Distributed Learning for Channel Allocation Over a Shared Spectrum

S. M. Zafaruddin, *Member, IEEE*, Ilai Bistritz, *Student Member, IEEE*, Amir Leshem, *Senior Member, IEEE*, and Dusit (Tao) Niyato, *Fellow, IEEE*

***Abstract*—Channel allocation is the task of assigning channels to users such that some objective (e.g., sum-rate) is maximized. In centralized networks such as cellular networks, this task is carried by the base station (BS) which gathers the channel state information (CSI) from the users and computes the optimal solution. In distributed networks such as ad-hoc and device-to-device (D2D) networks, no BS exists and conveying global CSI between users is costly or simply impractical. When the CSI is time varying and unknown to the users, the users face the challenge of both learning the channel statistics online and converging to a good channel allocation. This introduces a multi-armed bandit (MAB) scenario with multiple decision makers. If two or more users choose the same channel, a collision occurs and they all receive zero reward. We propose a distributed channel allocation algorithm that each user runs and converges to the optimal allocation while achieving an order optimal regret of $\mathcal{O}\left(\log T\right)$, where T denotes the length of time horizon. The algorithm is based on a carrier sensing multiple access (CSMA) implementation of the distributed auction algorithm. It does not require any exchange of information between users. Users need only to observe a single channel at a time and sense if there is a transmission on that channel, without decoding the transmissions or identifying the transmitting users. We demonstrate the performance of our algorithm using simulated LTE and 5G channels.**

***Index Terms*—Distributed channel allocation, Multiplayer multi-armed bandit, online learning, dynamic spectrum accesses, resource management.**

I. Introduction

Channel allocation in wireless communication is one of the fundamental management tasks. It has been widely studied for various wireless networks [1]–[5]. In the traditional centralized systems, Orthogonal Frequency Division Multiplexing Access (OFDMA) was investigated extensively to meet the high demand for efficient spectrum utilization. If users can be assigned to sub-channels efficiently, certain gains can be derived from the diversity of the channel. The main issue for the OFDMA systems is joint power and sub-carrier allocation in the downlink direction [6]–[9] and sub-carrier assignment in the uplink direction [10]–[12]. Due to the global view of the whole network, the centralized approach is able to obtain the optimal solution of a desired performance metric. The optimal channel allocation can be computed using the well-known Hungarian method [13].

However, there are some disadvantages that limit the practicality of the centralized approach such as significant signaling overhead, increased implementation complexity and higher latency in dealing with resource allocation problems. Moreover, emerging wireless networking paradigms such as cognitive radio networks, ad-hoc networks, and D2D communications are inherently distributed. A complete information about the network state is typically not available online, which makes the computation of optimal policies intractable for these networks. Hence, it is desirable to develop a distributed learning algorithm for dynamic spectrum access that can effectively adapt to general complex real-world settings in dense and heterogeneous wireless environments. Moreover, open sharing model employs spectrum sharing among peer users as the basis for managing a spectral band. Advocates of this model draw support from the phenomenal success of wireless services operating in the unlicensed industrial, scientific, and medical (ISM) radio band (e.g., WiFi). Centralized and distributed spectrum sharing strategies have been initially investigated to address technological challenges under this spectrum management model.

The center of the channel allocation task is the combinatorial optimization assignment problem. Solving the assignment problem distributedly is a major challenge that has received considerable attention. The famous auction algorithm [14] proposed a distributed method to solve the assignment problem where users send their bids to an auctioneer. In [15] a fully distributed version of the auction algorithm was suggested that exploits carrier sense multiple access (CSMA) in order to avoid the need for an auctioneer.

If the resource (channel) values are not known in advance by the users, they have to learn these values online. Learning the CSI in real-time comes at the expense of using the best known channels so far. This introduces the well-known trade

This research was supported by the ISF-NRF Joint research Program, under grant ISF 2277/16 and by ISF 1644/18. This work was supported in part by WASP/NTU M4082187 (4080), Singapore MOE Tier 1 under Grant 2017-T1-002-007 RG122/17, and Singapore NRF2015-NRF-ISF001-2277. S. M. Zafaruddin was partially funded by the Israeli Planning and Budget Committee (PBC) post-doctoral fellowship (2016-2018). Part of this research to be presented at the 20th IEEE International Workshop on Signal Processing Advances in Wireless Communications (SPAWC), Cannes, France, July 2019.

S. M. Zafaruddin was with the Faculty of Engineering, Bar-Ilan University, Ramat Gan 5290002, Israel (e-mail: smzafar@biu.ac.il). Currently, he is with the Department of Electrical and Electronics Engineering, BITS Pilani, Pilani-333031, Rajasthan, India (email: syed.zafaruddin@pilani.bits-pilani.ac.in).

I. Bistritz was with the Faculty of Engineering, Bar-Ilan University, Ramat Gan 5290002, Israel. Currently, he is with the Department of Electrical Engineering, Stanford University, Stanford, CA, 94305 (e-mail: bistritz@stanford.edu).

A. Leshem is with the Faculty of Engineering, Bar-Ilan University, Ramat Gan 5290002, Israel (e-mail: leshema@biu.ac.il).

Dusit (Tao) Niyato is with School of Computer Science and Engineering (SCSE), Nanyang Technological University, Singapore 639798. (e-mail: dniyato@ntu.edu.sg).

off between exploration and exploitation that is captured by the multi-player multi-armed bandit (MAB) problem. In this case, there are several decision makers facing this problem, and when two or more choose the same channel, they receive zero reward. Similarly to other MAB problems, the performance is measured by the expected difference between the actual sum of rewards and the sum of rewards that could have been achieved if the users had perfect knowledge of the CSI. However, as opposed to classical MAB problems, the interaction between the users significantly complicates the learning aspects of the problem. To address that, deep reinforcement learning and Q-learning methods have been proposed for these problems [16]–[18], and shown to perform well for small-size models. However, for large-scale networks these methods perform poorly since the number of states of the learning algorithm increases exponentially in the number of users.

In [19], the auction algorithm [14] was used as a basis for a distributed algorithm that achieves an expected sum regret of $\mathcal{O}(\log T)$, where T denotes the length of time horizon. However, since it relies on [14], this algorithm requires communication between users in order to exchange the bids and determine the winning user in each auction. To implement this algorithm, users need to know which user transmitted on which channel. In this manner, they can use their public channel choices as a signaling method. In practice, this knowledge requires that users decode at least part of the transmission to identify the ID of the transmitting users. Besides being computationally demanding, this might be highly non-trivial when multiple users transmit on the same channel and all their IDs need to be decoded from the mixture.

In this paper, we overcome this requirement by proposing a distributed algorithm termed Online Auction based Learning Algorithm (OALA) that relies on [15] instead of [14]. The algorithm in [15] assumes that the CSI is known. It also uses a continuous back-off time and assumes no tied bids. We lift all these assumptions in our novel MAC protocol. Our protocol achieves an expected sum regret of $\mathcal{O}(\log T)$, but in contrast to [19], only requires each user to sense the channel that the user is using and detect if there are other transmissions on this channel. Users do not need to know which user transmitted on which channel or how many of them did. Therefore, our algorithm offers the same order optimal performance as [19] but with dramatically simpler implementation.

A. Related Works

Developing multi-armed bandit (MAB)-based methods for solving dynamic spectrum allocation (DSA) problems is a relatively new research direction, motivated by recent developments of MAB in various other fields, and many works have been done in this direction recently. A couple of these works [20]–[23] considered a cognitive radio scenario where a set of channels can be either free or occupied by a primary user that interferes all secondary users. A generalized scenario was considered in [24]–[26], where the channel qualities are not binary, but still all users have the same vector of channel qualities. Recently, the case of a full channel allocation scenario where different users have different channel qualities (a matrix of channel qualities) was considered in [27], and later improved in [19], by the same authors, to have an order optimal sum-regret of $\mathcal{O}\left(\log T\right)$.

Recently, it has been shown in [28] (which improved [29]) that achieving a sum-regret of near-$\mathcal{O}\left(\log T\right)$ is possible even without communication between users and with a matrix of expected rewards. The algorithm in [28] is general but has a slow convergence rate in T. It can be regarded as a multi-channel ALOHA protocol where only the collision indicator is available. In this paper, we adopt a more practical and advanced communication approach using CSMA and achieve an order optimal sum-regret of $\mathcal{O}\left(\log T\right)$. Our algorithm still does not require any communication between users, and each device only needs to sense a single channel at a time (instead of simultaneously all of them as in [19]). It is made possible by adding assumptions that are always valid from a practical perspective: the expected rewards i.e. Quality of Service (QoS) are integer multiplications of a common resolution $\Delta_{\min}$, and a device can choose not to transmit on any channel and instead only to sense a single channel of its choice. Our algorithm is much easier and less costly to implement than that of [19] and has a much better convergence time than that of [28].

The literature on distributed channel allocation without learning, where the CSI is assumed to be known, is vast and we can only cover part of it here. Recently there has been growing interest in distributed spectrum optimization for frequency selective channels, where the assignment problem arises. However, most of the work done in this field relies on explicit exchange of CSI. Several suboptimal approaches that do not require information sharing have been suggested [30]–[33]. In [30], a greedy approach to the channel assignment problem was introduced. In [31] and [32], the use of opportunistic carrier sensing was combined with the Gale-Shapley algorithm for stable matching [34] to provide a fully distributed stable channel assignment. This solution basically achieves the greedy channel assignment and analysis of this technique for Rayleigh fading channels was done in [33].

Game theory is often used to design distributed channel allocation algorithms [35]–[42]. In [35] the channel assignment problem was formulated as a many-to-one matching game under the limitation that each primary channel can only be assigned to one secondary user. In [36], an algorithm was proposed based on a game with utility design that leads to an asymptotically optimal performance in all Nash equilibrium. In [37] the spectrum sharing problem between D2D pairs and multiple co-located cellular networks was formulated as a Bayesian non-transferable utility overlapping coalition formation game. Nash bargaining solutions for channel allocation were considered in [38]–[40], and distributed allocation using multichannel ALOHA and potential games was considered in [41], [42].

The auction algorithm has been extensively used to solve a variety of assignment problems. It gets its name from operating similarly to an auction. As in this paper and many others, the auction algorithm may have nothing to do with actual auctions that rely on economic and game-theoretic principles, as was done in [43]–[46]. In [47] the auction algorithm was used to solve the channel assignment problem for the uplink, using the base station as the auctioneer. In [48] a distributed auction algorithm with shared memory was used for switch scheduling. In [49] it was shown that a modification of the auction algorithm is equivalent to max product belief propagation. However, all these modified auction algorithms require a base station or shared memory, which prevents them from being fully distributed. In addition, all these algorithms, including [15] that is being used here, assume that the CSI is known to the users. Our algorithm generalizes the distributed CSMA auction algorithm [15] to an online learning framework.

B. Outline

The paper is organized as follows: Section II describes the system model and our network assumptions. Section III discusses our novel MAC protocol and explains the details of the algorithm. Section IV analyzes the exploration and auction phases of our algorithm, and provides theoretical performance guarantees. Section V provides simulation results of our algorithm on practical LTE channels, along with a performance comparison. Section VI concludes the paper.

II. System Model

We consider an ad hoc network with a set of transmitter-receiver pairs (links) $\mathcal{N} = \{1, \ldots, N\}$ and a set of channels $\mathcal{K} = \{1, \ldots, K\}$, where $K \geq N$. Each channel consists of several OFDMA sub-carriers, and each link uses a single channel. In the case of more users than channels ($N > K$), a combined OFDMA-TDMA (time division multiple access) can be used instead in order to have enough resources for all users. However, since this is a trivial consequence of our analysis which only complicates the notation, we choose to focus on the case of a single time slot without TDMA. The number of channels K is chosen by the protocol designer to be large enough to support N links in an environment with outside interferers where some of the channels can be very poor and practically unavailable. The identity and number of subcarriers that constitute each channel can also be optimized with respect to the typical channels used by the significant interferes. Links may use multiple-input multiple-output (MIMO) transmission, with different capabilities for each link.

We consider that the links are located in a geographical proximity in an area that typically includes other coexisting networks nearby. This is relevant, for example, for WiFi networks and Internet of Things (IoT) networks. As a result, each receiver experiences alien interference from the transmission of other users. Due to the geometry of the links and the different channels used by different interferers, the average interference is different for each receiver in our network. A toy example of our network with $K = N = 6$ is depicted in Fig. 1. The channel used by each link is indicated by the color of the arrow between its transmitter and receiver. Outside the area of the network there are four major interferers that use four of the six available channels. In this example, links successfully avoid using channels with significant interference at their receiver side. This outside transmissions can be constant over time or bursty, and may overlap any part of the subcarriers used by a particular link. In addition, the fading of the channel may cause significant changes to the channel gains of the subcarriers.

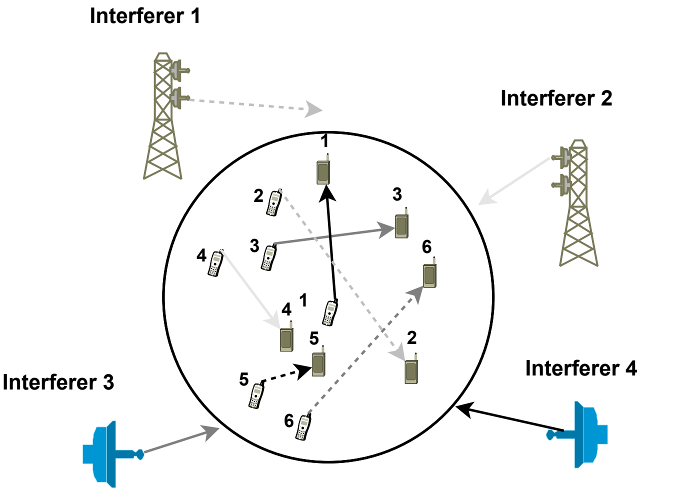


Fig. 1. System model of a network with $K = N = 6$ with four major interferers outside the area of the network.

As any modern device, the transmitter and receiver of each link adopt techniques such as adaptive beamforming and modulation together with interleaving and coding for fading channels in order to provide a stable (on average) and reliable communication for the users. However, since the channel statistics and the interference pattern are initially unknown, each link needs to learn them online as fast as possible in order to deduce which QoS that it can support.

In practical systems, while the quality of the physical conditions is continuous, the number of available transmission schemes (e.g., coding and modulation scheme: 64QAM + 3/4 rate LDPC, BPSK + 1/2 rate Turbo code, and so on) is always discrete. It is noted that the instantaneous QoS (which can be regarded as an SNR) is continuous and measured on a finer scale such that the average value will be accurate. Thus, the measured QoS is an instantaneous QoS of a single time slot and the quantized QoS is the QoS that the device can support on average over time. Each link estimates the physical conditions to identify the best transmission scheme supported by the channel. Following this reasoning, we assume that there is some resolution for the supported QoS (e.g., 100 Kbps), we denote by $\Delta_{\min}$. The supported QoS set is $\mathcal{Q} \triangleq \{Q_1, \ldots, Q_M\}$ where for each i, $Q_i = l_i \Delta_{\min}$ for a non-negative integer l_i and $0 < Q_1 < \cdots < Q_M$. The QoS experienced by link n using channel i is denoted by $Q_{n,i}$. A value in this set may represent the weighted quality of a combination of parameters, e.g., 1 Mbps for Internet, 256 Kbps for voice and 10 Mbps for video. In general, different links have a subset of different possible QoS values from $\mathcal{Q}$ due to different capabilities, e.g., number of transmitting and receiving antennas. Being part of the standard of the protocol, we assume that the parameters $\Delta_{\min}$ and $\Delta_{\max} = Q_M - Q_1$ are known to all devices.

We assume that the time is slotted and indexed by t, such that in each time slot, L OFDM symbols are transmitted. The number of OFDM symbols per time slot L can be designed to match the coherence time of the channel, such that the CSI typically changes every time slot. Hence, we assume a fast-fading scenario where the coherence time is proportional to an OFDM symbol duration. The links are active for a total of T time slots, where T is unknown in advance by the links. We assume that each link can sense a single channel at each time slot, which is the channel they use, and detect whether other links are transmitting on this channel. The chosen channel of link n at time t is denoted by $a_n(t)$. Naturally, links can choose not to transmit at all at a given time slot, which is denoted $a_n(t) = 0$. Non-transmitting links can still sense transmissions on a single chosen channel. In each time slot t, each link measures the instantaneous QoS $q_{n,i}(t)$ by using a finer resolution than that of $\mathcal{Q}$, in order for the estimation of the average to be accurate.

The instantaneous QoS $q_{n,i}(t)$ captures the fast-fading of the channel, bursty interference and other measurement noise. It is naturally measured at the receiver by measuring the instantaneous SNR at a given channel. Since the transmitter adapts its transmission scheme to provide stable and reliable communication over time, the rapidly changing $q_{n,i}(t)$ do not determine how good a channel is, but only their average does. The instantaneous QoS needs to be continuous (or at least with a high resolution) such that the measurements are accurate leading to a good estimate for the average QoS that the device can support. The independence assumption for different n is justified since the fading patterns of different devices, typically located a few meters apart, are independent [50]. In fact, this independence assumption even holds for different antennas of the same MIMO transceiver [51]. Note that QoS $\{Q_{n,i}\}$ can be similar for close by users that are affected by the same alien interferer. The distribution of $q_{n,i}(t)$ is bounded since $Q_1 \leq q_{n,i}(t) \leq Q_M$, and can be either discrete or continuous due to arbitrarily fine measurements.

While our approach is fully distributed, we assume that the links in the network are synchronized, so all devices use the same reference clock. This is typical to many distributed communications networks, e.g. WiFi and any practical slotted CSMA or ALOHA. Common synchronization techniques use global positioning system (GPS) or any other beacon that can be broadcasted to the devices in the network on a much slower time scale, or converge to a global synchronization using local connections and consensus protocols [52].

We model $q_{n,i}(t)$ as i.i.d. sequence in time, independent for different n or i. The distribution of $q_{n,i}(t)$ is bounded since $Q_1 \leq q_{n,i}(t) \leq Q_M$, and can be either discrete or continuous due to arbitrarily fine measurements.

Define the set of links that are transmitting on channel i at time t by

$$\mathcal{N}_i(t) = \{n \,|\, a_n(t) = i\}. \tag{1}$$

Define the no-collision indicator of channel i at time t by

$$\eta_i(t) = \begin{cases} 0 & \left|\mathcal{N}_i(t)\right| > 1 \\ 1 & o.w. \end{cases}. \tag{2}$$

The instantaneous reward of link n at time t from transmitting on channel a_n is

$$r_{n,a_n}(t) = q_{n,a_n}(t)\,\eta_{a_n}(t). \tag{3}$$

The theoretical guarantee of our algorithm is formulated using the well-known notion of regret, defined as follows.

Definition 1. The total regret is defined as the random variable

$$R = \sum_{t=1}^{T}\sum_{n=1}^{N} Q_n^* - \sum_{t=1}^{T}\sum_{n=1}^{N} q_{n,a_n(t)}(t)\,\eta_{a_n(t)}(t). \tag{4}$$

The value Q_n^* is the expectation of the QoS of the channel that link n is assigned to:

$$a^* = \arg\max_{a_1,\ldots,a_N}\sum_{n=1}^{N} Q_{n,a_n}. \tag{5}$$

The expected total regret $\bar{R} \triangleq E\{R\}$ is the average of (4) over the randomness of the rewards $\{r_{n,i}(t)\}_t$ that dictate the random channel choices $\{a_n(t)\}$.

III. Protocol Description

We design a novel MAC protocol where each link runs distributedly in order to maximize the total sum of QoS (over all users). In the original auction algorithm, an auctioneer is needed to collect the bids and compute the highest bidder. Such an auctioneer is not available in a distributed wireless network. The algorithm in [15] exploits the CSMA mechanism to bypass the need for an auctioneer and by doing that, implements the auction algorithm distributedly. For this purpose, links compute a continuous back-off time that is decreasing with their bid. The highest bidder for a particular channel is simply the first link which accesses this channel. Since we assume all links can sense the channel that they choose, all links will agree on which link is the highest bidder for their channel. Note that we do not analyze selfish links, but consider devices that are programmed to run our designed MAC protocol in a cooperative manner. This is the way that most MAC protocols operate.

The key advantage of the proposed OALA algorithm is that it only requires from each receiver to sense if there are transmissions on a single channel, which is a basic requirement. We assume that all links are at a sensing distance from each other (a fully-connected network). As is common in CSMA systems, this assumption can be relaxed using request to send or clear to send (RTS/CTS) protocols where the RTS/CTS messages are much shorter and have higher priority in transmission. For simplicity of exposition, we ignore this aspect. However, as opposed to [19], the links do not know which transmission belongs to which link. This is the scenario in practice with wireless links located in close enough proximity. In our protocol, links do not need to distinguish between the transmission of other links, which may require decoding an ID for each link. Moreover, it can be extremely computationally demanding in practice to separate colliding transmissions and discern the IDs involved. Sensing a single channel at a time instead of all the K channels is another major advantage of our algorithm over [19]. We also note that the computational complexity of running Algorithm 1 for each device is $\mathcal{O}(K)$, since maximization over a K-sized vectors is required.

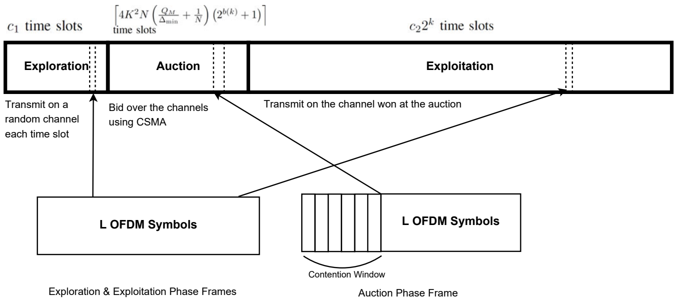

c_1 time slots
$\left\lceil 4K^2N\left(\frac{Q_M}{\Delta_{\min}}+\frac{1}{N}\right)\left(2^{b(k)}+1\right)\right\rceil$ time slots
$c_2 2^k$ time slots
Exploration
Auction
Exploitation
Transmit on a random channel each time slot
Bid over the channels using CSMA
Transmit on the channel won at the auction
L OFDM Symbols
L OFDM Symbols
Contention Window
Exploration & Exploitation Phase Frames
Auction Phase Frame

Fig. 2. The k-th packet of Algorithm 1: Online auction based learning algorithm.

We divide the T time slots into packets with a dynamic length, one starting immediately after the other. Each packet is further divided into three phases: exploration, auction and exploitation. The following subsections describe these three phases, for the k-th packet.

A. Exploration Phase

The exploration phase has a length of c_1 time-slots in each packet, and is used for estimating the expected reward in each channel, as given in the exploration phase of Algorithm 1. During this phase, links choose channels uniformly and independently at random. The estimated values are artificially dithered in order to avoid ties in the subsequent auction phase. This amounts to adding small artificial uniformly distributed noise. Collisions can be excluded from each link's estimation since they result in zero reward. If c_1 is large enough, then the probability that the estimation leads to a wrong optimal solution in the auction phase is "very small". This phase adds a $\mathcal{O}(\log T)$ to the expected total regret.

1) Unknown T and N: The exploration phase does not require the links to know the total number of links N or the total duration of transmission T. Hence, links cannot use a single long enough exploration phase at the beginning, since they want the exploration error probability to be designed according to T and N. The packet structure in Fig. 2 maintains the required balance. In each packet, only a constant number c_1 of time slots is dedicated to exploration, but the estimation of the k-th exploration phase uses all the previous exploration phases.

2) Dither: The estimated QoS of the channels is needed for the next auction phase to converge to the optimal allocation. However, due to its distributed nature, ties cannot be arbitrarily broken. Hence, the exploration phase needs to output accurate enough estimates that guarantee that there will be no ties in the bids in the auction algorithm. For that purpose, after the estimation of the expected QoS is completed, artificial dither noise is added to the estimated values. This dither values are generated in advance independently and uniformly at random on a small interval.

B. Auction Phase

The auction phase has a length of $\left\lceil 4K^2N\left(\frac{Q_M}{\Delta_{\min}}+\frac{1}{N}\right)\left(2^{b(k)}+1\right)\right\rceil$ time slots in the k-th packet, which is the convergence time of the distributed auction algorithm, as dictated by Lemma 5 (given in the next section). In this phase, the links run the distributed auction of [15] on the estimated expected rewards using $b(k)$ bits for the quantized back-off time. In the distributed auction algorithm, links iteratively update their virtual bids $\{B_{n,i}\}$ and bid on the different channels. In order to do that, they keep track of the "maximum profit" or the best channel and the second "maximum profit" channel. They then bid only on their best channel. In general, the parameter ε captures the tradeoff between the accuracy of the solution and the convergence time. For our purpose, we choose it to be small enough such that the distributed auction always converges to the exact optimal solution. Intuitively, this bidding mechanism allows links to distributedly agree which of them will contribute a larger term to the sum-rate (the objective), since their bid for this link will eventually be higher. For example, a link for which the best and second best channels are similar will not insist on the best channel by raising the bid much, as can be seen in (9) of Algorithm 1. To overcome the fact that there is no auction-manager, the distributed auction maps the bid of each link to a delay (CSMA back-off) of accessing the channel. This allows the wireless channel to manage the auction. The link with the highest bid for a channel accesses the channel first, which can be detected by all other links. The link who won the bid (accessed the channel first) changes its state to "assigned", and all other links that bid on this channel become "unassigned". Eventually, all links will be assigned, and the assignment is guaranteed to be optimal. A complete analysis of the distributed auction algorithm appears in [15].

In this paper we modify the distributed auction algorithm such that it can be employed in an online learning setting. This involves running the algorithm on dithered estimated QoS instead of the actual ones, and distributedly agreeing on a fine enough quantization of the CSMA back-off time. We suggest a collision resolution step that begins after the distributed auction algorithm. During the collision resolution, all links that collided over a channel, due to an insufficient quantization resolution, signal on channel 1. If some transmission is detected on channel 1 by all other links, they can deduce that the quantization of the back-off time was insufficient and they increase the resolution.. We emphasize that with a continuous back-off time, no ties in the bids are possible, so this problem did not arise in [15].

1) Quantization of CSMA back-off time: The multi-armed bandit problem uses a discrete time axis. Hence, a continuous back-off time as used in [15] is not possible. From a practical perspective, links cannot implement a truly continuous delay but a quantized one. With integer quantized delays, it is possible that two links use the same delay for the same channel although their continuous bids are different. In this case, they cannot agree on which of them won the bid and got the channel. It is clear that for a fine enough quantization, these bidding collisions will be avoided. However, due to the distributed nature of the problem, links do not know in advance what is considered a fine enough quantization. We propose a collision resolution algorithm that increases the quantization bits, described in step 3 in the Auction phase in Algorithm

1. Links coordinate their quantization by employing a “voting turn” that only uses the fact that all links can sense a single channel of their choice. In this special time slot, links listen to channel 1 which is used to signal if a collision occurred for some of the links.

2) Convergence time of $b(k)$*:* The function $b\left(k\right)$ converges to a constant that is independent of k. In practice, it is easy to guarantee that $b\left(0\right)$ is already large enough, but the designer can shorten the convergence time by starting from smaller $b\left(0\right)$ values and let the algorithm find the minimal $b\left(k\right)$ necessary.

C. Exploitation Phase

The exploitation phase has a length of $c_2 2^k$ time slots for some constant c_2. During this phase, the links transmit on the channel that they are allocated in the auction phase. If the exploration phase provided an accurate enough estimation of the QoS and the CSMA back-off time uses enough bits for quantization, then this phase adds no regret to the expected total regret since the links use the optimal allocation.

In the following, we discuss the length of the time slots and the overall communication overhead required for implementing the proposed protocol:

Remark 1 (Length of time slot T**):** In the problem formulation, the length of the time slot is not specified. This is done in order to keep the theoretical framework identical to other multi-armed bandits algorithms and measure the regret using the same scale. However, when implementing OALA (as given in Algorithm 1) in practice, there is no need to assume that all time slots are of equal length. In particular, the time slots used to implement the CSMA back-off time can be much shorter than time slots that are used to transmit a frame of L OFDM symbols. The result, depicted in Fig. 2, is the well-known structure of a CSMA frame, like that used in WiFi. At the beginning of the k-th frame, a contention window of $2^{b(k)}$ short slots is used, followed by the transmission over the chosen channel, for a period of L OFDM symbols. During the exploration and exploitation phases, no contention window is required, which makes the overhead of the contention window negligible compared to T.

Remark 2 (Communication overhead): The fact that the exploitation phase takes an exponential number of time slots does not mean that it takes a longer time in practice. In fact, it only means that the lengths of the exploration and auction phases are much shorter. Note that T is finite and can be set by the designer. Therefore, even the last (longest) exploitation phase can still consist of just a couple of thousands of OFDM symbols, which amounts to only a few milliseconds. From a practical point of view, this is the desirable packet structure since the actual transmission takes the vast majority of the OFDM symbols while the equivalents of the synchronization header do not cause a significant overhead. The overhead caused by the exploration and auction phases is naturally measured by the sum of regrets as in (4).

The description of our OALA algorithm is given in Algorithm 1 and the structure of the k-th packet of the protocol is depicted in Fig. 2.

Algorithm 1 Online Auction based Learning Algorithm (OALA)

```
Initialization Choose $\varepsilon < \frac{\Delta_{\min}}{4K}$. Set $V_{n,i}\left(0\right) = 0$ and $s_{n,i}\left(0\right) = 0$ for all $i$ and $b\left(0\right) = 8$.
  1) Dither Values — Generate $u_{n,i}$ for each $i$, independently and uniformly distributed over $\left[-\frac{\Delta_{\min}}{8N}, \frac{\Delta_{\min}}{8N}\right]$.
For $t = 1, \ldots, T$ (which determines $k$) do
A. Exploration Phase — For the next $c_1$ time slots
  1) Choose a channel $i \in [1,..,K]$ uniformly at random.
  2) Receive the reward $r_{n,i}(t)$. Update $V_{n,i}\left(t\right) = V_{n,i}\left(t-1\right) + \eta_i\left(t\right)$ and $s_{n,i}\left(t\right) = s_{n,i}\left(t-1\right) + r_{n,i}\left(t\right)$, where $\eta_i$ and $r_{n,i}$ are defined in (2) and (3), respectively.
  3) Create a dithered estimation of $Q_{n,i}$ by computing $Q^k_{n,i} = \frac{s_i(t)}{o_i} + u_{n,i}$ for $i = 1, \ldots, K$.
B. Auction Phase —set state unassigned and $B_{n,i} = 0, \forall i$.
For the next $\left\lceil 4K^2 N \left(\frac{Q_M}{\Delta_{\min}} + \frac{1}{N}\right)\left(2^{b(k)} + 1\right)\right\rceil$ time slots
Each auction iteration do
  1) If unassigned then
    a) Calculate its own maximum profit:
       $$\gamma_n = \max_i \left(Q^k_{n,i} - B_{n,i}\right) \quad (6)$$
    b) Calculate its own second maximum profit:
       $$\tilde{i}_n = \arg\max_k \left(Q^k_{n,i} - B_{n,i}\right) \quad (7)$$
       $$w_n = \max_{i \neq \tilde{i}_n} \left(Q^k_{n,i} - B_{n,i}\right) \quad (8)$$
    c) Update the bid for its best channel $\tilde{i}_n$:
       $$B_{n,\tilde{i}_n} = B_{n,\tilde{i}_n} + \gamma_n - w_n + \varepsilon \quad (9)$$
  2) During the next $2^{b(k)}$ time slots —Sense the channel $\tilde{i}_n$ after a back-off time of
       $$\tau_n = f_{b(k)}\left(B_{n,\tilde{i}_n}\right) \quad (10)$$
     time slots, where $f_{b(k)}$ is a quantization of some decreasing function $f$ (e.g., $f\left(x\right) = 2^{b(k)} - x$) using $b\left(k\right)$ bits, such that $0 \leq \tau_n \leq 2^{b(k)}$.
    a) If the channel is not busy set state to assigned and to unassigned otherwise.
  3) Collision Resolution — In the $\tau_{\max} = 2^{b(k)} + 1$ time slot
    a) Transmit over channel 1 if it is assigned a channel with a collision.
    b) If links sense a transmission on channel 1, then they update $b\left(k+1\right) = b\left(k\right) + 1$.
End
C. Exploitation Phase — for the next $c_2 2^k$ time slots
  a) Transmit over the channel assigned at the end of the auction phase.
End
```

IV. Regret Performance Analysis

In this section, we analyze the performance of the exploration and auction phases of the proposed protocol and provide our main result: The expected sum regret is an order optimal regret of $\mathcal{O}(\log T)$.

A. Exploration Phase — Estimation of the QoS

In this subsection, we analyze the performance of the exploration phase and its contribution to the expected sum-regret. The distributed algorithm of [15] assumes each device knows its CSI, or the possible QoS each channel supports. Our algorithm lifts this assumption by working on online estimations of the CSI (or QoS) instead. Each link obtains these estimations by randomly exploring the different K channels and averaging the instantaneous measurements of the QoS of each channel.

The following lemma characterizes the required estimation accuracy of the exploration phase, taking into account the dither noise.

Lemma 1 (Accuracy of Exploration Phase)**.** *Denote the dithered estimations of the expected QoS values in packet k by $\left\{Q_{n,i}^k\right\}$. Assume that $\left|Q_{n,i}^k - Q_{n,i} - u_{n,i}\right| \leq \Delta$ for each link n and channel i for some positive Δ. If $\Delta < \frac{3\Delta_{\min}}{8N}$, then*

$$\arg\max_{a_1,\ldots,a_N} \sum_{n=1}^{N} Q_{n,a(n)} = \arg\max_{a_1,\ldots,a_N} \sum_{n=1}^{N} Q_{n,a(n)}^k. \qquad (11)$$

Proof: The proof follows from the fact that if $Q_{n,i}^k$ and $Q_{n,i}$ are close enough for every i and n, then the optimal assignment on $\left\{Q_{n,i}^k\right\}$ and $\left\{Q_{n,i}\right\}$ must be identical. For details see Appendix A. ■

The following lemma concludes this subsection by providing an upper bound for the probability that the estimation for packet k failed. The fact that this error probability exponentially vanishes with k, allows us to limit the number of exploration time slots to c_1, keeping the overhead caused by the exploration phase negligible.

Lemma 2 (Exploration Error Probability)**.** *Denote the dithered estimations of the expected QoS values in packet k by $\left\{Q_{n,i}^k\right\}$. If the length of the exploration phase satisfies $c_1 \geq K \max\left\{\frac{81}{2}K, \frac{128}{9}\left(\frac{\Delta_{\max}}{\Delta_{\min}}\right)^2 N^2\right\}$, then after the k-th packet, we have*

$$P_{e,k} \triangleq \Pr\left(\max_{n,i}\left|Q_{n,i}^k - Q_{n,i}\right| > \frac{3\Delta_{\min}}{8N}\right) \leq 3NKe^{-k}. \qquad (12)$$

Proof: The proof uses Hoeffding's bound on both $\left|Q_{n,i}^k - Q_{n,i}\right|$ and the number of samples of $Q_{n,i}$ without collision. For details see Appendix B. ■

B. Auction Phase — Converging to the Optimal Allocation

As discussed in Section III-B1, we use the collision resolution algorithm to increase the quantization bits in order to avoid collision. Another issue to be resolved is when the continuous bids of two links m and n are identical, $B_{n,i} = B_{m,j}$. Since there is no auctioneer, the links cannot agree on an arbitrary tie braking without communication. Hence, identical bids can prevent the CSMA auction algorithm from converging to the optimal solution. In order to avoid this problem, the auction phase uses a noisy version of the estimated expected rewards from the exploration phase. This noise is an artificial dither added by the links independently such that the probability for identical bids will be zero.

Lemma 3. *After the k-th exploration phase we have $\Pr\left(B_{n,i} = B_{m,j}\right) = 0$ for any $n \neq m$ and any i, j.*

Proof: Due to the continuous (uniform) distribution of $u_{n,i}^k$ and $u_{m,j}^k$, for any $m \neq n$ and i, j, the probability that $Q_{n,i}^k = \frac{s_i(t)}{o_i} + u_{n,i}^k = Q_{m,j}^k = \frac{s_j(t)}{o_j} + u_{m,j}^k$ is zero. Since any bid $B_{n,i}$ is a linear combination of rewards and ε, also the probability that at a certain iteration of the auction algorithm $B_{n,i} = B_{m,j}$ is zero. ■

We emphasize that Lemma 3, and Lemma 4 (as given below), only help to show (in Lemma 5) that Algorithm 1 eventually converges to the optimal solution. Links start transmitting data from the first packet, using a possibly suboptimal allocation in the exploitation phase. Hence, Algorithm 1 is likely to perform well much before convergence to the optimal allocation occurred. Nevertheless, our simulations in Section VI suggest that convergence to the optimal allocation occurs very fast, already in the first or the second packet.

Lemma 4. *Algorithm 1 converges to some final value b_f, i.e., there exists a k_0 such that $b(k) = b_f$ for all $k > k_0$.*

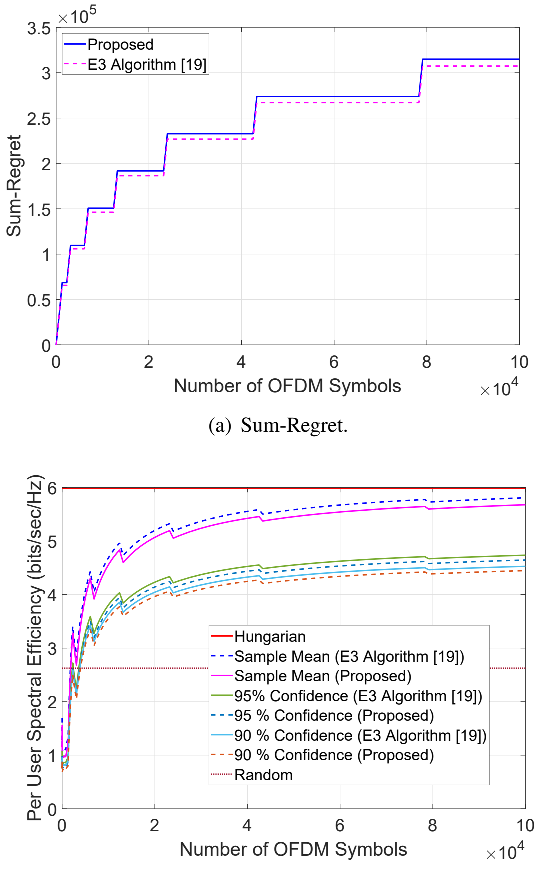
Proof: Consider two different bids $B_{n,i} \neq B_{m,j}$ of two different links $n \neq m$, and assume that after quantization to $b(k)$ bits we have $f_{b(k)}\left(B_{n,i}\right) = f_{b(k)}\left(B_{m,j}\right)$. In this case, links will detect a collision after the auction phase and will increase the number of bits used for quantization. Since $B_{n,i} - B_{m,j}$ is a sum of rewards and some multiplication of ε, for large enough $b(k) = b^*$, we have $f\left(B_{n,i}\right) \neq f\left(B_{m,j}\right)$ for any m, n, i, j such that $n \neq m$ and $B_{n,i} \neq B_{m,j}$. Hence, $b(k)$ will not increase above b^*, since collisions between $B_{n,i} \neq B_{m,j}$ cannot occur with $b(k) = b^*$. Collisions from identical bids $B_{n,i} = B_{m,j}$ do not occur simply because their probability is zero, as shown in Lemma 3. ■

Lemma 5 (Auction Convergence)**.** *Assume that $b(k') = b_f$ for all $k' > k$. If the k-th exploration phase succeeded, then the k-th auction phase converges to an allocation $a_1, \ldots, a_N$ such that $\left|\sum_{n=1}^{N} Q_{n,a_n}^k - \max_{a_1,\ldots,a_N} \sum_{n=1}^{N} Q_{n,a_n}^k\right| \leq \varepsilon$ with less than $\frac{KN}{\varepsilon_k}\left(Q_M + \frac{\Delta_{\min}}{8N}\right)2^{b(k)}$ time slots with probability 1. If $\varepsilon < \frac{\Delta_{\min}}{4K}$, then the auction phase converges to $\arg\max_{a} \sum_{n=1}^{N} Q_{n,a_n}$.*

Proof: The proof follows from the convergence and performance guarantees proven in [15] together with Lemma 1. For details see Appendix C. ■

Our main Theorem is formulated as follows.

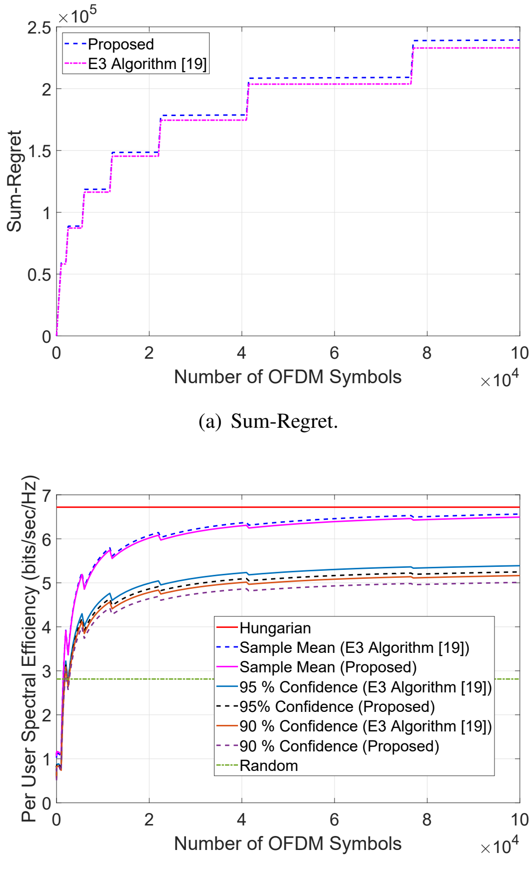
Theorem 1 (Main Result)**.** *Assume that the instantaneous QoS $\left\{q_{n,i}(t)\right\}_t$ are independent in n and i.i.d in time t, with expectations $Q_{n,i} \in \left\{Q_1, \ldots, Q_M\right\}$ such that $Q_i = l_i \Delta_{\min}$ for*



(a) Sum-Regret.

(b) Spectral Efficiency.

Fig. 3. Performance evaluation over i.i.d. Rayleigh fading channel. Simulation parameters are: $N = K = 10$, explore length= 800 OFDM symbols, and auction length = 500 OFDM symbols.



(a) Sum-Regret.

(b) Spectral Efficiency.

Fig. 4. Performance evaluation over LTE fading channel. Simulation parameters are: $N = K = 10$, explore length= 800 OFDM symbols, and auction length = 500 OFDM symbols.

a non-negative integer l_i and a positive $\Delta_{\min}$, and $Q_1 < \ldots < Q_M$. Denote $\Delta_{\max} = Q_M - Q_1$. Let each link run Algorithm 1 with $\varepsilon < \frac{\Delta_{\min}}{4K}$ and an exploration phase of length

$$c_1 \geq K \max\left\{\frac{81}{2}K, \frac{128}{9}\left(\frac{\Delta_{\max}}{\Delta_{\min}}\right)^2 N^2\right\}. \tag{13}$$

Then, the expected sum of regrets is $\bar{R} \sim \mathcal{O}(\log T)$.

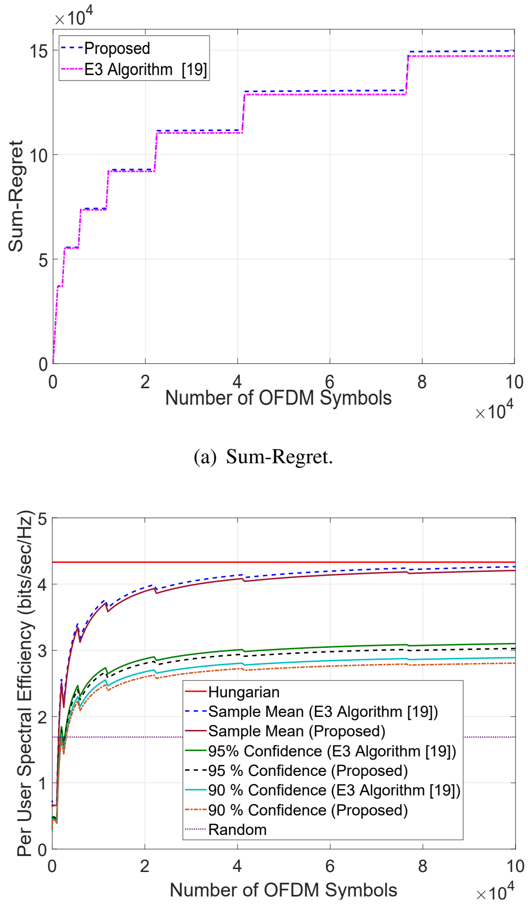
Proof: Lemma 5 shows that if the exploration phase succeeds and enough bits are used for the CSMA back-off quantization, then the exploitation phase contributes no regret to the sum of regret. Moreover, Lemma 2 upper bounds the error probability of the exploration phase. The bound implies that it decreases exponentially with k. The proof follows by bounding from above the expected regret using these two facts. For details see Appendix D. ■

V. Simulation Results

In this section, we demonstrate the performance of Algorithm 1 using computer simulations. We compare the proposed OALA algorithm (by simulating Algorithm 1) with the centralized Hungarian method, random channel selection and the E3 algorithm in [19]. The Hungarian method requires some central entity to know the CSI of all users. Requiring much less information, the E3 algorithm assumes that each user can decode the channel chosen by all other users. The proposed algorithm requires significantly less information — each user only needs to sense whether there is a transmission on a given channel. The role of the simulations is to show that despite our much stricter information constraints, our algorithm performs as well as the E3 algorithm and the optimal Hungarian algorithm. The comparison with the random channel selection assures that an algorithm that does not strive to converge to the optimal allocation performs very badly. This serves to show that the problem is far from being degenerated or trivial.

We verify our algorithm under various network scenarios consisting of different path losses and fading environments. The channel is divided into N sub-channels and we use $N = K = 10$. The transmit power spectral density (PSD) is fixed at 12dBm/Hz for each user. The users are assumed to be moving at a speed of 3km/h. We used a transmission duration of $T = 10^5$ time slots, with a single OFDM symbol per time slot ($L = 1$). Our transmission packet (see Fig. 2) has exploration phase of 800 OFDM symbols and an auction phase of 500 OFDM symbols. Each experiment consists of averaging 1000 independent realizations.

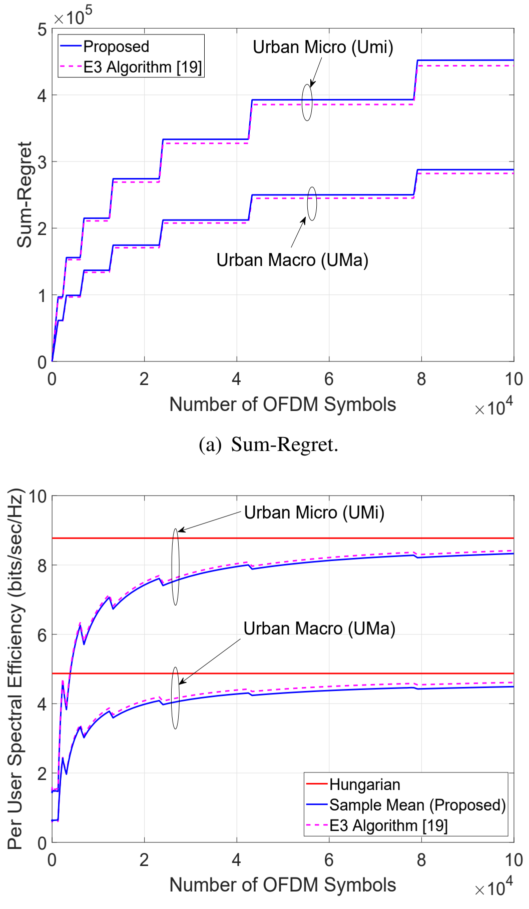
First, we consider an ad hoc network of N links uniformly distributed in a disk of radius 500 m. The central carrier



(a) Sum-Regret.

(b) Spectral Efficiency.

Fig. 5. Performance evaluation over LTE fading channel with alien interference. Simulation parameters: $N = K = 10$, explore length= 500 OFDM symbols, and auction length = 500 OFDM symbols.



(a) Sum-Regret.

(b) Spectral Efficiency.

Fig. 6. Performance evaluation over 5G fading channel. Simulation parameters: $N = K = 10$, explore length= 800 OFDM symbols, and auction length = 500 OFDM symbols.

frequency was 2 GHz with a per-user transmission bandwidth of 200 KHz. The path loss is computed using path loss exponent of $\alpha = 4$. We consider two types of channel models: i.i.d. Rayleigh fading channel and the extended pedestrian A model (EPA) of the LTE standard with 9 random taps. In Fig. 3a, we compare the sum-regret performance of our algorithm to that of the E3 algorithm [19] under an i.i.d. Rayleigh fading channel. It is evident that the performance of both algorithms is essentially identical, despite the fact that our algorithm uses no communication between users in contrast to the E3 algorithm. Both algorithms have an expected sum-regret increasing as $\log T$. Both converge to the optimal allocation quite rapidly. In Fig. 3b, we present the spectral efficiency of both algorithms together with the 90% and 95% outage (worst realizations). It can be seen that the performance of the OALA algorithm is similar to that of the E3 algorithm. It also shows that the proposed algorithm approaches the optimal performance within a few packets, which is much better than a random selection and behaves very similarly in all realizations. We have repeated the above experiment for the more realistic scenario of LTE channels in Fig. 4. This again confirms that our performance is identical to that of the E3 algorithm with significantly reduced communication and sensing overheads.

Next, we demonstrate the performance of the proposed algorithm in the presence of alien interference for LTE channels in Fig. 5. In this scenario, we consider four interferers that use four out of $K = 10$ available channels. These interfering nodes are randomly located outside the network disk and within a distance of 500 m from the annular region of the disk. It can be seen from the right graph in Fig. 5 that the spectral efficiency is reduced by ~2 bits/sec/Hz. However, the proposed algorithm achieves the optimal performance within few thousand symbols similar to the interference-free case, as shown in Fig. 4. This scenario again confirms that our performance is identical to that of the E3 algorithm.

Finally, we consider a 5G system consisting of pathloss, short-term fading, and long-term shadowing. We compute path loss from empirical models of urban macro (UMa) in the distance range of 45m to 1429m and urban micro-street canyon (UMi-SC) in the distance range of 19m to 272m [53], [54]. The shadowing factor is 6dB and 7.8dB the UMa and UMi-SC models, respectively. The fading channel consists of tapped delay line (TDL-A) model with 23 taps and a delay spread of 100ns. The central carrier frequency is 6GHz with a per-user transmission bandwidth of 720KHz. The results in Fig. 6 demonstrate that the all the realistic channel phenomena we simulate do not prevent the proposed algorithm from quickly converging to the optimal solution.

The simulations in this section provide additional solid support that our algorithm offers the same performance as [19] with significantly less overhead. Another important finding is that the exploration phase from a theoretical perspective needs to satisfy (13) can be much shorter in practice. This happens since in real scenarios, a correct estimation of all expected QoS is unnecessary, as most of them never appear during the iterations of the distributed auction algorithm. Hence, the learning overhead of our algorithm is negligible in practice.

VI. Conclusions

In this paper, we presented a novel online auction based learning algorithm for channel allocation over wireless channels where links initially have no estimation for the statistics of the channels. Learning the statistics of the channels in real-time (exploration) comes at the expanse of using the best known channels (exploitation). The scenario is modelled by a multi-player multi-armed bandit problem where a collision occurs if two or more links transmit on the same channel. We proved that our algorithm achieves the optimal order of regret — $\mathcal{O}(\log T)$. Our algorithm is based on a distributed auction algorithm that uses CSMA to avoid the need for an auctioneer (base station). In contrast to the state-of-the-art algorithms, our algorithm requires neither centralized management nor any communication between devices, which makes it very relevant to cognitive ad-hoc networks. Our algorithm only requires sensing a single channel at each time slot, which is K times less than the state-of-the-art algorithms. Only a detection of whether there are transmissions on this channel is required, and no decoding and demixing operations are needed to discern which user chose which channel. From a practical point of view, this results in a significant complexity reduction of the physical layer design. Simulations show that our algorithm performs very well on realistic LTE and 5G channels.

Appendix A
Proof of Lemma 1: Accuracy of Exploration Phase

Proof: Recall that $\Delta_{\min} = \min\limits_{i \neq j} |Q_i - Q_j|$. For all n and i we have $Q^k_{n,i} = Q_{n,i} + z_{n,i} + u_{n,i}$ such that $|u_{n,i}| \leq \frac{\Delta_{\min}}{8N}$, and we assume that $|z_{n,i}| \leq \Delta$. In the perturbed assignment problem, an optimal assignment $a^1 \in \arg \max\limits_{a_1,\ldots,a_N} \sum_{n=1}^N Q_{n,a_n}$ performs at least as well as

$$\sum_{n=1}^N Q^k_{n,a^1_n} = \sum_{n=1}^N \left(Q_{n,a^1_n} + z_{n,a^1_n} + u_{n,a^1_n}\right) \geq \sum_{n=1}^N Q_{n,a^1(n)} - \left(\Delta + \frac{\Delta_{\min}}{8N}\right) N. \quad (14)$$

Any non-optimal assignment a performs at most as well as

$$\sum_{n=1}^N Q^k_{n,a_n} \leq \sum_{n=1}^N \left(Q_{n,a^2_n} + z_{n,a_n} + u_{n,a_n}\right) \leq \sum_{n=1}^N Q_{n,a^2_n} + \left(\Delta + \frac{\Delta_{\min}}{8N}\right) N, \quad (15)$$

where a^2 is an assignment with the second best objective. For any two allocations $a \neq a'$ with a different sum of QoS we have

$$\sum_{n=1}^N Q_{n,a_n} - \sum_{n=1}^N Q_{n,a'_n} \geq \Delta_{\min}. \quad (16)$$

We conclude that for any non optimal a

$$\sum_{n=1}^N Q^k_{n,a^1_n} - \sum_{n=1}^N Q^k_{n,a_n} \underset{(a)}{\geq} \left(\sum_{n=1}^N Q_{n,a^1_n} - \sum_{n=1}^N Q_{n,a^2_n}\right) - \left(2\Delta + \frac{\Delta_{\min}}{4N}\right) N \underset{(b)}{\geq} \frac{3\Delta_{\min}}{4} - 2\Delta N \underset{(c)}{>} 0 \quad (17)$$

where (a) follows from (14) and (15), (b) from (16) and (c) holds for $\Delta < \frac{3\Delta_{\min}}{8N}$. ■

Appendix B
Proof of Lemma 2: Exploration Error Probability

Proof: After the k-th exploration phase, the number of samples that are used for estimating the expected QoS is $T_e(k) = c_1 k$. Let $A_{n,i}(t)$ be the indicator that is equal to one if only link n chose channel i at time slot t. Also define $V_{n,i}(t) \triangleq \sum_\tau A_{n,i}(\tau)$, which is the number of times that link n has used channel i with no collision, up to time slot t and define $V_{\min} = \min\limits_{n,i} V_{n,i}(t)$. Recall that $\Delta_{\max} = Q_M - Q_1$ and define the estimation error of channel i for link n by

$$\xi_{n,i} \triangleq \left| \frac{1}{V_{n,i}(t)} \sum A_{n,i}(\tau) r_{n,i}(\tau) - Q_{n,i} \right|, \quad (18)$$

Denote by E the event in which there exists a link n that has $\xi_{n,i} \geq \Delta$ for some channel i. We have

$$\Pr(E | V_{\min} = v) = \Pr\left(\bigcup_{i=1}^K \bigcup_{n=1}^N \{\xi_{n,i} \geq \Delta \,|\, V_{\min} = v\}\right) \underset{(a)}{\leq} NK \max_{n,i} \Pr(\xi_{n,i} \geq \Delta \,|\, V_{\min} = v) \underset{(b)}{\leq} 2NKe^{-\frac{2\Delta^2}{\Delta^2_{\max}} v}. \quad (19)$$

where (a) follows by taking the union bound over all links and channels and (b) from using Hoeffding's inequality for bounded variables [55]. Since the exploration phase consists of uniform and independent arm choices we have

$$\Pr(A_{n,i}(t) = 1) = \frac{1}{K}\left(1 - \frac{1}{K}\right)^{N-1}. \quad (20)$$

Therefore

$$\Pr\left(V_{\min} < \frac{T_e\left(k\right)}{4K}\right) = \Pr\left(\bigcup_{i=1}^{K}\bigcup_{n=1}^{N}\left\{V_{n,i}\left(t\right) \leq \frac{T_e\left(k\right)}{4K}\right\}\right)$$
$$\underset{(a)}{\leq} NK\Pr\left(V_{1,1}\left(t\right) \leq \frac{T_e\left(k\right)}{4K}\right)$$
$$\underset{(b)}{\leq} NKe^{-2\frac{1}{K^2}\left(\left(1-\frac{1}{K}\right)^{N-1}-\frac{1}{4}\right)^2 T_e(k)} \underset{(c)}{\leq} NKe^{-\frac{2}{81K^2}T_e(k)}, \tag{21}$$

where (a) follows from the union bound, (b) from Hoeffding's inequality for Bernoulli random variables and (c) since $K \geq N$ and $\left(1-\frac{1}{K}\right)^{K-1} - \frac{1}{4} \geq e^{-1} - \frac{1}{4} > \frac{1}{9}$. We conclude that

$$P_{e,k} \leq \Pr\left(E\right) =$$
$$\sum_{v=0}^{T_e(k)} \Pr\left(E|V_{\min} = v\right)\Pr\left(V_{\min} = v\right) \leq \sum_{v=0}^{\left\lfloor\frac{T_e(k)}{4K}\right\rfloor} \Pr\left(V_{\min} = v\right)$$
$$+ \sum_{\left\lceil\frac{T_e(k)}{4K}\right\rceil+1}^{T_e(k)} \Pr\left(E|V_{\min} = v\right)\Pr\left(V_{\min} = v\right)$$
$$\leq \Pr\left(V_{\min} < \frac{T_e\left(k\right)}{4K}\right) + \Pr\left(E|\ V_{\min} \geq \frac{T_e\left(k\right)}{4K}\right)$$
$$\underset{(a)}{\leq} 2NKe^{-\frac{\Delta^2 c_1}{2K\Delta_{\max}^2}k} + NKe^{-\frac{2c_1 k}{81K^2}}, \tag{22}$$

where (a) follows from (19) and (21). We choose $\Delta = \frac{3\Delta_{\min}}{8N}$ and $c_1 = K\max\left\{\frac{81}{2}K, \frac{128}{9}\left(\frac{\Delta_{\max}}{\Delta_{\min}}\right)^2 N^2\right\}$ to obtain

$$P_{e,k} \leq 2NKe^{-\frac{9\Delta_{\min}^2 c_1}{128K\Delta_{\max}^2 N^2}k} + NKe^{-\frac{2c_1 k}{81K^2}} \leq 3NKe^{-k}. \tag{23}$$

■

Appendix C
Proof of Lemma 5: Auction Convergence

Proof: In [15, Lemma 3] it is shown that the number of iterations I_{auc} of the distributed auction algorithm with ε is bounded by

$$I_{\text{auc}} \leq KN + \frac{K}{\varepsilon}\sum_{n=1}^{N} Q_{n,i}^k \underset{(a)}{\leq} KN + \frac{KN}{\varepsilon}\left(Q_M + \frac{\Delta_{\min}}{8N}\right), \tag{24}$$

where (a) follows since $Q_{n,i}^k \leq Q_M + \frac{\Delta_{\min}}{8N}$ for all n and i. Note that each iteration of the auction phase takes $2^{b(k)} + 1$ time slots. If the k-th exploration phase succeeded we have $\max\limits_{n,i}\left|Q_{n,i}^k - Q_{n,i}\right| < \frac{3\Delta_{\min}}{8N}$. For any two allocation $a \neq a'$ with a different sum of QoS we have

$$\left|\sum_{n=1}^{N} Q_{n,a_n} - \sum_{n=1}^{N} Q_{n,a'_n}\right| \geq \Delta_{\min} \tag{25}$$

Hence

$$\left|\sum_{n=1}^{N} Q_{n,a_n}^k - \sum_{n=1}^{N} Q_{n,a'_n}^k\right| =$$
$$\left|\sum_{n=1}^{N} Q_{n,a_n}^k - \sum_{n=1}^{N} Q_{n,a_n} + \sum_{n=1}^{N} Q_{n,a_n}\right.$$
$$\left. - \sum_{n=1}^{N} Q_{n,a'_n} + \sum_{n=1}^{N} Q_{n,a'_n} - \sum_{n=1}^{N} Q_{n,a'_n}^k\right|$$
$$\underset{(a)}{\geq} \left|\sum_{n=1}^{N} Q_{n,a_n} - \sum_{n=1}^{N} Q_{n,a'_n}\right| - \left|\sum_{n=1}^{N} Q_{n,a_n}^k - \sum_{n=1}^{N} Q_{n,a_n}\right|$$
$$-\left|\sum_{n=1}^{N} Q_{n,a'_n}^k - \sum_{n=1}^{N} Q_{n,a'_n}\right| \underset{(b)}{>} \Delta_{\min} - \frac{3\Delta_{\min}}{4} \geq \frac{\Delta_{\min}}{4}, \tag{26}$$

where (a) follows from the reverse triangle inequality and (b) from (25) and $\max\limits_{n,i}\left|Q_{n,i}^k - Q_{n,i}\right| < \frac{3\Delta_{\min}}{8N}$.

Denote by $\tilde{a}$ the allocation after the convergence of the auction phase. If $\varepsilon < \frac{\Delta_{\min}}{4K}$, then Theorem 1 in [15] guarantees that

$$\left|\sum_{n=1}^{N} Q_{n,\tilde{a}_n}^k - \max_{a_1,\ldots,a_N}\sum_{n=1}^{N} Q_{n,a_n}^k\right| < \frac{\Delta_{\min}}{4K}N \leq \frac{\Delta_{\min}}{4} \tag{27}$$

which, by (26), is only possible if

$$\sum_{n=1}^{N} Q_{n,\tilde{a}_n}^k = \max_{a_1,\ldots,a_N}\sum_{n=1}^{N} Q_{n,a_n}^k \underset{(a)}{=} \arg\max_{a_1,\ldots,a_N}\sum_{n=1}^{N} Q_{n,a_n} \tag{28}$$

where (a) follows from Lemma 1 since we assume that the k-th exploration phase succeeded. ■

Appendix D
Proof of Theorem 1: Main Result

Proof: Denote the number of packets that start within T time slots by E. Let k_0 be the index of a sufficiently large epoch. We compute the expected total regret as follows:

$$\bar{R} \leq \underbrace{\sum_{k=1}^{k_0}\bar{R}_k}_{\bar{R}_0} + \sum_{k=k_0+1}^{E}\bar{R}_k, \tag{29}$$

where $\bar{R}_k$ is the expected total regret of epoch k and $\bar{R}_0$ is a constant with respect to T. Denote by $P_{e,k}$ the error probability of the exploration of epoch k. In Lemma 5, we prove that if the exploration phase succeeded and the number of quantization bits $b\left(k\right)$ for the CSMA delay is large enough, then the auction phase is guaranteed to converge to the optimal solution of (5) for any $\varepsilon < \frac{\Delta_{\min}}{4K}$. This optimal allocation is played in the exploitation phase, which adds no additional regret to the total regret. We prove in Lemma 2 that if (13) holds then, $P_{e,k} \leq 3NKe^{-k}$. Hence, we obtain that for large enough k such that $b\left(k\right)$ is sufficiently large that

$$
\begin{aligned}
\bar{R}_k \le \left(c_1 + 4K^2 N\left(\frac{Q_M}{\Delta_{\min}} + \frac{1}{N}\right)\left(2^{b(k)} + 1\right)\right) NQ_M \\
+ 3NKc_2\left(\frac{2}{e}\right)^k NQ_M \\
\le 2\left(c_1 + 4K^2 N\left(\frac{Q_M}{\Delta_{\min}} + \frac{1}{N}\right)\left(2^{b_f} + 1\right)\right) NQ_M \qquad (30)
\end{aligned}
$$

for some constant b_f. We conclude that

$$
\begin{aligned}
\bar{R} &\le \bar{R}_0 + \sum_{k=k_0+1}^{E} \bar{R}_k \\
&\underset{(a)}{\le} \bar{R}_0 + 2\left(c_1 + 4K^2 N\left(\frac{Q_M}{\Delta_{\min}} + \frac{1}{N}\right)\left(2^{b_f} + 1\right)\right) NQ_M E \\
&\underset{(b)}{\le} \bar{R}_0 + 2\left(c_1 + 4K^2 N\left(\frac{Q_M}{\Delta_{\min}} + \frac{1}{N}\right)\left(2^{b_f} + 1\right)\right) \\
&\qquad \times NQ_M \log_2\left(\frac{T}{c_2} + 2\right) \qquad (31)
\end{aligned}
$$

where in (a) we used the fact that completing the last epoch to be a full epoch only increases $\bar{R}_k$. In (b) we used $T > \sum_{k=1}^{E-1} c_2 2^k \ge c_2\left(2^E - 2\right)$, which yields $E \le \log_2\left(\frac{T}{c_2} + 2\right)$. ■

References

[1] I. Katzela and M. Naghshineh, "Channel assignment schemes for cellular mobile telecommunication systems: A comprehensive survey," *IEEE Communications Surveys Tutorials*, vol. 3, no. 2, pp. 10–31, Second 2000.

[2] S. Chieochan, E. Hossain, and J. Diamond, "Channel assignment schemes for infrastructure-based 802.11 WLANs: A survey," *IEEE Communications Surveys Tutorials*, vol. 12, no. 1, pp. 124–136, First 2010.

[3] G. Ku and J. M. Walsh, "Resource allocation and link adaptation in LTE and LTE advanced: A tutorial," *IEEE Communications Surveys Tutorials*, vol. 17, no. 3, pp. 1605–1633, thirdquarter 2015.

[4] E. Z. Tragos, S. Zeadally, A. G. Fragkiadakis, and V. A. Siris, "Spectrum assignment in cognitive radio networks: A comprehensive survey," *IEEE Communications Surveys Tutorials*, vol. 15, no. 3, pp. 1108–1135, Third 2013.

[5] M. E. Tanab and W. Hamouda, "Resource allocation for underlay cognitive radio networks: A survey," *IEEE Communications Surveys Tutorials*, vol. 19, no. 2, pp. 1249–1276, Secondquarter 2017.

[6] C. Y. Wong, R. S. Cheng, K. B. Lataief, and R. D. Murch, "Multiuser OFDM with adaptive subcarrier, bit, and power allocation," *IEEE Journal on Selected Areas in Communications*, vol. 17, no. 10, pp. 1747–1758, Oct 1999.

[7] Z. Shen, J. G. Andrews, and B. L. Evans, "Adaptive resource allocation in multiuser OFDM systems with proportional rate constraints," *IEEE Transactions on Wireless Communications*, vol. 4, no. 6, pp. 2726–2737, Nov 2005.

[8] S. Sadr, A. Anpalagan, and K. Raahemifar, "Radio resource allocation algorithms for the downlink of multiuser OFDM communication systems," *IEEE Communications Surveys Tutorials*, vol. 11, no. 3, pp. 92–106, rd 2009.

[9] S. Huberman, C. Leung, and T. Le-Ngoc, "Dynamic spectrum management (DSM) algorithms for multi-user xDSL," *IEEE Communications Surveys Tutorials*, vol. 14, no. 1, pp. 109–130, First 2012.

[10] L. Gao and S. Cui, "Efficient subcarrier, power, and rate allocation with fairness consideration for OFDMA uplink," *IEEE Transactions on Wireless Communications*, vol. 7, no. 5, pp. 1507–1511, May 2008.

[11] J. Huang, V. G. Subramanian, R. Agrawal, and R. Berry, "Joint scheduling and resource allocation in uplink OFDM systems for broadband wireless access networks," *IEEE Journal on Selected Areas in Communications*, vol. 27, no. 2, pp. 226–234, Feb 2009.

[12] E. Yaacoub and Z. Dawy, "A survey on uplink resource allocation in OFDMA wireless networks," *IEEE Communications Surveys Tutorials*, vol. 14, no. 2, pp. 322–337, Second 2012.

[13] C. Papadimitriou and K. Steiglitz, *Combinatorial Optimization: Algorithms and Complexity*. Dover, 1998.

[14] D. P. Bertsekas, "The auction algorithm: A distributed relaxation method for the assignment problem," *Annals of Operations Research*, vol. 14, no. 1, pp. 105–123, Dec 1988.

[15] O. Naparstek and A. Leshem, "Fully distributed optimal channel assignment for open spectrum access," *IEEE Transactions on Signal Processing*, vol. 62, no. 2, pp. 283–294, Jan 2014.

[16] U. Challita, L. Dong, and W. Saad, "Proactive resource management in LTE-U systems: A deep learning perspective," *arXiv preprint arXiv:1702.07031*, June 2017.

[17] S. Wang, H. Liu, P. H. Gomes, and B. Krishnamachari, "Deep reinforcement learning for dynamic multichannel access in wireless networks," *IEEE Transactions on Cognitive Communications and Networking*, vol. 4, no. 2, pp. 257–265, June 2018.

[18] O. Naparstek and K. Cohen, "Deep multi-user reinforcement learning for distributed dynamic spectrum access," *IEEE Transactions on Wireless Communications*, vol. 18, no. 1, pp. 310–323, Jan 2019.

[19] N. Nayyar, D. Kalathil, and R. Jain, "On regret-optimal learning in decentralized multiplayer multiarmed bandits," *IEEE Transactions on Control of Network Systems*, vol. 5, no. 1, pp. 597–606, March 2018.

[20] O. Avner and S. Mannor, "Concurrent bandits and cognitive radio networks," in *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, Nancy, France, 14-18 Sept, 2014, pp. 66–81.

[21] A. Anandkumar, N. Michael, A. K. Tang, and A. Swami, "Distributed algorithms for learning and cognitive medium access with logarithmic regret," *IEEE Journal on Selected Areas in Communications*, vol. 29, no. 4, pp. 731–745, April 2011.

[22] K. Liu and Q. Zhao, "Cooperative game in dynamic spectrum access with unknown model and imperfect sensing," *IEEE Transactions on Wireless Communications*, vol. 11, no. 4, pp. 1596–1604, April 2012.

[23] K. Liu and Q. Zhao, "Indexability of restless bandit problems and optimality of Whittle index for dynamic multichannel access," *IEEE Transactions on Information Theory*, vol. 56, no. 11, pp. 5547–5567, Nov 2010.

[24] S. Vakili, K. Liu, and Q. Zhao, "Deterministic sequencing of exploration and exploitation for multi-armed bandit problems," *IEEE Journal of Selected Topics in Signal Processing*, vol. 7, no. 5, pp. 759–767, Oct 2013.

[25] J. Rosenski, O. Shamir, and L. Szlak, "Multi-player bandits—a musical chairs approach," in *33rd International Conference on Machine Learning*, New York, USA, 19-24 June, 2016, pp. 155–163.

[26] K. Liu and Q. Zhao, "Distributed learning in multi-armed bandit with multiple players," *IEEE Transactions on Signal Processing*, vol. 58, no. 11, pp. 5667–5681, Nov 2010.

[27] D. Kalathil, N. Nayyar, and R. Jain, "Decentralized learning for multiplayer multiarmed bandits," *IEEE Transactions on Information Theory*, vol. 60, no. 4, pp. 2331–2345, April 2014.

[28] I. Bistritz and A. Leshem, "Game of thrones: Fully distributed learning for multi-player bandits," arXiv:1810.11162 (v3), Feb 2019.

[29] ——, "Distributed multi-player bandits-a game of thrones approach," in *32nd Advances in Neural Information Processing Systems*, Vancouver Canada, 2-8 Dec, 2018, pp. 7222–7232.

[30] H. Kwon, S. Kim, and B. G. Lee, "Opportunistic multi-channel CSMA protocol for OFDMA systems," *IEEE Transactions on Wireless Communications*, vol. 9, no. 5, pp. 1552–1557, May 2010.

[31] Y. Yaffe, A. Leshem, and E. Zehavi, "Stable matching for channel access control in cognitive radio systems," in *2010 2nd International Workshop on Cognitive Information Processing*, Elba, Italy, 14-16 June 2010, pp. 470–475.

[32] A. Leshem, E. Zehavi, and Y. Yaffe, "Multichannel opportunistic carrier sensing for stable channel access control in cognitive radio systems," *IEEE Journal on Selected Areas in Communications*, vol. 30, no. 1, pp. 82–95, Jan 2012.

[33] O. Naparstek and A. Leshem, "Bounds on the expected optimal channel assignment in Rayleigh channels," in *2012 IEEE 13th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, Cesme, Turkey, 17-20 June 2012, pp. 294–298.

[34] D. Gale and L.Shapley, "College admissions and the stability of marriage," *The American Mathematical Monthly*, vol. 69, no. 1, pp. 9–15, 1962.

[35] R. Mochaourab, B. Holfeld, and T. Wirth, “Distributed channel assignment in cognitive radio networks: Stable matching and walrasian equilibrium,” *IEEE Transactions on Wireless Communications*, vol. 14, no. 7, pp. 3924–3936, July 2015.

[36] I. Bistritz and A. Leshem, “Game theoretic dynamic channel allocation for frequency-selective interference channels,” *IEEE Transactions on Information Theory*, vol. 65, no. 1, pp. 330–353, Jan 2019.

[37] Y. Xiao, K. C. Chen, C. Yuen, Z. Han, and L. A. DaSilva, “A bayesian overlapping coalition formation game for device-to-device spectrum sharing in cellular networks,” *IEEE Transactions on Wireless Communications*, vol. 14, no. 7, pp. 4034–4051, July 2015.

[38] A. Leshem and E. Zehavi, “Cooperative game theory and the Gaussian interference channel,” *IEEE Journal on Selected Areas in Communications*, vol. 26, no. 7, pp. 1078–1088, Sept 2008.

[39] Z. Han, Z. Ji, and K. J. R. Liu, “Fair multiuser channel allocation for OFDMA networks using Nash bargaining solutions and coalitions,” *IEEE Transactions on Communications*, vol. 53, no. 8, pp. 1366–1376, Aug 2005.

[40] A. Leshem and E. Zehavi, “Smart carrier sensing for distributed computation of the generalized Nash bargaining solution,” in *2011 17th International Conference on Digital Signal Processing (DSP)*, July 2011, pp. 1–5.

[41] K. Cohen, A. Leshem, and E. Zehavi, “Game theoretic aspects of the multi-channel ALOHA protocol in cognitive radio networks,” *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 11, pp. 2276–2288, Nov 2013.

[42] K. Cohen and A. Leshem, “Distributed game-theoretic optimization and management of multichannel ALOHA networks,” *IEEE/ACM Transactions on Networking*, vol. 24, no. 3, pp. 1718–1731, June 2016.

[43] J. Sun, E. Modiano, and L. Zheng, “Wireless channel allocation using an auction algorithm,” *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 5, pp. 1085–1096, May 2006.

[44] Z. Han, R. Zheng, and H. V. Poor, “Repeated auctions with Bayesian nonparametric learning for spectrum access in cognitive radio networks,” *IEEE Transactions on Wireless Communications*, vol. 10, no. 3, pp. 890–900, March 2011.

[45] A. Mukherjee and H. M. Kwon, “General auction-theoretic strategies for distributed partner selection in cooperative wireless networks,” *IEEE Transactions on Communications*, vol. 58, no. 10, pp. 2903–2915, Oct 2010.

[46] H. B. Chang and K. C. Chen, “Auction-based spectrum management of cognitive radio networks,” *IEEE Transactions on Vehicular Technology*, vol. 59, no. 4, pp. 1923–1935, May 2010.

[47] K. Yang, N. Prasad, and X. Wang, “An auction approach to resource allocation in uplink OFDMA systems,” *IEEE Transactions on Signal Processing*, vol. 57, no. 11, pp. 4482–4496, Nov 2009.

[48] M. Bayati, B. Prabhakar, D. Shah, and M. Sharma, “Iterative scheduling algorithms,” in *IEEE INFOCOM 2007 - 26th IEEE International Conference on Computer Communications*, Barcelona, Spain, May 2007, pp. 445–453.

[49] M. Bayati, D. Shah, and M. Sharma, “Max-product for maximum weight matching: Convergence, correctness, and LP duality,” *IEEE Transactions on Information Theory*, vol. 54, no. 3, pp. 1241–1251, March 2008.

[50] H. Zhu and J. Wang, “Radio resource allocation in multiuser distributed antenna systems,” *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 10, pp. 2058–2066, Oct. 2013.

[51] F. Rusek, D. Persson, B. K. Lau, E. G. Larsson, T. L. Marzetta, O. Edfors, and F. Tufvesson, “Scaling up MIMO: opportunities and challenges with very large arrays,” *IEEE Signal Processing Magazine*, vol. 30, no. 1, pp. 40–60, Jan. 2013.

[52] O. Simeone, U. Spagnolini, Y. Bar-Ness, and S. H. Strogatz, “Distributed synchronization in wireless networks,” *IEEE Signal Processing Magazine*, vol. 25, no. 5, pp. 81–97, Sept 2008.

[53] J. Meredith, “Study on channel model for frequency spectrum above 6 GHz,” 3GPP TR 38.900, Jun, Tech. Rep., 2016.

[54] S. Sun, T. S. Rappaport, S. Rangan, T. A. Thomas, A. Ghosh, I. Z. Kovacs, I. Rodriguez, O. Koymen, A. Partyka, and J. Jarvelainen, “Propagation path loss models for 5G urban micro-and macro-cellular scenarios,” in *2016 IEEE 83rd Vehicular Technology Conference (VTC Spring)*. Nanjing, China, 15-18 May 2016, pp. 1–6.

[55] W. Hoeffding, “Probability inequalities for sums of bounded random variables,” *Journal of the American Statistical Association*, vol. 58, no. 301, pp. 13–30, 1963.

S. M. Zafaruddin (M’12) received his Ph.D. degree in electrical engineering from the Indian Institute of Technology Delhi in 2013. From 2012 to 2015, he was with Ikanos Communications (now Qualcomm), Bangalore, India, working directly with the CTO Office, Red Bank, New Jersey, USA, where he was involved in research and development for xDSL systems. From 2015 to 2018, he was with the faculty of engineering at Bar-Ilan University, Ramat Gan, Israel, as a post-doctoral researcher working on signal processing for wireline and wireless communications. The Council for Higher Education, Israel, awarded him the Planning and Budgeting Commission Fellowship for Outstanding Post-Doctoral Researchers from China and India (2016—2018). He is currently a faculty member in the Department of Electrical and Electronics Engineering, BITS-Pilani, Pilani Campus. His current research interests include signal processing and machine learning for wireless and wireline communications, distributed signal processing, and resource allocation algorithms. He is an Associate Editor of IEEE Access.

Ilai Bistritz (S’16) received the B.Sc. degree (magna cum laude) and the M.Sc. degree (summa cum laude) in electrical engineering from Tel-Aviv University, Israel, in 2012 and 2016, respectively. Currently, he is pursuing Ph.D. in Electrical Engineering from the Stanford University, USA. His main research interest is game theory for distributed optimization with a focus on using stochastic tools to analyze games on networks and design distributed algorithms.

Amir Leshem (M’98—SM’06) received the B.Sc. degree (cum laude) in mathematics and physics, the M.Sc. degree (cum laude) in mathematics, and the Ph.D. degree in mathematics from the Hebrew University, Jerusalem, Israel, in 1986, 1990, and 1998, respectively. He is currently a Professor and one of the founders of the Faculty of Engineering, Bar-Ilan University, where he is also the Head of the Signal Processing Track. From 2003 to 2005, he was the Technical Manager of the U-BROAD Consortium developing technologies to provide 100 Mbps and beyond over copper lines. His main research interests include multichannel wireless and wireline communication, applications of game theory to dynamic and adaptive spectrum management of communication networks, array and statistical signal processing with applications to multiple element sensor arrays and networks, wireless communications, radio-astronomical imaging and brain research, set theory, and logic and foundations of mathematics. He was an Associate Editor of the IEEE Transactions on Signal Processing from 2008 to 2011. He was the Leading Guest Editor for special issues on signal processing for astronomy and cosmology in the IEEE Signal Processing Magazine and the IEEE Journal of Selected Topics.

Dusit Niyato (M’09—SM’15—F’17) is currently a professor in the School of Computer Science and Engineering, Nanyang Technological University. He received his B.Eng. from King Mongkut’s Institute of Technology Ladkrabang, Thailand, in 1999 and his Ph.D. in electrical and computer engineering from the University of Manitoba, Canada, in 2008. His research interests are in the area of energy harvesting for wireless communication, the Internet of Things, and sensor networks.