Energy-Aware Hybrid RF-VLC Multiband Selection in D2D Communication: A Stochastic Multiarmed Bandit Approach

Sherief Hashima, *Senior Member, IEEE*, Mostafa M. Fouda, *Senior Member, IEEE*, Sadman Sakib, Zubair Md. Fadlullah, *Senior Member, IEEE*, Kohei Hatano, Ehab Mahmoud Mohamed, *Member, IEEE*, and Xuemin Shen, *Fellow, IEEE*

***Abstract*—To handle the exponentially growing service expectations from mobile users and circumvent the band switching slow rate, device-to-device (D2D) communication is receiving much research attention in the Internet of Things (IoT). While the emerging D2D nodes can support heterogeneous frequency bands [radio frequency (RF) including 2.4 GHz/5 GHz wireless local area network (WLAN), 38-GHz millimeter wave (mmWave), and visible light communication (VLC)], the physical constraints (e.g., blocking) require the user devices to dynamically switch between the bands in order to avoid the loss of connectivity and throughput degradation. In this article, we investigate an effective online link selection in hybrid RF-VLC scenarios for direct user data handling. First, we model the multiband selection issue as a multiarmed bandit (MAB) problem. The source/relay node acts as a player who gambles to maximize its long-term feedback/reward via selecting suitable arms, i.e., available bands (WLAN, mmWave, or VLC). Then, we propose an online, energy-aware band selection (EABS) methodology by leveraging three theoretically guaranteed MAB techniques [upper confidence bound (UCB), Thompson sampling (TS), and minimax optimal stochastic strategy (MOSS)] to derive optimal band selection policies. Based on these adopted policies, we propose three algorithms, namely, EABS-UCB, EABS-TS, and EABS-MOSS, to implement the EABS strategy, respectively. Extensive simulations demonstrate our proposed algorithms' superior performance compared to the traditional link selection schemes regarding energy efficiency, average throughput, and convergence rate. In particular, EABS-MOSS emerges as the best algorithm as it exhibits near-optimal performance due to its flexibility to both stochastic and adversarial environments.**

***Index Terms*—Band selection, channel selection, device-to-device communication (D2D), millimeter wave (mmWave), minimax optimal stochastic strategy (MOSS), multiarmed bandit (MAB), radio frequency (RF), Thompson sampling (TS), upper confidence interval (UCB), wireless local area network (WLAN).**

Manuscript received 8 October 2021; accepted 19 March 2022. Date of publication 24 March 2022; date of current version 7 September 2022. This work was supported by the JSPS KAKENHI under Grant JP19H04174 and Grant JP21K14162. *(Corresponding author: Sherief Hashima.)*

Sherief Hashima is with the Computational Learning Theory Team, RIKEN-Advanced Intelligent Project, Fukuoka 819-0395, Japan, and also with the Department of Engineering and Scientific Equipment, Egyptian Atomic Energy Authority, Cairo 13759, Egypt (e-mail: sherief.hashima@riken.jp).

Mostafa M. Fouda is with the Department of Electrical and Computer Engineering, College of Science and Engineering, Idaho State University, Pocatello, ID 83209 USA, and also with the Department of Electrical Engineering, Faculty of Engineering at Shoubra, Benha University, Cairo 11629, Egypt (e-mail: mfouda@ieee.org).

Sadman Sakib is with the Department of Computer Science, Lakehead University, Thunder Bay, ON P7B 5E1, Canada (e-mail: ssak2921@lakeheadu.ca).

Zubair Md. Fadlullah is with the Department of Computer Science, Lakehead University, Thunder Bay, ON P7B 5E1, Canada, and also with Thunder Bay Regional Health Research Institute, Thunder Bay, ON P7B 7A5, Canada (e-mail: zubair.fadlullah@lakeheadu.ca).

Kohei Hatano is with the Faculty of Arts and Science, Kyushu University, Fukuoka 819-0395, Japan, and also with the Computational Learning Theory Team, RIKEN-AIP, Fukuoka 819-0395, Japan (e-mail: hatano@inf.kyushu-u.ac.jp).

Ehab Mahmoud Mohamed is with the Electrical Engineering Department, College of Engineering, Prince Sattam Bin Abdulaziz University, Wadi Addwasir 11991, Saudi Arabia, and also with the Electrical Engineering Department, Aswan University, Aswan 81542, Egypt (e-mail: ehab_mahmoud@aswu.edu.eg).

Xuemin Shen is with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON N2L 3G1, Canada (e-mail: sshen@uwaterloo.ca).

Digital Object Identifier 10.1109/JIOT.2022.3162135

I. Introduction

THE NEXT-GENERATION (5G+/6G) mobile communication networks are anticipated to exploit a plethora of frequency bands to satisfy the exponentially growing bandwidth-demand of mobile users while meeting the Internet of Things (IoT) and machine-to-machine (M2M) service requirements. While the Giga/Tera Hertz (GHz/THz) spectra, such as 38-GHz millimeter wave (mmWave) and 400–800 THz visible light communication (VLC), are being considered to dramatically increase the 5G+/6G capacity, these high frequencies are adversely impacted by fast channel fading, signal attenuation, shadowing and blocking effect of walls and other obstacles, limited range, and so forth. Therefore, the legacy radio frequency (RF) band (i.e., 2.4/5.25-GHz WLAN) and the sub-6 GHz mmWave are also considered to ensure connectivity when the channels of the higher frequency bands experience poor conditions. Incorporating Hybrid RF-VLC systems into 5G+/6G networks and IoT systems, particularly with device-to-device (D2D) communications [1], as shown in Fig. 1, can enhance the system performance. For instance, by switching between the multibands RF (Wi-Fi and mmWave) and VLC, the D2D pairs can avoid service outage which may otherwise occur due to substantial interference and full blockage across an entire band. However, the optimized link selection using the best band and its corresponding channel in D2D communications, emerges as a difficult-to-formulate research problem.

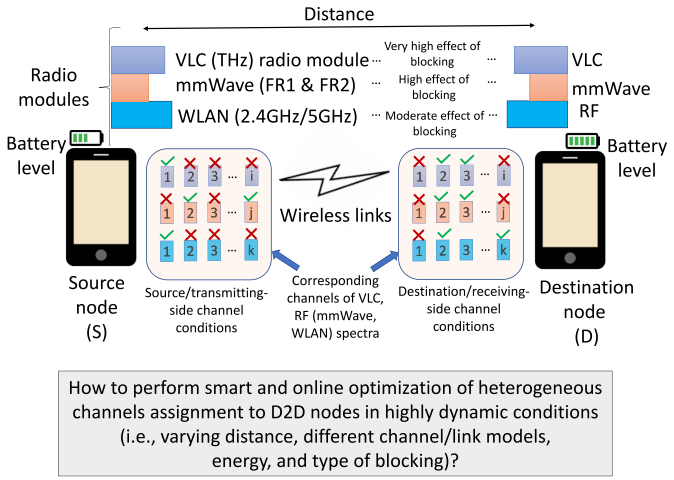

Fig. 1. Considered system model and problem from a high level demonstrating our main objective to carry out online selection of the most suitable band and corresponding channel in a dynamically varying D2D environment. For simple depiction, only a pair of source (transmitting) and destination (receiving) devices is generalized that can be easily extended to multiple D2D relay nodes.

In this article, we introduce a formal formulation of this optimization problem of multiband selection targeting the selection of RF or VLC for every D2D pair. This formulated problem requires full channel knowledge and is identified to be computationally hard. Therefore, we stress on the need to reformulate this original optimization problem in a relaxed form so that acceptable quality solutions (i.e., suboptimal band/channel selection decisions) could be derived at real time at the resource-constrained D2D nodes without the need for the full channel knowledge.

In this vein, we explore the state-of-the-art machine learning techniques, particularly the reinforcement learning (RL) family with online-learning and self-deciding capabilities [2], [3], to overcome the challenges involved in the multiband selection considering both blocking and transmitting device's battery levels within the selection operation. Our investigation leads to reformulate the original band selection optimization problem into a stochastic multiarmed bandit (MAB) game, a specific type of RL whereby an agent attempts to maximize its long-term rewards via online learning to solve the well-known "exploitation-exploration dilemma" [4]–[6]. Our stochastic MAB model depends on the reward independence of each arm (i.e., a frequency band and any of its corresponding channel). After deciding an action selection, a stochastic reward is revealed for only the chosen action.

To solve the formulated RF-VLC band selection MAB problem for individual D2D pairs, we present several online, energy-aware learning algorithms. Unlike contemporary machine/deep learning techniques, our proposed algorithms do not require prior training plus they are theoretically guaranteed, and thereby can improve the overall system complexity and overheads. The key contributions of this article are summarized as follows.

1) The problem of hybrid RF-VLC band selection is formulated as a stochastic MAB where the transmitting device attempts to maximize its long-term reward via appropriate band selection by considering the impact of blockage.
2) To the best of our knowledge, our research is the first work in the literature to present a unified RF-VLC channel model which considers blockage loss of all frequency bands comprehensively. Furthermore, we present a general formula for the blockage loss, which can be extended for any multiband frequencies used by the same transmitting (TX) and receiving (RX) nodes.
3) We leverage three theoretically guaranteed MAB-based schemes, i.e., upper confidence bound (UCB), Thompson sampling (TS), and minimax optimal stochastic strategy (MOSS) in solving the formulated problem to provide the best band selection at real-time between individual D2D pairs. Our applied MAB techniques to solve our formulated communication problem hinge upon the fundamental theory that ascertain bounded performance guarantees. Through our applied and comparative MAB-based, online algorithms, we demonstrate how to efficiently assign the channels in a distributed fashion, without intervention from a centralized oracle.
4) Based on the utilized band, we also incorporate energy-aware band selection (EABS) using the UCB, TS, and MOSS schemes. Accordingly, we present three online, energy-aware, fast-converging algorithms, namely, EABS-UCB, EABS-TS, and EABS-MOSS, respectively, to consider the source's consumed energies according to the chosen transmission band. Our proposed algorithms simulate the real scenarios where the energy cost of selecting any band is carefully reflected.
5) Our proposed algorithms are compared with one other and with traditional band selection techniques for performance analysis and evaluation. Extensive computer-based simulation results are reported to verify the superior band selection performance of our proposed algorithms in contrast with the traditional approaches.

The remainder of this article is organized as follows. Section II surveys the relevant research work. Section III describes the utilized system model, including the used RF-VLC channel models and the unified channel and blockage model considered in our scenario. Section IV formulates an optimization problem of hybrid RF-VLC band selection and then provides a reformulation of that original problem as an MAB game. In Section V, we present our proposed EABS-UCB, EABS-TS, and EABS-MOSS algorithms. Section VI provides the detailed simulation results and discussion by comparing our proposed algorithms with contemporary band selection methods in terms of average throughput and convergence rate. Finally, Section VII offers concluding remarks and open research issues. Table I summarizes the utilized abbreviations and mathematical symbols.

II. Related Work

This section summarizes the relevant research work from the perspective of RF-VLC systems and learning-based multiband selection strategies in the next generation communication networks.

A. Hybrid RF-VLC Systems

While legacy RF bands may offer nominal data rates over wider coverage areas, 38-GHz mmWave and 400–800 THz

TABLE I
LIST OF ACRONYMS AND NOTATIONS USED IN THIS WORK

mmWave	Millimeter-wave	RF	Radio Frequency	WLAN	Wireless Local Area Network
D2D	Device-to-device	5G	Fifth generation	5G+	Beyond fifth generation
EABS	Energy-aware band selection	6G	Sixth generation	MAB	Multi armed bandit
ML	Machine-learning	DNN	Deep Neural network	DL	Deep Learning
CNN	Convolutional Neural network	AP	Access point	Li-Fi	Light-Fidelity
VLC	Visible light communication	UCB	Upper Confidence bound	TS	Thomson Sampling
MOSS	Mini max optimal stochastic strategy	GHz	Giga hertz	THz	Terahertz
TX	Transmitter	RX	Receiver	QoS	Quality of service
RL	Reinforcement-Learning	IoT	Internet of things	SINR	Signal to interference plus noise ratio
BF	Beanforming	LoS	Line of sight	NLoS	Non line of sight
D-NOMA	Distributed non orthogonal multiple access	EE	Energy Efficiency	SE	Spectral efficiency
S	Source	D	Destination	L, NL	LoS, NLoS
P_D^w,P_S^w	Received and transmitted WLAN powers	P_L^w	Reference WLAN path loss	γ_w	WLAN based path loss exponent
ψ_w	WLAN based log normal shadowing	σ_w	WLAN based standard deviation	P_D^m	Received mmWave power at D
P_S^m	mmWave source power	x	S and D Separation distance	P_D^m	mmWave recieved power
Λ_S, Λ_D	S and D mmWave BF gains	σ_m^l	mmWave Standard deviation	$PL^m(x)$	mmWave path loss
η_m^l	mmWave path loss	γ_m^l	mmWave path loss exponent	χ_m^l	log-normal shadowing
H_{VLC}	VLC channel gain	A_R	Optical detector's physical area	θ_v	VLC incidence angle
ϕ_v	irradiance angle	θ, θ_{-3dB}	Azimuth angle, -3dB beamwidth	Λ_0	Maximum antenna gain
g	Lambertian model's order	$T_s(\phi_v)$	Optical filter's gain	$g(\phi_v)$	Optical concentrator's gain
P_D^v	Received optical power	P_S^v	Transmitted optical power	T_D	Data transmission time
N	Number of available channels	W_n	Channel bandwidth	$T_{h,n}$	Overhead time
$\Xi_{S,n}(t)$	S's standing energy	Ξ_{th}	Threshold energy	$\Gamma_n(t)$	SE bps/Hz

VLC bands provide much higher data rates at the expense of limited communication range [7], [8]. Therefore, hybrid RF-VLC systems have been recently conceptualized to exploit the large capacity of VLC links and improved RF coverage in next-generation networks [7], [8]. Readers may find a detailed survey on hybrid RF-VLC systems in [9]. A stochastic geometry model for coverage and rate analysis regarding a typical user in a large indoor coverage area in hybrid RF-VLC networks was designed in [10]. On the other hand, the work in [11] envisioned an RF-VLC link selection scheme for data delivery in an indoor scenario with Quality of Service (QoS) guarantees by considering an intermittently active data source. A practical hybrid WLAN-VLC system, which aggregates wireless local area network (WLAN) and VLC downlinks, and shares the WLAN uplink instead of utilizing separate VLC uplink channels for indoor usage, was presented in [12]. A system integrating optical communication based on light fidelity (Li-Fi) and mmWave was proposed by a coauthor of our work in [13]. That work also suggested a mmWave beamforming training according to the Li-Fi localization. Furthermore, researchers of [14] considered a heterogeneous RF-VLC-based industrial network system to satisfy diverse QoS demands in IoT applications. The up/down-link resource management was modeled as a Markov decision process in that system. However, the aforementioned research work took into consideration neither the co-existence of mmWave with WLAN and VLC bands nor multiband D2D scenarios, which are essential 5G+/6G use-cases.

B. Learning-Based Multiband Selection Techniques

Various learning techniques exist for multiband or multichannel selection in diverse wireless network scenarios, which include cellular networks, IoT, D2D, drone cells, and so forth [15]–[18]). Among these, the relevant research work involving the hybrid bands is categorized into supervised learning and online learning as described in the remainder of this section.

1) Deep Learning-Based Wireless Resource Selection Techniques: A deep learning model, for WiFi-VLC band selection in a D2D scenario, was presented in [19]. Their solution was based on a deep neural network (DNN), which provides an initial band selection decision. Based on the DNN's output, a fast heuristic algorithm was proposed to improve the band selection decision further. However, this approach is limited due to the base station-based centralized implementation, suitable only for specific indoor scenarios. On the other hand, several coauthors of our work envisioned a deep learning-based predictive channel selection model to solve the dynamic channel conditions problem in multiband relay networks in [20] and [21]. They customized a convolutional neural network (CNN) model to predict the channels belonging to heterogeneous frequency bands to achieve the best signal-to-interference-plus-noise-ratio (SINR) value. Moreover, they envisioned deep learning-based controlled and smart channel assignment policies. However, they did not consider the VLC channel model in their work. Moreover, the aforementioned deep learning models require extensive training data sets, which may not be available until the roll-out of 5G+ systems with hybrid RF-VLC bands.

2) Online Learning-Based Wireless Resource Selection Techniques: The existing online learning techniques, dominated by reinforcement/deep RL [22]–[24], typically consider dynamic multichannel access problems whereby several correlated channels are assumed to follow an unknown joint Markov

model. An RL-based prediction and selection algorithm for idle channels is proposed to sense industrial big spectrum data in [25]. Also, In [26], an adaptable decentralized joint power and channel allocation scheme is proposed for better power performance in D2D unlicensed pairs. Furthermore, MAB has gained much interest recently to formulate the exploration–exploitation tradeoff in solving severe existing wireless radio resource allocation problems and enhancing the overall system performance in an online manner. For instance, different MAB problems were considered for D2D neighbor discovery and selection [27], [28], gateway unmanned aerial vehicle (UAV) selection [29], concurrent beamforming (BF) [30], and beam alignment [31], [32]. A brief tutorial on bandit problems and its recent applications in the wireless communication domain was presented in [4]. Vora and Kang [33] improved the throughput of 5G systems by directly applying the TS algorithm. In [34], MAB selection approaches were employed to determine the access point (AP) configuration, which maximizes the long-term throughput. A distributed nonorthogonal multiple access-based MAB (D-NOMA-MAB) method was proposed in [35] to address channel and power selection problems jointly. However, the aforementioned MAB-based work did not consider a hybrid RF-VLC band selection problem.

Only a few researchers recently started to consider MAB-based models for real-time resource scheduling in specific wireless network scenarios. For instance, in [36], the combined optimization of bandwidth, power, and user association problem in a hybrid RF-VLC system was considered by employing a deep Q-learning algorithm via targeting optimal policy learning. On the other hand, a context-aware transfer learning network selection algorithm, sensitive to the features of traffic, networks, and network load distribution for indoor RF-VLC heterogeneous scenarios, was designed in [37] and [38]. However, these research works did not consider a hybrid RF-VLC band selection problem in D2D communications where the online band selection issue leveraging MAB is likely to be impacted by the frequency-dependent blockage. There is a lack of a unified system model in the existing literature that is needed to generalize the channels and blockage effect across the heterogeneous wireless resources, namely, 2.4/5.25-GHz WLAN, 38-GHz mmWave, and 400–800 THz VLC bands, before an online multiband channel selection may be effectively considered.

III. System Model

D2D communications is a key enabling technology for realizable IoT applications/services in future 5G+/6G systems [39]. Hence, our considered system model, consisting of the RF and VLC link models, and the unified blockage model, is described in details in this section. For simplicity, Fig. 1 depicts only a single D2D pair, i.e., a source (S) node and its corresponding destination (D) node. Note that this can be generalized to k pairs distributed within a certain area in the considered D2D scenario. Each node in the D2D network is assumed to be equipped with multiple bands, each of which consists of multiple channels, to send, receive, and forward/relay data packets. Assume that the available channels belong to any of the RF (2.4/5.25-GHz WLAN, 38-GHz (WiGig) mmWave) or 400–800 THz VLC bands. Both V2X and V2V is a considerable application scenarios to our system model, which called Terahertz (THz)-assisted 6G-V2X [40].

A. RF Links Models

The RF communications considered in our system model consist of WLAN bands operating at 2.4 and 5.25 GHz, and the mmWave band operating at 38 GHz. These heterogeneous channels/links are modeled below.

1) Unified Link Model for WLAN: For a unified link model for both 2.4and 5-GHz WLAN systems, we first employ the model used in [41], and then provide the empirical parameters to specify either links. Let us consider the source and receiver nodes to be separated by a distance x. Then, the received WLAN power, P_D^w (expressed in dBm) in the WLAN channel model is formulated as in [41]

$$P_D^w[\text{dBm}] = P_S^w[\text{dBm}] - P_{L_0}^w - 10\gamma_w \log_{10}(x) - \psi_w \quad (1)$$

where P_S^w, P_D^w, and $P_{L_0}^w$ denote the transmitted power of the source device, received power at the destination node, and reference path-loss, respectively. γ_w represents the path loss exponent while $\psi_w \backsim \mathcal{N}(0, \sigma_w)$ denotes the log-normal shadowing with zero mean and σ_w standard deviation. The utilized parameters for 5.25-GHz WLAN system are $P_{L_0}^w = 47.2$, $\gamma_w = 2.32$, $\sigma_w = 6$ dB [42]. On the other hand, the 2.4-GHz WLAN system parameters are $\gamma_w = 2$, $P_{L_0}^w = 41.8$, and $\sigma_w = 2.15$ dB [43].

2) mmWave Links Model: To construct the mmWave links model, the mmWave received power (P_D^m) by considering BF gain and blockage effect [41] is given by

$$P_D^m = \mu(\mathbb{P}_{\text{LoS}}(x)) P_S^m \Lambda_S \Lambda_D / PL^m(x) \quad (2)$$

where $\mu(\mathbb{P}_{\text{LoS}}(x))$ is a Bernoulli random variable (RV) reflecting the blockage effect with the parameter $\mathbb{P}_{\text{LoS}}(x)$, which indicates the distance-dependent Line-of-Sight (LoS) probability. P_S^m denotes the mmWave source power while Λ_S and Λ_D represent the source (S) and destination (D) BF gains. $PL^m(x)$, expressed in dB, indicates the distance-dependent path loss as follows:

$$10\log_{10}\big(PL^m(x)\big) = PL_0^m + 10\gamma_m^l \log_{10}(x) + \chi_m^l \quad (3)$$

where PL_0^m, γ_m^l, and $\chi_g^l \backsim \mathcal{N}(0, \sigma_m^l)$ represent the mmWave path loss at a reference distance x_0, path loss exponent, and log-normal shadowing with zero mean and standard deviation of σ_m^l, respectively. From (3), we consider $l \in \{L, NL\}$, where L and N indicate LoS and non (NLoS), respectively. Only the LoS path is taken into account since this is typically dominant in mmWave communications as the gain of the LoS path can be 20-dB higher than those of NLoS [44]. Therefore, for simplicity of notation, the suffix l is removed in the remainder of this article unless otherwise specified. For Λ_S and Λ_D, the 2-D, steerable antenna model with a Gaussian main loop profile, described in [45], is utilized in our system model, that

is expressed as

$$\Lambda(\theta) = \Lambda_0 e^{-4\ln(2)\left(\frac{\theta}{\theta_{-3\ \mathrm{dB}}}\right)^2}, \ \Lambda_0 = \left(\frac{1.6162}{\sin\left(\frac{\theta_{-3\ \mathrm{dB}}}{2}\right)}\right)^2 \quad (4)$$

where $\theta, \theta_{-3\ \mathrm{dB}}$, and Λ_0 denote the azimuth angle, -3-dB beamwidth, and the maximum antenna gain, respectively.

B. VLC Links Model

For constructing the VLC channels and links in our considered heterogeneous system, we utilize the Lambertian model due to its suitability for the light-emitting diode (LED) transmitters [9], [11]. The channel gain, based on LED with the Lambertian patterns (considering the LED beam with solid as well as the maximum half-angles, the angle between the source-receiver line and beam axis, and the angle between the source-receiver line and receiver normal), is adopted similar to most existing research work [9]. The LoS channel gain for indoor VLC employing the Lambertian patterns is modeled as

$$H_{\mathrm{VLC}} = \frac{(g+1)A_R}{2\pi x^2} T_s(\phi_v) g(\phi_v) \cos^m \theta_v \cos\phi_v \ \forall \phi_v < \phi_c \quad (5)$$

where A_R denotes the optical detector's physical area. θ_v and ϕ_v define the incidence and irradiance angles, respectively. Next, ϕ_c represents the width of the field vision at the receiver while $(g = -([\ln(2)]/[\ln(\cos\theta_{v\max})]))$ identifies the order of the Lambertian model. Here, $T_s(\phi_v)$ and $g(\phi_v)$ indicate the gains of the optical filter and concentrator, respectively. Then, the received optical power, P_D^v, can be derived from the transmitted optical power, P_S^v, as

$$P_D^v = H_{\mathrm{VLC}} P_S^v. \quad (6)$$

C. Generalized Blocking Model

We leverage a unified blocking model that accounts for all the RF and VLC frequencies in the remainder of the section. Our generalized blocking model is inspired by the work in [46]. Experimental measurements conducted in an urban environment with quasi-static settings under small-sized ($5.07 \times 1.69 \times 1.93$ m^3) and large-sized ($7.01 \times 2.04 \times 2.63$ m^3) blockers have been considered in that work. Hence, motivated by the blockage loss analysis figures in [46], our generalized frequency-dependent blocking model is mathematically modeled as [46]

$$\text{Blockage loss[dB]} = \beta_a + \alpha_a \log\left(1 + \frac{f_{c,n}}{1\ \mathrm{GHz}}\right) \quad (7)$$

where α_a denotes the slope parameter, β_a represents the intercept of the line plotted via linear regression, and a indicates the index of the blocker type (i.e., small or large). This blockage loss can be amended to the path loss formula as a function of the operated frequency in addition to the blocker type.

IV. Problem Formulation

The hybrid RF-VLC band selection optimization problem in D2D scenario targets maximizing the long-term average throughput of the D2D linkage while considering the surrounding environment blocking and the battery consumption of the source device upon the selected band. To formally present the problem, by generalizing the system model described in Section III, we assume an undirected bi-partite graph, $G = (U, C, E)$ with where $U = \{u_1, \ldots, u_K\}$ and $C = \{c_1, \ldots, c_N\}$ denote the disjoint sets of users and channels, respectively; and $E = U \times C = \{e_{11}, \ldots, e_{\mathrm{KN}}\}$ represents a set of edges. Thus, an edge, e_{ij}, connects the vertices $u_i \in U$ and $c_j \in C$. Let x_{ij} be the weight for the edge e_{ij}. Furthermore, the cardinality of the sets U and C are given by $|U| = K$ and $|C| = N \leq \sum_{i=1}^{m} n_i$, respectively; where m and n_i denote the number of frequency bands (RF-VLC) and the number of channels in band i. Thus, N represents the total number of available channels across all the bands. Supposing that $K \leq N$, the target is to find a vertex-to-vertex matching with the maximum sum of weights such that no two vertices in U can be matched to any two vertices belonging to $|C|$ to prevent wireless resource (i.e., frequency band and corresponding channel) allocation conflict. The problem can then be formulated as

$$\arg\max_{i,j} \ \sum_{i=1}^{m}\sum_{j=1}^{n_i} x_{ij}\mathbb{E}\big(\psi_{ij}(t)\big) \quad (8a)$$

$$\text{s.t.} \quad \Xi_{s,ij}(t) > \Xi_{\mathrm{th}} \ \ \forall s \in \{\text{transmitting nodes}\} \quad (8b)$$

$$\sum_{i=1}^{m} n_i \leq N \quad (8c)$$

$$\sum_{i=1}^{m}\sum_{j=1}^{n_i} x_{ij} = 1 \quad (8d)$$

$$x_{ij} \in \{0, 1\} \quad (8e)$$

where $\psi_{ij}(t)$ indicates the throughput in bps of the S-D D2D link at time t. Note that t refers to the required time for establishing D2D link within a certain frequency band m. $x_{i,j}$ is a decision variable-based upon which the optimal band and its corresponding channel is to be selected to maximize the aggregated, expected throughput, denoted by $\mathbb{E}(\psi_{ij}(t))$. We express $\psi_{ij}(t)$ as

$$\psi_{ij}(t) = \frac{W_{i,j} T_D \Gamma_{i,j}(t)}{U(t) * T_{h,ij} + T_D} \quad (9)$$

where $W_{i,j}$ refers to the channel bandwidth of band i and its corresponding channel j. T_D refers to the data transmission time while $T_{h,ij}$ denotes the overhead time between the S-D (source destination) nodes pair according to the pulled frequency band. As mentioned earlier in Section III, this can be extended to a pair of relay nodes, or a relay-destination node also. Without any loss of generality for the conventional band selection method, $U(t) = N$ for $\forall t$. This is due to its policy of searching all the available bands before selection. $\Gamma_{i,j}(t)$ is the link spectral efficiency (SE) in bps/Hz upon the chosen band/frequency at a time t, which can be mathematically described as

$$\Gamma_{i,j}(t) = \log_2\left(1 + \frac{P_D^{i,j}(t)}{N_0 + I(t)}\right) \quad (10)$$

where $P_D^{i,j}(t)$ is the received power at destination at a time t according to the selected band, N_0 is the noise power at D, I is the interference from nearby devices that utilize the

same frequency. In this article, we consider the interference issued from the two WLAN channels only. The mmWave and VLC systems are directional ones, hence their interference are negligible with respect to the random noise. $\Xi_{s,i,j}(t)$ represents the standing energy in joule of the source device S at time t upon the utilized channel j of band i, and Ξ_{th} is the energy threshold after which the source devices cannot establish D2D links, and saves its power for its main activity.

The objective function (8a) is subject to several constraints, which are now described. For each transmitting node s, constraint (8b) indicates that the standing energy of device s should be larger than a threshold, Ξ_{th}. Next, constraint (8c) states that the total number of channels across all the considered frequency bands should be equal to or larger than the sensed, available number of channels, N. Then, constraints (8d) and (8e) signify that only one band and its corresponding channel can be selected at a time for a relay node, and that $x_{i,j}$ is a binary decision variable.

In order to solve the problem in (8a), *a centralized oracle with complete information*, i.e., the network-wide information of all bands and channel conditions for each node (source, destination, or relay), is required. Furthermore, the problem can be shown to be computationally hard for a large number of decision variables, and may be difficult to compute in polynomial time let alone real time. Therefore, we need to reformulate the objective of choosing the most appropriate band that maximizes the throughput of the entire link L_{S-D} using a distributed, online learning technique for every D2D node. This needs to be carried out by considering the utilized bands' blocking effects, which is variable from one band to another upon the nature of utilized frequencies. Motivated by this, we convert the problem [in (8a)] into an online, MAB problem [10]. Let us consider each channel j in a band i that belongs to the set C to be an arm. Note that unknown distributions are considered for arm selection affecting the reward (i.e., throughput). If we were to have the knowledge of the distribution $f_{i,j}$, we should always pull the arm that has the maximum mean $\mu_{i,j}$. Therefore, the optimal arm that should be selected is

$$\{i,j\}^* = \arg\max_{i,j}\ \mu_{i,j}. \tag{11}$$

In the following section, we develop algorithmic strategies to explore and identify the optimal arm, i.e., the optimal channel in a frequency band without sacrificing the reward completely during exploration. The policy of calculating $\Xi_{s,i,j}(t)$ is explained in detail in Section V.

V. Proposed Method

Similar to the typical bandit description, our formulated bandit has $[N] = 1, 2, 3, \ldots, n$ arms (available channels with different frequencies). During each round $t \in T$ (time horizon), selecting a channel n returns a reward $rw_{n,t}$ defined as the throughput. These rewards are RVs, independently and identically sampled, according to a specific probability distribution. However, in our formulated problem, pulling an arm n during every trial t is associated with a cost $c_{n,t}$, which is the estimated battery consumption of the source if band n were to be selected for transmission. This is because of the changed consumption values with respect to the utilized frequency during each round. The battery update equation of the transmitting node according to the selected band is given by

$$\Xi_{S,n^*_{\text{MAB}}}(t) = \Xi_{S,n^*_{\text{MAB}}}(t-1) - \frac{P^n_S L_{\text{Data}}}{W_{n^*_{\text{MAB}}}\Gamma_{n^*_{\text{MAB}}}(t)} \tag{12}$$

where MAB reflects the utilized policy (e.g., EABS-UCB, EABS-TS, or EABS-MOSS), $E_{S,n^*_{\text{MAB}}}(t-1)$ is the remaining energy of the source at the previous round $(t-1)$, and $P^n_S L_{\text{Data}}/W_{n^*_{\text{MAB}}}\Gamma_{n^*_{\text{MAB}}}(t)$ reflects the energy expenditure to transfer the desired data of L_{Data} bits using a data rate of $W_{n^*_{\text{MAB}}}\Gamma_{n^*_{\text{MAB}}}(t)$ bps via the chosen band n^*_{MAB}. For a fixed length of data, the source energy consumption depends on the utilized frequency in terms of the transmission bandwidth (W_n) and transmission power $(P_{S,n})$. In the remainder of this section, we propose three EABS algorithms based on UCB, TS, and MOSS MAB-based policies, respectively.

Practically TS is often achieves a smaller regret than many UCB-based algorithms [47], [48]. In addition, TS is simple and easy to implement. Despite these advantages, the theoretical analysis of TS algorithms has not been established until the past decade plus it is suitable for stochastic environments only. Moreover, MOSS is adaptable for both adversarial and stochastic bandit settings. The Three MAB techniques are guaranteed theoretically and can be applied in different scenarios.

A. EABS-UCB Algorithm

UCB (optimism against uncertainty) is one of the famous stochastic bandit algorithms that achieves a lower regret bound with the possibility of finite time analysis [49]. The UCB policy balances the exploration–exploitation tradeoff via collecting more information of the environment. It first focuses on exploration by attempting each arm once; then it concentrates on exploitation by choosing the action with the largest reward. For each arm, there is a lower and upper boundary (confidence interval). UCB name comes from upper bound concern where the bandit player is trying to obtain the arm with the maximum long-term reward. Our proposed EABS-UCB algorithm follows the original UCB policy with reflecting the battery consumption due to the chosen band for a realistic decision-making scenario. In our customized EABS-UCB algorithm, during every trial $t > N$, satisfying the following maximization equation is selected for constructing the D2D link:

$$n^*_{\text{EABS-UCB}}(t) = arg\max_n \left\{ \bar{\psi}_n(t) + \sqrt{\frac{\varrho \ln\ t}{M_{n,t}}} - \frac{x_n}{\Xi_{S,n}(t)} \right\} \tag{13}$$

where $\bar{\psi}_n(t)$ is the average throughput obtained from transmission band n till time t. $M_{n,t}$ is the number of times n has been picked till time t and ϱ is a tunable parameter for exploration and usually equals 2 [49]. The new term $(x_n/[\Xi_{S,n}(t)]$ is appended to the standard UCB formula to consider the battery consumption of the source relative to its distance from the destination due to a selected band/arm n at time t. Hence, for a fixed L_{Data} value in (12), the residual energy of the source is

Algorithm 1: EABS-UCB

Input: $t=0$, $\bar{\psi}_n(t)=0$, $M_{n,t}=0$, Ξ_{th}, $\Xi_{S,n}(t=1)$, $1\le n\le N$, $1\le t\le T$.
Initialization: pick each band n once in the first $t=N$ trials and update the Source remaining energies $\Xi_{S,n}(t)$ for $\forall n$ using eq. (12).
for $t=N+1:T$ **do**
  if $\Xi_{S,n}(t)>\Xi_{th}$ *for any* $n\in N$ **then**
    1. Choose the most appropriate channel index $n^*_{UCB}(t)$ for establishing the D2D connection using EABS-UCB eq. (13).
    2. Obtain $\psi_{n^{*(t)}_{UCB}}$ using eq. (9).
    3. $M_{n^{*,t}_{UCB}}=M_{n^{*,t-1}_{UCB}}+1$.
    4. $\bar{\psi}_{n^{*,t}_{UCB}}=\frac{1}{M_{n^{*,t}_{UCB}}}\sum_{j=1}^{M_{nUCB^{*,t}}}\psi_{n^{*,j}_{UCB}}$.
    5. Update the remaining energy $\Xi_{S,n^*_{UCB}}(t)$ using eq. (12).
  end
end

Algorithm 2: EABS-TS

Input: $t=0$, $\bar{\psi}_n(t)=0$, $M_{n,t}=0$, $\sigma^2_{n,t}=1$, Ξ_{th}, $\Xi_{S,n}(t=1)$, $1\le n\le N$, $1\le t\le T$.
Initialization: Pull each band n once in the first $t=N$ steps and update the remaining energies $\Xi_{S,n}(t)$ for $\forall n$ using eq. (12).
for $t=N+1:T$ **do**
  if $\Xi_{S,n}(t)>\Xi_{th}$ *for any* $n\in N$ **then**
    1. Apply the proposed EABS-TS to discover the channel index $n^*_{TS}(t)$ most appropriate for establishing the D2D connection at time t using eq. (14).
    2. Obtain $\psi_{n^*_{TS}(t)}$ using eq. (9).
    3. $M_{n^*_{TS},t}=M_{n^*_{TS},t-1}+1$.
    4. $\bar{\psi}_{n^*_{TS},t}=\frac{1}{M_{n^*_{TS},t}}\sum_{j=1}^{M_{n^*_{TS},t}}\psi_{n^*_{TS},j}$.
    5. $\sigma^2_{n^*_{TS},t}=\frac{1}{M_{n^*_{TS},t}+1}$.
    6. Update the remaining energy of the selected device upon the chosen band $\Xi_{S,n^*_{TS}}(t)$ using eq. (12).
  end
end

strongly related to the utilized frequency band n^*_{UCB} plus the distance from the destination x_n. This also impacts the achievable data rate $W_n\Gamma_{n^*_{\mathrm{UCB}}}(t)$ in (12). Algorithm 1 summarizes the main steps of the EABS-UCB algorithm. First, the algorithm tries all arms N one time and observe its reward with updating the source remaining energy $\forall n$ using (12). For the remaining of the time horizon T, the best arm is selected according to (13) with updating different parameters upon the chosen band. The game ends when the source remaining energy falls below a threshold value. Although simple idea of EABS-UCB, it is suitable for stochastic environments only and have lower performance than its similar-based TS version [50].

B. EABS-TS Algorithm

TS follows a Bayesian strategy, where posterior distributions are constructed for the gained rewards based on a predefined probabilistic model. TS achieves good empirical performance with guarantees even better than those warranted by UCB, especially when the said model highly matches the actual distribution of the rewards [50]. At the beginning of the TS algorithm, prior distributions are constructed for rewards based on parameter initialization of the said model. Then, the TS policy keeps track of the posterior distributions of the rewards based on the collected rewards during the learning process, which is achieved by updating the parameters of the probabilistic models [50], [51]. This is conducted through sampling each arm (channel of a particular band in our case) from the posterior distribution and determining the arm with the maximum expected reward. Due to the additive white Gaussian noise (AWGN), our reward distribution is assumed to be Gaussian. Therefore, we apply the Gaussian-based TS with a structure similar to the one used in [34]. Thus, a random sample is decided for each n based on the normal distribution function and selecting the arm with a maximum reward with applying energy constraint conditions. In the EABS-TS algorithm, the new term $(x_n/[\Xi_{S,n}(t)])$ is introduced to the central TS equation as follows:

$$n^*_{\mathrm{EABS-TS}}(t)=\arg\max_n\left\{\theta_n(t)-\frac{x_n}{\Xi_{S,n}(t)}\right\}$$

$$\theta_n(t)\sim\mathcal{N}\left(\bar{\psi}_n(t),\frac{1}{M_{n,t}+1}\right) \quad (14)$$

where $\mathcal{N}(\bar{\psi}_n(t),[1/(M_{n,t}+1)])$ referes to a normal distribution with $\bar{\psi}_n(t)$ mean and $[1/(M_{n,t}+1)]$ variance.

Algorithm 2 highlights the main steps of our proposed EABS-TS algorithm, where the Gaussian distribution is reflected for the output throughput rewards calculated by each band, i.e., $\mathcal{N}(\bar{\psi}_n(t),[1/(M_{n,t}+1)])$. An earlier distribution is initialized for the rewards by initializing $\bar{\psi}_n(t)$ and $\sigma^2_{n,t}$ as stated in Algorithm 2. Afterward, the samples $\theta_n(t)$ are sampled from these distributions, and the pillar band n^*_{TS} having the most extreme sample value is chosen to be played using (14). The sample of each arm is taken by the player (i.e., each transmitting D2D node), which selects the maximum arm and updates other variables for the forthcoming rounds. Upon the continuous updated, the rewards' posterior distributions are improved within the learning phase, resulting in better band/channel selections over the time horizon T.

C. EABS-MOSS Algorithm

MOSS is a mutated MAB arm selection policy that is adaptable to all associated bandit variants or subproblems (e.g., stochastic and adversarial setups) [51], [52]. The confidence interval of MOSS is designed to assign the time horizon T fairly within the number of arms N and the number of pulls each arm is picked. Once the suboptimal arms are more frequently selected, their correlated confidence interval narrows down, thereby demonstrating not only satisfactory exploration but also rapid convergence to the optimal arm. In our proposed EABS-MOSS, at each time $t>N$, the source/transmitting device, S, follows the maximization formula to select the best suitable link for transmission as expressed below

$$n^*_{\mathrm{EABS-MOSS}}(t)=\arg\max_n\left\{\bar{\psi}_n(t)+\sqrt{\frac{\max\left(\log\left(\frac{t}{M_{n,t}}\right),0\right)}{M_{n,t}}}-\frac{x_n}{\Xi_{S,n}(t)}\right\} \quad (15)$$

Algorithm 3: EABS-MOSS

Input: $t = 0$, $\bar{\psi}_n(t) = 0$, $M_{n,t} = 0$, Ξ_{th}, $\Xi_{S,n}(t = 1)$, $1 \le n \le N$, $1 \le t \le T$.
Initialization: Pull each band n once in the first $t = N$ steps and update the remaining energies $\Xi_{S,n}(t)$ for $\forall n$ using eq. (12).
for $t = N + 1 : T$ **do**
  if $\Xi_n(t) > \Xi_{th}$ *for any* $n \in N$ **then**
    1. Apply the proposed EABS-MOSS to discover the device index $n^*_{EABS-MOSS}(t)$ most appropriate for establishing the D2D connection at time t using eq. (15).
    2. Obtain $\psi_{n^*_{MOSS}(t)}$ using eq. (9).
    3. $M_{n^*_{MOSS},t} = M_{n^*_{MOSS},t-1} + 1$.
    4. $\bar{\psi}_{n^*_{MOSS},t} = \frac{1}{M_{n^*_{MOSS},t}} \sum_{j=1}^{M_{n^*_{MOSS},t}} \psi_{n^*_{MOSS},j}$.
    5. Update the remaining energy of source $\Xi_{S,n^*_{MOSS}}(t)$ using eq. (12).
  end
end

where $\bar{\psi}_n(t)$ denotes the average reward/throughput collected from channel n at time t. $M_{n,t}$ defines the total number of times channel n has been chosen until time t. The expression $(x_n/[\Xi_{S,n}(t)])$ is annexed to the regular MOSS equation to reflect the battery consumption of the source related to its distance from the destination and the chosen band. The steps of our proposed EABS-MOSS are summarized in Algorithm 3. All arms are pulled for once after initializing different parameters. The remaining energy due to each arm selection is reflected using (12). Afterward for the remaining trials of T, i.e., $N < t < T$, the best arm is selected according to the EABS-MOSS policy using (15) with updating different algorithm parameters. Finally, the algorithm chooses the best band at the last trial of T.

Readers interested in the regret bound derivation for each algorithm may refer to the work in [51] and [53]. Regarding the timing complexity, for EABS-UCB, the time complexity per trial is $O(N)$. For EABS-TS, time complexity per trial is $O(N)$ also, where a random sample is obtained from a Gaussian distribution by employing the standard Box-Muller transform [54]. Also, the timing complexity of EABS-MOSS algorithms is $O(N)$. In the following section, we evaluate the performance of these algorithms with and without energy awareness to carry out a comprehensive comparative performance analysis to verify the most suitable algorithm among the afore-mentioned three with the same computational complexity. In other words, we aim to identify the best solution without degrading the computational complexity of $O(N)$.

VI. Performance Evaluation

In this section, we evaluate the performance of our proposed algorithms, i.e., EABS-UCB, EABS-TS, and EABS-MOSS in a systematic manner in two different setups. The first setup investigates the impact of the adopted arm selection policies on the throughput and convergence performance in a nonenergy-awareness scenario. The original MAB policies (i.e., UCB, TS, and MOSS) are implemented without the newely added term to the multiband/channel selection problem named BS-UCB, BS-TS, and BS-MOSS, for clarity and ease of notation. Their performances are also benchmarked against the optimal, conventional, and random multiband/channel selection schemes for performance evaluation. The optimal selection technique directly selects the ideal (i.e., the best) channel with the complete knowledge of a centralized oracle as described in Section IV. The conventional scheme uses brute force to discover the best channel that requires a large search space and also requires complete information regarding the channel quality across all the bands. This consumes a considerable time. On the other hand, the random scheme arbitrarily picks a channel during each round t, without any channel quality consideration.

TABLE II
Considered Simulation Parameters

Simulation parameter	Value
N	4, i.e. (5.25,2.4,38,10^5) GHz
$W_1, f_{c,1}, P^1_S, T_{h,1}$	40MHz, 5.25GHz, 20mW, 3.6μsec
$W_2, f_{c,2}, P^2_S, T_{h,2}$	20MHz, 2.4GHz, 20mW, 3.6μsec
$W_3, f_{c,3}, P^3_S, T_{h,3}$	100MHz, 38GHz, 10mW, 0.28msec
$W_4, f_{c,4}, P^4_S, T_{h,4}$	20 MHz, 10^5GHz, 20mW, 3.6μsec
$\Xi_{S,n}(t = 1), \Xi_{th}$	uniform [0.01 $\rightarrow$ 1], 1%
x_0, x	5 m, [10 - 100] m
$\gamma^{LoS}_m, PL^m_0, \theta_{-3dB}$	2.3, 82.02, 20^o [55]
$A_R, \theta_v, \phi_v, \phi_c$	$1m^2$, 10^o, 60^o, 70^o
$T_s(\phi_v), g(\phi_v)$	1, 1.5
T, L_{Data}, T_D	1000, 1TB, 0.1 S
$\alpha_{small}, \beta_{small}$	2.6, 3 [46]
$\alpha_{large}, \beta_{large}$	3.6, 7.7 [46]

The second setup reflects the energy-awareness option where our proposed EABS-UCB, EABS-TS, and EABS-MOSS algorithms are compared in terms of both throughput and energy efficiency (EE) metrics, as well as convergence speed against previously mentioned three traditional selection methods for a comprehensive performance evaluation.

Table II lists the various parameters utilized in our conducted simulations in both setups. Without losing generality, we tried four arms/channels. The first and second channels are WIFI ones with 5.25 and 2.4-GHz central frequencies and bandwidths of 40 and 20 MHz, respectively. Both have same transmit power of 20 mW. The third arm is a mmWAve-based one with 38-GHz operating frequency, 100-MHz bandwidth, and 10-mW transmit power. The fourth arm is a VLC-based one with 10^5-GHz central frequency, 20-MHz bandwidth, and 20-mW transmit power. Other main parameters for these different channels and blockers sizes are detailed in Table II. We reconfirm our adoption of quasi-static blocking scenario in [46] for urban setup. As indicated earlier, the utilized performance indicators for the first scenario consist of throughput and convergence rates while the second setup comprises an additional performance metric in terms of the average EE. EE satisfies the average throughput over energy expenditure per selected band in bit/sec/joule that is expressed as

$$\text{EE} = \frac{1}{N} \sum_{n=1}^{N} \psi_n(T) / \big(\Xi_{S,n}(t = 1) - \Xi_{S,n}(t = T)\big) \quad (16)$$

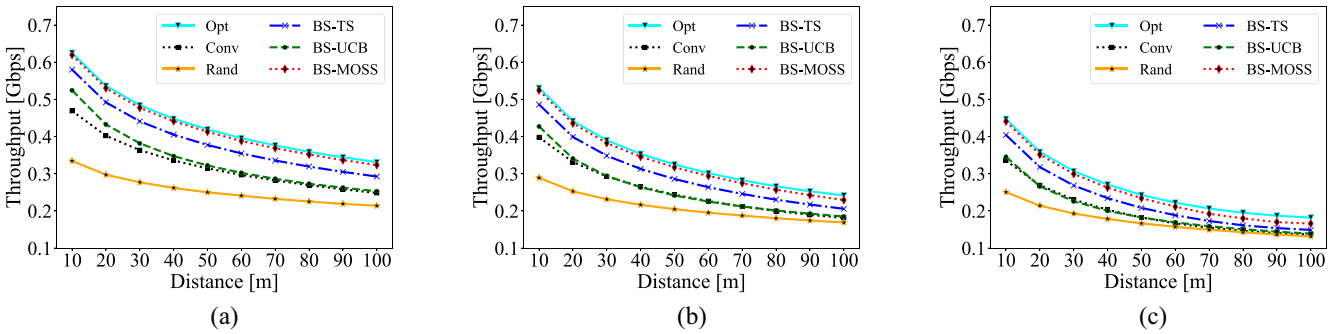

Fig. 2. Comparison of the conventional and MAB-based algorithms for assessing the throughput performance, without energy-awareness, for varying distances between transmitting and receiving D2D nodes and different blocking scenarios. (a) No blocking scenario. (b) Small blocking scenario. (c) Large blocking scenario.

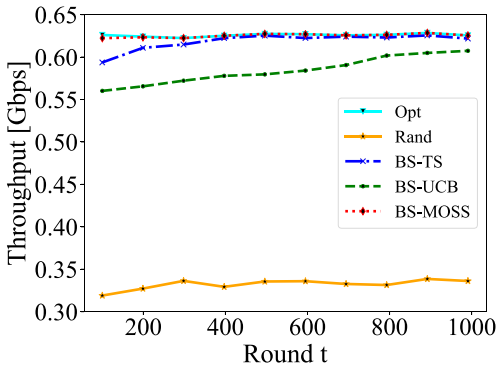

Fig. 3. Convergence performances of the MAB-based algorithms when energy-awareness option is not used.

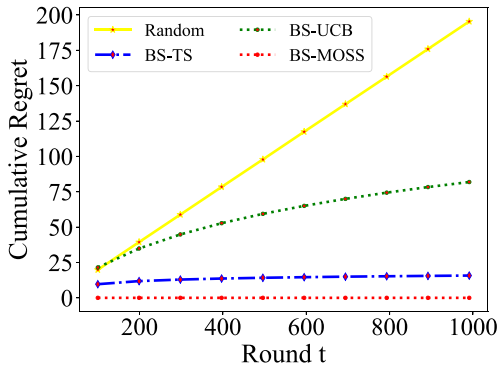

Fig. 4. Cumulative regret of the BS-MAB-based algorithms at 10-m separation distance and no blocking.

where $\Xi_{S,n}(t=1)$ denotes the initial energy (i.e., battery level) of S (the source/transmitting node) while $\Xi_{S,n}(t=T)$ represents its final energy after utilizing the nth $\in 1, 2, \ldots, N$ channel/arm.

A. Nonenergy-Aware Results

First, the results obtained in the first scenario are reported in Figs. 2 and 3. Fig. 2 illustrates the average throughput over multiple runs for varying distances (in meters) in terms of three different blocking scenarios. The average throughput comparisons for no blocking, small blocking, large blocking conditions are displayed in Fig. 2(a)–(c), respectively. Due to the blocking, the average throughput among the three blocking cases experiences a decreasing trend as the blocking is increased from none to large blockers. In each of the adopted blocking scenarios, the proposed MAB-based techniques outperform the conventional and random methods for selecting the most suitable channels in a multiband channel selection task. For all the considered distances, the proposed MAB-based MOSS, TS, and UCB achieve 97.23%, 88.32%, and 76.76% of the ideal throughput, respectively. On the other hand, the conventional techniques could not achieve such high efficiency according to the obtained simulation outcomes. The conventional brute-force and random approaches achieve only 75% and 62.24% of the ideal throughput, respectively. Thereby, the encouraging performance over multiple runs ensures the effectiveness of the proposed MAB-based techniques. Also, among the three MAB-based techniques (i.e., without the energy-awareness incorporation), the MOSS policy approach outperforms all other techniques due to its ability to select the most suitable channels in dynamic environments.

Next, the convergence rate for the leveraged MAB-based band selection techniques, while not considering the energy-awareness option, are presented in Fig. 3 and compared with the optimal, conventional, and random methods by considering 10 m of S-D distance at $T = 1000$. The best performance for the proposed approach is noted over 500 runs. At the round, t = 300, the proposed MAB-based MOSS technique reaches 99% convergence compared to the optimal scheme. TS and UCB-based policies obtained the convergence of 98.78% and 93.54%, respectively, at t = 300. These results demonstrate that these proposed policies can converge promptly even before emerging half of the total rounds (T) when the energy-awareness option is not enabled.

Next, in Fig. 4, we demonstrate the cumulative regret incurred during the progressive rounds for the baseline random implementation and the three MAB variants without energy consideration. As demonstrated in the results, BS-MOSS outperforms all the other methods and attains substantially low and stable regret values over increasing time rounds.

B. Energy-Aware Results

In the remainder of the section, we describe the simulation results for the second scenario where the energy-awareness is

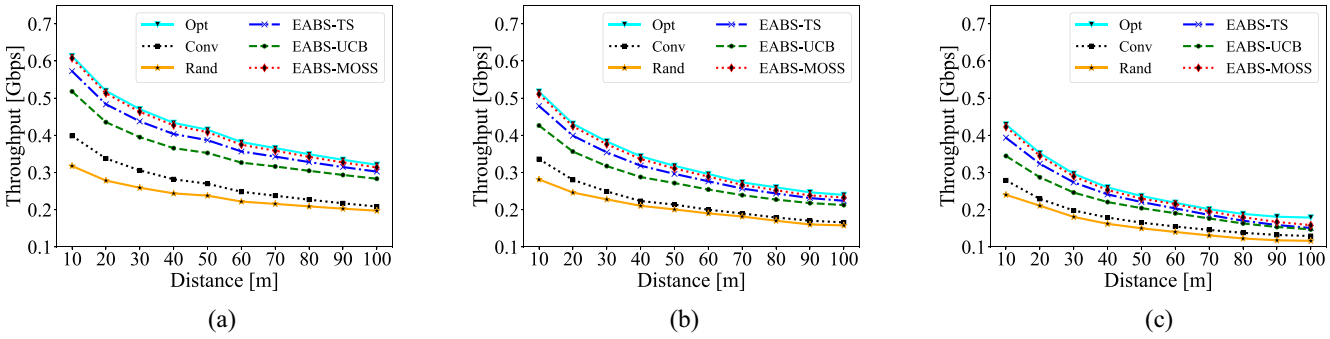

Fig. 5. Comparison of the conventional and MAB-based algorithms for assessing the throughput performance, with energy-awareness, for varying distances between transmitting and receiving D2D nodes and different blocking scenarios. (a) No blocking case. (b) Small blocking scenario. (c) Large blocking scenario.

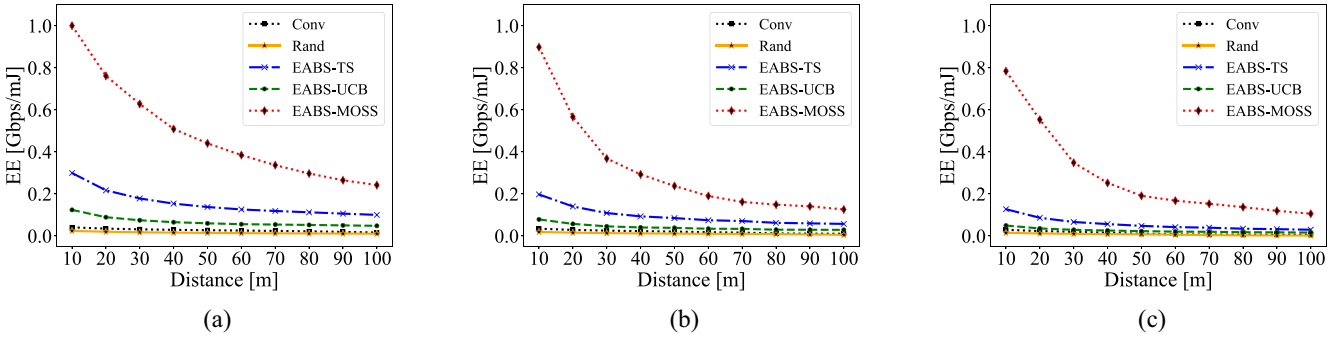

Fig. 6. Comparison of the conventional and MAB-based algorithms for assessing the EE performance, with energy-awareness, for varying distances between transmitting and receiving D2D nodes and different blocking scenarios. (a) No blocking scenario. (b) Small blocking scenario. (c) Large blocking scenario.

incorporated in our proposed EABS variants, namely, EABS-UCB, EABS-TS, and EABS-MOSS. The average throughput comparisons of these proposed algorithms with the traditional techniques are demonstrated in Fig. 5. The energy threshold for the considered transmitting device is set to 1% as listed in Table II. From Fig. 5, we can observe that the throughput demonstrated a downward trend as the blocking increased. The three proposed EABS variants demonstrate superior performance to the conventional and random selection techniques in all three blocking scenarios. In contrast with the optimal scenario, the EABS-MOSS, EABS-TS, and EABS-UCB achieve up to 98%, 92%, and 84% of the ideal throughput, respectively. On the other hand, the conventional brute-force method and random technique achieved only 59% and 66% of the ideal throughput. The random channel selection strategy manifests the worst performance with a steep decrease in the average throughput with an increasing distance because of the randomness in the chosen channel which might encounter a poor channel state. The conventional brute-force technique also illustrates poor performance due to searching over all options while deciding the channel to utilize.

Next, Fig. 6 exhibits the viability of our proposed EABS algorithms in terms of EE in Gbps/mJ over varying S-D distances (in meters). Here, Fig. 6(a)–(c) demonstrate the outcomes in the considered no-blocking, small blocking, and large blocking scenarios, respectively. The EE for all schemes illustrated a downward trend as the distance increased, due to the path loss. In all three considered blocking scenarios, the EABS-MOSS approach outperforms all other techniques due

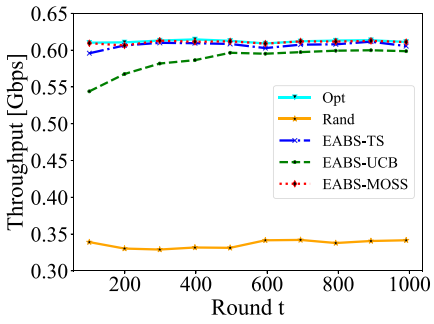

Fig. 7. Comparison of the convergence performances of the proposed EABS algorithms.

to its better performance and appropriate selection of channels with appropriate consideration of the consumed energy of the transmitting D2D node. Meanwhile, EABS-TS and EABS-UCB also displayed encouraging performances across all distances in terms of EE. However, due to the changeable selection policy in an energy-constrained situation, the performances of the conventional brute-force and random selection techniques were much worse than our proposed algorithms. The random channel selection technique yields the worst EE performance across all distances. Thus, the encouraging experimental results demonstrate that our proposed EABS algorithm(s), particularly EABS-MOSS, can be regarded as the most viable channel selection method(s) in multiband, multichannel D2D network environments.

Finally, in Fig. 7, a comparative convergence analysis of our proposed EABS-UCB, EABS-TS, and EABS-MOSS

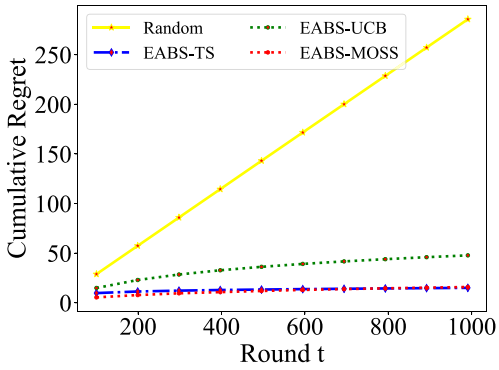

Fig. 8. Cumulative Regret of the EABS-MAB-based algorithms at 10-m separation distance and no blocking.

algorithms is conducted in terms of the average throughput at the S-D distance of 10 m and $T = 1000$. The proposed EABS-MOSS technique clearly outperforms other methods that reflects its fast convergence and reasonably accurate performance. At the time round, $t = 300$, EABS-MOSS attains a convergence rate of 99.98%, whereas EABS-TS and EABS-UCB result in around 99.4% and 94.9% of the optimal performance. On the other hand, the random selection technique obtains only a 53.65% convergence rate of the optimal performance.

Finally, in Fig. 8, we demonstrate the performance of the cumulative regret observed over the increasing time rounds for the compared methods with energy awareness. Compared to the nonenergy-aware scenario depicted in Fig. 4 where the MOSS-based MAB variant emerged as the best method, in the energy-aware scenario, EABS-MOSS and EABS-TS both provide satisfactory performances by achieving significantly low and stable regret values over the increasing time rounds up to 1000.

VII. Conclusion

In this article, motivated by IoT-enabled applications and services using D2D communications, we considered the problem of assigning the best wireless resources at real time to resource-constrained yet bandwidth-hungry D2D nodes, which are equipped with multiple bands and channels ranging from radio frequency or RF (e.g., 2.4-GHz WLAN, 5.25-GHz WLAN, and 38-GHz mmWave) to ultrahigh-frequency VLC technology operating at 400 to 800 THz. We formulated the best band selection optimization problem using energy-aware stochastic MAB techniques to increase the network throughput with the constraint of energy expenditure of the D2D nodes. Hence, our proposed algorithms, i.e., EABS-UCB, EABS-TS, and EABS-MOSS, reflected the transmitting D2D node's battery usage according to the selected transmission band. For each of the proposed EABS variants, we reported extensive simulation results to evaluate their performances. The performance evaluation clearly demonstrated that our proposed EABS-MOSS algorithm emerged as the most viable technique for finding the best bands/channels at an individual D2D transmitting node while preserving its energy-efficiency requirement. Future research directions are directed toward multiplayers scenario extension, where the interference will play a dominant role. Moreover, investigations of similar problems in time-varying environments like high speed communications where adversarial bandits is more suitable. Finally, investigations of such a hard problem in UAV communications with a real application such as disaster area coverage will be more beneficial.

References

[1] M. M. Salim, D. Wang, H. A. E. A. Elsayed, Y. Liu, and M. A. Elaziz, "Joint optimization of energy-harvesting-powered two-way relaying D2D communication for IoT: A rate–energy efficiency tradeoff," *IEEE Internet Things J.*, vol. 7, no. 12, pp. 11735–11752, Dec. 2020.

[2] M. E. Morocho-Cayamcela, H. Lee, and W. Lim, "Machine learning for 5G/B5G mobile and wireless communications: Potential, limitations, and future directions," *IEEE Access*, vol. 7, pp. 137184–137206, 2019.

[3] M. Zolanvari, M. A. Teixeira, L. Gupta, K. M. Khan, and R. Jain, "Machine learning-based network vulnerability analysis of industrial Internet of Things," *IEEE Internet Things J.*, vol. 6, no. 4, pp. 6822–6834, Aug. 2019.

[4] S. Maghsudi and E. Hossain, "Multi-armed bandits with application to 5G small cells," *IEEE Wireless Commun.*, vol. 23, no. 3, pp. 64–73, Jun. 2016.

[5] S. Chen, Y. Tao, D. Yu, F. Li, B. Gong, and X. Cheng, "Privacy-preserving collaborative learning for multiarmed bandits in IoT," *IEEE Internet Things J.*, vol. 8, no. 5, pp. 3276–3286, Mar. 2021.

[6] S. Misra, S. P. Rachuri, P. K. Deb, and A. Mukherjee, "Multiarmed-bandit-based decentralized computation offloading in fog-enabled IoT," *IEEE Internet Things J.*, vol. 8, no. 12, pp. 10010–10017, Jun. 2021.

[7] H. Chowdhury and M. Katz, "Cooperative data download on the move in indoor hybrid radio-optical WLAN-VLC hotspot coverage," *Trans. Emerg. Telecommun. Technol.*, vol. 25, no. 6, pp. 666–677, 2014.

[8] M. Ayyash *et al.*, "Coexistence of WiFi and LiFi toward 5G: Concepts, opportunities, and challenges," *IEEE Commun. Mag.*, vol. 54, no. 2, pp. 64–71, Feb. 2016.

[9] H. Abuella *et al.*, "Hybrid RF/VLC systems: A comprehensive survey on network topologies, performance analyses, applications, and future directions," 2020, *arXiv:2007.02466*.

[10] H. Tabassum and E. Hossain, "Coverage and rate analysis for co-existing RF/VLC downlink cellular networks," *IEEE Trans. Wireless Commun.*, vol. 17, no. 4, pp. 2588–2601, Apr. 2018.

[11] M. Hammouda, S. Akın, A. M. Vegni, H. Haas, and J. Peissig, "Link selection in hybrid RF/VLC systems under statistical queueing constraints," *IEEE Trans. Wireless Commun.*, vol. 17, no. 4, pp. 2738–2754, Apr. 2018.

[12] Z. Li *et al.*, "Design and implementation of a hybrid RF-VLC system with bandwidth aggregation," in *Proc. 14th Int. Wireless Commun. Mobile Comput. Conf. (IWCMC)*, 2018, pp. 194–200, doi: 10.1109/IWCMC.2018.8450350.

[13] A. M. Nor and E. M. Mohamed, "Li-Fi positioning for efficient millimeter wave beamforming training in indoor environment," *Mobile Netw. Appl.*, vol. 24, pp. 517–531, Apr. 2019.

[14] H. Yang, A. Alphones, W.-D. Zhong, C. Chen, and X. Xie, "Learning-based energy-efficient resource management by heterogeneous RF/VLC for ultra-reliable low-latency industrial IoT networks," *IEEE Trans. Ind. Informat.*, vol. 16, no. 8, pp. 5565–5576, Aug. 2020.

[15] J. Huang, C.-X. Wang, H. Chang, J. Sun, and X. Gao, "Multi-frequency multi-scenario millimeter wave MIMO channel measurements and modeling for B5G wireless communication systems," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 9, pp. 2010–2025, Sep. 2020.

[16] X. Zhang, D. Guo, K. An, G. Zheng, S. Chatzinotas, and B. Zhang, "Auction-based multichannel cooperative spectrum sharing in hybrid satellite-terrestrial IoT networks," *IEEE Internet Things J.*, vol. 8, no. 8, pp. 7009–7023, Apr. 2021.

[17] M. Islam, S. Sharmin, F. N. Nur, M. A. Razzaque, M. M. Hassan, and A. Alelaiwi, "High-throughput link-channel selection and power allocation in wireless mesh networks," *IEEE Access*, vol. 7, pp. 161040–161051, 2019.

[18] X. Zhao, L. Li, S. Geng, H. Zhang, and Y. Ma, "A link-based variable probability learning approach for partially overlapping channels assignment on multi-radio multi-channel wireless mesh information-centric IoT networks," *IEEE Access*, vol. 7, pp. 45137–45145, 2019.

[19] M. Najla, P. Mach, and Z. Becvar, "Deep learning for selection between RF and VLC bands in device-to-device communication," *IEEE Wireless Commun. Lett.*, vol. 9, no. 10, pp. 1763–1767, Oct. 2020.
[20] S. Sakib, T. Tazrin, M. M. Fouda, Z. M. Fadlullah, and N. Nasser, "A deep learning method for predictive channel assignment in beyond 5G networks," *IEEE Netw.*, vol. 35, no. 1, pp. 266–272, Jan./Feb. 2021.
[21] S. Sakib, T. Tazrin, M. M. Fouda, Z. M. Fadlullah, and N. Nasser, "An efficient and lightweight predictive channel assignment scheme for multiband B5G-enabled massive IoT: A deep learning approach," *IEEE Internet Things J.*, vol. 8, no. 7, pp. 5285–5297, Apr. 2021.
[22] Y. S. Nasir and D. Guo, "Multi-agent deep reinforcement learning for dynamic power allocation in wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 10, pp. 2239–2250, Oct. 2019.
[23] S. Wang and T. Lv, "Dynamic multichannel access for 5G and beyond with fast time-varying channel," in *Proc. IEEE Int. Conf. Commun. (ICC)*, 2020, pp. 1–6.
[24] Z. Wang, T. Zhang, Y. Liu, and W. Xu, "Caching placement and resource allocation for AR application in UAV noma networks," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, 2020, pp. 1–6.
[25] X. Liu, C. Sun, M. Zhou, C. Wu, B. Peng, and P. Li, "Reinforcement learning-based multislot double-threshold spectrum sensing with Bayesian fusion for industrial big spectrum data," *IEEE Trans. Ind. Informat.*, vol. 17, no. 5, pp. 3391–3400, May 2021.
[26] R. Yin, Z. Wu, S. Liu, C. Wu, J. Yuan, and X. Chen, "Decentralized radio resource adaptation in D2D-U networks," *IEEE Internet Things J.*, vol. 8, no. 8, pp. 6720–6732, Apr. 2021.
[27] S. Hashima, K. Hatano, E. Takimoto, and E. M. Mohamed, "Neighbor discovery and selection in millimeter wave D2D networks using stochastic MAB," *IEEE Commun. Lett.*, vol. 24, no. 8, pp. 1840–1844, Aug. 2020.
[28] S. Hashima, K. Hatano, H. Kasban, and E. M. Mohamed, "Wi-Fi assisted contextual multi-armed bandit for neighbor discovery and selection in millimeter wave device to device communications," *Sensors*, vol. 21, no. 8, p. 2835, 2021.
[29] E. M. Mohamed, S. Hashima, A. Aldosary, K. Hatano, and M. A. Abdelghany, "Gateway selection in millimeter wave UAV wireless networks using multi-player multi-armed bandit," *Sensors*, vol. 20, no. 14, p. 3947, 2020.
[30] E. M. Mohamed, S. Hashima, K. Hatano, H. Kasban, and M. Rihan, "Millimeter-wave concurrent beamforming: A multi-player multi-armed bandit approach," *Comput. Mater. Continua*, vol. 65, no. 3, pp. 1987–2007, 2020.
[31] I. Aykin, B. Akgun, M. Feng, and M. Krunz, "MAMBA: A multi-armed bandit framework for beam tracking in millimeter-wave systems," in *Proc. IEEE INFOCOM Conf. Comput. Commun.*, 2020, pp. 1469–1478, doi: 10.1109/INFOCOM41043.2020.9155408.
[32] I. Chafaa, E. V. Belmega, and M. Debbah, "Adversarial multi-armed bandit for mmWave beam alignment with one-bit feedback," in *Proc. 12th EAI Int. Conf. Perform. Eval. Methodol. Tools*, New York, NY, USA, 2019, pp. 23–30, doi: 10.1145/3306309.3306315.
[33] A. Vora and K.-D. Kang, "Throughput enhancement via multi-armed bandit in heterogeneous 5G networks," in *Proc. IEEE 88th Veh. Technol. Conf. (VTC-Fall)*, 2018, pp. 1–5, doi: 10.1109/VTCFall.2018.8690592.
[34] F. Wilhelmi, C. Cano, G. Neu, B. Bellalta, A. Jonsson, and S. Barrachina-Muñoz, "Collaborative spatial reuse in wireless networks via selfish multi-armed bandits," *Ad Hoc Netw.*, vol. 88, pp. 129–141, May 2019.
[35] Z. Tian, J. Wang, J. Wang, and J. Song, "Distributed NOMA-based multi-armed bandit approach for channel access in cognitive radio networks," *IEEE Wireless Commun. Lett.*, vol. 8, no. 4, pp. 1112–1115, Aug. 2019.
[36] S. Shrivastava, B. Chen, C. Chen, H. Wang, and M. Dai, "Deep Q-network learning based downlink resource allocation for hybrid RF/VLC systems," *IEEE Access*, vol. 8, pp. 149412–149434, 2020.
[37] Z. Du, C. Wang, Y. Sun, and G. Wu, "Context-aware indoor VLC/RF heterogeneous network selection: Reinforcement learning with knowledge transfer," *IEEE Access*, vol. 6, pp. 33275–33284, 2018.
[38] C. Wang, G. Wu, Z. Du, and B. Jiang, "Reinforcement learning based network selection for hybrid VLC and RF systems," in *Proc. Int. Conf. Smart Mater. Intell. Manuf. Autom.*, vol. 173, 2018, p. 3014, doi: 10.1051/matecconf/201817303014.
[39] Y. Li, Y. Liang, Q. Liu, and H. Wang, "Resources allocation in multicell D2D communications for Internet of Things," *IEEE Internet Things J.*, vol. 5, no. 5, pp. 4100–4108, Oct. 2018.
[40] M. Noor-A-Rahim *et al.*, "6G for vehicle-to-everything (V2X) communications: Enabling technologies, challenges, and opportunities," 2020, *arXiv:2012.07753*.
[41] E. M. Mohamed, M. A. Abdelghany, and M. Zareei, "An efficient paradigm for multiband WiGig D2D networks," *IEEE Access*, vol. 7, pp. 70032–70045, 2019.
[42] X. Gao *et al.*, "Large-scale characteristics of 5.25 GHz based on wideband MIMO channel measurements," *IEEE Antennas Wireless Propag. Lett.*, vol. 6, pp. 263–266, 2007.
[43] S. Kaddouri, M. E. Hajj, G. Zaharia, and G. E. Zein, "Indoor path loss measurements and modeling in an open-space office at 2.4 GHz and 5.8 GHz in the presence of people," in *Proc. IEEE 29th Annu. Int. Symp. Pers. Indoor Mobile Radio Commun. (PIMRC)*, 2018, pp. 1–7, doi: 10.1109/PIMRC.2018.8580695.
[44] N. Wei, X. Lin, and Z. Zhang, "Optimal relay probing in millimeter-wave cellular systems with device-to-device relaying," *IEEE Trans. Veh. Technol.*, vol. 65, no. 12, pp. 10218–10222, Dec. 2016.
[45] S. Wu, R. Atat, N. Mastronarde, and L. Liu, "Improving the coverage and spectral efficiency of millimeter-wave cellular networks using device-to-device relays," *IEEE Trans. Commun.*, vol. 66, no. 5, pp. 2251–2265, May 2018.
[46] M. Boban *et al.*, "Multi-band vehicle-to-vehicle channel characterization in the presence of vehicle blockage," *IEEE Access*, vol. 7, pp. 9724–9735, 2019.
[47] O. Chapelle and L. Li, "An empirical evaluation of thompson sampling," in *Advances in Neural Information Processing Systems*, vol. 24. Red Hook, NY, USA: Curran Assoc., Inc., 2011.
[48] S. Wang and W. Chen, "Thompson sampling for combinatorial semi-bandits," in *Proc. 35th Int. Conf. Mach. Learn.*, vol. 80, 2018, pp. 5114–5122.
[49] P. Auer, N. Cesa-Bianchi, and P. Fischer, "Finite-time analysis of the multiarmed bandit problem," *Mach. Learn.*, vol. 47, no. 2, pp. 235–256, 2002.
[50] S. Agrawal and N. Goyal, "Analysis of thompson sampling for the multi-armed bandit problem," in *Proc. 25th Annu. Conf. Learn. Theory*, vol. 23, 2012, pp. 1–26.
[51] S. Agrawal and N. Goyal, "Near-optimal regret bounds for thompson sampling," *J. ACM*, vol. 64, no. 5, pp. 1–24, Sep. 2017.
[52] S. Hashima, K. Hatano, E. Takimoto, and E. M. Mohamed, "Minimax optimal stochastic strategy (MOSS) for neighbor discovery and selection in millimeter wave D2D networks," in *Proc. 23rd Int. Symp. Wireless Pers. Multimedia Commun. (WPMC)*, 2020, pp. 1–6, doi: 10.1109/WPMC50192.2020.9309495.
[53] J.-Y. Audibert and S. Bubeck, "Minimax policies for adversarial and stochastic bandits," in *Proc. 22nd Annu. Conf. Learn. Theory (COLT)*, 2009, pp. 217–226.
[54] N. Balakrishnan, H. Y. So, and X. J. Zhu, "On Box–Muller transformation and simulation of normal record data," *Commun. Stat. Simul. Comput.*, vol. 45, no. 10, pp. 3670–3682, 2016.
[55] T. S. Rappaport *et al.*, "Millimeter wave mobile communications for 5G cellular: It will work!" *IEEE Access*, vol. 1, pp. 335–349, 2013.

Sherief Hashima (Senior Member, IEEE) received the B.Sc. degree (Hons.) in electronics and communication engineering from Tanta University, Tanta, Egypt, in 2004, the M.Sc. degree from Menoufiya University, Shibin Al Kawm, Egypt, in 2010, and the Ph.D. degree from Egypt-Japan University of Science & Technology, Alexandria, Egypt, in 2014.

He has been a Postdoctoral Researcher with the Computational Learning Theory Team, RIKEN AIP, Fukuoka, Japan, since July 2019. He has been working as an Associate Professor with the Engineering and Scientific Equipment Department, Nuclear Research Center, Egyptian Atomic Energy Authority, Cairo, Egypt, since 2014. From January 2018 to June 2018, he was a Visiting Researcher with the Center for Japan-Egypt Cooperation in Science and Technology, Kyushu University, Fukuoka. His research interests include wireless communications, machine learning, online learning, 5G, B5G, and 6G systems, image processing, millimeter waves, nuclear instrumentation, and Internet of Things.

Dr. Hashima is a technical committee member in many international conferences and a reviewer in many international conferences, journals, and transactions. He is a member of AAAI.

Mostafa M. Fouda (Senior Member, IEEE) received the Ph.D. degree in information sciences from Tohoku University, Sendai, Japan, in 2011.

He is currently an Assistant Professor with the Department of Electrical and Computer Engineering, Idaho State University, Pocatello, ID, USA. He also holds the position of an Associate Professor with Benha University, Banha, Egypt. He has served as an Assistant Professor with Tohoku University. He was a Postdoctoral Research Associate with Tennessee Technological University, Cookeville, TN, USA. He has more than 90 publications in international conferences, journal papers, and book chapters. His research interests include cyber security, machine learning, blockchain, the IoT, 6G networks, smart healthcare, and smart grid communications.

Dr. Fouda is also a reviewer in several IEEE TRANSACTIONS and magazines. He is an Editor of IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY and an Associate Editor of IEEE ACCESS. He has served on the technical committees of several IEEE conferences.

Sadman Sakib received the B.Sc. degree in CSE from the Ahsanullah University of Science and Technology, Dhaka, Bangladesh, in 2017. He is currently pursuing the Graduate degree with the Department of Computer Science, Lakehead University, Thunder Bay, ON, Canada.

His research interests are in the field of intelligent computing/communication systems for mobile health applications for providing better health outcomes in northern Ontario and beyond.

Mr. Sakib was the Winner of the Best Student's Poster Award of the Canadian Institutes of Health Research category in the graduate conference held at the Graduate Student Conference—Poster Presentation, Research and Innovation Week, Lakehead University, 2020. He also received an internship opportunity through the Mitacs Accelerate research grant in 2020.

Zubair Md. Fadlullah (Senior Member, IEEE) received the Ph.D. degree in information sciences from Tohoku University, Sendai, Japan, in 2011.

He is currently an Associate Professor with the Computer Science Department, Lakehead University, Thunder Bay, ON, Canada, and a Research Chair of the Thunder Bay Regional Health Research Institute, Thunder Bay. He was an Associate Professor with the Graduate School of Information Sciences, Tohoku University, Sendai, from 2017 to 2019. His main research interests are in the areas of emerging communication systems, such as 5G new radio and beyond, deep learning applications on solving computer science and communication system problems, UAV-based systems, smart health technology, cyber security, game theory, smart grid, and emerging communication systems.

Dr. Fadlullah received several best paper awards at conferences, including IEEE/ACM IWCMC, IEEE GLOBECOM, and IEEE IC-NIDC. He is currently an Editor of IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, *IEEE Network Magazine*, IEEE ACCESS, IEEE OPEN JOURNAL OF THE COMMUNICATIONS SOCIETY, and *Ad Hoc & Sensor Wireless Networks* journal. He is a Senior Member of IEEE Communications Society.

Kohei Hatano received the Ph.D. degree from Tokyo Institute of Technology, Tokyo, Japan, in 2005.

He is currently an Associate Professor with Research and Development Division, Kyushu University Library, Fukuoka, Japan. He is also the Leader of the Computational Learning Theory Team, RIKEN AIP, Fukuoka. His research interests include machine learning, computational learning theory, online learning, and their applications.

Ehab Mahmoud Mohamed (Member, IEEE) received the B.E. and M.E. degrees in electrical engineering from South Valley University, Qena, Egypt, in 2001 and 2006, respectively, and the Ph.D. degree in information science and electrical engineering from Kyushu University, Fukuoka, Japan, in 2012.

From 2013 to 2016, he has joined Osaka University, Suita, Japan, as a Specially Appointed Researcher. Since 2017, he has been an Associate Professor with Aswan University, Tingar, Egypt. He has also been an Associate Professor with Prince Sattam Bin Abdulaziz University, Al-Kharj, Saudi Arabia, since 2019. His current research interests include 5G, B5G and 6G networks, cognitive radio networks, millimeter-wave transmissions, Li-Fi technology, MIMO systems, and underwater communication.

Dr. Mohamed is a technical committee member of many international conferences and a reviewer of many international conferences, journals, and transactions. He is the General Chair of the IEEE ITEMS'16 and IEEE ISWC'18.

Xuemin (Sherman) Shen (Fellow, IEEE) received the Ph.D. degree in electrical engineering from Rutgers University, New Brunswick, NJ, USA, in 1990.

He is currently a University Professor with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON, Canada. His research focuses on resource management in interconnected wireless/wired networks, wireless network security, social networks, smart grid, and vehicular *ad hoc* and sensor networks.

Dr. Shen received the R. A. Fessenden Award from IEEE, Canada, in 2019, the Award of Merit from the Federation of Chinese Canadian Professionals (Ontario) presents in 2019, the James Evans Avant Garde Award from the IEEE Vehicular Technology Society in 2018, the Joseph LoCicero Award in 2015, the Education Award from the IEEE Communications Society in 2017, and the Technical Recognition Award from Wireless Communications Technical Committee in 2019 and AHSN Technical Committee in 2013. He has also received the Excellent Graduate Supervision Award from the University of Waterloo in 2006 and the Premier's Research Excellence Award from the Province of Ontario, Canada, in 2003. He served as the Technical Program Committee Chair/Co-Chair for the IEEE Globecom'16, the IEEE Infocom'14, the IEEE VTC'10 Fall, and the IEEE Globecom'07, the Symposia Chair for the IEEE ICC'10, and the Chair for the IEEE Communications Society Technical Committee on Wireless Communications. He is the elected IEEE Communications Society Vice President for Technical and Educational Activities, a Vice President for Publications, a Member-at-Large on the Board of Governors, a Chair of the Distinguished Lecturer Selection Committee, a member of IEEE Fellow Selection Committee. He was/is the Editor-in-Chief of the IEEE INTERNET OF THINGS JOURNAL, IEEE NETWORK, *IET Communications*, and *Peer-to-Peer Networking and Applications*. He is a Registered Professional Engineer of Ontario, Canada, an Engineering Institute of Canada Fellow, a Canadian Academy of Engineering Fellow, a Royal Society of Canada Fellow, a Chinese Academy of Engineering Foreign Fellow, and a Distinguished Lecturer of the IEEE Vehicular Technology Society and Communications Society.