Fair Coexistence of Heterogeneous Networks: A Novel Probabilistic Multi-Armed Bandit Approach

Zhiwu Guo* Chicheng Zhang† Ming Li* Marwan Krunz*
* Electrical and Computer Engineering, The University of Arizona, Tucson, Arizona, 85721, USA
† Computer Science, The University of Arizona, Tucson, Arizona, 85721, USA
Email: zhiwuguo@arizona.edu, chichengz@cs.arizona.edu, {lim, krunz}@arizona.edu

***Abstract*—The licensed spectrum of cellular networks has become increasingly crowded, leading to the standardization of LTE licensed assisted access (LTE-LAA) and 5G NR-U for deployment in unlicensed bands such as 5 GHz. To coexist harmoniously with other unlicensed wireless technologies like WiFi, LAA and 5G NR-U enforce listen-before-talk (LBT) protocol. This paper proposes methods to enhance the overall spectrum efficiency and fairness of each coexisting heterogeneous link. To improve the overall spectrum efficiency, we propose enabling concurrent transmissions of multiple links. Motivated by the need for fair coexistence of heterogeneous networks with concurrent transmissions, we formulate a variant of the multi-armed bandit (MAB) problem that finds a probabilistic transmission strategy to maximize the minimum link throughput. We propose the Fair Probabilistic Explore-Then-Commit (FP-ETC) algorithm, which achieves the expected regret of $O\left(T^{\frac{2}{3}}(K\log T)^{\frac{1}{3}}\right)$. We compare FP-ETC with existing MAB algorithms via extensive simulations, and the results show that FP-ETC significantly outperforms the baseline algorithms.**

***Index Terms*—Online learning, probabilistic multi-armed bandit, max-min fairness, explore-then-commit.**

I. Introduction

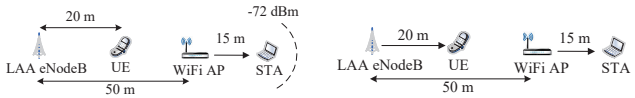
Due to the limited availability of licensed spectrum for cellular networks, 3GPP has standardized the usage of 5 GHz unlicensed bands for LTE-LAA [1]. Later on, the utilization of both 5 GHz and 6 GHz unlicensed spectrum was extended for 5G NR-U [2]. For medium access control (MAC), both LTE-LAA and 5G NR-U adopt listen-before-talk (LBT) to coexist harmoniously with other wireless technologies (e.g., WiFi) in these unlicensed bands. LBT is similar to carrier sense multiple access with collision avoidance (CSMA/CA) adopted in WiFi networks. LBT and CSMA/CA enforce clear channel assessment (CCA), which utilizes energy detection (ED) to determine if a channel is occupied or clear.

These collision avoidance-based MAC protocols have led to inefficient spectrum utilization [3], [4]. Fairness is also an important issue in heterogeneous network coexistence such as LAA/WiFi. There exists an asymmetry between the ED thresholds of LAA and WiFi links. Specifically, WiFi devices have an ED threshold of -62 dBm [5], while LAA devices have a lower ED threshold of -72 dBm [1]. The different ED thresholds, combined with other differences in contention and transmission parameters (such as transmission durations), can result in unfair coexistence between LAA and WiFi networks [4].

This research was supported in part by NSF (grants # 2229386 and 1822071) and by the Broadband Wireless Access & Applications Center (BWAC). Any opinions, findings, conclusions, or recommendations expressed in this paper are those of the author(s) and do not necessarily reflect the views of NSF.

A. Motivation

Concurrent transmissions of multiple heterogeneous links have been proposed to improve the spectrum utilization. The capture effect [6], multiple-input and multiple-output (MIMO) [7], and successive interference cancellation (SIC) [8], [9] can all be leveraged to recover the interfered signals. We use a toy example in Fig. 1 to demonstrate the potential throughput improvement, where the set of concurrently transmitting links is defined as *concurrent transmission set* (CTS). The transmission power of WiFi AP and LAA eNodeB are both set as 23 dBm [1], [5]. We simulate the normalized throughput of both LAA and WiFi links. The capture effect [6] is considered in LAA UE and WiFi STA. In Fig. 1(a), we observe that the total normalized throughput of both networks under collision avoidance is approximately 0.98 (0.58 for WiFi and 0.4 for LAA links). In Fig. 1(b), we allow concurrent transmissions for both links, and the total normalized throughput is increased from 0.98 to 1.27 with concurrent transmissions, resulting in an improved total throughput.



(a) Collision avoidance, $thr_L = 0.40, thr_W = 0.58$.

(b) Concurrent transmission, $thr_L = 0.33, thr_W = 0.94$

Fig. 1: Toy example of two-link LAA/WiFi coexistence.

While concurrent transmissions can improve the overall throughput, optimizing the transmission strategy for fairness objectives, such as max-min fairness, is not straightforward. For example, in Fig. 1(b), always transmitting both LAA and WiFi links concurrently does not achieve optimal max-min fairness since the minimum link throughput is only 0.33. To improve fairness, one can propose a transmission strategy where LAA transmits alone with a probability of 0.379, and LAA and WiFi concurrently transmit with a probability of 0.621. This optimized transmission strategy results in a minimum link throughput of 0.58, an improvement from the previous strategy.

In practice, the transmission success probabilities (and throughput) of each link within each CTS are unknown a priori. While offline training can be used to collect such information, it can incur significant overhead and delay due to the large number of CTSes [9]–[11]. Therefore, an online learning approach is preferred, where the link success probabilities are learned while (concurrent) transmissions happen, and optimal transmission decisions are made on-the-fly to maximize certain performance objective (e.g., total or minimum throughput).

B. Related Work and Challenges

The stochastic multi-armed bandit (MAB) problem [12] is a classic problem for sequential decision-making in an uncertain environment. The goal of the decision-maker is to maximize the expected cumulative reward. The MAB model has been widely applied in practice, such as cognitive radio networks [13] and resource allocation [14]. However, the basic MAB model has limited efficiency in a combinatorial setting, where there can be a large number of arms (e.g., exponential) and multiple arms can be played simultaneously. To address these challenges, the combinatorial multi-armed bandit (CMAB) [15], [16] has been proposed. In CMAB, the decision-maker plays a super arm, which consists of multiple individual arms, in each round. The rewards of the selected individual arms are then observed.

However, neither the basic MAB or CMAB can solve our problem. The basic MAB model overlooks the fact that multiple links can be simultaneously transmitted. And unlike the CMAB model, in our problem, the reward (throughput) of each CTS is not a linear combination of the rewards of each individual link. Furthermore, the reward of a link in one CTS is not related to the corresponding link's reward in another CTS, as they form different coexistence topologies. Thus, there does not exist any correlation for the rewards across different link sets/combinations.

In addition, several works have considered fairness in bandit algorithms. For example, the FAIRBANDIT algorithm, proposed in [17], plays all arms with equal probability until they can be distinguished with a high degree of confidence. Other works, such as [18], [19], propose fair MAB algorithms that ensure each arm is pulled at least a pre-specified fraction of the time. However, these values are difficult to determine in practice, since the algorithm does not know which CTS performs better than others in advance. Another work [20] proposes Maxmin-UCB, which integrates max-min fairness into the UCB algorithm. However, for our problem, using Maxmin-UCB will lead to poor performance, as not all links are active in each CTS in our problem, and Maxmin-UCB does not utilize a probabilistic strategy (instead it identifies the best arm).

C. Contributions

Our main contributions are summarized as follows:

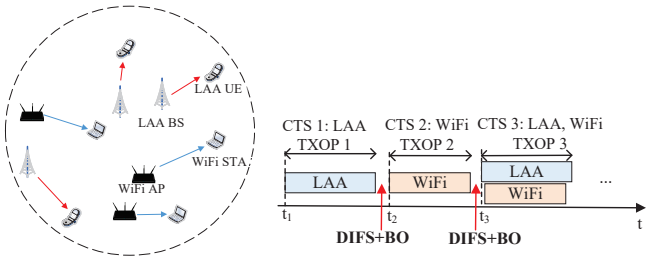
(1) Motivated by the fair coexistence of heterogeneous networks with concurrent transmissions, we formulate a variant of the multi-armed bandit problem that finds a probabilistic transmission strategy to maximize the minimum link throughput. We also define a novel notion of regret under this setting.

(2) We propose the Fair Probabilistic Explore-Then-Commit (FP-ETC) algorithm to solve the above probabilistic multi-armed bandit problem. We also obtain a sublinear regret upper bound for the proposed algorithm.

(3) We conduct extensive simulations to evaluate the effectiveness of the proposed algorithm. We compare FP-ETC with three baselines, namely UCB and ETC for maximizing the total CTS throughput, and Maxmin-UCB. Simulation results show that FP-ETC significantly outperforms the baseline algorithms. Finally, we show that our proposed algorithm is also applicable to many other real-world applications, such as energy harvesting [20], scheduling in wireless networks [21], and etc.

II. System Model and Problem Formulation

We consider a multi-link LTE-LAA/WiFi coexistence model [9], [22] in which multiple LAA and WiFi links share the same 5 GHz unlicensed band in the same area, as depicted in Fig. 2(a). Without loss of generality, we focus on the downlink scenario for each link, i.e., the transmission from the LAA BS to UE and from the WiFi AP to STA. Unlike the state-of-the-art collision avoidance-based MAC protocols such as CSMA/CA and LBT, we assume that the concurrent transmissions are allowed, and all links are situated in the same sensing domain, meaning that every transmitter can sense all others' transmissions. Furthermore, we assume that there are no hidden terminals.



(a) (b)

Fig. 2: (a) System model; (b) Illustrative example of contention-based MAC protocol implementation.

To model the concurrent transmissions of the coexisting links, we define *concurrent transmission set (CTS)* as follows:

Definition 1 (CTS [9])**.** *A concurrent transmission set (CTS) is a set of links with overlapping transmissions.*

Remark 1: For the scenario of N coexisting LAA/WiFi links, there are $K = 2^N - 1$ different combinations of non-empty sets. Thus, there are $2^N - 1$ possible CTSes. For example, for $N = 2$, there are three CTSes: $\text{CTS}_1 = \{\text{link } 1\}$, $\text{CTS}_2 = \{\text{link } 2\}$, $\text{CTS}_3 = \{\text{link } 1, \text{link } 2\}$. As empirically indicated in [9], in a collision domain, it is reasonable to have $N \leq 5$. For large N, to reduce the number of CTSes, we can divide N links into multiple collision domains and allocate orthogonal resources to different collision domains.

Next, we describe the setup of our problem. For the scenario of N coexisting LAA/WiFi links, let $\mathcal{N} = \{1, 2, \cdots, N\}$ be the link set, there are $K = 2^N - 1$ CTSes. Denote each CTS as an arm. We illustrate the online decision-making process for CTS selection in Alg. 1. Note that K can be any arbitrary positive integer, depending on real-world applications. Therefore, Alg. 1 is generic and can be applied to other applications. In $t = 1, 2, \cdots, T$, a decision maker picks CTS a to transmit. It then receives reward $r(a, l, t)$ for any link $l \in C_a$, where C_a is the link set of CTS a, $r(a, l, t)$ is randomly sampled from an unknown distribution $\mathcal{D}_{a,l}$. We assume that the reward for each CTS is i.i.d. over time.

Algorithm 1 Online decision-making for CTS selection

1: **Parameters**: $K = 2^N - 1$ CTSes and T transmission periods (both are known); reward distribution $\mathcal{D}_{a,l}$ (unknown), $l \in C_a, a \in [K]$.
2: **for** $t = 1$ to T **do**
3: A decision maker picks CTS a to transmit.
4: Observe reward $r(a, l, t) \sim \mathcal{D}_{a,l}$ for any link $l \in C_a$.
5: **end for**

To facilitate the optimization of fairness objectives, we define the transmission strategy as a probabilistic selection of CTSes (instead of a fixed choice) as follows:

Definition 2 (CTS Selection Vector)**.** *CTS selection vector is denoted as* $\mathbf{p} = (p_1, \ldots, p_K)$*, where* $K = 2^N - 1$*,* $p_i (1 \leq i \leq K)$ *represents the probability of CTS*$_i$ *being selected in each transmission period.*

In this paper, we aim to find the optimal $\mathbf{p}$ considering the fairness objective of *maximizing the minimum link throughput*. Denote $g(a, l)$ as the true mean of successful decoding probability of link $l \in C_a$. $\forall a \in [K], l \in C_a$, if $g(a, l)$ is known, one can obtain the optimal $\mathbf{p}$ by solving the following max-min optimization problem:

$$\begin{aligned} \text{Opt-min}: \max_{\mathbf{p}} \quad & f(\mathbf{p}) \\ \text{s.t.} \quad & 0 \leq p_a \leq 1, a \in [K], \\ & \sum_{a, a \in [K]} p_a = 1, \end{aligned} \tag{1}$$

where $f(\mathbf{p}) = \min_{l \in \mathcal{N}} \{\sum_{a \in [K]} \big(p_a \times g(a, l)\big)\}$, p_a is the a-th element of $\mathbf{p}$. For an objective comparison, we also present Opt-total, which aims at *maximizing the total throughput*:

$$\begin{aligned} \text{Opt-total}: \max_{\mathbf{p}} \quad & f_2(\mathbf{p}) \\ \text{s.t.} \quad & \text{same constraint with Opt-min}, \end{aligned} \tag{2}$$

where $f_2(\mathbf{p}) = \sum_{a \in [K]} p_a \sum_{l \in C_a} g(a, l)$.

However, $\forall a \in [K], l \in C_a, g(a, l)$ is unknown to the decision maker in advance. Hence, it needs to explore all CTSes and learns to obtain an accurate estimation of $g(a, l), l \in C_a, a \in [K]$. Denote $\hat{g}(a, l, t)$ as the empirical mean of successful decoding probability of link $l \in C_a$ until time t. We present how to obtain $\hat{g}(a, l, t)$ as follows. For CTS a, if the transmission of link $l \in C_a$ is successful, i.e., an ACK is observed at time t, $r(a, l, t) = 1$, otherwise $r(a, l, t) = 0$. For any link $l' \notin C_a$, namely, link l' does not transmit in CTS a, we let $r(a, l', t) = 0$ for generalization purpose. Denote $n_t(a)$ as the number of times that CTS a has been transmitted until t, which can be represented as $n_t(a) = \sum_{\tau=1}^{t} \mathbf{1}\{a_\tau = a\}$, where a_τ is the CTS index transmitted in transmission round τ. Then, $\hat{g}(a, l, t)$ can be presented as follows:

$$\hat{g}(a, l, t) = \frac{1}{n_t(a)} \sum_{\tau=1}^{t} \mathbf{1}\{a_\tau = a\} r(a, l, \tau). \tag{3}$$

As previously mentioned, the decision maker explores all CTSes to learn a more accurate $\hat{g}(a, l, t)$, while also exploiting the currently-known information to make the best action. Both exploration and exploitation incur a loss compared to the best action. We call this loss *regret*, which measures how much the decision maker regrets not knowing the best action in advance. The goal of the decision maker is to minimize the incurred regret. However, unlike classic maximization problems, Opt-min is a max-min optimization problem, which requires a redefinition of the regret. Inspired by the definition of regret (i.e., Equation (1.1) of [23]), the regret of opt-min is defined as:

$$R_T = \min_{l \in \mathcal{N}} \sum_{t=1}^{T} r(b_t, l, t) - \min_{l \in \mathcal{N}} \sum_{t=1}^{T} r(a_t, l, t), \tag{4}$$

where $b_1, \ldots, b_T$ is a sequence of CTSes drawn i.i.d. from $\mathbf{p}^*$, and $a_1, \ldots, a_T$ is the sequence of CTSes chosen by the decision maker. Accordingly, define the expected regret to be

$$\mathbb{E}[R_T] = \mathbb{E}\left[\min_{l \in \mathcal{N}} \sum_{t=1}^{T} r(b_t, l, t)\right] - \mathbb{E}\left[\min_{l \in \mathcal{N}} \sum_{t=1}^{T} r(a_t, l, t)\right], \tag{5}$$

where the expectation is with respect to (1) the choices of $b_1, \ldots, b_T$; (2) the random rewards drawn from the environment; (3) the random choices of $a_1, \ldots, a_T$ selected by the decision maker.

Remark 2: There are two ways to implement the overlapping transmissions of CTS in each transmission period. The simplest way is to implement it as a slotted MAC protocol, such as time division multiple access (TDMA) or duty-cycle MAC protocols [24]. More practically, it can be implemented as a contention-based MAC protocol, such as CSMA/CA, LBT. Regardless of the implementation, the decision maker makes decision and learns $\hat{\mathbf{p}}$ in each transmission period. For the slotted MAC protocol implementation, the decision maker conducts Alg. 1 in a centralized manner. For the contention-based MAC protocol implementation, each link can make distributed decision using Alg. 1. For such a case, we provide an illustrative example of concurrent transmission protocol under LAA/WiFi coexistence in Fig. 2(b). In this example, time is divided into multiple transmission opportunities (TXOPs), with DIFS and backoff (BO) between different TXOPs. In each TXOP, each link conducts contentions to access the shared channel. Based on the learned $\hat{\mathbf{p}}$, other links will

calculate a probability to determine whether they should access the channel with the ongoing transmissions. For instance, assume $\hat{\mathbf{p}} = (p_1, p_2, p_3)$, and the LAA link wins contention and transmit at t_1. The WiFi link obtains the probability of concurrently transmitting with the LAA link as $\frac{p_3}{p_3+p_1}$.

III. Fair Probabilistic Explore-then-commit Algorithm

In this section, we introduce the Fair Probabilistic Explore-Then-Commit (FP-ETC) algorithm for the proposed probabilistic MAB setting. After that, we analyze its regret under Opt-min [1].

The procedure of FP-ETC algorithm is outlined in Alg. 2, where the input m is a fixed positive integer, K is the number of CTSes. If $t \leq mK$, the algorithm is in the exploration phase (i.e., seeking better options) as shown from Step 3 to Step 6. Specifically, the FP-ETC algorithm first transmits each CTS in a round robin fashion m times, it then updates the empirical mean of successful decoding probability of all links under the corresponding CTS. Once $t > mK$, the algorithm enters the exploitation phase (i.e., staying with the currently-known best option) starting from Step 8. In Step 8, the minimum link throughput is maximized to obtain the estimated $\hat{\mathbf{p}}$. After that, FP-ETC sticks to the currently-known best option (i.e., $\hat{\mathbf{p}}$) and sample out a CTS a_t based on categorical distribution of $\hat{\mathbf{p}}$ as shown in Step 9. The sampled CTS is transmitted in Step 10.

Note that Alg. 2 is utilized to determine the transmission policy of CTS in each transmission period and it interacts with line 3 of Alg. 1.

Algorithm 2 Fair Probabilistic Explore-Then-Commit (FP-ETC)

1: **Input** : Positive integers m, K.
2: **for** $t = 1$ to T **do**
3: **if** $t \leq mK$ **then**
4: $a_t = t \bmod K + 1$.
5: Transmit CTS$_{a_t}$ in transmission period t
6: Update $\hat{g}(a_t, l, t), l \in C_{a_t}$.
7: **else**
8: $\hat{\mathbf{p}} = \arg\max_{\mathbf{p}} \quad \min_{l\in\mathcal{N}}\{\sum_{a\in[K]} \big(p_a \hat{g}(a, l, mK)\big)\}$
9: Sample out a CTS a_t based on categorical distribution of $\hat{\mathbf{p}}$.
10: Transmit CTS$_{a_t}$ in transmission period t
11: **end if**
12: **end for**

Remark 3: If we treat our problem as a MAB problem with each CTS as an arm and the reward of each arm defined as the summation of rewards for all links in the corresponding arm. Directly applying the standard ETC algorithm [12] cannot address Opt-min as the standard ETC algorithm was designed without considering fairness for each link. The reason why FP-ETC can address the fairness requirement is that it adds a new layer (i.e., $\mathbf{p}$) on top of CTSes, treating $\hat{\mathbf{p}}$ as a virtual arm. The action space for FP-ETC is continuous, which presents the first challenge in analyzing the algorithm. To bridge the gap between $\hat{\mathbf{p}}$ and CTSes, we first present a CTS-level concentration bound (Lemma 1) using the Hoeffding inequality. We then obtain a corresponding concentration bound with respect to any $\mathbf{p}$ (Lemma 2).

[1]The FP-ETC algorithm under Opt-total is reduced to traditional ETC algorithm, for space limitation, we omit the analysis in this paper.

A. Concentration Bounds

Lemma 1. *$\forall$ CTS a, $l \in C_a$, define event $E_{a,l}$: $\big|\hat{g}(a, l, mK) - g(a, l)\big| \leq \sqrt{\frac{2\log(T)}{m}}$, where m is the number of rounds that each CTS is transmitted in the exploration phase, K is the number of CTSes, T is the total number of transmission rounds. Then $Pr(E_{a,l}) \geq 1 - \frac{2}{T^4}$.*

Lemma 1 is a direct application of the Hoeffding inequality (Theorem A.1 of [12] by setting $\alpha = 2, \beta = 1$), given $r(a, l, t) \in [0, 1]$. It shows that the estimated $\hat{g}(a, l, mK)$ concentrates around its true mean $g(a, l)$ after the exploration phase.

As previously mentioned, $\hat{\mathbf{p}}$ is viewed as an arm in FP-ETC. Obtaining CTS-level concentration bound is insufficient to derive the regret bound of FP-ETC; to address this, we present the concentration bound for any $\mathbf{p}$ as follows.

Lemma 2 (Concentration Bound for Any $\mathbf{p}$)**.** *Define function $f(\mathbf{p}) = \min_{l\in\mathcal{N}}\{\sum_{a\in[K]} \big(p_a \times g(a, l)\big)\}$, where p_a is the a-th element of $\mathbf{p}$. Under FP-ETC, $\forall$ $\mathbf{p}$, $Pr\Big(\Big|\hat{f}(\mathbf{p}) - f(\mathbf{p})\Big| \leq \sqrt{\frac{2\log(T)}{m}}\Big) \geq 1 - \frac{2NK}{T^4}$, where N is the number of coexisting links, $\hat{f}(\mathbf{p}) = \min_{l\in\mathcal{N}}\{\sum_{a\in[K]} \big(p_a \times \hat{g}(a, l, mK)\big)\}$.*

Proof outline: we first rewrite $f(\mathbf{p}) = \min_{l\in\mathcal{N}} h_l(\mathbf{p})$ and bound $\Big|h_l(\mathbf{p}) - \hat{h}_l(\mathbf{p})\Big|$ for a given link $l \in \mathcal{N}$. After that, we prove that $\Big|\hat{f}(\mathbf{p}) - f(\mathbf{p})\Big|$ can be bounded by utilizing Lemma 3. The detailed proof of Lemma 2 and Lemma 3 are shown in Section VII-A of Appendix.

B. Upper Bound on Regret

After obtaining the concentration bound with regard to any $\mathbf{p}$. We are ready to upper bound the regret of FP-ETC. We state the results in Theorem 1.

Theorem 1. *Under FP-ETC, the average regret $\mathbb{E}[R_T]$ defined in Eq. (5) is upper bounded by $O\big(T^{\frac{2}{3}}(K\log T)^{\frac{1}{3}}\big)$.*

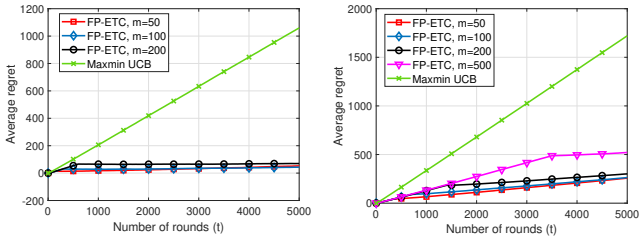
Proof outline: We first utilize Lemma 2 to upper bound $R_T^f = \sum_{t=1}^{T}[f(\mathbf{p}^*) - f(\mathbf{p_t})]$, where $\mathbf{p_t}$ is the $\mathbf{p}$ vector at transmission round t. However, there is still a gap between R_T^f and R_T of Eq. (4). To bridge the gap, we make use of Hoeffding inequality, union bound, and Lemma 4. The detailed proof of Theorem 1 is shown in Section VII-B of Appendix. Lemma 4 is presented in Section VII-C of Appendix.

IV. Simulation Results

In this section, we evaluate the performance of the PF-ETC algorithm via simulations under LAA/WiFi coexistence

scenarios, where SIC is enabled in each receiver to cancel possible interference and the successful decoding SINR threshold is set to 10 dB. Rayleigh channel is considered for each link. Due to space limitation, we only present the results for $N = 2$ and $N = 3$ coexisting links, as other scenarios have similar observations. There are 3 CTSes and 7 CTSes for the scenario of $N = 2$ and $N = 3$, respectively. We use the average regret defined in Eq. (5), minimum link throughput, Jain fairness index (JFI) [25] as the performance metrics to measure our proposed algorithm and baselines. Specifically, the throughput is normalized and represents the effective channel utilization. Let x_i be the throughput of link $i \in \mathcal{N}$, $JFI(x_1, \cdots, x_N) = \frac{(\sum_{i=1}^{N} x_i)^2}{N \times \sum_{i=1}^{N} x_i^2}$, which ranges from $\frac{1}{N}$ (worst case) to 1 (best case), it is maximum when all links have the same throughput. We simulate 500 randomized LAA/WiFi coexistence topology, where all nodes in each topology are uniformly distributed in a 100×100 m^2 area.

First, we show the average regret of FP-ETC for different choices of m and compare FP-ETC with the Maxmin UCB algorithm [20]. These results are presented in Fig. 3. We observe that the average regret of Maxmin UCB increases linearly with t, which is due to their algorithm selecting CTS_{a_t} in transmission round t according to the following rule: $a_t = \arg\max_{a \in [K]} [\min_{l \in \mathcal{N}} \hat{g}(a, l, t-1) + \sqrt{\frac{2 \log T}{n_{t-1}(a)}}]$. The Maxmin UCB algorithm aims to identify one best CTS instead of a probabilistic CTS selection strategy. On the other hand, FP-ETC can select a combination of different CTSes to maximize the reward and satisfy the max-min fairness at the same time, which is not achievable in Maxmin UCB. As a result, FP-ETC has much lower regret than Maxmin UCB. The probabilistic MAB algorithm is actually a generalized version of the corresponding basic MAB algorithm. It is also expected that higher m generally results in higher average regret as FP-ETC incurs significant regret during the exploration phase.

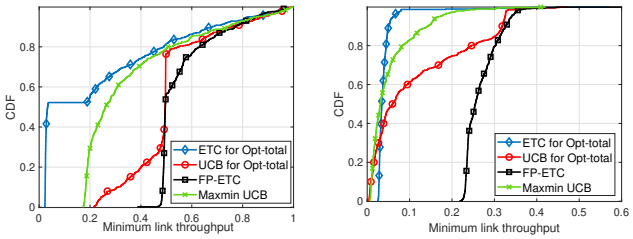


(a) $N = 2$ (b) $N = 3$

Fig. 3: Average regret vs the number of rounds t.

Next, we compare FP-ETC with three baselines: UCB for Opt-total, ETC for Opt-total, and Maxmin UCB [20], where Opt-total is defined in Eq. (2). The UCB for Opt-total selects CTS_{a_t} in transmission round t according to the following rule: $a_t = \arg\max_{a \in [K]} [\sum_{l \in C_a} \hat{g}(a, l, t-1) + |C_a| \sqrt{\frac{2 \log T}{n_{t-1}(a)}}]$, $|C_a|$ is the number of links in CTS a. m is set to 100 in ETC algorithms. The minimum link throughput and Jain fairness index (JFI) are compared for these algorithms.

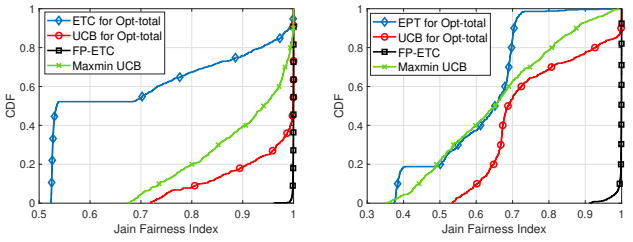
The cumulative distribution functions (CDFs) of the minimum link throughput for the four aforementioned algorithms are presented in Fig. 4. The simulation is based on 500 randomized LAA/WiFi coexistence topologies with $T = 5000$. As we can see, FP-ETC achieves the highest minimum link throughput among the four algorithms. This is attributed to the additional CTS selection vector $\mathbf{p}$ in FP-ETC, which allows for tuning the selection probability of each CTS to satisfy the max-min fairness requirement.



(a) $N = 2$ (b) $N = 3$

Fig. 4: CDF of minimum link throughput.

Fig. 5 shows the CDFs of JFI for the four aforementioned algorithms. It is clear that FP-ETC has the highest JFI compared to the other algorithms. It is worth noting that FP-ETC almost guarantees the same throughput for all links under all LAA/WiFi coexisting topologies, which is demonstrated by the fact that all JFI values of FP-ETC are close to 1. This result confirms that FP-ETC is an effective solution to achieve the max-min fairness requirement.



(a) $N = 2$ (b) $N = 3$

Fig. 5: CDF of Jain fairness index (JFI).

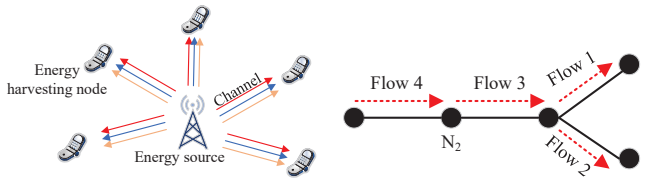
V. Other Applications

In this section, we discuss other practical applications of our proposed probabilistic MAB approach. Abstractly, the proposed framework is applicable to scenarios with two features: (1) a non-additive set consisting of multiple individual elements is explored in each round; (2) the decision maker's goal is to optimize the fairness of the individual elements. In this section, we present two more applications as examples: energy harvesting in wireless networks [20], [26], scheduling in wireless networks [21].

A. Energy Harvesting in Wireless Networks

Consider the scenario depicted in Fig. 6(a), where an energy source wirelessly charges 5 nodes. The energy source

divides its available bandwidth for energy transmission into 3 channels. At each time slot, the energy source can transmit at most a fixed amount of power on one of the channels. The amount of energy harvested by any node is stochastic and independent in each channel. The goal of the energy source is to *select a channel* to maximize the minimum average energy harvested by any node. However, the harvested energy for different nodes in one channel is not additive and cannot be exploited by other channels since difference channels exhibit different propagation characteristics even for the same energy harvesting node. Let s_{ij} be the indicator of allocating channel i to energy harvesting node j, $j \in \mathcal{J}$, where $\mathcal{J}$ is defined as the set of all energy harvesting nodes. To apply the proposed framework to this problem, we can map $S_i = \{s_{ij} | \forall j \in \mathcal{J}\}$ to one CTS in Alg. 1, and the energy source makes a decision in each round.



(a) (b)

Fig. 6: (a) Example of energy harvesting [20]; (b) Example of wireless scheduling [21]

B. Scheduling in Wireless Networks

Consider the following example of scheduling in wireless networks shown in Fig. 6(b). There are four flows, and each flow can be either a single hop link or a set of multi-hop links. Flows that share common nodes with other flows are considered as contending flows, which means they cannot transmit packets simultaneously. For instance, flows 3 and 4 are contending flows as they share a common node N_2. The goal of the scheduler is to maximize the minimum average throughput of each node. To apply the proposed framework to this problem, we can map a set of non-contending flows to one CTS in Alg. 1, and the scheduler makes a decision in each round.

VI. Conclusions and Future Work

In this paper, we study the design of online learning (MAB) algorithms with max-min fairness. As a motivating application, we aim at maximizing the minimum link throughput for concurrent transmissions of heterogeneous links. To solve this problem, we propose a novel probabilistic MAB framework, where we learn a probabilistic CTS selection strategy. We develop a Fair Probabilistic Explore-Then-Commit (FP-ETC) algorithm, which achieves the average regret of $O\left(T^{\frac{2}{3}}(K\log T)^{\frac{1}{3}}\right)$. Simulation results show that FP-ETC can effectively achieve max-min fairness, compared with existing MAB algorithms.

As future work, we plan to develop a more efficient probabilistic MAB algorithm that can achieve a lower regret. Additionally, we will consider other fairness metrics such as proportional fairness or treat fairness as a constraint.

VII. Appendix

A. Proof of Lemma 2

Proof. First, we define event $E = \cap_{a\in[K],l\in\mathcal{N}} E_{a,l}$, where $E_{a,l}$ is defined in Lemma 1. Using the property of union bound, $Pr(E) \geq 1 - \sum_{\forall a,l} Pr(\bar{E}_{a,l}) \geq 1 - \sum_{a\in[K]} \sum_{l\in C_a} \frac{2}{T^4} = 1 - \sum_{a\in[K]} |C_a| \frac{2}{T^4} \geq 1 - \sum_{a\in[K]} N \frac{2}{T^4} = 1 - \frac{2NK}{T^4}$. For the remainder of the proof, we condition on event E happening.

For any $\mathbf{p}$, given $f(\mathbf{p}) = \min_{l\in\mathcal{N}}\{\sum_{a\in[K]} \big(p_a \times g(a,l)\big)\}$, the estimation of $f(\mathbf{p})$ at the end of exploration phase (i.e., $t = mK$) is $\hat{f}(\mathbf{p}) = \min_{l\in\mathcal{N}}\{\sum_{a\in[K]} \big(p_a \times \hat{g}(a,l,mK)\big)\}$.

For the convenience of the proof, for every $l \in \mathcal{N}$, we define $h_l(\mathbf{p}) = \sum_{a\in[K]} \big(p_a \times g(a,l)\big)$ and $\hat{h}_l(\mathbf{p}) = \sum_{a\in[K]} \big(p_a \times \hat{g}(a,l,mK)\big)$, then we can easily obtain the relationship of $f(\mathbf{p})$ and $h_l(\mathbf{p})$, $\hat{f}(\mathbf{p})$ and $\hat{h}_l(\mathbf{p})$, respectively, which are $f(\mathbf{p}) = \min_{l\in\mathcal{N}} h_l(\mathbf{p})$ and $\hat{f}(\mathbf{p}) = \min_{l\in\mathcal{N}} \hat{h}_l(\mathbf{p})$. We first bound $\left|h_l(\mathbf{p}) - \hat{h}_l(\mathbf{p})\right|$ for every $l \in \mathcal{N}$:

$$\begin{aligned}
\left|h_l(\mathbf{p}) - \hat{h}_l(\mathbf{p})\right| &= \left|\sum_{a\in[K]} \big(p_a \times g(a,l)\big) - \sum_{a\in[K]} \big(p_a \times \hat{g}(a,l,mK)\big)\right| \\
&= \left|\sum_{a\in[K]} p_a\big(g(a,l) - \hat{g}(a,l,mK)\big)\right| \\
&\leq \left|\sum_{a\in[K]} p_a \sqrt{\frac{2\log(T)}{m}}\right| \text{ (if event } E \text{ happens)} \\
&= \sqrt{\frac{2\log(T)}{m}} \text{ (since } \sum_{a\in[K]} p_a = 1),
\end{aligned} \tag{6}$$

Eq. (6) shows that for every $l \in \mathcal{N}$, $\left|h_l(\mathbf{p}) - \hat{h}_l(\mathbf{p})\right|$ is upper-bounded by $\sqrt{\frac{2\log(T)}{m}}$ if event E happens.

Next, we utilize Eq. (6) to further bound $\left|\hat{f}(\mathbf{p}) - f(\mathbf{p})\right|$ for any $\mathbf{p}$. To do so, we need another Lemma, which is presented as follows.

Lemma 3. *Denote a sequence* $A = \big(a(l)\big)_{l\in\mathcal{N}}$ *and a sequence* $B = \big(b(l)\big)_{l\in\mathcal{N}}$, *if* $\forall l, a(l) \leq b(l)$, *then* $\min_{l\in\mathcal{N}} a(l) \leq \min_{l\in\mathcal{N}} b(l)$.

The proof of Lemma 3 is straightforward. Let $\arg\min_{l\in\mathcal{N}} b(l) = l^*$. We have $\min_{l\in\mathcal{N}} b(l) = b(l^*) \geq a(l^*) \geq \min_{l\in\mathcal{N}} a(l)$. This completes the proof of Lemma 3.

Given Lemma 3, we can take $a(l) = \hat{h}_l(\mathbf{p}) - \sqrt{\frac{2\log(T)}{m}}, l \in \mathcal{N}$ and $b(l) = h_l(\mathbf{p}), l \in \mathcal{N}$. According to Eq. (6), for every $l \in \mathcal{N}$, $a(l) \leq b(l)$. Therefore,

$$\min_{l\in\mathcal{N}}\{\hat{h}_l(\mathbf{p}) - \sqrt{\frac{2\log(T)}{m}}\} \leq \min_{l\in\mathcal{N}} h_l(\mathbf{p}). \tag{7}$$

Similarly, taking $a(l) = h_l(\mathbf{p}), l \in \mathcal{N}$ and $b(l) = \hat{h}_l(\mathbf{p}) + \sqrt{\frac{2\log(T)}{m}}, l \in \mathcal{N}$. According to Eq. (6), for every $l \in \mathcal{N}$, $a(l) \leq b(l)$. Therefore,

$$\min_{l\in\mathcal{N}} h_l(\mathbf{p}) \leq \min_{l\in\mathcal{N}}\{\hat{h}_l(\mathbf{p}) + \sqrt{\frac{2\log(T)}{m}}\}. \tag{8}$$

Combine Eq. (7) and Eq. (8), we can obtain

$$\min_{l\in\mathcal{N}} h_l(\mathbf{p}) - \sqrt{\frac{2\log(T)}{m}} \leq \min_{l\in\mathcal{N}} \hat{h}_l(\mathbf{p}) \leq \min_{l\in\mathcal{N}} h_l(\mathbf{p}) + \sqrt{\frac{2\log(T)}{m}}, \tag{9}$$

which is equivalently

$$\left|\hat{f}(\mathbf{p}) - f(\mathbf{p})\right| \leq \sqrt{\frac{2\log(T)}{m}}. \tag{10}$$

□

B. Proof of Theorem 1

Proof. Firstly, we define a clean event $\xi := \{\forall \mathbf{p}, \left|\hat{f}(\mathbf{p}) - f(\mathbf{p})\right| \leq \sqrt{\frac{2\log(T)}{m}}\}$, where $\hat{f}(\mathbf{p})$ and $f(\mathbf{p})$ are defined in Lemma 2. According to the Lemma 2, $Pr(\xi) \geq 1 - \frac{2NK}{T^4}$. We also define a bad event $\bar{\xi}$, which is the complement of ξ. Denote the optimal $\mathbf{p}^* = \arg\max_{\mathbf{p}} \quad f(\mathbf{p})$.

We analyze the regret under event ξ and $\bar{\xi}$, respectively.

We first analyze event ξ. If FP-ETC chooses $\mathbf{p}$, where $\mathbf{p} \neq \mathbf{p}^*$. Under event ξ, we have $f(\mathbf{p}) + \sqrt{\frac{2\log(T)}{m}} > \hat{f}(\mathbf{p}) > \hat{f}(\mathbf{p}^*) \geq f(\mathbf{p}^*) - \sqrt{\frac{2\log(T)}{m}}$. Re-arranging the terms, it follows that

$$f(\mathbf{p}^*) - f(\mathbf{p}) \leq 2\sqrt{\frac{2\log(T)}{m}}. \tag{11}$$

Therefore, under event ξ, each round in the exploitation phase of FP-ETC contributes at most $2\sqrt{\frac{2\log(T)}{m}}$ regret. In each round of the exploration phase, FP-ETC trivially contributes at most regret of 1 for each link. Thus, under event ξ,

$$\begin{aligned} R_T^f = \sum_{t=1}^{T}[f(\mathbf{p}^*) - f(\mathbf{p_t})] &\leq mK + (T - mK)2\sqrt{\frac{2\log(T)}{m}} \\ &\leq mK + 2T\sqrt{\frac{2\log(T)}{m}}, \end{aligned} \tag{12}$$

where $\mathbf{p_t}$ is the $\mathbf{p}$ vector at round t. When $t \leq mK$, $\mathbf{p_t}$ is a standard basis vector, with element 1 indicating that the corresponding CTS is selected in the exploration phase. The total regret in the exploration phase of FP-ETC is mK as there are K CTSes and each CTS is played m times.

Until now, we have upper bounded R_T^f under event ξ happens. However, there is still a gap between R_T^f and R_T of Eq. (4). Observe that the first term in Eq. (4) is $\min_{l\in\mathcal{N}} \sum_{t=1}^{T} r(b_t, l, t)$. Applying the Hoeffding inequality, we can easily know that $\forall l \in \mathcal{N}$, at least with probability $1 - \delta$,

$$\left|\sum_{t=1}^{T} r(b_t, l, t) - T\sum_{a\in[K]} p_a^* g(a, l)\right| \leq \sqrt{\frac{T}{2}\log(\frac{2}{\delta})}. \tag{13}$$

Without loss of generality, we set $\delta = T^{-2}$. Therefore,

$$\left|\min_{l\in\mathcal{N}}\sum_{t=1}^{T} r(b_t, l, t) - T\min_{l\in\mathcal{N}}\sum_{a\in[K]} p_a^* g(a, l)\right| \leq \sqrt{T\log(2T)}. \tag{14}$$

Note that $\mathbf{p_t}$ does not change when $t > mK$. Applying the Hoeffding inequality, we know that $\forall l \in \mathcal{N}$, at least with probability $1 - \frac{1}{T^2}$,

$$\begin{aligned} &\left|\sum_{t=mK+1}^{T} r(a_t, l, t) - \sum_{t=mK+1}^{T}\sum_{a\in[K]} p_{t,a} g(a, l)\right| \\ &\leq \sqrt{(T - mK)\log(2T)}. \end{aligned} \tag{15}$$

When $t \leq mK$, the reward for each link is upper bounded by 1. Combine $t \leq mK$ and $t > mK$ together, we obtain that $\forall l \in \mathcal{N}$, at least with probability $1 - \frac{1}{T^2}$,

$$\begin{aligned} &\left|\sum_{t=1}^{T} r(a_t, l, t) - \sum_{t=1}^{T}\sum_{a\in[K]} p_{t,a} g(a, l)\right| \\ &\leq mK + \left|\sum_{t=mK+1}^{T} r(a_t, l, t) - \sum_{t=mK+1}^{T}\sum_{a\in[K]} p_{t,a} g(a, l)\right| \\ &\leq mK + \sqrt{(T - mK)\log(2T)}, \end{aligned} \tag{16}$$

Therefore,

$$\begin{aligned} &\left|\min_{l\in\mathcal{N}}\sum_{t=1}^{T} r(a_t, l, t) - \min_{l\in\mathcal{N}}\sum_{t=1}^{T}\sum_{a\in[K]} p_{t,a} g(a, l)\right| \\ &\leq mK + \sqrt{(T - mK)\log(2T)}. \end{aligned} \tag{17}$$

Define event $\eta := \{\forall l \in \mathcal{N}, \text{Eq. (13) and Eq. (16) hold}\}$. Event $\bar{\eta}$ is the complement of η. Using union bound, we can know that event η happens at least with probability $1 - \frac{2N}{T^2}$.

Combining Eq.(14) and Eq (17) together, when event $\xi \cap \eta$ happens, R_T of Eq. (4) can be upper bounded, which is

$$\begin{aligned} R_T &= \min_{l\in\mathcal{N}}\sum_{t=1}^{T} r(b_t, l, t) - \min_{l\in\mathcal{N}}\sum_{t=1}^{T} r(a_t, l, t) \\ &\leq T\min_{l\in\mathcal{N}}\sum_{a\in[K]} p_a^* g(a, l) - \min_{l\in\mathcal{N}}\sum_{t=1}^{T}\sum_{a\in[K]} p_{t,a} g(a, l) \\ &+ \sqrt{T\log(2T)} + mK + \sqrt{(T - mK)\log(2T)} \\ &\leq Tf(\mathbf{p}^*) - f(\sum_{t=1}^{T}\mathbf{p_t}) + mK + 2\sqrt{T\log(2T)} \\ &\overset{(c)}{\leq} Tf(\mathbf{p}^*) - \sum_{t=1}^{T} f(\mathbf{p_t}) + mK + 2\sqrt{T\log(2T)} \\ &= R_T^f + mK + 2\sqrt{T\log(2T)} \\ &\leq 2mK + 2T\sqrt{\frac{2\log(T)}{m}} + 2\sqrt{T\log(2T)}, \end{aligned} \tag{18}$$

where (c) is because of Lemma 4 of Section VII-C.

Recall that m was given in advance in FP-ETC algorithm. Therefore, we can choose m to minimize the right-hand side of Eq. (18). Since the first two terms (i.e., $2mK$ and $2T\sqrt{\frac{2\log(T)}{m}}$) are monotonically increasing and monotonically

decreasing with respect to m. We can set m so that the two terms are approximately equal. By solving it, we obtain $m = O\big((\frac{T}{K})^{\frac{2}{3}}(\log T)^{\frac{1}{3}}\big)$. Plug it into Eq. (18), we have $R_T \leq O\big(T^{\frac{2}{3}}(K\log T)^{\frac{1}{3}}\big)$, where $O\big(T^{\frac{1}{2}}(\log T)^{\frac{1}{2}}\big)$ is neglected as it has a lower order than $O\big(T^{\frac{2}{3}}(K\log T)^{\frac{1}{3}}\big)$.

Using union bound, $Pr(\xi \cap \eta) \geq 1 - \frac{2N}{T^2} - \frac{2NK}{T^4}$, averaging all the events, then $\mathbb{E}[R_T]$ of Eq. (5) is

$$\begin{aligned}\mathbb{E}\left[R_T\right] \leq & \mathbb{E}\left[R_T I(\xi \cap \eta)\right] + \mathbb{E}\left[R_T I(\overline{\xi \cap \eta})\right] \\ & \leq O\big(T^{\frac{2}{3}}(K\log T)^{\frac{1}{3}}\big) + \mathbb{E}\left[T \cdot I(\overline{\xi \cap \eta})\right] \\ & \leq O\big(T^{\frac{2}{3}}(K\log T)^{\frac{1}{3}}\big) + T \cdot (\frac{2N}{T^2} + \frac{2NK}{T^4}) \\ & \leq O\big(T^{\frac{2}{3}}(K\log T)^{\frac{1}{3}}\big),\end{aligned} \tag{19}$$

where the last term $T \cdot (\frac{2N}{T^2} + \frac{2NK}{T^4})$ is neglected since it is the order of T^{-1}.

□

C. Lemma 4 and Its Proof

Lemma 4. *Given* $f(\mathbf{p}) = \min_{l\in\mathcal{N}}\{\sum_{a\in[K]}\big(p_a \times g(a,l)\big)\}$, $f(\sum_{t=1}^{T}\mathbf{p_t}) \geq \sum_{t=1}^{T} f(\mathbf{p_t})$.

Proof. Denote $l^* = \arg\min_{l\in\mathcal{N}}\sum_{a\in[K]}\big(\sum_{t=1}^{T} p_{t,a} \times g(a,l)\big)$, where $p_{t,a}$ is the a-th element of $\mathbf{p_t}$, we have

$$\begin{aligned} f(\sum_{t=1}^{T}\mathbf{p_t}) &= \min_{l\in\mathcal{N}}\{\sum_{a\in[K]}\big(\sum_{t=1}^{T} p_{t,a} \times g(a,l)\big)\} \\ &= \sum_{a\in[K]}\big(\sum_{t=1}^{T} p_{t,a} \times g(a,l^*)\big) \\ &= \sum_{t=1}^{T}\sum_{a\in[K]}\big(p_{t,a} \times g(a,l^*)\big) \\ &\geq \sum_{t=1}^{T}\min_{l\in\mathcal{N}}\{\sum_{a\in[K]}\big(p_{t,a} \times g(a,l)\big)\} \\ &= \sum_{t=1}^{T} f(\mathbf{p_t}). \end{aligned} \tag{20}$$

□

REFERENCES

[1] 3GPP, "Feasibility study on licensed-assisted access to unlicensed spectrum," Standard (TR) 36.889, V13.0.0, 2015.

[2] 3GPP, "Study on NR-based access to unlicensed spectrum," Standard (TR) 36.889, V16.0.0, 2018.

[3] G. Bianchi, "Performance analysis of the IEEE 802.11 distributed coordination function," *IEEE Journal on selected areas in communications*, vol. 18, no. 3, pp. 535–547, 2000.

[4] Y. Gao, X. Chu, and J. Zhang, "Performance analysis of LAA and WiFi coexistence in unlicensed spectrum based on Markov chain," in *IEEE Global Communications Conference (GLOBECOM)*. IEEE, 2016, pp. 1–6.

[5] IEEE, "Part 11: Wireless LAN medium access control (MAC) and physical layer (PHY) specifications," IEEE Standard 802.11, 2012.

[6] J. Lee, W. Kim, S.-J. Lee, D. Jo, J. Ryu, T. Kwon, and Y. Choi, "An experimental study on the capture effect in 802.11a networks," in *Proceedings of the second ACM international workshop on Wireless network testbeds, experimental evaluation and characterization*, 2007, pp. 19–26.

[7] S. Yun and L. Qiu, "Supporting WiFi and LTE coexistence," in *IEEE Conference on Computer Communications (INFOCOM)*, 2015, pp. 810–818.

[8] Y. Yan, P. Yang, X.-Y. Li, Y. Zhang, J. Lu, L. You, J. Wang, J. Han, and Y. Xiong, "WizBee: Wise ZigBee coexistence via interference cancellation with single antenna," *IEEE Transactions on Mobile Computing*, vol. 14, no. 12, pp. 2590–2603, 2015.

[9] Z. Guo, M. Li, and M. Krunz, "Exploiting successive interference cancellation for spectrum sharing over unlicensed bands," *IEEE Transactions on Mobile Computing*, pp. 1–18, 2023.

[10] Z. Guo, M. Li, and Y. Xiao, "Enhancing LAA/Wi-Fi coexistence via concurrent transmissions and interference cancellation," in *IEEE International Symposium on Dynamic Spectrum Access Networks (DySPAN)*. IEEE, 2019, pp. 1–10.

[11] Y. Gao and S. Roy, "Achieving proportional fairness for LTE-LAA and Wi-Fi coexistence in unlicensed spectrum," *IEEE Transactions on Wireless Communications*, vol. 19, no. 5, pp. 3390–3404, 2020.

[12] A. Slivkins *et al.*, "Introduction to multi-armed bandits," *Foundations and Trends® in Machine Learning*, vol. 12, no. 1-2, pp. 1–286, 2019.

[13] 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, 2011.

[14] M.-J. Youssef, V. V. Veeravalli, J. Farah, C. A. Nour, and C. Douillard, "Resource allocation in NOMA-based self-organizing networks using stochastic multi-armed bandits," *IEEE Transactions on Communications*, vol. 69, no. 9, pp. 6003–6017, 2021.

[15] W. Chen, Y. Wang, and Y. Yuan, "Combinatorial multi-armed bandit: General framework and applications," in *International conference on machine learning*. PMLR, 2013, pp. 151–159.

[16] Y. Gai, B. Krishnamachari, and R. Jain, "Combinatorial network optimization with unknown variables: Multi-armed bandits with linear rewards and individual observations," *IEEE/ACM Transactions on Networking*, vol. 20, no. 5, pp. 1466–1478, 2012.

[17] M. Joseph, M. Kearns, J. H. Morgenstern, and A. Roth, "Fairness in learning: Classic and contextual bandits," *Advances in neural information processing systems*, vol. 29, 2016.

[18] V. Patil, G. Ghalme, V. Nair, and Y. Narahari, "Achieving fairness in the stochastic multi-armed bandit problem," *The Journal of Machine Learning Research*, vol. 22, no. 1, pp. 7885–7915, 2021.

[19] F. Li, J. Liu, and B. Ji, "Combinatorial sleeping bandits with fairness constraints," *IEEE Transactions on Network Science and Engineering*, vol. 7, no. 3, pp. 1799–1813, 2019.

[20] D. Ghosh, A. Verma, and M. K. Hanawal, "Learning and fairness in energy harvesting: A maximin multi-armed bandits approach," in *International Conference on Signal Processing and Communications (SPCOM)*. IEEE, 2020, pp. 1–5.

[21] L. Tassiulas and S. Sarkar, "Maxmin fair scheduling in wireless networks," in *Proceedings of Twenty-First Annual Joint Conference of the IEEE Computer and Communications Societies*, vol. 2. IEEE, 2002, pp. 763–772.

[22] M. Mehrnoush, V. Sathya, S. Roy, and M. Ghosh, "Analytical modeling of Wi-Fi and LTE-LAA coexistence: Throughput and impact of energy detection threshold," *IEEE/ACM Transactions on Networking*, vol. 26, no. 4, pp. 1990–2003, 2018.

[23] S. Bubeck, N. Cesa-Bianchi *et al.*, "Regret analysis of stochastic and nonstochastic multi-armed bandit problems," *Foundations and Trends® in Machine Learning*, vol. 5, no. 1, pp. 1–122, 2012.

[24] Y. Li, W. Ye, and J. Heidemann, "Energy and latency control in low duty cycle MAC protocols," in *IEEE Wireless Communications and Networking Conference*, vol. 2. IEEE, 2005, pp. 676–682.

[25] A. B. Sediq, R. H. Gohary, R. Schoenen, and H. Yanikomeroglu, "Optimal tradeoff between sum-rate efficiency and Jain's fairness index in resource allocation," *IEEE Transactions on Wireless Communications*, vol. 12, no. 7, pp. 3496–3509, 2013.

[26] D. W. K. Ng and R. Schober, "Max-min fair wireless energy transfer for secure multiuser communication systems," in *IEEE Information Theory Workshop (ITW 2014)*. IEEE, 2014, pp. 326–330.