electronics

MDPI

Article

Cost-Aware Bandits for Efficient Channel Selection in Hybrid Band Networks

Sherief Hashima [1,2,*,†], Kohei Hatano [1,3,†], Mostafa M. Fouda [4,†], Zubair M. Fadlullah [5,6,†] and Ehab Mahmoud Mohamed [7,8,†]

1 Computational Learning Theory Team, RIKEN-Advanced Intelligence Project (AIP), Fukuoka 819-0395, Japan; hatano@inf.kyushu-u.ac.jp
2 Engineering Department, NRC, Egyptian Atomic Energy Authority, Cairo 13759, Egypt
3 Faculty of Arts and Science, Kyushu University, Fukuoka 819-0395, Japan
4 Department of Electrical and Computer Engineering, Idaho State University, Pocatello, ID 83209, USA; mfouda@ieee.org
5 Department of Computer Science, Lakehead University, Thunder Bay, ON P7B 5E1, Canada; zubair.fadlullah@lakeheadu.ca
6 Thunder Bay Regional Health Research Institute (TBRHRI), Thunder Bay, ON P7B 7A5, Canada
7 Electrical Engineering Department, College of Engineering, Prince Sattam Bin Abdulaziz University, Wadi Addwasir 11991, Saudi Arabia; ehab_mahmoud@aswu.edu.eg
8 Electrical Engineering Department, Faculty of Engineering, Aswan University, Aswan 81542, Egypt
* Correspondence: sherief.hashima@riken.jp; Tel.: +81-070-8531-1053
† These authors contributed equally to this work.

Abstract: Recently, hybrid band communications have received much attention to fulfil the exponentially growing user demands in next-generation communication networks. Still, determining the best band to communicate over is a challenging issue, especially in the dynamic channel conditions in multi-band wireless systems. In this paper, we manipulate a practical online-learning-based solution for the best band/channel selection in hybrid radio frequency and visible light communication (RF/VLC) wireless systems. The best band selection difficulty is formulated as a multi-armed bandit (MAB) with cost subsidy, in which the learner (transmitter) endeavors not only to increase his total reward (throughput) but also reduce his cost (energy consumption). Consequently, we propose two hybrid band selection (HBS) algorithms, named cost subsidy upper confidence bound (CSUCB-HBS) and cost subsidy Thompson sampling (CSTS-HBS), to efficiently handle this problem and obtain the best band with high throughput and low energy consumption. Extensive simulations confirm that CSTS-/CSUCB-HBS outperform the naive TS/UCB and heuristic HBS approaches regarding energy consumption, energy efficiency, throughput, and convergence speed.

Keywords: WiGig; MABs; cost subsidy; VLC; RF

Citation: Hashima, S.; Hatano, K.; Fouda, M.M.; Fadlullah, Z.M.; Mohamed, E.M. Cost-Aware Bandits for Efficient Channel Selection in Hybrid Band Networks. *Electronics* **2022**, *11*, 1782. https://doi.org/10.3390/electronics11111782

Academic Editors: Bouziane Brik, Junaid Ahmed Khan and Guangjie Han

Received: 1 May 2022
Accepted: 31 May 2022
Published: 3 June 2022

1. Introduction

Recently, hybrid band (HB) systems, mainly radio frequency/visible light communication (RF/VLC) systems, have received attention as an attractive paradigm to boost ultra-high-capacity wireless applications with enhanced connectivity via multi-band standards [1]. Nevertheless, those HB systems encounter fast channel fading, signal attenuation, acute blockage effects, etc. [2]. Still, the optimal hybrid band selection (HBS) of such systems is difficult to model due to the various multi-frequency bands with variable dynamic channel status. Additionally, HB systems should attain satisfying arrangements as soon as possible in order to maintain a high quality of service (QoS) in various application environments. Implementing efficient HBS is not straightforward because of the rapidly adapting environments where their performance relies on a plethora of parameters [3]. As a result, there is a strong direction in the wireless world to use machine learning (ML) instead of classic HBS protocols based on predefined rules or heuristics, where learning from the past is a powerful direction.

1.1. Related Work

Previous formulations of the HBS optimization problem targeting the selection of RF (Wi-Fi/WiGig) or VLC for every transmitter-receiver (Tx-Rx) pair have required full channel knowledge; besides, such formulations are computationally difficult. Former deep learning (DL)-aided solutions such as deep neural networks (DNNs) [4] and convolutional NNs (CNNs) [5] require huge training data sets, which might be hard to obtain till the roll-out of B5G/6G systems with HBs. In [6], a deep Q-learning algorithm is employed to solve the joint optimization problem in HB systems. In [5,7], the coauthors of this manuscript planned a DL-assisted prospective channel assignment framework to tackle the changeable channel environment difficulty in hybrid-band relay systems. However, they neglected the VLC band and considered RF bands only.

Current online-learning-enhanced solutions mostly influenced by reinforcement learning (RL) and multi-armed bandits (MAB) [8–12] mainly consider dynamic multi-channel access problems. The authors in [13] propose a distributed joint power allocation and channel assignment approach for improved power performance in D2D pairs. In addition, in [14,15], a context-aware transfer learning network selection technique for indoor RF-VLC systems was developed. However, previous work neither found an efficient practical solution nor considered the optimal HBS problem. MAB is an ML technique, where the learner tries to maximize the cumulative rewards and minimize the cost through online learning to handle the famous "exploitation–exploration compromise" [16]. Lately, bandits have been leveraged for vast wireless communication problems, including D2D communications [17–20], relay probing [21], aerial-enabled communications [22], wireless sensor networks (WSNs) [23], resource allocation [24,25], reconfigurable intelligent surface (RIS) enhanced communications [26,27], and millimeter Wave beamforming [28–30]. Furthermore, the authors of this paper recently contributed to the HBS problem using energy-aware MAB solutions [31–33] via subtracting the cost from the reward in the exploration part only. However, our main focus was to only handle the exploration–exploitation dilemma, which might cause more energy consumption to obtain the highest cumulative payoff. Hence, in this paper we extend the HBS problem to optimize both cost and payoff metrics and avoid causing exorbitant costs. Specifically, we anticipate the cost in both the exploitation and exploration parts of the bandit formula. This is more applicable especially in applications that have changed types of cost and payoff. Moreover, practical energy-aware HBS applications target both reward maximization (i.e., throughput) and cost minimization (battery life) due to their limited capacity.

1.2. Paper Contributions

In this paper, we attempt to manage the costs (energy consumption) of the HBS problem, where the learner is able to tolerate losing a small amount from the highest payoff (calculated as the payoff that the naive MAB approach could achieve in the absence of costs). This can be modeled as multi-objective optimization, which is important to investigate using online learning solutions, especially advanced proper MAB methods. This paper's contributions are highlighted as follows:

- We aim to relax/reformulate the HBS multi-objective optimization problem and obtain acceptable solutions (i.e., sub-optimal HBS decisions) in real time without prior channel knowledge using online learning techniques that easily handle blockage and energy consumption during the selection process.
- We reformulate the HBS optimization problem into a cost subsidy multi-armed bandit (CS-MAB) that accounts for the cost during selection in both exploitation and exploration terms.
- We propose CS—upper confidence bound (CSUCB-HBS) and CS—Thompson sampling (CSTS-HBS) algorithms and evaluate their performance compared with the ordinary MAB techniques (UCB and TS), and traditional HBS (i.e., conventional and random choice).

- Simulation results indicate the superior performance of our proposed CS-MAB techniques over classical MAB methods, especially CSTS-HBS, which exhibits better performance than CSUCB-HBS and others.

The paper is organized as follows: Section 2 discusses the considered HBS system model including the formulas of utilized channel models. Section 3 overviews the HBS optimization problem formulation. Section 4 discusses the proposed CS-HBS algorithms. Section 5 discusses the simulation results, followed by the paper's conclusion in Section 6.

2. System Model

Figure 1 presents the HB design under consideration, whereby a pair of Tx/Rx devices are presented for simplicity. The Tx/Rx pair might consist of base stations, mobile terminals or D2D relays, mounted by hybrid RF (i.e., WLAN/WiGig)/VLC frequency bands. The RF channel is IEEE 802.11ac/n (WiFi in 5.25 GHz and 2.4 GHz), and we leverage the linkage formula of [19]. The WiFi received power can be expressed as

$$P^F_{R_x}[dBm] = P^F_{T_x}[dBm] - P^F_{L_0} - 10\alpha^F \log_{10}(r) - \psi^F, \quad (1)$$

where $P^F_{R_x}$, $P^F_{T_x}$, and $P^F_{L_0}$ reflect the received power at Rx, the transmitted power from Tx, and the referenced path loss, respectively. α^F is the WiFi path loss exponent, while $\psi^F \backsim \mathcal{N}(0, \sigma^F)$ is the log-normal shadowing with zero mean and σ^F standard deviation. Tx–Rx separation distance is denoted by r.

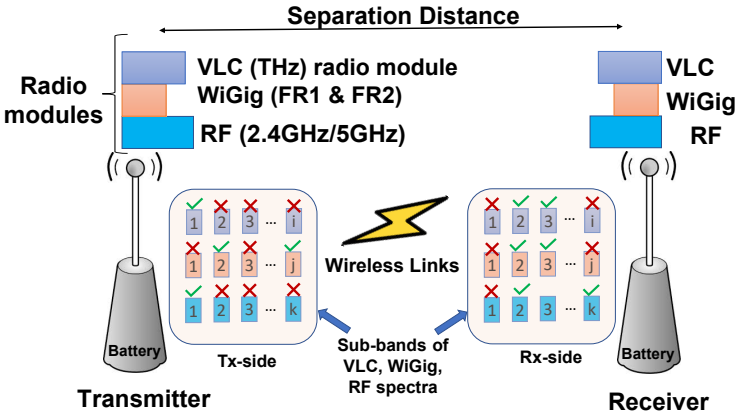


Figure 1. Hybrid band system model: How to self optimize hybrid channels in fluctuating channel conditions (distance, energy level, and blocking?).

For the WiGig linkage model, we leveraged the 38 GHz band [19]. The WiGig delivered power ($P^G_{R_x}$) with accounting both beamforming (BF) and blocking influences [19] formulated by

$$P^G_{R_x} = \mu(\mathbb{P}_{LOS}(r)) P^G_{T_x} \Lambda_{T_x} \Lambda_{R_x} / PL^G(r), \quad (2)$$

where $\mu(\mathbb{P}_{LOS}(r))$ is a random variable (RV) that follows Bernoulli distribution and reflects the blockage influence via proximity-dependent line-of-sight (LOS) likelihood denoted by $\mathbb{P}_{LOS}(r)$. $P^G_{T_x}$ is the Tx's WiGig transmitted power. The Tx and Rx BF gains are Λ_{T_x} and Λ_{R_x}, respectively. The distance-based path loss, $PL^G(r)$, is expressed in dB by:

$$10\log_{10}\left(PL^G(r)\right) = PL^G_0 + 10\alpha^G \log_{10}(r) + \chi^G, \quad (3)$$

where PL^G_0, α^G, and $\chi^{G,L} \backsim \mathcal{N}\left(0, \sigma^{2^G}\right)$ are the WiGig path loss at a benchmark distance r_0, WiGig path loss exponent, and WiGig-based log-normal shadowing having zero mean and variance of σ^{2^G}, respectively. We take into account only the LOS path because of its

dominant nature [34]. Λ_{T_x} and Λ_{R_x} are the 2D steerable antenna model with a Gaussian main loop profile, which can be written as [19]:

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

where θ, θ_{-3dB}, and Λ_0 reflect the azimuth angle, half power beamwidth, and the largest total radiated power by the antenna, respectively.

Regarding VLC linkage formulation, we employ the Lambertian model suitable for light-emitting diode (LED) transmitters [1], where the LoS channel gain for indoor VLC using Lambertian layouts is calculated as follows:

$$H_{VLC} = \frac{(g+1)A_R}{2\pi r^2} G_{T_x}(\phi_{vlc}) g(\phi_{vlc}) \cos^m \theta_{vlc} \cos \phi_{vlc} \forall \phi_{vlc} < \phi_c, \tag{5}$$

where A_R is the optical detector's physical area. θ_{vlc} and ϕ_{vlc} are the angles of incidence and radiance, respectively. ϕ_c identifies the Rx field vision's width and $(g = -\frac{ln(2)}{ln(\cos \theta_{vmax})})$ reflects the Lambertian model's order. The gains of the optical filter and concentrator are $G_{T_x}(\phi_{vlc})$ and $g(\phi_{vlc})$, respectively. Hence, the delivered optical power, $P_{R_x}^{vlc}$, is expressed as:

$$P_{R_x}^{vlc} = H_{vlc} P_{T_x}^{vlc}. \tag{6}$$

Finally, regarding the hybrid band blockage model, we leveraged the urban frequency-dependent blockage model given in [35]. This model is almost static, where the Tx is fixed and the Rx moves away and the blocker (small/large car) is located at a fixed distance from the Tx. Therefore, the generalized frequency-dependent blockage formulation can be expressed as [35]:

$$Blockingloss[dB] = \beta_b + \gamma_b \log(1 + \frac{f_{c,n}}{1 \text{ GHz}}), \tag{7}$$

where γ_b and β_b define the slope and the intercept of the linear relationship, and b reflects the blocker dimensions (i.e., small or large). To estimate the blocking loss, the above equation was utilized whilst changing the blocker size and the utilized frequency.

3. Problem Formulation

The aforementioned HBS problem in our HB system is formulated as follows:

$$\arg\max_{m,n} \sum_{m=1}^{M} \sum_{n=1}^{N_m} x_{mn} \mathbb{E}(\psi_{mn}(t)) \tag{8a}$$

$$s.t.$$

$$\arg\min_{m,n} \mathcal{C}_{T_x,mn}(t) \tag{8b}$$

$$\Xi_{T_x,mn}(t) > \Xi_{th}, \tag{8c}$$

$$\sum_{m=1}^{M} N_m \leq J \tag{8d}$$

$$\sum_{m=1}^{M} \sum_{n=1}^{N_m} x_{mn} = 1, \tag{8e}$$

$$x_{mn} \in \{0,1\}, \tag{8f}$$

where M defines the number of heterogeneous frequency bands and N_m is the channels count in each band m. J defines all the available channels over whole bands. $\psi_{mn}(t)$ is the Tx-Rx linkage throughput at time t using channel n of band m. $\Xi_{T_x,mn}(t)$ is the remaining energy (in Joules) of Tx at time t upon the used channel n of band m, and Ξ_{th} is the threshold

level energy at which the transmitter will not be capable of setting up wireless linkages and preserve its energy for its core purpose. x_{mn} is a decision variable for choosing the best band and its related channel that maximizes the throughput $\mathbb{E}(\psi_{mn}(t))$ expressed as:

$$\psi_{mn}(t) = \frac{B_{mn} T_D \Gamma_{mn}(t)}{U(t) T_{h,mn} + T_D}, \tag{9}$$

where T_D is the data transmission duration while $T_{h,mn}$ is the Tx-Rx time overhead due to the operating frequency. B_{mn} is the bandwidth, and $\Gamma_{mn}(t)$ is the spectral efficiency (SE) in bps/Hz upon the selected band/frequency at t, given by

$$\Gamma_{mn}(t) = \log_2(1 + \frac{P_{R_x}^{mn}(t)}{N_0 + I(t)}), \tag{10}$$

where $P_{R_x}^{mn}(t)$ is the Rx received power at time t due to the chosen band, N_0 is the Rx's noise power, and $I(t)$ is the interference from surrounding equipments that use that frequency. Only the interference from the two Wifi channels is considered here. The WiGig and VLC systems have negligible interference compared to the random noise as they are directional systems. As a result, there is a trade-off between investigating alternative frequency bands/channels and keeping the selected band to obtain the maximal throughput while reducing energy usage and so extending the battery lifespan. Such a trade-off can be addressed using sequential self-decision-making online learning algorithms. Therefore, we handle the above problem via reformulating it as a CS-MAB to not only deliver ultra-fast decisions under the influence of variability and dynamic blocking, but also to save the battery life of the transmitter. MAB is a confidence algorithmic approach that perfectly manipulates the exploration–exploitation compromise [16,17,29]. The main CS-MAB layout includes N probable actions (i.e., arms) to decide the best one within T trials. In each trial $t \in T$, selecting an arm, the learner fetches a payoff from the picked arm and consumes a cost due to the choice of that arm [16]. The player tries not only to maximize his reward but also to minimize his cost consumption. The energy consumption of Tx according to the selected band, $\mathcal{C}_{T_x,N_{MAB}^*}(t)$ given in (8b), is expressed as follows:

$$\mathcal{C}_{T_x,N_{MAB}^*}(t) = \frac{P_{T_x}^N L_D}{B_{N_{MAB}^*} \Gamma_{N_{MAB}^*}(t)}, \tag{11}$$

where *MAB* refers to the applied MAB technique (e.g., CSUCB, CSTS, UCB, and TS) and $C_{T_x,N_{MAB}^*}(t)$ is the consumed energy to transmit a datum of L_D bits with a transmission speed of $B_{N_{MAB}^*} \Gamma_{N_{MAB}^*}(t)$ bps.

The objective function (8a) contains several constraints. For each Tx, constraint (8b) indicates that the selected band should have low power consumption. Constraint (8c) states that Tx's standing energy should be greater than a certain level, Ξ_{th}. The following constraint (8d) indicates that the overall number of channels over all the frequency bands analyzed should be equal to or less than the observed accessible number of channels, J. Afterwards, constraints (8e) and (8f) indicate that a relay node can only choose one band and its associated channel at a time, and $x_{m,n}$ is a binary decision variable. We should always draw the arm with the lowest cost subsidized from a of the maximum mean reward $\mu_{m,n}$. Note that $0 \leq a \leq 1$. If $a = 0$, then the priority is the highest reward arm regardless cost and if $a = 1$, the priority will be the lower-cost arm regardless its reward value. Hence, the optimal multi-objective arm is the cheapest arm of the high-quality arms that achieves

$$\{m,n\}^* = \arg\max_{m,n} (1-a)\mu_{m,n} \; S.T \; \arg\min_{m,n} \mathcal{C}_{m,n}. \tag{12}$$

4. Envisioned CS-HBS Methods

In this section, we discuss our planned CSUCB-HBS and CSTS-HBS schemes. Both methods modify the naive UCB and TS stochastic bandit schemes, respectively, to be cost subsidy, where the Tx can tolerate a small loss from the largest payoff (subsidy, i.e., the amount of payoff the Tx may forgo to refine costs). Since UCB and TS are the most common stochastic MAB schemes, we present the cost subsidy concept of both of them to check their performance regarding the HBS issue.

4.1. CSUCB-HBS Algorithm

Our proposed CSUCB-HBS algorithm makes use of the UCB strategy, which provides optimism under unknown channel constrains through heterogeneous bands. Rather than using the same confidence bound for each arm in every trial, the UCB strategy selects the best arm via a conservative estimate [16]. Changed from the original UCB, CS-UCB first chooses the highest reward according to the subsidy parameter, and then investigates the cheapest arm from the feasibility set to play with it. This can be expressed as:

$$\mu_N^{score}(t) = \bar{\psi}_N(t) + \sqrt{\frac{2ln\ t}{\rho_{N,t}}}, \tag{13a}$$

$$m_t = \arg\max_N \mu_N^{score}(t), \tag{13b}$$

$$F(t) = \{N : \mu_N^{score}(t) - (1-a)\mu_{m_t}^{score}(t) \geq 0\} \tag{13c}$$

$$N^*_{CSUCB-HBS}(t) = \arg\min_{N \in F(t)} \mathcal{C}_N; \tag{13d}$$

where $\bar{\psi}_N(t)$ is the average throughput collected from the transmission band N until time t. $\rho_{N,t}$ is a counter for arm N if it has been drawn until time t. Equation (13a) is the original UCB equation and its arm selection m_t policy in Equation (13b) without arm cost consideration [16]. $F(t)$ is the feasibility set according to subsidy factor a from reward. Equation (13d) is the CSUCB-selected cheapest arm within $F(t)$ computed from (11).

4.2. CSTS-HBS Algorithm

Similar to CSUCB-HBS, the CSTS-HBS makes use of the TS algorithm [16] Bayesian strategy through prior Gaussian reward assumption. After defining the maximum reward arm, the feasibility set is constructed according to the subsidy value a, and then the cheapest arm is drawn from this set. This can be mathematically expressed as:

$$N^*_{TS-HBS}(t) = arg\max_N\{\theta_n(t)\}, \theta_N(t) \sim \mathbb{N}(\bar{\psi}_N(t), \frac{1}{\rho_{N,t}+1}), \tag{14a}$$

$$F(t) = \{N : \theta_N^{score}(t) - (1-a)\theta^{score}_{N^*_{TS-HBS}}(t) \geq 0\} \tag{14b}$$

$$N^*_{CSTS-HBS}(t) = \arg\min_{N \in F(t)} \mathcal{C}_N; \tag{14c}$$

where $\mathbb{N}(\bar{\psi}_N(t), \frac{1}{\rho_{N,t}+1})$ is a normal distribution with $\bar{\psi}_N(t)$ mean and $\frac{1}{\rho_{N,t}+1}$ variance.

Algorithm 1 outlines the key steps of our suggested online HBS algorithms. First, we try each arm once (i.e., channel) ($N = N_{ch}$) and fetch its payoff. If the conditions $(N_{ch}+1) < t < T$ and $E_{T_x} > E_{th}$ are not fulfilled, no more steps are completed and the algorithm terminates. On the other hand, it extracts a better payoff (channel index) $N^*_{MAB}(t)$ in the time trial $t \in T$ based on the CSUCB-HBS and CSTS-HBS policies, which select the lowest-cost arm from $(1-a)$ of the maximum reward, respectively. Afterwards, the bandit-based parameters are upgraded and the remaining energy levels of the transmitting device are estimated taking into account the drown band/channel. As a result, when the Tx decides to send new data frames to Rx, the process begins again.

Algorithm 1: CSUCB/CSTS-HBS Algorithms

Result: Channel N drawn every trail $t \in [T]$.
Input: $t = 0$, $\bar{\psi}_N(t) = 0$, $\rho_{n,t} = 0$, $\Xi_{th}, \Xi_{S,N}(t = 1)$, $1 \leq N \leq J$, $1 \leq t \leq T$.

```
for t ∈ [J] do
    I_t = t;
    play arm I_t and observe reward ψ_{I_t};
    ρ_i(t+1) = ρ_i(t) + 1 {I_t = i} ∀ i ∈ [J]
    update Ξ_{T_x,i}(t) = Ξ_{T_x,i}(t) − C_{T_x,i}(t), {I_t = i} ∀ i ∈ [J]
end
for t ∈ [J+1 : T] do
    μ̂_i(t) ← ψ_i(t)/ρ_i(t), β_i(t) ← sqrt(2log(T)/ρ_i(t)), UCB: μ_i^score(t) ← min{μ̂_i(t) + β_i(t)},
    TS: sample μ_i^score(t) from posterior distribution, m_t = arg max_i μ_i^score(t);
      F(t) = {i : μ_i^UCB(t) − (1 − a)μ_{m_t}^score(t) ≥ 0 ; I_t = arg min_{i∈F(t)} C_i;
    Play arm I_t then obtain reward ψ_{I_t}(t); ρ_i(t+1) = ρ_i(t) + 1 {I_t = i} ∀ i ∈ [J]
    update Ξ_{T_x,I_t}(t) = Ξ_{T_x,I_t}(t) − C_{T_x,I_t}(t)
end
```

5. Results

This section investigates the performance of our proposed CSUCB-HBS and CSTS-HBS algorithms compared to the original UCB/TS [16] schemes and traditional (e.g., optimal, conventional, and random) HBS schemes. The optimal HBS is achieved through the instantaneous selection of the best channel, while the conventional HBS scheme investigates all available bands first and then decides which one to link with. Finally, the random HBS approach selects a random band without any channel quality curiosity. Table 1 lists the utilized simulation parameters including blockage information. The average throughout, convergence rate, and the normalized energy consumption per selected band/channel in mjoule are the evaluation metrics.

Table 1. Simulation parameters.

Simulation Parameters		Value
Number of channels		4 (WiFi 2.4 GHz, 5.25 GHz, WiGig 38 GHz, VLC 10^5 GHz)
T, L_D, E_{th}		1000, 1 TB, 1%
Operating frequencies of each channel		5.25, 2.4, 38, 10^5 GHz
BW		40, 20, 40, 20 MHz
r		{10–100} m
Blocking model [35]	Small blocker: {length, width, height}	{5.07, 1.69, 1.93} m
	Large blocker: {length, width, height}	{7.01, 2.04, 2.63} m

Figure 2 presents the energy consumption against different values of the cost subsidy parameter a for both CSUC-HBS/CSTS-HBS. At low a values, the energy consumption is increased, especially the CSUCB-HBS scheme due to the main focus of the maximum reward. At higher values of a, the energy consumption is reduced according to the algorithm selection policy.

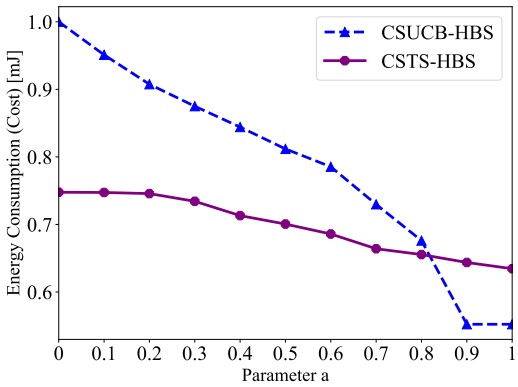
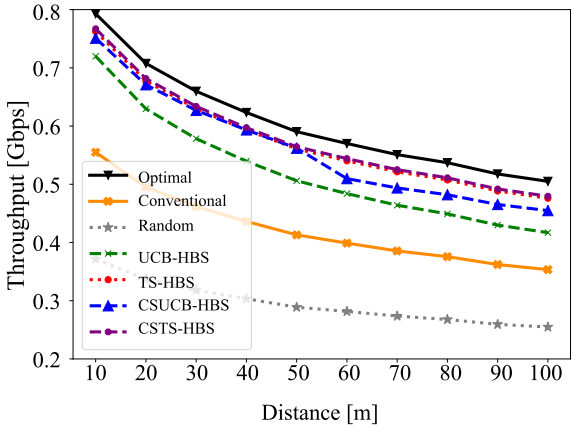


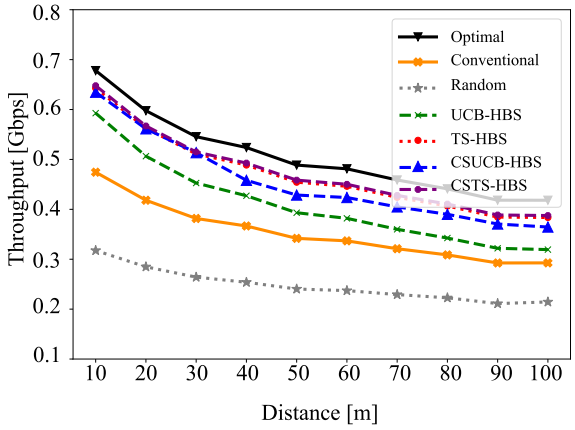
Figure 2. The energy consumption (cost) vs. subsidy parameter a at $r = 10$ and no blocking.

Figure 3 presents the throughput performance evaluation of our proposed CSUCB-HBS and CSTS-HBS algorithms at $a = 0.5$ with UCB, TS, optimal, conventional, and random HBS schemes at different r values at three changeable blockage types (i.e., none, small, and large). The attained reward, i.e., th average throughput, is reversely proportional to the blockage value which is apparent in the downward transition of the average throughput values due to blocked objects, i.e., no-blocking, small blocking, and large blocking, as illustrated in Figure 3a–c, respectively. Compared to the traditional and random HBS schemes, the envisioned CSUCB-HBS and CSTS-HBS techniques demonstrate promising performances (near optimum). The attained average throughput throughout all distance range is fairly close compared to the ideal case. The CSTS-HBS and CSUCB-HBS techniques deliver up to 99% and 80% average throughputs, respectively, within all separation distances values compared to the optimal case. Thus, CSTS-HBS arises as the best economic approach for the HBS difficulty due to not only the TS's Bayesian learning policy of TS but also its ability to save energy consumption. The second best performance is the CSUCB-HBS due to applying both upper-bound and cost subsidy policies. Nerveless, the random HBS attains only 60% of the average throughput, offering the worst performance due to the randomized selection policy without any considerations.

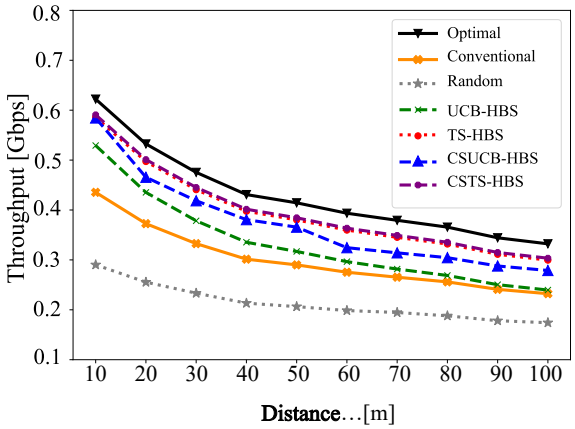


(**a**)

Figure 3. *Cont.*



(**b**)



(**c**)

Figure 3. Average throughput comparison of CSTS/CSUCB-HBS approaches vs. separation distances at distinct blocking layouts. (**a**) No blockage. (**b**) Small blockage. (**c**) Large blockage.

Figure 4 presents the energy consumption of the HBS schemes at no blocking and $a = 0.7$. Our proposed CSTS-HBS algorithm outperformed all the other schemes due to the proper channel selection strategy while minimizing the consumed energy of the transmitter. As r is incremented, the energy consumption of all traditional HBS methods are increased relatively with an increase in distance, whereas CSTS-HBS/CSUCB-HBS methods persisted with considerably superior low consumption performance. Still, the conventional HBS technique offers the highest energy consumption due to attempting all the available bands in every trial without energy awareness, opposite to our proposed MAB policies. Hence, the promising practical results enhance the affordability of the proposed CSTS-HBS then CSUCB-HBS schemes in HBS in B5G/6G systems.

The convergence curves of the proposed CSUCB-HBS and CSTS-HBS schemes are drawn in Figure 5 at $r = 10$ m. The results show that the suggested CSTS-HBS method exhibits a superior performance over all the trials t. As t approaches 400, the proposed CSTS-HBS scheme converges to 99.5% of the optimal throughput due to Bayesian and cost subsidy strategies. Furthermore, the CSUCB-HBS reaches 97.2% due to the upper bound and cost subsidy conceptualizations. This result demonstrates the practicality of our envisioned methods for HBS in B5G/6G systems with prolonging the battery life time.

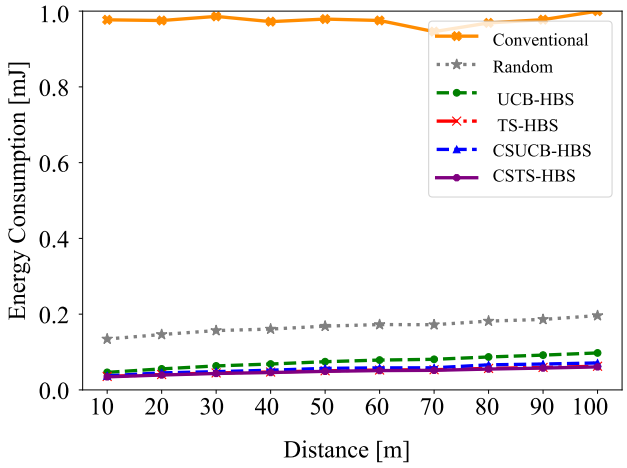


Figure 4. Energy consumption vs. r without considering blockage.

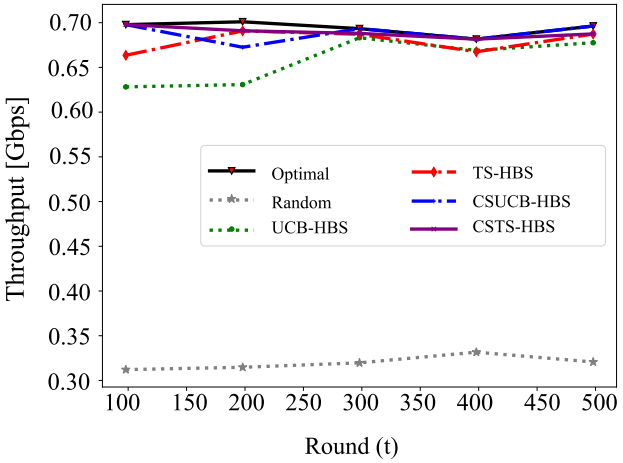


Figure 5. Convergence Rate Evaluation.

Finally, Figure 6 exhibits the viability of our envisioned HBS schemes in terms of energy efficiency (defined as the average payoff (throughput) over energy expenditure per chosen band in bit/sec/joule [32,33]) in Gbps/mJ over distinct Tx-Rx distances. Here, Figure 6a–c preview the energy efficiency performance for no-blocking, small blocking, and large blocking layouts, respectively. For all approaches, the energy efficiency is inversely related to the separation distance because of the path loss effect. For all the plotted blocking layouts, the CSTS-HBS method outperforms other approaches due to its better performance and appropriate channel selection policy, reflecting the lowest energy consumption. Meanwhile, TS-HBS, CSUCB-HBS, and UCB-HBS displayed encouraging energy efficiency performances over all separation distances, respectively. However, due to the traditional choice strategies, the random and conventional approaches show much worse performances due to randomization and attempting whole available bands, respectively. Therefore, these promising experimental results ensure that our suggested CSUCB/TS-HBS approaches, especially CSTS-HBS, are the most feasible HBS methods for multi-band wireless communication networks.

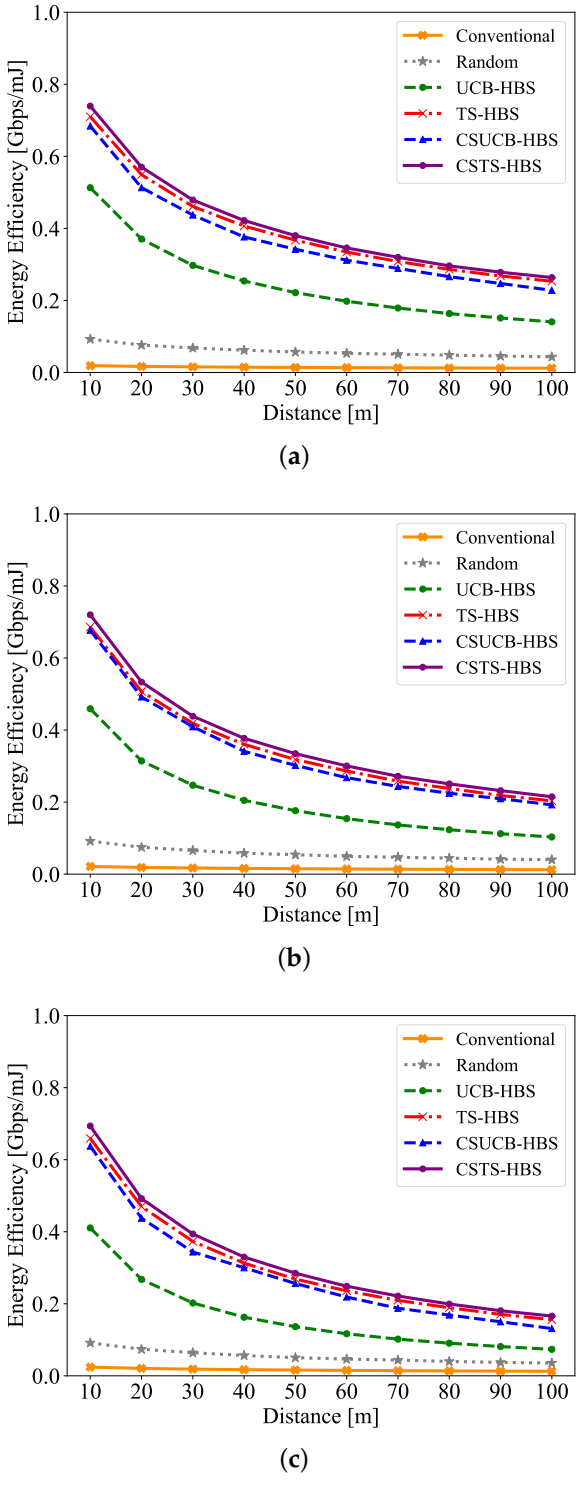


Figure 6. Energy efficiency performance of CSTS/CSUCB-HBS approaches vs. separation distances at distinct blocking layouts. (**a**) No blockage. (**b**) Small blockage. (**c**) Large blockage.

6. Conclusions and Outlook

As heterogeneous band wireless communication systems are being introduced, we examined the significant problem of efficient, realistic online channel allocation in this paper. Furthermore, we emphasize the necessity to design an online HBS technique. Hence, the HBS problem is restructured as a stochastic cost subsidy MAB, where each band selection consumes energy from the player/Tx. This paper investigated a realistic solution to the optimal band decision problem in hybrid (RF/VLC) systems by introducing two cost

subsidy MAB algorithms, i.e., CSUCB, and CSTS-HBS, which outperform the primary UCB and TS implementations, respectively. Especially, CSTS-HBS confirmed not only the near-optimal performance but also a faster convergence. The results of our extensive simulation highlight the near-optimal performance of the advised CSTS-HBS algorithm. Future work includes a multi-Tx-Rx scenario extension and further real data investigations throughout hybrid network analysis.

Author Contributions: All authors contributed equally in this paper. All authors have read and agreed to the published version of the manuscript.

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

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Acknowledgments: This work was supported by JSPS KAKENHI Grant Numbers JP19H04174 and JP21K14162.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Abuella, H.; Elamassie, M.; Uysal, M.; Xu, Z.; Serpedin, E.; Qaraqe, K.A.; Ekin, S. Hybrid RF/VLC Systems: A Comprehensive Survey on Network Topologies, Performance Analyses, Applications, and Future Directions. *IEEE Access* **2021**, *9*, 160402–160436. [CrossRef]
2. Chen, Y.; Ai, B.; Niu, Y.; He, R.; Zhong, Z.; Han, Z. Resource Allocation for Device-to-Device Communications in Multi-Cell Multi-Band Heterogeneous Cellular Networks. *IEEE Trans. Veh. Technol.* **2019**, *68*, 4760–4773. [CrossRef]
3. Mughal, B.; Fadlullah, Z.M.; Fouda, M.M.; Ikki, S. Allocation Schemes for Relay Communications: A Multiband Multichannel Approach Using Game Theory. *IEEE Sens. Lett.* **2022**, *6*, 7500104. [CrossRef]
4. Najla, M.; Mach, P.; Becvar, Z. Deep Learning for Selection Between RF and VLC Bands in Device-to-Device Communication. *IEEE Wirel. Commun. Lett.* **2020**, *9*, 1763–1767. [CrossRef]
5. Sakib, S.; Tazrin, T.; Fouda, M.M.; Fadlullah, Z.M.; Nasser, N. A Deep Learning Method for Predictive Channel Assignment in Beyond 5G Networks. *IEEE Netw.* **2021**, *35*, 266–272. [CrossRef]
6. Shrivastava, S.; Chen, B.; Chen, C.; Wang, H.; Dai, M. Deep Q-Network Learning Based Downlink Resource Allocation for Hybrid RF/VLC Systems. *IEEE Access* **2020**, *8*, 149412–149434. [CrossRef]
7. Sakib, S.; Tazrin, T.; Fouda, M.M.; Fadlullah, Z.M.; Nasser, N. An Efficient and Light-weight Predictive Channel Assignment Scheme for Multi-Band B5G Enabled Massive IoT: A Deep Learning Approach. *IEEE Internet Things J.* **2021**, *8*, 5285–5297. [CrossRef]
8. Bakri, S.; Brik, B.; Ksentini, A. On using reinforcement learning for network slice admission control in 5G: Offline vs. online. *Int. J. Commun. Syst.* **2021**, *34*, 1987–2007. [CrossRef]
9. Nasir, Y.S.; Guo, D. Multi-Agent Deep Reinforcement Learning for Dynamic Power Allocation in Wireless Networks. *IEEE J. Sel. Areas Commun.* **2019**, *37*, 2239–2250. [CrossRef]
10. Wang, S.; Lv, T. Dynamic Multichannel Access for 5G and Beyond with Fast Time-Varying Channel. In Proceedings of the ICC 2020—2020 IEEE International Conference on Communications (ICC), Dublin, Ireland, 7–11 June 2020; pp. 1–6. [CrossRef]
11. Wang, Z.; Zhang, T.; Liu, Y.; Xu, W. Caching Placement and Resource Allocation for AR Application in UAV NOMA Networks. In Proceedings of the GLOBECOM 2020—2020 IEEE Global Communications Conference, Taipei, Taiwan, 7–11 December 2020; pp. 1–6. [CrossRef]
12. Hashima, S.; Fadlullah, Z.M.; Fouda, M.M.; Mohamed, E.M.; Hatano, K.; ElHalawany, B.M.; Guizani, M. On Softwarization of Intelligence in 6G Networks for Ultra-Fast Optimal Policy Selection: Challenges and Opportunities. *IEEE Netw.* **2022**, 1–9. [CrossRef]
13. Yin, R.; Wu, Z.; Liu, S.; Wu, C.; Yuan, J.; Chen, X. Decentralized Radio Resource Adaptation in *D2D-U* Networks. *IEEE Internet Things J.* **2021**, *8*, 6720–6732. [CrossRef]
14. Du, Z.; Wang, C.; Sun, Y.; Wu, G. Context-Aware Indoor VLC/RF Heterogeneous Network Selection: Reinforcement Learning With Knowledge Transfer. *IEEE Access* **2018**, *6*, 33275–33284. [CrossRef]
15. Wang, C.; Wu, G.; Du, Z.; Jiang, B. Reinforcement learning based network selection for hybrid VLC and RF systems. In *MATEC Web of Conferences*; EDP Sciences: Les Ulis, France, 2018; Volume 173, p. 03014. [CrossRef]
16. Lattimore, T. *Bandit Algorithms*; Cambridge University Press: Cambridge, UK, 2020.
17. Hashima, S.; ElHalawany, B.M.; Hatano, K.; Wu, K.; Mohamed, E.M. Leveraging Machine-Learning for D2D Communications in 5G/Beyond 5G Networks. *Electronics* **2021**, *10*, 169. [CrossRef]

18. Hashima, S.; Hatano, K.; Takimoto, E.; Mohamed, E.M. Neighbor Discovery and Selection in Millimeter Wave D2D Networks Using Stochastic MAB. *IEEE Commun. Lett.* **2020**, *24*, 1840–1844. [CrossRef]
19. Hashima, S.; Hatano, K.; Kasban, H.; Mohamed, E.M. Wi-Fi Assisted Contextual Multi-Armed Bandit for Neighbor Discovery and Selection in Millimeter Wave Device to Device Communications. *Sensors* **2021**, *21*, 2835. [CrossRef]
20. Hashima, S.; Hatano, K.; Takimoto, E.; Mohamed, E.M. Minimax Optimal Stochastic Strategy (MOSS) For Neighbor Discovery and Selection In Millimeter Wave D2D Networks. In Proceedings of the 2020 23rd International Symposium on Wireless Personal Multimedia Communications (WPMC), Okayama, Japan, 19–26 October 2020; pp. 1–6. [CrossRef]
21. Mohamed, E.M.; Hashima, S.; Hatano, K.; Aldossari, S.A.; Zareei, M.; Rihan, M. Two-Hop Relay Probing in WiGig Device-to-Device Networks Using Sleeping Contextual Bandits. *IEEE Wirel. Commun. Lett.* **2021**, *10*, 1581–1585. [CrossRef]
22. Mohamed, E.M.; Hashima, S.; Aldosary, A.; Hatano, K.; Abdelghany, M.A. Gateway Selection in Millimeter Wave UAV Wireless Networks Using Multi-Player Multi-Armed Bandit. *Sensors* **2020**, *20*, 3947. [CrossRef]
23. Hashima, S.; Mohamed, E.M.; Hatano, K.; Takimoto, E. WiGig Wireless Sensor Selection Using Sophisticated Multi Armed Bandit Schemes. In Proceedings of the 2021 Thirteenth International Conference on Mobile Computing and Ubiquitous Network (ICMU), Tokyo, Japan, 17–19 November 2021; pp. 1–6. [CrossRef]
24. Barrachina-Muñoz, S.; Chiumento, A.; Bellalta, B. Multi-Armed Bandits for Spectrum Allocation in Multi-Agent Channel Bonding WLANs. *IEEE Access* **2021**, *9*, 133472–133490. [CrossRef]
25. Zuo, J.; Joe-Wong, C. Combinatorial Multi-armed Bandits for Resource Allocation. In Proceedings of the 2021 55th Annual Conference on Information Sciences and Systems (CISS), Baltimore, MD, USA, 24–26 March 2021. [CrossRef]
26. Mohamed, E.M.; Hashima, S.; Hatano, K. Energy Aware Multi-Armed Bandit for Millimeter Wave Based UAV Mounted RIS Networks. *IEEE Wirel. Commun. Lett.* **2022**. [CrossRef]
27. Mohamed, E.M.; Hashima, S.; Hatano, K.; Aldossari, S.A. Two-Stage Multiarmed Bandit for Reconfigurable Intelligent Surface Aided Millimeter Wave Communications. *Sensors* **2022**, *22*, 2179. [CrossRef]
28. Mohamed, E.M.; Hashima, S.; Hatano, K.; Kasban, H.; Rihan, M. Millimeter-Wave Concurrent Beamforming: A Multi-Player Multi-Armed Bandit Approach. *Comput. Mater. Contin.* **2020**, *65*, 1987–2007. [CrossRef]
29. ElHalawany, B.M.; Hashima, S.; Hatano, K.; Wu, K.; Mohamed, E.M. Leveraging Machine Learning for Millimeter Wave Beamforming in Beyond 5G Networks. *IEEE Syst. J.* **2021**, 1–12. [CrossRef]
30. Hashima, S.; Hatano, K.; Kasban, H.; Rihan, M.; Mohamed, E.M. Multiagent Multi-Armed Bandit Techniques for Millimeter Wave Concurrent Beamforming. In Proceedings of the 2020 8th International Japan-Africa Conference on Electronics, Communications, and Computations (JAC-ECC), Virtual, 14–15 December 2020; pp. 56–59. [CrossRef]
31. Fouda, M.; Hashima, S.; Sakib, S.; Fadlullah, Z.; Hatano, K.; Shen, X. Optimal Channel Selection in Hybrid RF/VLC Networks: A Multi-Armed Bandit Approach. *IEEE Trans. Veh. Technol.* **2022**. [CrossRef]
32. Hashima, S.; Fouda, M.M.; Sakib, S.; Fadlullah, Z.M.; Hatano, K.; Mohamed, E.M.; Shen, X. Energy-Aware Hybrid RF-VLC Multi-Band Selection in D2D Communication: A Stochastic Multi-Armed Bandit Approach. *IEEE Internet Things J.* **2022**. [CrossRef]
33. Hashima, S.; Fouda, M.M.; Fadlullah, Z.M.; Mohamed, E.M.; Hatano, K. Improved UCB-based Energy-Efficient Channel Selection in Hybrid-Band Wireless Communication. In Proceedings of the 2021 IEEE Global Communications Conference (GLOBECOM), Madrid, Spain, 7–11 December 2021; pp. 1–6. [CrossRef]
34. Wei, N.; Lin, X.; Zhang, Z. Optimal Relay Probing in Millimeter-Wave Cellular Systems with Device-to-Device Relaying. *IEEE Trans. Veh. Technol.* **2016**, *65*, 10218–10222. [CrossRef]
35. Boban, M.; Dupleich, D.; Iqbal, N.; Luo, J.; Schneider, C.; Müller, R.; Yu, Z.; Steer, D.; Jämsä, T.; Li, J.; et al. Multi-Band Vehicle-to-Vehicle Channel Characterization in the Presence of Vehicle Blockage. *IEEE Access* **2019**, *7*, 9724–9735. [CrossRef]