [4]. The selection of measurements for the validation of UCB$_1$ and UCB$_1$-M algorithms includes amateur channels of 3 kHz each located from 14.07 MHz to 14.14 MHz (see Fig. 2), and recorded when this amateur band was heavily occupied.

B. Analysed Approaches for Performance Evaluation

UCB$_1$ and UCB$_1$-M performances are evaluated from a radio communication point of view in a CR context instead of the metrics typically utilised in the ML community. It is worth highlighting that in a ML context, the final goal is to find and
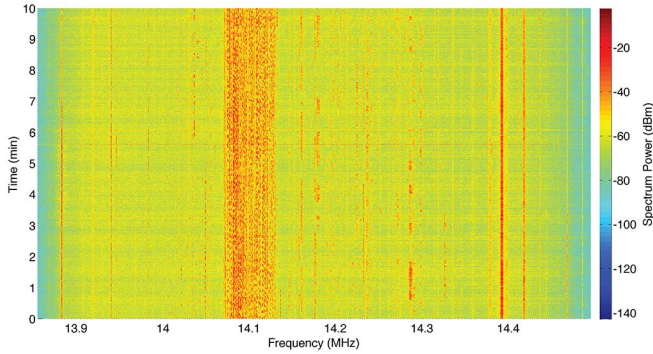
Fig. 2. Example of a measurement of the High Activity scenario (acquired at weekends) of the HF database [4].

to subsequently exploit the best channel in terms of availability, whereas in a CR context the goal is to find and to exploit any possible available channel to efficiently use the spectrum opportunities. Thus, the selection of the best available channel is not as important as the selection of any available channel at each time slot to establish the communication link. Considering the CR context, new performance metrics for RL algorithms were presented in [26] in terms of transmission opportunities effectively obtained and utilised for data transmission. These new metrics are based on the percentage of time that the RL algorithm selects an available channel, named as 'percentage of successful trials', which can be also seen as the successful transmission rate obtained by the RL algorithm.

In this paper, the performance evaluations of UCB_1 and UCB_1-M are based on the percentage of successful trials achieved within a set of channels. A 'successful transmission rate' is defined as the percentage of transmission opportunities effectively detected and used to transmit by a secondary user during the whole testing time period. The following approaches have been defined to facilitate comparisons:

- *Uniform random selection*: the first question that we should answer is: 'Is it worth applying RL for decision making in the HF environment?' If so, 'how much can we improve the performances compared to a non intelligent approach?' Thus, the successful transmission rate achieved by a uniform random channel selection in each set of channels is computed to establish a reference level where no learning is used.
- *Best channel*: from the set of evaluated channels, it is the channel with the longest availability time within the overall testing time (10 minutes), which is considered as the maximum performance achievable under the ML perspective.
- *Worst channel*: it is the channel with the shortest availability time within the overall testing time from the set of evaluated channels.
- *Best opportunistic selection*: Since the algorithm is evaluated for a CR application, the performances of UCB_1 and UCB_1-M algorithms will be compared with a kind of genie-aided policy, i.e. a policy with a perfect knowledge of the environment's behaviour, where all spectrum holes within the set of evaluated channels are used for transmission achieving a more efficient use of the band. In this

approach, the system will use any detected free channel, independently of being or not the channel with the longest availability time.

As previously stated, the UCB_1 algorithm can select one or multiple channels (UCB_1-M) at each time slot. The sections that follow will analyse the behaviour of this algorithm using single and multiple channel selection strategies.

C. The Exploration vs. Exploitation Dilemma in Single-Channel UCB_1

Two parameters must be set before executing the UCB_1 algorithm: the total number N of channels considered within a set, e.g. the number of arms in a MAB approach, and the exploitation-exploration factor α. The value of α is chosen depending on the environment's conditions in terms of activity or presence of primary users, whereas N is determined by the final system restrictions such as the number of channels to explore or the limited number of assigned channels. In this section, we will focus on how the performance of the UCB_1 algorithm with a single-channel selection varies according to N and α parameters, i.e. the exploration vs. exploitation dilemma.

The *best opportunistic selection* approach, i.e. the genie-aided policy, is used for numerical comparisons of the UCB_1 algorithm performance with the other approaches given in Section III-B. This approach is selected as a reference of the best possible performance because it has prior knowledge of the band activity. Therefore, the percentage of successful transmission rate achieved by each approach after 10 minutes of testing time (T_{TEST}) is computed with respect to the successful transmission rate of the *best opportunistic selection* after T_{TEST}, i.e. the maximum achievable successful transmission rate. As stated previously in Table I, the maximum testing time is about 10 minutes according to T_{TEST} of the HF database in [4]. This time period corresponds to 300 samples of 2 seconds that could not be considered as a large enough amount of samples from a ML perspective which allow for convergence. However, it has been stated and experimentally verified in [25] that in an OSA context UCB_1 learning is sufficient after a small number of time slots.

Fig. 3 depicts the obtained results when comparing the UCB_1 algorithm and the *best opportunistic selection* for $\alpha \in [0, 5]$, and sets of $N = \{4, 8, 16\}$ channels. Note that for different values of N the best performance is reached at $\alpha = 0.4$. When $N = 4$, UCB_1 can achieve a 81% of the successful transmission rate achieved by *best opportunistic selection* for $\alpha \in [0.4, 1.8]$, and it remains close to this maximum for $\alpha > 1.8$. Nevertheless, when the number of channels within a set increases to $N = 8$ or $N = 16$, the number of channels that UCB_1 has to explore also increases. Thus, if the exploitation-exploration factor α is high, the UCB_1 algorithm will tend to explore more than to exploit the channels that have been previously labelled as available. This fact explains why, when $N = 8$, the best performance (77%) is achieved for $\alpha \in [0.3, 0.4]$, whereas it reaches 79% for $N = 16$ and $\alpha \in [0.2, 0.4]$.

Best channel selection, *uniform random selection*, and *worst channel selection* approaches are also compared in terms of mean percentage of achieved successful transmission rate with
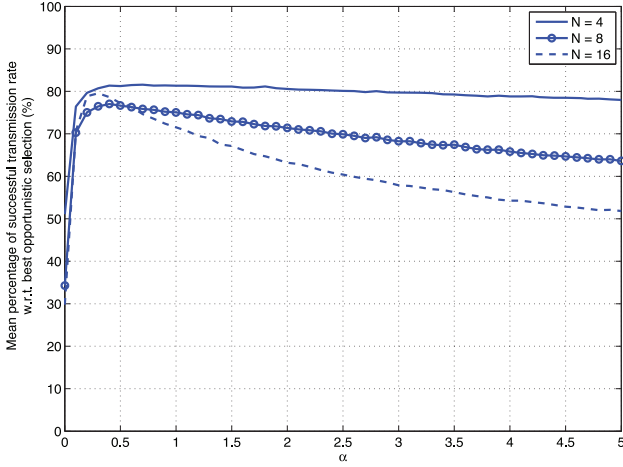
Fig. 3. UCB_1: Mean percentage of successful transmission rate with respect to best opportunistic selection approach vs. α with $N = \{4, 8, 16\}$.

Fig. 4. Percentage of improvement using UCB_1 with respect to uniform random selection approach vs. α with $N = \{4, 8, 16\}$.

TABLE II

SINGLE-CHANNEL SELECTION: MEAN PERCENTAGE OF SUCCESSFUL TRANSMISSION RATE ACHIEVED BY EACH APPROACH WITH RESPECT TO BEST OPPORTUNISTIC SELECTION APPROACH

	$N = 4$	$N = 8$	$N = 16$
UCB_1 ($\alpha = 0.4$)	81%	77%	79%
Best channel selection	83%	81%	88%
Uniform random selection	53%	39%	35%
Worst channel selection	27%	9%	3%

respect to the *best opportunistic selection* approach in Table II. Both UCB_1 and *best channel selection* approach improve on in these terms the performance of *uniform random selection* and *worst channel selection* approaches.

Another metric that reveals how the UCB_1 algorithm performs in a CR context is the mean percentage of improvement in the successful transmission rate with respect to a uniform random selection of channels, since it also reflects the improvement of applying learning in OSA for HF environment. This is analysed in Fig. 4 with $\alpha \in [0, 5]$, and with sets of $N = \{4, 8, 16\}$ channels. The mean percentage of improvement when the size of the set of channels is small ($N = 4$) remains close to 178% for any value of α within the interval $[0.4, 5]$, which is nearly twice as much as that with respect to *uniform random selection* approach. However, the maximum percentage of improvement is reached at $\alpha = 0.4$, with 228% for $N = 8$ and 245% for $N = 16$. This behaviour exhibits the same trend as obtained performances in terms of mean percentage of successful transmission rate with respect to *best opportunistic selection* approach in Fig. 3. If there is a small set of channels, then there are fewer to explore and the variability in the exploitation-exploration factor α does not strongly influence UCB_1 performance. Otherwise, when the size of the set of channels increases, the UCB_1 algorithm will tend to explore more as α increases, thus, it results in a decrease in the percentage of improvement with respect to a uniform random selection.

Note that in Fig. 3 and Fig. 4 the exploration vs. exploitation dilemma is highly dependent on the size of the set of channels N. As N increases, the sensitivity of α is higher, the UCB_1
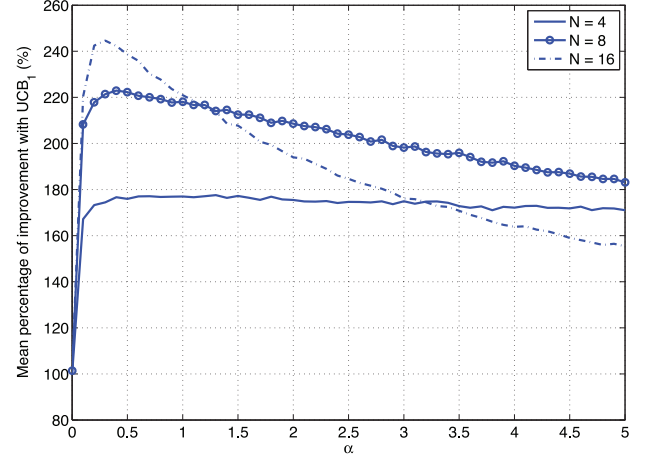
algorithm is forced to explore more channels and the transmission opportunities are reduced. Due to this fact, in Fig. 3 and Fig. 4, as N increases, the interval of values of α where the best performance is reached is also reduced at α close to 0.4.

D. The Exploration vs. Exploitation Dilemma in Multi-Channel UCB_1 (UCB_1-M)

This Subsection analyses how the UCB_1 algorithm with multiple-channel selection (UCB_1-M), known as UCB_1-multiple plays, behaves. The same performance metrics and strategies followed in Subsection III-C are used for the validation of the UCB_1 algorithm with multi-channel selection. An evaluation of the mean percentage of successful transmission rate achieved by each approach defined in Subsection III-B compared to the best possible performance, i.e. *best opportunistic selection* approach, is performed. Unlike the UCB_1 algorithm in the previous analysis where two parameters were defined: N and α, a third parameter is introduced here which is the number of channels M (being $M \leq N$) that will be selected by the algorithm to transmit.

The obtained results for the UCB_1-M algorithm in the HF environment in terms of mean percentage of successful transmission rate with respect to *best opportunistic selection* are depicted in Fig. 5. These results are represented for different values of $\alpha \in [0, 5]$, and ratios of $M/N = \{1/8, 1/4, 1/2\}$ with $12 \leq N \leq 30$. When the ratio M/N is high ($M/N = 1/2$), i.e. a higher percentage of channels of N can be jointly selected by the algorithm, the maximum percentage of successful transmission rate achieved by UCB_1-M is equal to 88% of the achieved successful transmission rate achieved by *best opportunistic selection* for values of α within the studied interval. Therefore, we can conclude that the UCB_1-M performance does not strongly depend on α for $M/N = 1/2$. However, when M/N decreases because there are fewer channels for data transmission that can be chosen by the UCB_1-M algorithm ($M/N = 1/4$ and $M/N = 1/8$), the maximum performance of UCB_1-M algorithm is achieved at $\alpha \in [0.2, 0.4]$ when $M/N = 1/8$, and $\alpha = 0.4$ if $M/N = 1/4$. It is worth noting that this behaviour is identical for single-channel selection UCB_1 in Fig. 3 when
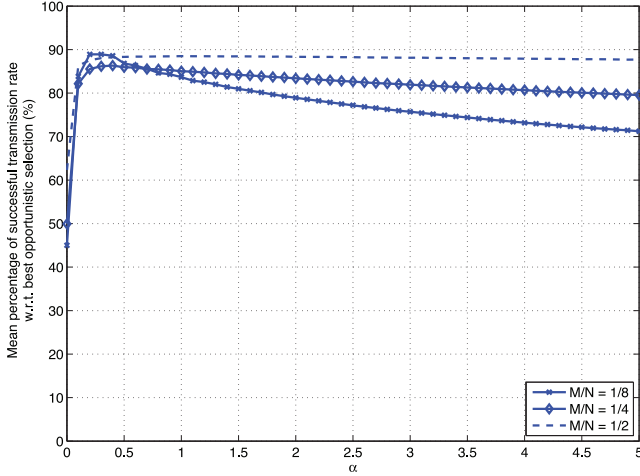
Fig. 5. UCB$_1$-M: Mean percentage of successful transmission rate with respect to best opportunistic selection approach vs. α with $M/N = \{1/8, 1/4, 1/2\}$.

Fig. 6. Percentage of improvement using UCB$_1$-M with respect to uniform random selection approach vs. α with $M/N = \{1/8, 1/4, 1/2\}$.

TABLE III
MULTI-CHANNEL SELECTION: MEAN PERCENTAGE OF SUCCESSFUL
TRANSMISSION RATE ACHIEVED BY EACH APPROACH COMPARED
W.R.T. BEST OPPORTUNISTIC SELECTION

	$M/N = 1/8$	$M/N = 1/4$	$M/N = 1/2$
UCB$_1$-M ($\alpha = 0.4$)	88%	86%	88%
Best channel selection	100%	92%	92%
Uniform random selection	48%	52%	64%
Worst channel selection	4%	14%	37%

the size of the set of channels N increases. In this case, the UCB$_1$-M algorithm has a larger set of channels to explore and if α increases, it is forced to explore instead of exploiting the channels previously labelled as available, and consequently will lose transmission opportunities. This phenomenon is shown in Fig. 5 for $\alpha = 5$, where the mean percentage of successful transmission rate achieved by UCB$_1$-M with $M/N = 1/4$ decreases by 6% compared to its maximum at $\alpha = 0.4$, and it decreases by 17% compared to its maximum at $\alpha \in [0.2, 0.4]$ when $M/N = 1/8$.

In Table III, the performance of the UCB$_1$-M algorithm is also compared to the *best channel selection*, *uniform random selection*, and *worst channel selection* approaches. It can be observed that *best channel selection* performs similarly to the *best opportunistic selection* because when $M > 1$, these sets of M channels contain channels that are available during the whole testing time, therefore, both achieve the maximum performance. Furthermore, similar to UCB$_1$ results in Table II, both UCB$_1$-M and *best channel selection* improve on the performance of *uniform random selection* and *worst channel selection* approaches.

Finally, we present how performance improves when learning is applied to OSA for the HF spectrum. The mean percentage of improvement using UCB$_1$-M with respect to a uniform random selection of channels is shown in Fig. 6, and has been computed as the percentage of improvement in terms of achieved successful transmission rates for values of $\alpha \in$
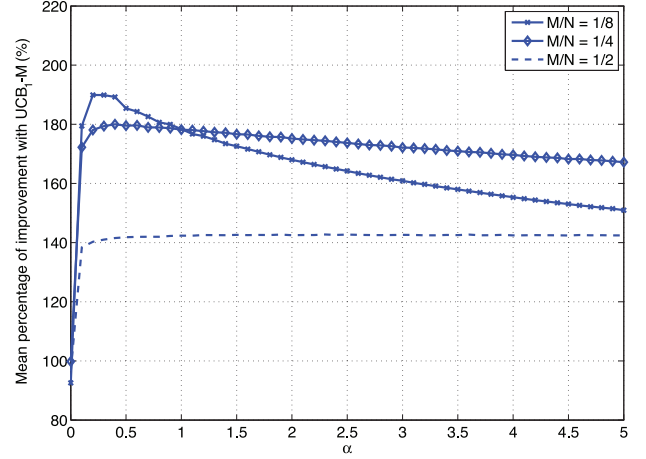
$[0, 5]$ and $M/N = \{1/8, 1/4, 1/2\}$. When the number of channels that can be selected by UCB$_1$-M algorithm is high, i.e. $M/N = 1/2$, UCB$_1$-M outperforms the uniform random selection by 140% when $\alpha \in [0.4, 5]$. As the M/N ratio decreases, the percentage of improvement increases. When $M/N = 1/4$ the percentage of improvement is equal to 180% for $\alpha \in [0.3, 0.7]$, whereas for $M/N = 1/8$, the maximum percentage of improvement is reached for $\alpha \in [0.3, 0.4]$ and is equal to 190%, being the successful transmission rate of UCB$_1$-M algorithm nearly three times higher than the successful transmission rate of the uniform random selection.

As seen in Fig. 5 and Fig. 6, the UCB$_1$-M algorithm performs best in terms of successful transmission rate when the exploitation-exploration factor α is close to 0.4 and $M/N = 1/8$. This will hereafter be exploited in the proposed joint combination of HMM and UCB$_1$ learning methods.

IV. A HYBRID LEARNING SOLUTION FOR COGNITIVE RADIOS IN THE HF ENVIRONMENT

Signalling is required to coordinate any transmission and reception processes, and this is essentially the role of link management. Both sides of the communication link can be interconnected for link management with an in-band or an out-band channel. CR systems do require link management, for instance a Cognitive Pilot Channel (CPC) has been proposed in [27]. This link management could be an in-band or out-band channel that informs secondary users about the channel status. In this paper, we also consider channel signalling to coordinate transmitter and receiver sides in HF communications. Due to HF communication rate limitation and its link instability, whatever the adopted signalling solution is, for instance CPC, or other suitable mechanisms (this is out of the scope of this paper), the efficient management of the amount of signalling information to be exchanged is highly recommended. As will be detailed in Section IV-B4, the channel signalling load of this proposal is affordable for use in this environment. It is worth noting that in this paper, channel signalling is defined as the amount of transmissions needed to inform the receiver about the selected channels for data transmission.

This aspect is significant and has not been previously considered in Sections III-C and III-D for the validation of UCB_1 and UCB_1-M algorithms in HF band.

The HMM based prediction model proposed in [4] was designed to make long-term predictions (1 minute duration) with low complexity in a specific HF channel. By making long-term predictions, channel signalling can be significantly decreased with respect to short-term prediction methods such as UCB since it is only updated once in a period of one minute. Furthermore, longer data transmissions can be established according to these long-term predictions. Nevertheless, multiple HMMs are required when we want to sense and predict multiple channels. In other words, if we wish to sense N channels, N parallel HMMs are required to obtain the predictions in the complete set of channels. It is worth mentioning that the HMM based prediction model was designed as a single-channel prediction model due to its simplicity, while one HMM based prediction model that learns multiple channels simultaneously is more complex than multiple HMM based prediction models working in parallel [4].

We propose a hybrid solution combining learning with HMM and reinforcement learning with UCB_1-M, whose goals are: (1) to decrease the amount of signalling information exchanged between transmitter and receiver, (2) to reduce the complexity of N HMM based models working in parallel, and (3) to adapt slots of data transmission to the behaviour of the environment. This hybrid solution can be developed by using modern HF systems with multi-carrier modulation, which support short delays and moderated bit rates, as proposed in [9], [10], and [11], combined with link management protocols with lower overhead than current HF standards.

A. Hybrid UCB-HMM System Description

The UCB_1-M and HMM learning methods run in parallel in the proposed scheme over two transmission slots; a short-term slot of 2 seconds with UCB_1-M (T_{UCB}), and a long-term slot using HMM predictions of 1 minute duration (T_{HMM}). The same data segmentation of the HF database in [4] (see Fig. 1) is hereafter considered.

Achievable data rates in the HF environment are very small, which make longer transmissions preferable when sending higher amounts of data. The proposed scheme is able to automatically adapt its configuration to the changes in channels' activity, which means that if there are plenty of available channels for transmission during a minute, the hybrid UCB-HMM system will switch to long-term transmission slots following the predictions of M HMMs working in parallel (M-HMMs), being $M < N$. Nevertheless, if the environment changes and most of the channels are unavailable or partially-available, it will transmit in a short-term slot following the decisions of the UCB_1-M algorithm. This is indeed a CR procedure that selects different cognitive engines according to the conditions of the environment, e.g. a metacognitive radio engine [20].

If only short-term predictions and transmissions are accomplished, the channel signalling used for link management between transmitter and receiver will increase. This increase occurs if UCB_1-M is only used in the system, as it has to predict

---

Algorithm 2. Hybrid UCB-HMM algorithm

Inputs: N, M, α

Init.: $M_{UCB} = M$, $M_{HMM} = 0$, slot = 0, $T_{UCB} = 2$ s, $T_{HMM} = 1$min, $ID_{UCB} = \varnothing$, $ID_{HMM} = \varnothing$

1: **loop**
2:     **if** $t = slot + T_{HMM}$ **then**
3:         $slot \leftarrow t$
4:         **if** $ID_{UCB} \neq \varnothing$ **then**
5:             Predict with M-HMMs
6:             $ID_{HMM} \leftarrow$ indices (channels predicted as free)
7:             $M_{HMM} \leftarrow$ size (ID_{HMM})
8:             $M_{UCB} \leftarrow M - M_{HMM}$
9:             M-HMMs **SEND** UCB_1-M: M_{UCB}, ID_{HMM}
10:         **end if**
11:         $ID_{UCB} \leftarrow$ indices (M highest $B_{t,k,T_k(t)}$)
12:         UCB_1-M **SENDS** M-HMMs: M, ID_{UCB}
13:     **end if**
14:     $t \leftarrow t + T_{UCB}$
15:     Compute UCB_1-M
16:     **TX** in M_{UCB} channels $\in N \setminus ID_{HMM}$
17:     **if** $ID_{UCB} \neq \varnothing$ **then**
18:         M-HMMs save sensing information of ID_{UCB} channels
19:         **if** $ID_{HMM} \neq \varnothing$ **then**
20:             **TX** in M_{HMM} channels $\in ID_{HMM}$
21:         **end if**
22:     **end if**
23: **end loop**

---

channels' activity every 2 seconds (T_{UCB}) and, therefore, it will require a higher number of signalling transmissions. However, in the proposed hybrid scheme, channel signalling will decrease when it switches to M-HMMs to transmit according to their predictions over a minute (T_{HMM}). For each channel predicted as available in the next minute by a HMM, the transmitter will send selected channel frequency to the receiver only once during this one minute slot, thereby reducing thirty times (T_{HMM}/T_{UCB}) the channel signalling and, therefore, increasing feedback information capacity. This reduction factor remains at a constant value throughout the entire execution of the proposed hybrid scheme since M-HMMs are trained before their use for prediction [4] and there is no added workload.

Furthermore, the hybrid UCB-HMM allows for a reduction in the complexity of N HMM working in parallel by combining two learning methods. This combination only requires M HMM with $M < N$ and UCB_1-M algorithm to select the M best channels in terms of availability from a larger set of N channels, which is not nearly as complex [5]. It will be shown that the best of both learning methods will be exploited in this proposal.

The hybrid UCB-HMM system is described in Algorithm 2 and Fig. 7, where a simplified diagram is depicted. During its execution, the UCB_1-M algorithm selects the M best channels with short-term predictions in T_{UCB} slots, and subsequently, these channels are learnt by M-HMMs to predict their activity in a long-term slot, T_{HMM}. Hence, M-HMMs' effort is optimised as M-HMMs are executed in the M best available
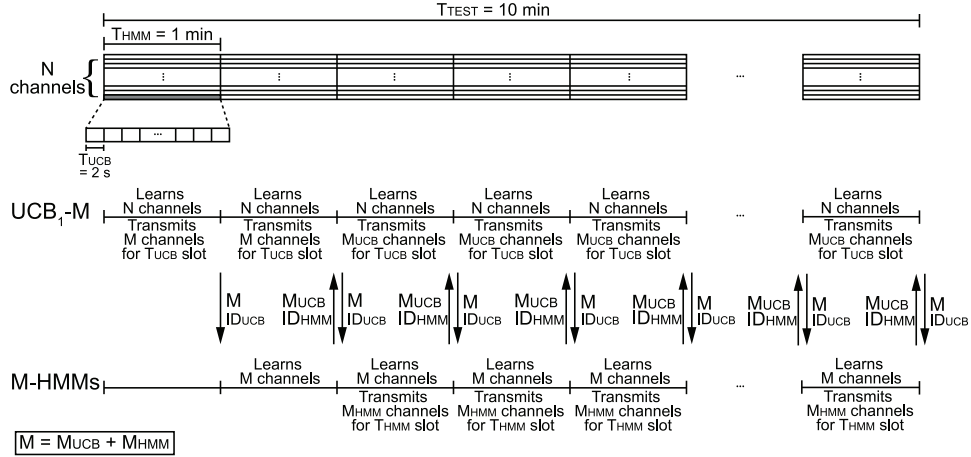
Fig. 7. Time execution of the hybrid UCB-HMM system with M HMM based prediction models and the UCB$_1$-M algorithm.

channels from the complete set of N channels, e.g. in those channels with the longest availability time during next T$_{HMM}$ to establish longer data transmissions.

Two sets are modified during the execution of the hybrid UCB-HMM system to avoid any conflict between both learning methods: ID$_{UCB}$, which contains the M best channels selected by UCB$_1$-M, and ID$_{HMM}$, which includes channels predicted by M-HMMs as available for the next T$_{HMM}$ slot. Both learning methods update and transmit their sets to each other at every T$_{HMM}$ according to their predictions as described in lines 9 and 12 of Algorithm 2.

There is an initialization phase where only UCB$_1$-M is learning during T$_{HMM}$ over N channels until it extracts which are the M best available channels (see line 11 of Algorithm 2 and Fig. 7). After the first T$_{HMM}$ slot, it transmits the M best channels in the set ID$_{UCB}$ to M-HMMs. During the second T$_{HMM}$ slot (see Fig. 7), each HMM will learn one of the M best channels to predict which of them will be available, partially available or unavailable for the next T$_{HMM}$ slot. While M-HMMs are learning, UCB$_1$-M continues its execution and transmits in at most M channels in T$_{UCB}$ slots. Once M-HMMs have learnt for a T$_{HMM}$ slot, ID$_{HMM}$ set is modified to contain those channels predicted as available for the next T$_{HMM}$ slot, and sent to UCB$_1$-M algorithm jointly with the maximum number of channels M$_{UCB}$ where UCB$_1$-M can transmit in the next T$_{HMM}$ slot (see lines 4-10 of Algorithm 2). At this instant (at the beginning of the third T$_{HMM}$ slot) both learning methods start transmitting in their corresponding channels, M-HMMs in M$_{HMM}$ channels from the set ID$_{HMM}$ and UCB$_1$-M in M$_{UCB}$ channels from the relative complement of ID$_{HMM}$ in N ($N \setminus$ ID$_{HMM}$). Thus, the hybrid UCB-HMM system always tries to transmit in the M maximum number of assigned channels. Note that M-HMMs execution is restricted to 8 minutes (T$_{TEST}$ - 2·T$_{HMM}$) due to the initialization phase of the proposed hybrid UCB-HMM system shown in Fig. 7.

B. Performance Evaluation

The performance of the proposed hybrid UCB-HMM system is hereafter evaluated by depicting the benefits of the combination of the UCB$_1$-M and HMM learning methods.
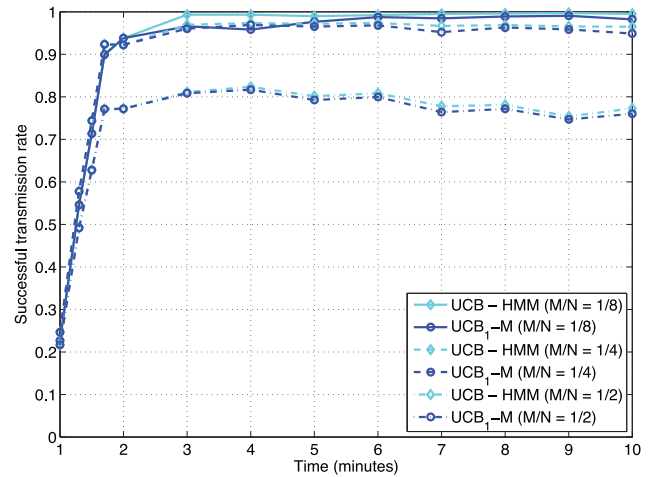

Fig. 8. Successful transmission rate of hybrid UCB-HMM system and UCB$_1$-M with $\alpha = 0.4$ and $M/N = \{1/8, 1/4, 1/2\}$.

1) Successful Transmission Rate: The validation of the UCB$_1$-M algorithm in Section III-D was carried out based on successful transmission rates, attaining its best performance for $\alpha = 0.4$ and $M/N = 1/8$. Fig. 8 depicts the comparison of the achieved successful transmission rate after each T$_{HMM}$ slot by the hybrid UCB-HMM system and the UCB$_1$-M algorithm with $\alpha = 0.4$ and $M/N = \{1/8, 1/4, 1/2\}$. As described in Fig. 7, T$_{TEST}$ is restricted to 10 minutes and long-term data transmissions according to M-HMMs predictions, which start after 2·T$_{HMM}$ slots (2 minutes). Note that, as seen in Fig. 8, the UCB$_1$-M and the hybrid UCB-HMM scheme learning gradually increases and so does the successful transmission rate from minute 1 to minute 1.7. At this point, both UCB$_1$-M and the proposed hybrid scheme are close to their maximum performances for $M/N = \{1/8, 1/4, 1/2\}$. The best successful transmission rate is achieved by the proposed hybrid UCB-HMM with $M/N = 1/8$, since it exploits 99% of the transmission opportunities from the third minute when M-HMMs start transmitting and the learning process is almost done (see Fig. 8 at minute 3). This successful transmission rate is 4% higher than that achieved when $M/N = 1/4$, and 20% higher than that achieved
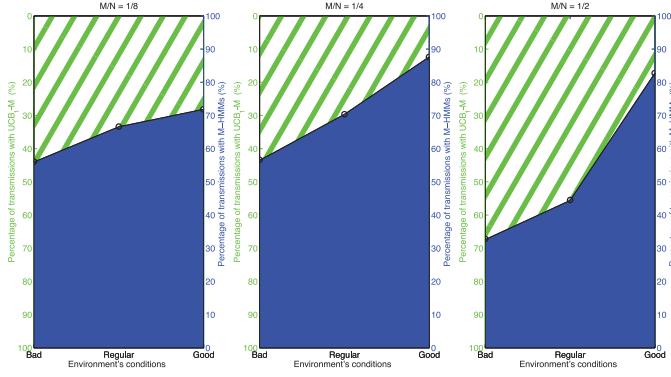
Fig. 9. Percentage of long-term (M-HMMs) and short-term (UCB$_1$-M) transmissions in proposed hybrid UCB-HMM system with $M/N = \{1/8, 1/4, 1/2\}$ vs. environment's conditions.

Fig. 10. Histogram of data transmission's slots according to UCB$_1$-M predictions in the proposed hybrid UCB-HMM system vs. slots' duration.

Fig. 11. Histogram of data transmission's slots according to M-HMMs predictions in the proposed hybrid UCB-HMM system vs. slots' duration.

when $M/N = 1/2$. Note that global performance of the proposed hybrid UCB-HMM is highly dependent on the initial selection of the M best channels made by the UCB$_1$-M algorithm and, as shown in Section III-D for UCB$_1$-M, the best performance is achieved for $\alpha = 0.4$ and $M/N = 1/8$.

2) Metacognitive Strategy: As previously stated, the proposed hybrid UCB-HMM solution exhibits the same behaviour as a metacognitive radio engine [20] since it changes its learning method, also known as a cognitive engine, based on changes in the environment. The proposed solution switches from short-term transmissions to long-term transmissions when there is an improvement in the environment's conditions as regards availability. Nevertheless, if the environment's conditions worsen, it will switch from long-term transmissions to short-term transmissions. It is noteworthy that this adaptability enhances the proposed hybrid UCB-HMM performance to meet its goals, namely to transmit data in long-term slots when the environment is favourable and to reduce the amount of channel signalling between transmitter and receiver.

Fig. 9 illustrates the transmission adaptability of the proposed hybrid UCB-HMM scheme for $M/N = \{1/8, 1/4, 1/2\}$. The percentage of channels used for transmission by each cognitive engine (UCB$_1$-M or M-HMMs) is represented versus the status of the environment in terms of availability. We define three different scenarios following the same classification of the HF database used in [4]: a 'good scenario' where most of the channels are completely available, a 'regular scenario' where most of the channels are partially available, and a 'bad scenario' where most of the channels are unavailable. Fig. 9 shows that an increase in the number of channels classified as *available* leads to a decrease in the percentage of channels used for short-term transmissions according to UCB$_1$-M predictions, whereas the percentage of channels transmitting in a long-term basis according to M-HMMs predictions increases. This increase in the percentage of channels used for long-term transmissions (or inversely the decrease in the percentage of channels used for short-term transmissions) when the environment's conditions improve is about 16% for $M/N = 1/8$, 31% for $M/N = 1/4$, and 50% for $M/N = 1/2$. It is shown that the highest increase in the percentage of channels used for long-term transmissions is reached when $M/N = 1/2$. It occurs because as M increases,
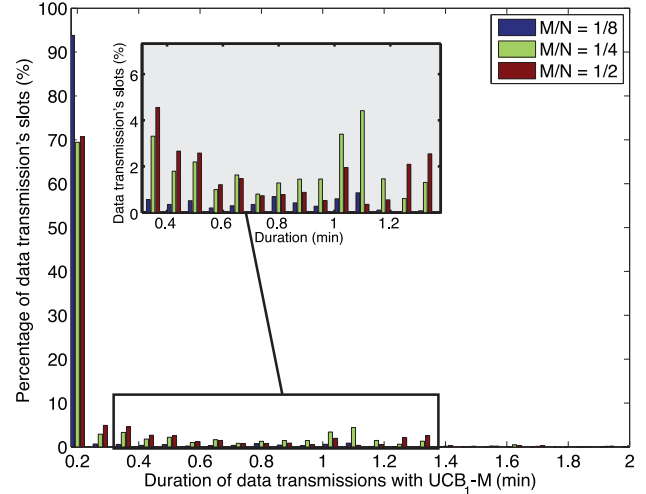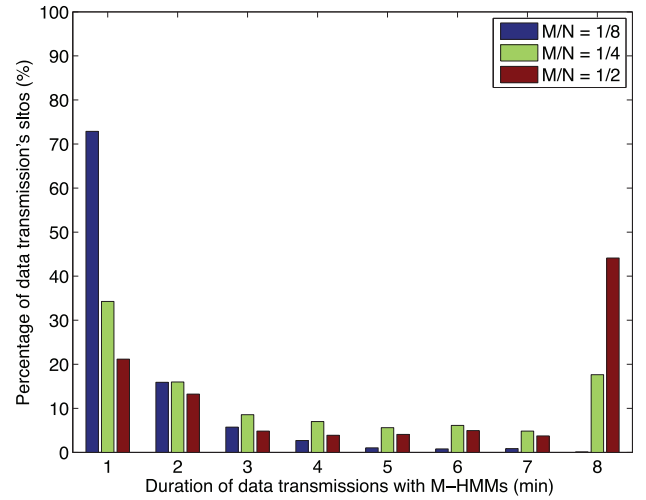
the proposed hybrid UCB-HMM has to transmit in a higher number of channels, and, in 'bad scenarios', it is more difficult to take advantage of all possible transmission opportunities within the total set of N channels. Thus, in 'bad scenarios' it switches to UCB$_1$-M for 68% of data transmissions in the short-term whereas in 'good scenarios' it reaches 83% of data transmissions in the long-term.

3) Analysis of the Duration of Data Transmission's Slots: The duration of data transmission's slots of each cognitive engine of the proposed hybrid UCB-HMM is also noteworthy. Fig. 10 and Fig. 11 show the histogram of data transmission's slots made by both learning methods when they are executed in this hybrid UCB-HMM scheme with respect to their slots' duration. These results include the performance of the proposed hybrid UCB-HMM with $M/N = \{1/8, 1/4, 1/2\}$.

Fig. 10 shows that, for $M/N = 1/8$, 94% of data transmission's slots according to UCB$_1$-M predictions have a duration of less than 0.2 minutes (12 seconds). As M/N increases to $M/N = 1/4$ and $M/N = 1/2$, i.e., the number of channels

for transmission increases, UCB_1-M transmission's slots also increase their duration to at most 2 minutes. Therefore, the percentage of transmissions' slots with a duration less than 12 seconds is reduced to 70%. This reduction occurs because the set of channels that can be chosen to transmit is larger and thus, the probability of having more available channels for longer periods also increases.

This effect can also be observed in Fig. 11, where the histogram of data transmission's slots according to M-HMMs predictions is depicted. When the value of the ratio M/N is small ($M/N = 1/8$), 73% of data transmission's slots have a length of only one minute according to the M-HMMs predictions. However, as M/N increases to 1/4 and 1/2, this percentage decreases to 34% when $M/N = 1/4$ and to 21% when $M/N = 1/2$, whereas the percentage of data transmission's slots of 8 minutes according to M-HMMs predictions without shifting to other channels increases to 18% when $M/N = 1/4$ and to 44% when $M/N = 1/2$. Therefore, in order to exploit longer data transmissions, M/N must be set to 1/2 or 1/4.

4) Channel Signalling Improvement: A reduction in the complexity of N parallel HMM based prediction models is not the only aim of the hybrid UCB-HMM system. A second objective is to simplify the link management between transmitter and receiver. This simplification can be seen as a reduction in the channel signalling due to the possibility of establishing long-term transmissions by using slots of $T_{HMM} = 1$ min with M-HMMs predictions instead of short-term transmissions during slots of $T_{UCB} = 2$ s with UCB_1-M predictions. A comparison of the proposed UCB-HMM system to the M parallel HMMs and UCB_1-M algorithms working separately and independently is shown in Fig. 12 with the aim of analysing the amount of required channel signalling. As mentioned previously, channel signalling is the amount of transmissions needed to inform the receiver about the selected channels for data transmission. Therefore, when M-HMMs are used for prediction and transmission, the selection of M channels for data transmission is updated and transmitted to the receiver once every T_{HMM} slot. Nevertheless, UCB_1-M requires more channel signalling because it updates and transmits the selection of M channels every T_{UCB} slots.

Fig. 12 shows that required channel signalling by the proposed hybrid UCB-HMM scheme with $M/N = \{1/8, 1/4, 1/2\}$ after T_{TEST} is higher than that of M-HMMs but lower than UCB_1-M. It is important to remark that required channel signalling by M-HMMs, hybrid UCB-HMM system, and UCB_1-M algorithm increases when M/N also does since the number of channels that can be used to transmit (M) is also increasing. However, even if the maximum number of channels to explore in the hybrid UCB-HMM system is considered ($N = 30$), the amount of signalling transmissions and channel signalling load required to coordinate transmitter and receiver are totally affordable. The most demanding configuration of the hybrid scheme is when $N = 30$ and $M/N = 1/2$, which allows simultaneous data transmissions in at most $M = 15$ channels. In this configuration, 5 bits are needed to identify each channel for the set of $N = 30$. The highest channel signalling load to send will be $5 \times M = 5 \times 15 = 75$ bits every 2 seconds, i.e.
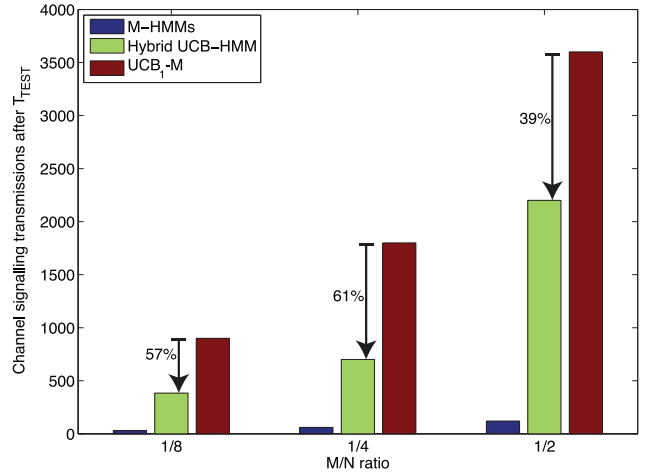

Fig. 12. Required channel signalling by M-HMMs, UCB_1-M and proposed hybrid UCB-HMM system vs. M/N ratio.

only when UCB_1-M is used to transmit in the short-term due to poor environment's conditions. If these conditions improve, less than M channels will be used in the short-term and the channel signalling load will decrease.

We computed the attained reduction in channel signalling by the proposed hybrid UCB-HMM with respect to that required by the UCB_1-M algorithm. It is worth mentioning that as UCB_1-M updates the selection of channels for data transmission every T_{UCB} slots, it is the learning method that requires the highest amount of channel signalling as defined in this paper. Fig. 12 shows that required channel signalling by the hybrid UCB-HMM system is equivalent to a reduction of 57% when $M/N = 1/8$, 61% when $M/N = 1/4$ and 39% when $M/N = 1/2$. It was shown in Fig. 8 that the proposed UCB-HMM system has similar successful transmission rates to UCB_1-M. However, as the proposed hybrid UCB-HMM system switches to long-term transmissions, required channel signalling is significantly reduced for all M/N ratios. This clearly demonstrates the benefits of the proposed hybrid UCB-HMM system compared to each learning method used separately.

In order to exploit hybrid UCB-HMM system's capability to transmit data for longer slots according to environment's conditions, a trade-off between achieved successful transmission rate, the duration of data transmissions, and the reduction of channel signalling must be found. The depicted results reveal that this trade-off can be accomplished by selecting $M/N = 1/4$. With this configuration, the proposed hybrid UCB-HMM achieves 95% successful transmission rate, and reduces by 61% the required channel signalling compared to that required by the UCB_1-M algorithm.

V. CONCLUSION

This paper has shown the feasibility of the Upper Confidence Bound (UCB) algorithm in the HF band. Both single-channel and multiple-channel versions of UCB_1 (UCB_1 and UCB_1-M) achieve the best trade-off between exploration and exploitation in the HF environment when the exploitation-exploration factor α is equal to 0.4, which means that UCB_1 and UCB_1-M

algorithms try to exploit channels previously labelled as *available* instead of trying to explore new channels. Furthermore, we propose a new hybrid scheme based on a metacognitive engine approach. This hybrid scheme combines two separate cognitive engines: the UCB_1-M algorithm and M parallel HMM based prediction models for the HF band (M-HMMs). The proposed hybrid UCB-HMM with $\alpha = 0.4$ and M/N ratio equal to 1/4 achieves its best performance: 95% of successful transmission rate is achieved as long as it automatically adapts its configuration to the changes in the environment. An 88% of data transmissions are established in the long-term (1 minute) according to M-HMMs predictions when the activity in the band decreases whereas 44% of the data transmissions are established in the short-term (2 seconds) according to UCB_1-M predictions when the activity in the band increases. We have also shown that the link management of cognitive radio can be significantly reduced by using the proposed hybrid UCB-HMM since long-term transmissions are established, and required channel signalling is reduced by 61% with respect to that required by the UCB_1-M algorithm. These results were obtained by evaluating the proposed hybrid system with real measurements from the HF band.

REFERENCES

[1] N. M. Maslin, *HF Communications: A Systems Approach*. New York, NY, USA: Pitman, 1987.

[2] *Interoperability and Performance Standards for Medium and High Frequency Radio Equipment*, Military Standard MIL-STD-188-141B, 1999.

[3] B. A. Fette, *Cognitive Radio Technology*. New York, NY, USA: Academic, 2009.

[4] L. Melián-Gutiérrez, S. Zazo, J. Blanco-Murillo, I. Pérez-Álvarez, A. García-Rodríguez, and B. Pérez-Díaz, "HF spectrum activity prediction model based on HMM for cognitive radio applications," *Phys. Commun.*, vol. 9, pp. 199–211, Dec. 2013.

[5] L. Melián-Gutiérrez, N. Modi, C. Moy, I. Pérez-Álvarez, F. Bader, and S. Zazo, "Upper confidence bound learning approach for real HF measurements," in *Proc. IEEE Int. Conf. Commun. Workshops (ICC)*, London, U.K., 2015, pp. 381–386.

[6] *STANAG 4285 Characteristics of 1200/2400/3600 Bits Per Second Single Tone Modulators/Demodulators for HF Radio Links*, 1989.

[7] *STANAG 4539 Technical Standards for Non-Hopping HF Communications Waveforms*, 2005.

[8] *STANAG 5066 Profile for HF Radio Communications* Ed. 2, 2008.

[9] I. Pérez-Álvarez, I. Raos, S. Zazo, E. Mendieta-Otero, H. Santana-Sosa, and J. Páez-Borrallo, "Interactive digital voice over HF," in *Proc. 9th Int. Conf. Ionospheric Radio Syst. Techn. (IRST'03)*, Bath, U.K., 2003, pp. 31–36.

[10] H. Santana-Sosa, S. Zazo-Bello, I. Pérez-Álvarez, I. Raos, E. Mendieta-Otero, and J. López-Pérez, "Validation of a HF spread spectrum multi-carrier technology through real-link measurements," *Eur. Trans. Telecommun.*, vol. 17, no. 6, pp. 651–657, 2006.

[11] I. Pérez-Álvarez, J. López-Pérez, S. Zazo-Bello, I. Raos, B. Pérez-Díaz, and E. Jiménez-Yguacel, "Real link of a high data rate OFDM modem: Description and performance," in *Proc. 11th Int. Conf. Ionospheric Radio Syst. Techn. (IRST)*, Edinburgh, U.K., 2009, pp. 1–5.

[12] H. Haralambous and H. Papadopoulos, "24-hour neural network congestion models for high-frequency broadcast users," *IEEE Trans. Broadcast.*, vol. 55, no. 1, pp. 145–154, Mar. 2009.

[13] O. Altinay, E. Tulunay, and Y. Tulunay, "Forecasting of ionospheric critical frequency using neural networks," *Geophys. Res. Lett.*, vol. 24, no. 12, pp. 1467–1470, Jun. 1997.

[14] A. Chu and D. Conn, "Neural network prediction of HF ionospheric propagation loss," *Electron. Lett.*, vol. 35, no. 20, pp. 1774–1776, Sep. 1999.

[15] *Interoperability and Performance Standards for Data Modems*, Military Standard MIL-STD-188-110C, 2011.

[16] Q. Zhao and B. Sadler, "A survey of dynamic spectrum access," *IEEE Signal Process. Mag.*, vol. 24, no. 3, pp. 79–89, May 2007.

[17] W. Furman and E. Koski, "Next generation ALE concepts," in *Proc. 11th Int. Conf. Ionospheric Radio Syst. Techn. (IRST)*, Edinburgh, U.K., 2009, pp. 1–5.

[18] E. Koski and W. Furman, "Applying cognitive radio concepts to HF communications," in *Proc. 11th Int. Conf. Ionospheric Radio Syst. Techn. (IRST)*, Edinburgh, U.K., 2009, pp. 1–6.

[19] T. Vanninen, T. Linden, M. Raustia, and H. Saarnisaari, "Cognitive HF–New perspectives to use the high frequency band," in *Proc. 9th Int. Conf. Cognit. Radio Oriented Wireless Netw. Commun. (CROWNCOM)*, Oulu, Finland, 2014, pp. 108–113.

[20] H. Asadi, H. Volos, M. Marefat, and T. Bose, "Metacognitive radio engine design and standardization," *IEEE J. Sel. Areas Commun.*, vol. 33, no. 4, pp. 711–724, Apr. 2015.

[21] R. Agrawal, "Sample mean based index policies with O(log n) regret for the multi-armed bandit problem," *Adv. Appl. Probab.*, vol. 27, no. 4, pp. 1054–1078, Dec. 1995.

[22] P. Auer, N. Cesa-Bianchi, and P. Fischer, "Finite-time analysis of the multiarmed bandit problem," *Mach. Learn.*, vol. 47, nos. 2–3, pp. 235–256, May 2002.

[23] H. Robbins, "Some aspects of the sequential design of experiments," *Bull. Amer. Math. Soc.*, vol. 58, no. 5, pp. 527–535, Sep. 1952.

[24] W. Jouini, D. Ernst, C. Moy, and J. Palicot, "Multi-armed bandit based policies for cognitive radio's decision making issues," in *Proc. 3rd Int. Conf. Signals Circuits Syst. (SCS)*, Medenine, Tunisia, 2009, pp. 1–6.

[25] C. Moy, "Reinforcement learning real experiments for opportunistic spectrum access," in *Proc. Karlsruhe Workshop Softw. Radios*, Karlsruhe, Germany, 2014, pp. 1–10.

[26] C. Robert, C. Moy, and C.-X. Wang, "Reinforcement learning approaches and evaluation criteria for opportunistic spectrum access," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Sidney, NSW, Australia, 2014, pp. 1508–1513.

[27] P. Cordier *et al.*, "E^2R cognitive pilot channel concept," in *Proc. IST Mobile Summit*, Mykonos, Greece, 2006, pp. 1–4.

Laura Melián-Gutiérrez received the M.Sc. in telecommunication engineering from the Universidad de Las Palmas de Gran Canaria (ULPGC), Las Palmas de Gran Canaria, Spain, in 2010, and the M.Sc. in communication systems and technologies from the Universidad Politécnica de Madrid (UPM), Madrid, Spain, in 2011. She was a Student Research Assistant at the Technological Centre for Innovation in Communications (CeTIC), Spain, from 2007 to 2010. Since 2011, she is a Research Assistant at the Institute for Technological Development and Innovation in Communications (IDeTIC), Spain, where she is pursuing the Ph.D. degree in the fields of signal processing and cognitive radio.

Navikkumar Modi received the B.Tech. degree in electrical, electronics, and communications engineering from U.V. Patel College of Engineering, Mehsana, India, in 2010, and the M.S. degree in communications and signal processing from Ilmenau University of Technology, Ilmenau, Germany, in 2013. He was a Student Research Assistant with Digital Broadcasting Research Laboratory, TU Ilmenau, Germany. He is currently pursuing the Ph.D. degree at CentraleSupelec (created in January 2015), formerly Supelec. He is studying machine learning algorithms for green cognitive radio systems in TEPN CominLabs Labex project.

Christophe Moy (M'06) received the M.Sc. degree in engineering, and the Ph.D. degree in electronics from the INSA (National Institute of Applied Sciences), Rennes, France, in 1995 and 1999, respectively. He then worked for six years in Mitsubishi Electric ITE focused on software radio systems and concepts. Since 2005, he has been working as a Professor with Supelec, now CentraleSupelec, since January 2015, in the IETR Laboratory of CNRS (UMR 6164): Institute of Electronics and Telecommunications of Rennes. His research interests include software radio, cognitive radio, and green radio. He addresses

heterogeneous design techniques for SDR, high-level design methodologies for cognitive management, decentralized decision making and learning for cognitive radio equipment and systems, and dynamic spectrum access. He is working in the following collaborative projects: French ANR WiNoCoD and SoGreen, CominLabs labex TEPN. He has been involved in several European collaborative projects: Networks of Excellence NEWCOM and NEWCOM++, EULER, and ER-phase2, ER. He has participated also in several French collaborative projects (ANR and Reseaux" cluster): Mopcom, Idromel, A3S, and SoftRF. He is currently the Head of the Communication Department of IETR, and a Member of Institute of Research and Technology B-COM, in Rennes.

Faouzi Bader (SM'07) received the Ph.D. degree (with Hons.) in telecommunications from Universidad Politécnica de Madrid (UPM), Madrid, Spain, in 2002. He joined the Centre Technologic de Telecomunicacions de Catalunya (CTTC), Barcelona, Spain, as an Associate Researcher, in 2002, and was nominated Senior Research Associate at the same institution, in 2006. Since June 2013, he is Associate Professor at CentraleSupélec, Rennes, France. His research interests include IMT-Advanced systems, system design, advanced multicarrier waveforms (OFDM(A), (non-)uniform multimode filter-based multicarrier schemes), and frequency allocation techniques in relay cognitive environment. He has been involved in several European projects from the 5th7th EC research frameworks, and from 2012 to 2013, he was nominated the General Co-Ordinator and the Manager of the EC-funded research project ICT EMPhAtiC project, Enhanced Multicarrier Techniques for Professional Ad-Hoc and Cell-Based Communications. He has authored over 120 papers in peer-reviewed journals and international conferences, more than 13 book chapters, and 3 books. He has served as a technical program committee member in major IEEE ComSoc and VTS conferences, and as the General Chair of the 11th edition of the ISWCS'2014 conference, and the Co-Chair of the ISWCS'2015 edition.

Iván Pérez-Álvarez (M'89) received the Telecom Engineering and Dr. Engineering degrees from the Universidad Politécnica de Madrid (UPM), Madrid, Spain, in 1990 and 2000, respectively. From 1989 to 1997, he worked for Europea de Comunicaciones S.A. and Telefónica Sistemas S.A., where he was a staff member of the Special Projects Department and worked with the Digital Signal Processing Group. He was involved in the design of HF/VHF/UHF digital communications systems for the Spanish Ministry of Defence. He joined the Universidad de Las Palmas de Gran Canaria (ULPGC), Las Palmas de Gran Canaria, Spain, as an Associate Professor, in 1998. From 2006 to 2010, he was the Head of the Centre for the Technological Centre for Innovation in Communications (CeTIC), and is currently the Vice-Director of the Institute for Technological Development and Innovation in Communications (IDeTIC). His research interests include signal processing for radio communications.

Santiago Zazo (M'12) received the Telecom Engineering and Dr. Engineering degrees from the Universidad Politécnica de Madrid (UPM), Madrid, Spain, in 1990 and 1995, respectively. From 1991 to 1994, he was with the University of Valladolid, Valladolid, Spain, and from 1995 to 1997, with the University Alfonso X El Sabio, Madrid, Spain. In 1998, he joined UPM as an Associate Professor of Signal Theory and Communications. His research interests include signal processing with applications to audio, communications and radar, and MIMO communications and wireless sensor networks from both physical layer and networking points of view. Since 1990, he has authored/coauthored more than 30 journal papers and about 150 conference papers. He has also led many private and national Spanish projects and has participated in five European projects, leading three of them. He is leading the HF Communication Group, UPM, with large activity in publications and technological transfer to civil and military applications in this field.