II. System Model

In this section, we describe our considered hybrid-band wireless system, which consists of the RF and VLC link models and the unified blockage model.

Each transmitting/receiving node, also referred to as the transmitter-receiver (Tx-Rx) pair, in the considered hybrid band wireless system is assumed to utilize multiple frequency bands operating at 2.4/5.25GHz WLAN, 38 GHz WiGig, and 400-800THz VLC. Each frequency band comprises numerous channels to transmit/receive and forward/relay data frames. For a unified link model for both 5.25GHz and 2.4GHz WLAN systems, we utilize the channel model used in [6]. Log-normal shadowing with zero mean is considered for both WLAN types, while a standard deviation of 6dB [7] and 2.15dB [8] are considered for the 5.25GHz and 2.4GHz WLAN systems, respectively. The reference path loss and path loss exponents in the 5.25GHz WLAN are set to 47.2 and 2.32, respectively, while these parameters for the 2.4GHz system are considered to be 41.8 and 2, respectively. Moreover, a 2-D (two-dimensional), steerable antenna model with a Gaussian main loop profile, described in [9], is adopted in the system model having -3dB beamwidth. Next, for constructing the VLC channels model in the hybrid wireless system, the Lambertian model [2], [10] is considered since it adequately incorporates light-emitting diode (LED) transmitters. The channel gain, based on LED with 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 standard), are adopted similar the consideration given in the contemporary research work [10]. Next, we adopt a unified blocking model that accounts for all the RF and VLC frequencies by considering the experimental parameters of the setup presented by [11] in an urban environment with a quasi-static environment under small-sized (5.07x1.69x1.93m^3) and large-sized (7.01x2.04x2.63m^3) blockers. Hence, motivated by the blockage loss analysis figures in [11], our generalized frequency-dependent blocking model is modeled as:

$$Blockageloss[dB] = \beta_a + \alpha_a \log(1 + \frac{f_{c,n}}{1\text{GHz}}), \tag{1}$$

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.

III. Problem Formulation

The pre-described MBS (multi-band/channel selection) problem in our considered hybrid wireless system is formulated as follows.

$$\arg\max_{i,j} \sum_{i=1}^{m}\sum_{j=1}^{n_i} x_{ij}\mathbb{E}(\psi_{ij}(t)) \tag{2a}$$

$$s.t.$$

$$\Xi_{\text{Tx},ij}(t) > \Xi_{th}\ , \forall s \in \{\text{transmitting nodes}\} \tag{2b}$$

$$\sum_{i=1}^{m} n_i \geq N \tag{2c}$$

$$\sum_{i=1}^{m}\sum_{j=1}^{n_i} x_{ij} = 1\ , \tag{2d}$$

$$x_{ij} \in \{0, 1\}\ , \tag{2e}$$

where m and n_i denote the number of heterogeneous frequency bands and the number of channels in band i, respectively. N indicates the total number of available channels across all the bands. $\psi_{ij}(t)$ indicates the throughput in bps of the transmitter-receiver link at time t utilizing channel j of band i. Next, x_{ij} denotes a decision variable, based on 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:

$$\psi_{ij}(t) = \frac{W_{ij}T_D\Gamma_{ij}(t)}{U(t) * T_{h,ij} + T_D}, \tag{3}$$

where W_{ij} 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 hybrid band Tx-Rx pair according to the selected frequency band. Without any loss of generality for the conventional MBS scheme, $U(t) = N$ for $\forall t$. This is due to its policy of searching all the available bands before selection. $\Gamma_{ij}(t)$ is the link spectral efficiency (SE) in bps/Hz upon the chosen band/frequency at time t, which can be expressed as,

$$\Gamma_{ij}(t) = \log_2(1 + \frac{P^{ij}_{\text{Rx}}(t)}{N_0 + I(t)}), \tag{4}$$

where $P^{ij}_{\text{Rx}}(t)$ denotes the received power at Rx at time t according to the selected band, N_0 indicates the noise power at Rx, and I refers to the interference from nearby devices that utilize the same frequency. In this paper, 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. Next, $\Xi_{\text{Tx},ij}(t)$ represents the residual energy (in Joules) 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 may no longer be able to establish wireless links, and therefore, is compelled to save its power for its main activity.

The objective function (2a) is subject to several constraints, which are now described. For each hybrid-band transmitting Tx, constraint (2b) indicates that the residual energy of device s should be larger than a threshold, Ξ_{th}. Next, constraint (2c) 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 (2d) and (2e) signify that only one band and its corresponding channel can be selected at a time for a relay link, and that x_{ij} is a binary decision variable, respectively.

To maximize the objective function (2a), *a centralized oracle with complete information*, i.e., the network-wide information of all bands and channel conditions for each link (Tx, Rx, 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_{Tx-Rx} using a distributed, online learning technique for every link. 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 eq. (2a)) into an online, multi-armed bandit model [12]–[15]. 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_{ij}, we should always pull the arm that has the maximum mean μ_{ij}. Therefore, the optimal arm that should be selected is:

$$\{ij\}^* = \arg\max_{ij} \ \mu_{ij}. \tag{5}$$

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

IV. Proposed MBS Schemes

Our formulated bandit model consists of $N = \{1, 2, 3, ..., n\}$ arms (free channels with heterogeneous frequencies). At each trial $t \in T$ (time horizon), choosing a channel n provides a throughput reward $rw_{n,t}$. Similar to stochastic MABs, these rewards are i.i.d. (independent and identically distributed) and sampled from a probability distribution, unknown to the player. Furthermore, in our setting, drawing an arm n within each trial t is associated with a budget cost $c_{n,t}$, defined as the consumed energy of the Tx/player if band n was chosen for data frames transmission. Each band has different energy consumption according to the utilized frequency. The energy update formula of the Tx according to the chosen band is as follows:

$$\Xi_{\mathrm{Tx},n^*_{MAB}}(t) = \Xi_{\mathrm{Tx},n^*_{MAB}}(t-1) - \frac{P^n_{\mathrm{Tx}} L_D}{W_{n^*_{MAB}} \Gamma_{n^*_{MAB}}(t)}, \tag{6}$$

where *MAB* reflects the utilized MAB scheme (e.g., RUCB, KLUCB, UCB, and TS), $E_{\mathrm{Tx},n^*_{MAB}}(t-1)$ is the remaining energy of the hybrid-band Tx at the previous trial $(t-1)$, and the term $P^n_{\mathrm{Tx}} L_D / W_{n^*_{MAB}} \Gamma_{n^*_{MAB}}(t)$ defines the energy consumption for transmitting the required data of L_D bits with a data rate of $W_{n^*_{MAB}} \Gamma_{n^*_{MAB}}(t)$ bps using the selected band n^*_{MAB}. For a fixed length of data, the energy consumption, for the hybrid band selection of Tx, relies on the employed frequency in terms of the transmission bandwidth (W_n) and transmission power ($P_{\mathrm{Tx},n}$). In the remainder of this section, we discuss our proposed EA-RUCB-MBS and EA-KLUCB-MBS.

A. EA-RUCB-MBS Algorithm

RUCB is a modified version of UCB [3], [4]. Unlike the classical UCB which keeps the exploration parameter constant, RUCB aims at evaluating the exploration parameter in a more efficient manner. RUCB randomizes the confidence interval, and selects the most ideal arm by substituting the exploration parameter by a random variable Z_t, where $Z_1, ...Z_T$ are i.i.d sampled from the same distribution, which is Gaussian in our case due to the effect of the additive white Gaussian noise (AWGN) channel. We consider a discrete distribution for Z on the interval $[L; U]$, supported on 20 points. According to our proposed EA-RUCB-MBS technique as shown in Algorithm 1, during every trial $t > N$, the best arm is selected as follows:

$$n^*_{RUCB-MBS}(t) = \\ arg\max_n \{\bar{\psi}_n(t-1) + \mathrm{Z}_t \sqrt{\frac{2ln\ t}{M_{n,t-1}}} - \frac{x_n}{\Xi_{\mathrm{Tx},n}(t)}\}, \tag{7}$$

where $\bar{\psi}_n(t)$ denotes the average throughput obtained from the transmission band n until time t. $M_{n,t-1}$ refers to the number of times n has been picked until time t. A new term, $\frac{x_n}{\Xi_{\mathrm{Tx},n}(t)}$, is appended to the standard RUCB formula to reflect the battery consumption of the hybrid-band transmitter relative to its distance from the hybrid-band receiver. Hence, for a fixed L_D value in eq. (6), the residual energy of the source is strongly related to the utilized frequency band plus the distance from the destination. This impacts the achievable data rate $W_n \Gamma_{n^*_{UCB}}(t)$. Algorithm 1 summarizes the main steps of our proposed EA-RUCB-MBS scheme.

B. EA-KLUCB-MBS Algorithm

While KLUCB belongs to the UCB family, it aims to provide a better regret bound than those of the original UCB1 and its variants [3], [5]. The idea behind KLUCB is to calculate the bounded Kullback–Leibler divergence of the average reward of each arm, and the arm characterized with the maximum divergence parameter is played. Our proposed EA-KLUCB-MBS criterion is presented in eq. (8), where the energy consumption term is added to the original KLUCB expression (assuming Gaussian distribution) as follows:

$$n^*_{KLUCB-MBS}(t) = arg\max_n \\ \left\{ sup\left\{ \mu_n(t) \in (0,1) : d\left(\bar{\psi}_n(t), \mu_n(t)\right) \leq \frac{f(t)}{T_{n,t}} \right\} - \frac{x_n}{\Xi_{\mathrm{Tx},n}(t)} \right\}, \tag{8}$$

where, $f(t) = log\ t + 3log\,(log\ (t))$ and $d(\mu_1, \mu_2) = 2(\mu_1 - \mu_2)^2$. Accordingly, eq. $2(\bar{\psi}_n(t) - \mu_n(t))^2 \leq \frac{f(t)}{T_{n,t}}$ is solved to find out the divergence parameter $\mu_n(t)$ of band n

Algorithm 1 EA-RUCB/KLUCB-MBS Algorithms

1: **Input**: $t = 0$, $\bar{\psi}_n(t) = 0$, $M_{n,t} = 0$, $\Xi_{th}, \Xi_{S,n}(t = 1)$, $1 \leq n \leq N$, $1 \leq t \leq T$.
2: **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. (6).
3: **While** $\Xi_{S,n}(t) > \Xi_{th}$ **for any** $n \in N$
4: **For** t=N+1:T
5: Choose the most appropriate channel index $n^*_{(R/KL)UCB}(t)$ for Tx-Rx using EA-(R/KL)UCB-MBS eq. (7), and eq. (8), respectively.
6: Obtain $\psi^*_{n_{(R/KL)UCB}}(t)$ using eq. (3).
7: $M_{n^*_{(R/KL)UCB},t} = M_{n^*_{(R/KL)UCB},t-1} + 1$
8: $\bar{\psi}_{n^*_{(R/KL)UCB}}(t) = \frac{1}{M_{n^*_{(R/KL)UCB},t}} \sum_{j=1}^{M_{n^*_{(R/KL)UCB},t}} \psi_{n^*_{(R/KL)UCB}}(j)$
9: Update the remaining energy of the chosen band $\Xi_{\mathrm{Tx},n^*_{(R/KL)UCB}}(t)$ using eq. (6).
10: **End For**
11: **End while**

at time t. Algorithm 1 summarizes the proposed EA-KLUCB-MBS scheme, where $\Xi_{\mathrm{Tx},n}$, Ξ_{th}, and N denote the inputs to the algorithm and the selected band at each round t, i.e., $n^*_{KLUCB-MBS}(t)$, is its output. For initialization, each band is selected at once, i.e., $M_{n,t} = 1$, $\forall n$, and its corresponding spectral efficiency $\psi_n(t)$ is obtained. Then, for each time t, $N + 1 \leq t \leq T$, a band is selected, $n^*_{KLUCB-MBS}(t)$, based on eq. (8) as long as its residual energy is larger than the specified threshold Ξ_{th}. After selecting the optimal band, its parameters $M_{n^*_{KLUCB},t}$, $\bar{\psi}_{n^*_{KLUCB}}(t)$, and $\Xi_{\mathrm{Tx},n^*_{KLUCB}}(t)$ are updated for the next round of band selection as given in Algorithm 1.

V. Simulation Results

In this section, we evaluate the performance of our proposed EA-RUCB-MBS and EA-KLUCB-MBS algorithms. First, non-energy-aware RUCB-MBS and KLUCB-MBS algorithms are compared with similar reference schemes, such as UCB-MBS and TS-MBS, as well as baseline schemes including the oracle-based optimal MBS and a random MBS. The optimal MBS decides the ideal (i.e., the best) channel/band from a centralized oracle as mentioned in section III. Thus, it can be regarded as a conventional technique, which searches all available channels and needs to have a good channel state information to choose the best one. This intuitively consumes a considerable amount of time. On the other hand, the random MBS arbitrarily selects one channel during each round t. Furthermore, a conventional MBS is considered for comparison in the conducted simulations. Then, we compare the energy-aware algorithms with energy consumption reflection during the channel selection process. The proposed algorithms are compared with EA-UCB-MBS, EA-TS-MBS, optimal MBS, conventional MBS, and random MBS for a comprehensive performance evaluation.

Without losing any generality, our simulation setup consists of four arms/ channels with three different bands (RF, mmWave, and VLC), i.e, $N = 4$. Each arm details in terms of $W_n, f_{c,n}, P^n_{\mathrm{Tx}}, T_{h,n}$ are (40MHz, 5.25GHz, 20mW, 3.6μsec), (20MHz, 2.4GHz, 20mW, 3.6μsec), (100MHz, 38GHz, 10mW, 0.28msec), and (20 MHz, 10^5GHz, 20mW, 3.6μsec), respectively. Furthermore, T, L_D, and T_D are set to 1000, 1TB, and 0.1 sec, respectively. The blocking parameters for the small blockers are considered to be $\alpha_{small} = 2.6$ and $\beta_{small} = 3$ while those for the large blockers are set to $\alpha_{large} = 3.6$ and $\beta_{large} = 7.7$, respectively, [11]. Ξ_{th} is set to be 1% x 10 to 100 meters, and x_0 is considered to be 5m. For the scenario without energy-awareness, the throughput and convergence are regarded as performance metrics. On the other hand, for the scenario with energy-awareness, the average energy-efficiency (EE) is additionally considered. Here, EE is expressed as the average throughput over the consumed energy per chosen band in bit/sec/Joule as follows:

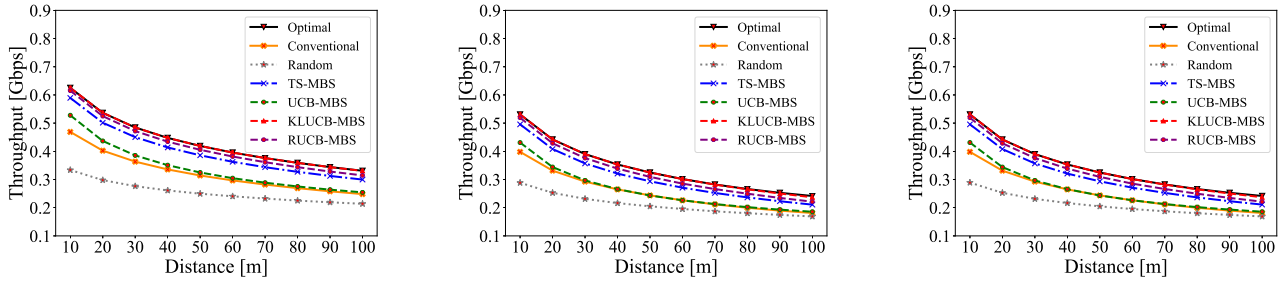
$$EE = \frac{1}{N} \sum_{n=1}^{N} \psi_n(T) \,/ \big(\Xi_{\mathrm{Tx},n}(t = 1) - \Xi_{\mathrm{Tx},n}(t = T)\big), \qquad (9)$$

where $\Xi_{\mathrm{Tx},n}(t = 1)$ denotes the initial energy (i.e., battery level) of Tx while $\Xi_{\mathrm{Tx},n}(t = T)$ represents its final energy after utilizing the $n_{th} \in 1, 2..., N$ channel/arm.

The throughput performance results for the scenario without energy-awareness are demonstrated in Fig. 1. The figure plots the average throughput over multiple runs for varying distances (in meters) in terms of three different blocking scenarios (no blockage, small blockage, and large blockage in Figs. 1(a), 1(b), 1(c), respectively). The throughput performance was found to be inversely related to the blocking. For all blockage cases, the proposed KLUCB-MBS and RUCB-MBS algorithms outperformed not only the reference UCB and TS schemes but also the conventional and random MBS schemes. While compared with the oracle-based optimal baseline, both our adopted approaches (i.e., KLUCB-MBS and RUCB-MBS) exhibited near-optimal performances. Also, the conventional MBS scheme was shown to offer a regular throughput performance due to the large overhead resulted from the brute-force beam training. Random MBS yielded the worst performance due to the arbitrary selection policy. In summary, for the scenario without energy-awareness, the KLUCB-MBS approach outperformed all other multi-band/channel selection strategies while RUCB-MBS emerged as the second best technique.

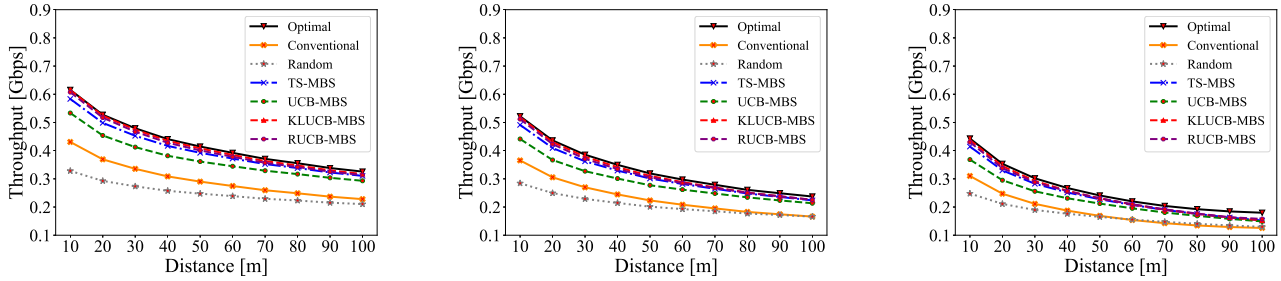
Next, for the scenario with energy-awareness, Fig. 2, demonstrates the throughput performance of the proposed MBS algorithms in case of various blocking situations. From the figure, it is evident that the throughput dropped with larger blocking environments. The two proposed EA-MBS variants exhibited superior performances to the conventional and random MBS schemes in all three blocking scenarios. The EA-KLUCB-MBS, EA-RUCB-MBS, EA-TS-MBS, and EA-UCB-MBS attained up to 99%, 98.5%, 92%, and 84% of the optimal MBS, respectively. However, the conventional MBS and random MBS delivered only 59% and 66% throughput performances, respectively. The random MBS policy revealed the most deficient performance with a sharp drop in the average throughput with

This is the author's version of an article that has been published in IEEE.
The final version of record is available at https://doi.org/10.1109/GLOBECOM46510.2021.9685996



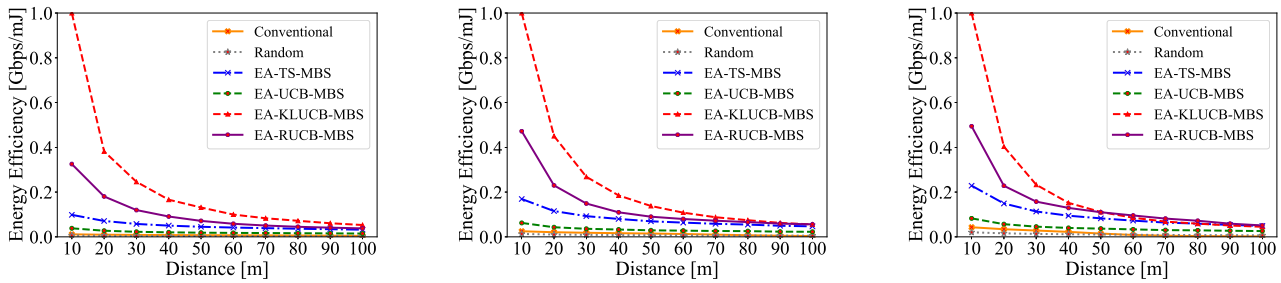
(a) No blockage. (b) Small blockage. (c) Large blockage.

Fig. 1. Comparison of RUCB-MBS and KLUCB-MBS against UCB-MBS, TS-MBS at distinct distances and different blocking types (no, small, large blocking).



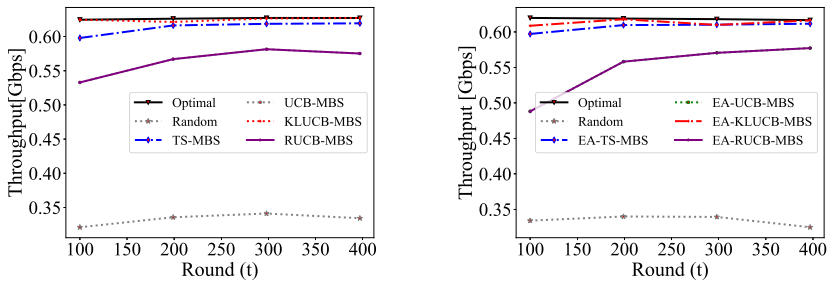
(a) No blockage. (b) Small blockage. (c) Large blockage.

Fig. 2. Average throughput comparison of EA-RUCB-MBS, EA-KLUCB-MBS algorithms Vs distinct distances at different blocking scenarios.



(a) No blockage. (b) Small blockage. (c) Large blockage.

Fig. 3. Energy-Efficiency (EE) comparison vs distinct distances at no, small, and large blocking situations.



(a) Scenario without energy-awareness and without blocking.

(b) Scenario with energy-awareness and without blocking.

Fig. 4. Convergence comparison for scenarios without and with energy-awareness for the considered MBSs.

the growing distance between the hybrid-band Tx-Rx pair. This happened due to the randomness in the chosen channel, which mostly encountered a weak channel state. The conventional MBS demonstrated the second worst throughput performance because of the increased overhead due to overall searching options required to determine the best channel for utilization.

For the scenarios with energy-awareness, we also evaluate the Energy-Efficiency (EE) performances in Fig. 3 in terms of the Gbps/mJ viability of our proposed EA-MBS algorithms across various hybrid-band Tx-Rx separation distances (in meters). Three blocking situations, i.e., no blocking, small blocking, and large blocking, are considered in Figs. 3(a), 3(b), and 3(c), respectively. The path loss is relatively related to the separation distance. Hence, the EE is impacted by the growing separation distance for all schemes. In all three considered blocking situations, the EA-KLUCB-MBS achieved a near-optimal EE performance due to its appropriate MBS policy with an adequate energy consumption. On the other hand, EA-

RUCB-MBS offered the second-best EE performance due to the efficiency of its randomized exploration parameter selection policy. The conventional and random MBS schemes were noticed to result in rather poor EE performances owing to their inadequate channel selection policies. From these results, it may be confirmed that our proposed EA-KLUCB-MBS and EA-RUCB-MBS algorithm(s) emerged as the most viable MBSs in the considered hybrid-band wireless system.

Next, we evaluate the convergence performances of the proposed and baseline MBSs in scenarios without and with energy-awareness. Fig. 4(a) demonstrates the convergence rate of the non-energy-aware MBS schemes (KLUCB-MBS, RUCB-MBS) with UCB-MBS, TS-MBS, optimal MBS, conventional MBS, and random MBS at 10 meters reference distance with $T = 1000$. KLUCB-MBS exhibited the fastest convergence while RUCB-MBS appeared to suffer from a much slower convergence similar to the classical UCB. This is because of the time taken by RUCB-MBS to define a suitable exploration parameter during every trial t. KLUCB-MBS scheme reaches 99% convergence compared to the optimal MBS. TS-MBS and RUCB/UCB-MBS achieved 98.78% and 93.54% of the optimal performance, respectively, at $t = 300$. Furthermore, for the energy-aware scenarios, the convergence performances of our proposed EA-RUCB-MBS and EA-KLUCB-MBS are shown in Fig. 4(b) in terms of the average throughput at 10 meters separation distance between hybrid-band Tx-Rx with $T = 1000$. The proposed EA-KLUCB-MBS technique outperforms the other methods by reflecting its quick convergence and fairly accurate performance. At $t = 300$, EA-KLUCB-MBS achieved a convergence rate of 99.98%, whereas EA-TS-MBS and EA-UCB-MBS achieved approximately 99.4% and 94.9% of the optimal performance, respectively. Moreover, the random MBS scheme could deliver only 53.65% of the optimal performance. EA-RUCB-MBS demonstrated a slower convergence to the legacy UCB due to the spent time in searching the best exploration parameter.

VI. Conclusion

As hybrid wireless communication systems operating at heterogeneous frequency bands emerge, we considered a fundamental problem of energy-efficient channel allocation in this paper. We described our system model based on the highly dynamic, unified channel conditions and formulated a resource allocation optimization problem that requires an oracle. Since an Oracle-based MBS strategy is impractical, we explained the need to design an online, Energy-Aware (EA)-MBS techniques. In this vein, the MBS problem is restructured as a stochastic MAB. In this paper, we investigate how to improve the performance of the UCB algorithm. Hence, we proposed an Energy-Aware Randomized UCB (EA-RUCB-MBS) algorithm which outperforms the original UCB and TS implementations. However, since EA-RUCB-MBS exhibits slow convergence, we developed another variant of UCB, referred to as Energy-Aware Kullback-Leibler UCB (EA-KLUCB-MBS), which demonstrates comparable performance with the optimal MBS solution with a much faster convergence. With extensive computer-based simulation results, we evaluated the near-optimal performance of the proposed EA-RUCB-MBS and EA-KLUCB-MBS algorithms. In the future, further algorithmic feasibility in terms of both complexity and a comprehensive log-regret analysis will be carried out in a theoretically manner to validate our simulation results and findings in this paper.

Acknowledgment

This work was supported by JSPS KAKENHI Grant Numbers JP19H04174 and JP21K14162, respectively.

References

[1] M. Ayyash*et al.*, "Coexistence of WiFi and LiFi toward 5G: concepts, opportunities, and challenges," *IEEE Communications Magazine*, vol. 54, no. 2, pp. 64–71, 2016, doi: 10.1109/MCOM.2016.7402263.

[2] M. Hammouda, S. Akın, A. M. Vegni, H. Haas, and J. Peissig, "Link selection in hybrid RF/VLC systems under statistical queueing constraints," *IEEE Transactions on Wireless Communications*, vol. 17, no. 4, pp. 2738–2754, 2018, doi: 10.1109/TWC.2018.2802937.

[3] P. Auer, N. Cesa-Bianchi, and P. Fischer, "Finite-time analysis of the multiarmed bandit problem," *Machine learning*, vol. 47, no. 2, pp. 235–256, 2002, doi: 10.1023/A:1013689704352.

[4] S. Vaswani, A. Mehrabian, A. Durand, and B. Kveton, "Old dog learns new tricks: Randomized UCB for bandit problems," in *Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics*, ser. Proceedings of Machine Learning Research, vol. 108. PMLR, 26–28 Aug. 2020, pp. 1988–1998.

[5] A. Garivier and O. Cappé, "The KL-UCB algorithm for bounded stochastic bandits and beyond," in *Proceedings of the 24th Annual Conference on Learning Theory*, ser. Proceedings of Machine Learning Research, vol. 19. PMLR, 9–11 Jun. 2011, pp. 359–376.

[6] S. Hashima, K. Hatano, H. Kasban, and E. Mahmoud 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, 2021.

[7] X. Gao, J. Zhang, G. Liu, D. Xu, P. Zhang, Y. Lu, and W. Dong, "Large-scale characteristics of 5.25 GHz based on wideband mimo channel measurements," *IEEE Antennas and Wireless Propagation Letters*, vol. 6, pp. 263–266, 2007, doi: 10.1109/LAWP.2007.897513.

[8] 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 *2018 IEEE 29th Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, 2018, doi: 10.1109/PIMRC.2018.8580695.

[9] 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 Transactions on Communications*, vol. 66, no. 5, pp. 2251–2265, 2018, doi: 10.1109/TCOMM.2017.2787990.

[10] H. Abuella, M. Elamassie, M. Uysal, Z. Xu, E. Serpedin, K. Qaraqe, and S. Ekin, "Hybrid RF/VLC systems: A comprehensive survey on network topologies, performance analyses, applications, and future directions," *ArXiv*, vol. abs/2007.02466, 2020.

[11] M. Boban, D. Dupleich, N. Iqbal, J. Luo, C. Schneider, R. Müller, Z. Yu, D. Steer, T. Jämsä, J. Li, and R. S. Thomä, "Multi-band vehicle-to-vehicle channel characterization in the presence of vehicle blockage," *IEEE Access*, vol. 7, pp. 9724–9735, 2019, doi: 10.1109/ACCESS.2019.2892238.

[12] S. Hashima, K. Hatano, E. Takimoto, and E. Mahmoud Mohamed, "Neighbor discovery and selection in millimeter wave D2D networks using stochastic MAB," *IEEE Communications Letters*, vol. 24, no. 8, pp. 1840–1844, 2020, doi: 10.1109/LCOMM.2020.2991535.

[13] 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, doi: 10.3390/s20143947.

[14] E. M. Mohamed, S. Hashima, K. Hatano, S. A. Aldossari, M. Zareei, and M. Rihan, "Two-hop relay probing in WiGig device-to-device networks using sleeping contextual bandits," *IEEE Wireless Communications Letters*, vol. 10, no. 7, pp. 1581–1585, 2021.

[15] S. Hashima, B. M. ElHalawany, K. Hatano, K. Wu, and E. M. Mohamed, "Leveraging machine-learning for D2D communications in 5G/Beyond 5G networks," *Electronics*, vol. 10, no. 2, 2021.