sensors

MDPI

Article

Online Learning Approach for Predictive Real-Time Energy Trading in Cloud-RANs

Wan Nur Suryani Firuz Wan Ariffin [1,*], Xinruo Zhang [2], Mohammad Reza Nakhai [3] and Hasliza A. Rahim [1,4,*] and R. Badlishah Ahmad [1,5]

[1] Faculty of Electronic Engineering Technology, Universiti Malaysia Perlis, Arau 02600, Malaysia; badli@unimap.edu.my
[2] School of Computer Science and Electronic Engineering, University of Essex, Wivenhoe Park, Colchester CO4 3SQ, UK; xinruo.zhang@essex.ac.uk
[3] Department of Informatics, Centre for Telecommunications Research, King's College London, Aldwych WC2B 4BG, UK; reza.nakhai@kcl.ac.uk
[4] Advanced Communication Engineering, Centre of Excellence (ACE), Universiti Malaysia Perlis, Kangar 01000, Malaysia
[5] Advanced Computing, Centre of Excellence (AdComp), Universiti Malaysia Perlis, Arau 02600, Malaysia
* Correspondence: suryanifiruz@unimap.edu.my (W.N.S.F.W.A.); haslizarahim@unimap.edu.my (H.A.R.)

Abstract: Constantly changing electricity demand has made variability and uncertainty inherent characteristics of both electric generation and cellular communication systems. This paper develops an online learning algorithm as a prescheduling mechanism to manage the variability and uncertainty to maintain cost-aware and reliable operation in cloud radio access networks (Cloud-RANs). The proposed algorithm employs a combinatorial multi-armed bandit model and minimizes the long-term energy cost at remote radio heads. The algorithm preschedules a set of cost-efficient energy packages to be purchased from an ancillary energy market for the future time slots by learning both from cooperative energy trading at previous time slots and by exploring new energy scheduling strategies at the current time slot. The simulation results confirm a significant performance gain of the proposed scheme in controlling the available power budgets and minimizing the overall energy cost compared with recently proposed approaches for real-time energy resources and energy trading in Cloud-RANs.

Keywords: cloud radio access network; combinatorial multi-armed bandit; online learning; energy trading

Citation: Wan Ariffin, W.N.S.F.; Zhang, X.; Nakhai, M.R.; Rahim, H.A.; Ahmad, R.B. Online Learning Approach for Predictive Real-Time Energy Trading in Cloud-RANs. *Sensors* **2021**, *21*, 2308. https://doi.org/10.3390/sensors210723087

Academic Editor: Maryline Chetto

Received: 8 December 2020
Accepted: 5 February 2021
Published: 25 March 2021

1. Introduction

Denser site deployment has been contemplated as a key enabling technology that can support the mushrooming of mobile data traffic and meet the demands of high-data-rate communications for next-generation wireless communication networks [1]. In contrast, conventional base stations (BSs) consume 80% of the electricity [2], as, in the BSs, all the radio and baseband processing functions are coordinated in the second-generation (2G) radio access network (RAN) architecture. Subsequently, the radio and baseband processing functions are divided into two separate nodes, i.e., remote radio head (RRH) and baseband processing unit (BBU), in the development of the third-generation (3G) and fourth-generation (4G) distributed radio access network (Distributed-RAN) architecture. Nevertheless, Distributed-RAN is incompetent in dealing with tremendous growths in data traffic to deliver high-bandwidth, low-latency, and cost-efficient services [3], and is incapable of supporting the demands of the quality of expectation (QoE) and quality of service (QoS) [4] for the fifth-generation (5G) of mobile communication systems. Cloud radio access networks (Cloud-RANs) have been regarded as a promising solution, owing to their superiority in reducing the capital expenditure (CAPEX) and operational expenditure (OPEX) of the network operators with the centralization and cloudification of BBUs and

their corresponding RRHs. Cloud-RANs can solve the limitations of the Distributed-RAN architecture in terms of expanding the network scalability, simplifying network management and maintenance, optimizing network performance, reducing energy consumption, and enhancing spectrum efficiency [3]. In a Cloud-RAN architecture, the conventional BSs are physically detached into two parts: BBUs, which are grouped as a cloud processing unit (CU) for designing all coordination and energy trading strategies, and the remaining RRHs, which are in charge of all radio frequency (RF) operations [5]. Even beamforming is designed in the CU; RRHs consume an enormous portion of electricity to amplify and transmit RF signals to users in order to satisfy their data-rate and energy requirements. However, due to the large number of densely deployed RRHs, with each serving a time-varying number of users in a highly dynamic wireless environment, the amount of energy demanded by the wireless network operators from the energy generation (EG) plants will be highly variable and statistically unknown over different times of the day. Equipping the RRHs with green energy technology that harvests energy from natural sources, such as wind and sunlight, to power next-generation mobile communication networks can significantly contribute to the reduction of the global carbon footprint [6]. However, the uncertain nature of renewable energy supply coupled with dynamic user energy demand necessitates the integration of green energy supply with the conventional grid to maximally benefit the network operator [7–15]. These random variations in electricity demand increase the OPEX of the energy generation process because the EG plants must maintain an instantaneous balance between the aggregate demand for electricity and the total power generated as a whole [7]. Hence, the operators need to routinely control the operation of the wireless network based on the well-known operating characteristics of the conventional EG plants. Deviation from the operating points of the EG plants to provide compensating variations in order to maintain the balance increases the total OPEX of the EG plants, which will, in turn, reflect on the OPEX of the network operators.

The operational time frame of the grid can be generally divided into regulation, load following, and unit commitment. During each one of these time frames, suitable reserved energy sources are dispatched to correct the imbalance between the generation and the demand. The EG sources reserved for load following, which are deployed on a slower time scale than the regulating frames, are used to accommodate for causes of variability and uncertainty, e.g., due to traffic energy demand and renewable energy generation, during the regular operation of the grid. Although the ramping and energy needed to follow the variations and uncertainties can be supplied by the ancillary energy markets, the insufficient ramping capability of the base low-cast conventional power plants can significantly inflate the price of the energy dispatched by expensive peaking EG units with fast ramp rates. Using conventional regulation units to compensate for uncertain abrupt ramps in energy demand is among the most expensive services. Hence, efficient control mechanisms are required to be developed for the flexibility in the EG fleet in order to maintain their cost-aware reliable operation under variability and uncertainty.

This paper focuses on designing an intelligent control mechanism for the steep ramps in energy demand in wireless cellular networks to minimize the long-term energy cost. We introduce an online learning approach for price-aware energy procurement at RRHs by supplying the load-following EG reserves within advance energy trading offers based on possible forthcoming variations and uncertainties in the energy demand. As the energy demand varies from low to peak values during different hours, the proposed strategy is designed to avoid paying for high peak-time energy cost by purchasing energy in advance at a lower off-peak price to reduce the OPEX. The proposed approach anticipates the future energy demand (surplus) at each RRH and prepares for purchasing (selling) the energy from the hour-ahead/day-ahead market (to the grid) before the actual demand (surplus) emerges. In this way, the EG units will have more time to regulate their electricity generation process according to the demand with slower ramp rates and, consequently, at lower prices.

1.1. Related Works

The authors in [7] first investigated the energy efficiency problem in a coordinated multipoint (CoMP) system powered by a smart gird. They formed the problem formulation for the proposed system as a simplified two-level Stackelberg game and concluded that such a design significantly reduces the OPEX. Equipping the end-user with renewable energy devices and accounting for the varying electricity price, the authors in [8] developed an energy trading algorithm to maximally benefit the network operator while satisfying the energy demand of end-users in a grid/renewable energy hybrid network. To take advantage of two-way energy trading with the grid and cooperative transmission, the authors in [9] proposed an aggregator-aided joint communication and energy cooperation strategy in the CoMP networks powered by both grid and renewable energy. In [10], the authors designed a joint real-time energy trading and cooperative transmission mechanism based on convex optimization techniques in a smart-grid-powered CoMP system. In [11], the authors studied energy trading in a more general setting, including trading among a set of storage units and the grid from the perspective of noncooperative game theory, and they proposed an algorithm that achieves at least one Nash equilibrium point. By assuming the availability of varying hourly profiles of the energy demand of base stations and renewable generation as well as the day-ahead knowledge of hourly varying electricity prices, the authors of [12] minimized the electricity bill in cellular base stations powered jointly by a smart grid and locally harvested solar energy. The authors of [13] integrated the CoMP system with a simultaneous wireless information and power transfer (SWIPT) concept and proposed a joint energy trading and partial cooperation design based on sparse beamforming, accounting for limited-capacity backhaul links in a green Cloud-RAN by minimizing the instantaneous energy cost without integrating reinforcement learning. The authors of [14] investigated the optimal power flow problem for smart micro-grids in a distributed manner and adopted an alternating direction method of multipliers to ensure the global optimum of the semidefinite programming (SDP) problem. It can be perceived that an abstract idea of the combinatorial multi-armed bandit (CMAB) approach was firstly tackled in [15] by introducing two iterative energy trading algorithms to search for a set of cost-efficient energy packages in ascending and descending order of package sizes and assuming invariability of wireless channel circumstances. Consequently, the study in [16] proposed a CMAB approach for energy trading in the cellular network to support the unpredictable wireless channel conditions to further lessen the total energy cost over a finite time horizon.

1.2. Main Contributions

This paper's main contributions to real-time energy resource and energy trading in Cloud-RAN environments are summarized as follows:

- A joint energy trading and clustering technique to account for limited-capacity backhaul links in a green Cloud-RAN with a SWIPT system was proposed in [13]. However, their design was based on myopic optimization of semidefinite programming (SDP) (i.e., minimizing the instantaneous energy cost for the current time only) without any learning process for future demand provisioning. Furthermore, their proposed design cannot cope with the time-varying system dynamics, since they considered no temporal dynamic of the energy demand and cost over time and provided no solution for the look-ahead energy purchase decisions.
- In contrast to [15], this paper develops a combinatorial upper confidence bound (CUCB) algorithm as a prescheduling mechanism to maintain cost-aware reliable operation in CRANs to handle the variability and uncertainty of both the electrical generation and the intrinsic characteristics of the cellular communication system. This paper predicts the best possible combination of energy packages to be purchased for the next time slot by exploring the rewards of new combinations of energy packages within given trials at the current time slot and exploiting the past captured informa-

tion on rewards of super arms from the previous time slots to optimize long-term averaged rewards.

- Differently from the system model proposed in [16], this paper considers a downlink Cloud-RAN with SWIPT, where the RRHs concurrently transfer satisfied data beams to information users and requested energy beams to active energy users. Furthermore, this paper also integrates a sparse beamforming technique to iteratively remove the cooperative links between the RRHs and the active information users based on the renewable power budgets and front-haul link capacity limitations at the individual RRHs. The clustering technique has been confirmed to enhance energy efficiency and decrease the total energy cost of the RRH in pragmatic Cloud-RANs [17]. In contrast to their CMAB approach, this paper estimates the imminent energy demands by dynamically deciding on an optimal set of super arms by exploring all of the possible minimal combinatorial energy packages to be purchased from the day-ahead market, thus diminishing the risk of regret factors.

This work's novel contribution is the development of a sequential learning algorithm that adaptively tracks the temporal variations of energy demands and makes predictive decisions on look-ahead energy purchases in dynamically changing environments with unknown statistics to asymptotically minimize the time-averaged overall energy cost in the long run. The proposed algorithm anticipates the future energy demands of the distributed RRHs in the Cloud-RAN and schedules these demands by invoking the various power plants well in advance so that higher energy prices at peak demand times are curtailed. The proposed algorithm does not require any other description of usage patterns or statistical distribution of stochastic events. It performs foresighted optimization based on online learning during the operation. It only uses the past captured data on averaged accumulated rewards for predicting the energy consumption at the next period based on the proposed strategy.

1.3. Organization and Notations

The rest of this paper is structured as follows. The system model for the downlink Cloud-RAN with SWIPT and the energy management model are introduced in the Section 2. In Section 3, the problem of real-time collaborative energy trading at an individual time frame is formulated and then transformed into a numerically tractable form. The predictive energy trading strategy is proposed in Section 4. Numerical simulation results are interpreted in Section 5. Finally, Section 6 summarizes the proposed work.

Notation 1. *w, $\mathbf{w}$, and $\mathbf{W} \succeq 0$, respectively, denote a scalar w, a vector $\mathbf{w}$, and a positive semidefinite matrix $\mathbf{W}$. $\mathbb{C}^{n\times m}$, $(.)^H$, $tr(.)$, and $\mathbb{E}$ indicate the sets of n-by-m dimensional complex matrices, the complex conjugate transpose operators, the trace operators, and the expected value, respectively. $\|.\|_p$ represents the ℓ_p-norm of a vector and $\|.\|_0$ denotes the number of non-zero entries in the vector. Notice that the duration of a time frame is normalized to one and the normalized energy unit, i.e., Js^{-1}, is assumed in this paper. Therefore, in this paper, the terms "power" and "energy" are mutually interchangeable.*

2. System Model

Consider a downlink transmission Cloud-RAN with SWIPT from N RRHs towards K_i information users (IUs), K_e active energy users (EUs), and $K_e^{[\text{idle}]}$ idle EUs, respectively, over a shared bandwidth. Notice that the active EUs located within the energy-serving area of an RRH can exploit the energy-carrying signals directly from that particular RRH. In contrast, the idle EUs located outside any energy-serving area of the RRHs can only scavenge energy from the ambient radio frequency signals for self-sustainability [13]. Each RRH is equipped with M antennas, and the individual IUs and EUs have one single antenna. Based on perfect knowledge of channel state information (CSI), the CU coordinates all the resource management and energy trading strategies for the RRHs and administers all the IUs' data to the corresponding RRHs finite-capacity front-haul links. Remark that,

under the perfect CSI assumption, all the channel properties of the downlink Cloud-RAN communication links, i.e., path loss, scattering, fading, shadowing, etc., are assumed to be perfectly known at both the IU and EU terminals.

Let $\mathcal{L}_b = \{1, \cdots, N\}$, $\mathcal{L}_i = \{1, \cdots, K_i\}$, $\mathcal{L}_e = \{1, \cdots, K_e\}$, and $\mathcal{L}_e^{[\text{idle}]} = \{1, \cdots, K_e^{[\text{idle}]}\}$ denote, respectively, the set of indexes of the RRHs, the IUs, the active EUs, and the idle EUs. The amount of energy flow in this paper depends on the data-rate requirements by the IUs, the wireless energy transfer requirements by the active EUs, and the harvested energy requirements from the environment by the idle EUs, whereas the amount of data flow depends only on the IUs. Let us divide the long-term period T into discrete time slots, indexed as $\mathcal{T} = \{1, \cdots, T\}$, and define $\mathcal{F} = \{1, \cdots, F\}$ and $\mathcal{K} = \{1, \cdots, K\}$ as the set of indexes of the frames within a time slot and the set of indexes of the learning trials within a frame, respectively. The channel is assumed to vary across frames, but remains invariant within each frame. This paper proposes an online learning algorithm that iteratively alternates between designing the overall transmission strategy using convex optimization and preparing for future energy demand from the day-/hour-ahead market via online learning, i.e., a CMAB approach, to avoid steep ramps in the energy generation plant and to minimize the long-term energy cost.

2.1. Energy Management Model

Similarly to [13], it is assumed that at least one renewable energy generator, i.e., solar panel or/and wind turbine, is installed in the vicinity of each RRH. In this setup, none of the RRHs are equipped with any frequently rechargeable storage devices. Furthermore, bidirectional energy trading with the primary grid is enabled at the individual RRHs. Thus, the RRHs can purchase energy in the day-/hour-ahead market during off-peak hours at a lower price and/or in the spot market during peak hours at a higher price, and the surplus energy can also be sold back to the grid at an agreed-upon price. Let $B_n^{[\text{spot}]}$, $B_n^{[\text{ahead}]}$, S_n, and E_n denote, at time slot t, $t \in \mathcal{T}$, the amount of real-time energy purchases from the spot-market for the n-th RRH to cover an instantaneous energy shortage, the amount of look-ahead energy purchases from the day-/hour-ahead market at the end of previous time slot "t $-$ 1", the amount of surplus energy to be traded back to the primary grid, and the amount of renewable energy generation at the n-th RRH, respectively. In addition, let $P_n^{[\text{Tx}]}$ be the total transmit power and $P_n^{[\text{circ}]}$ be the total power consumption of the hardware circuits at the n-th RRH. Furthermore, in any frame, the total energy consumption at the n-th RRH, i.e., $P_n^{[\text{total}]}$, is constrained as

$$P_n^{[\text{total}]} = P_n^{[\text{Tx}]} + P_n^{[\text{circ}]} = B_n^{[\text{spot}]} + B_n^{[\text{ahead}]} - S_n + E_n. \tag{1}$$

By viewing from the perspective of supply and demand, let us assume $\pi^{[\text{spot}]} \geq \pi^{[\text{ahead}]} \geq \pi^{[\text{sell}]} \geq \pi^{[\text{renew}]}$, where $\pi^{[\text{spot}]}$, $\pi^{[\text{ahead}]}$, $\pi^{[\text{sell}]}$, and $\pi^{[\text{renew}]}$ denote the price of purchasing (selling) per unit energy of $B_n^{[\text{spot}]}$, $B_n^{[\text{ahead}]}$, S_n, and generating per unit energy of E_n (by averaging the capital expenses and OPEX of renewable devices over their lifetime), respectively. Then, the cumulative energy cost procured by the n-th RRH at the k-th trial, $k \in \mathcal{K}$ of the frame f, $f \in \mathcal{F}$ at the time slot t, $t \in \mathcal{T}$, i.e., $B_n^{[\text{total}]}(k)$, is given by

$$\begin{aligned} B_n^{[\text{total}]}(k) \quad &= \pi^{[\text{spot}]} B_n^{[\text{spot}]}(k) + \pi^{[\text{ahead}]} B_n^{[\text{ahead}]}(k) \\ &\quad - \pi^{[\text{sell}]} S_n(k) + \pi^{[\text{renew}]} E_n(k), \ \forall n \in \mathcal{L}_b. \end{aligned} \tag{2}$$

2.2. Downlink Transmission Model

Let $\mathbf{w}_i = [\mathbf{w}_{1i}^H, \cdots, \mathbf{w}_{Ni}^H]^H \in \mathbb{C}^{MN \times 1}$ and $\mathbf{v}_e = [\mathbf{v}_{1e}^H, \cdots, \mathbf{v}_{Ne}^H]^H \in \mathbb{C}^{MN \times 1}$ be defined, respectively, as the set of indexes of the beamforming vector from all RRHs towards the i-th IU, $i \in \mathcal{L}_i$ and the e-th active EU, $e \in \mathcal{L}_e$, where $\mathbf{w}_{ni} \in \mathbb{C}^{M \times 1}$ and $\mathbf{v}_{ne} \in \mathbb{C}^{M \times 1}$ represent the beamformer from the n-th RRH to the i-th IU and the e-th active EU, respectively. In addition, let $\mathbf{h}_i = [\mathbf{h}_{1i}^H, \cdots, \mathbf{h}_{Ni}^H]^H \in \mathbb{C}^{MN \times 1}$ denote the set of indexes of the channel vector

between all RRHs and the i-th IU, where $\mathbf{h}_{ni} \in \mathbb{C}^{M\times 1}$ denotes the channel vector from the n-th RRH to the i-th IU. Accordingly, the signal collected at the i-th IU, $i \in \mathcal{L}_i$, can be expressed as the summation of the dedicated information-carrying signal, the inter-user interference induced by other non-devoted information beams, the interference provoked by the energy-carrying signals assigned to all active EUs, and the additive white Gaussian noise at the i-th IU as

$$y_i = \mathbf{h}_i^H \mathbf{w}_i s_i^{[\mathrm{IU}]} + \sum_{\substack{j\neq i \\ j\in\mathcal{L}_i}} \mathbf{h}_i^H \mathbf{w}_j s_j^{[\mathrm{IU}]} + \sum_{e\in\mathcal{L}_e} \mathbf{h}_i^H \mathbf{v}_e s_e^{[\mathrm{EU}]} + n_i. \tag{3}$$

Due to the fact that energy beams carry no information, only the data of IUs will be delivered via the front-haul links. Without loss of generality, $\mathbb{E}(s_i^{[\mathrm{IU}]}) = \mathbb{E}(s_e^{[\mathrm{EU}]}) = 1$ is assumed, and the signal-to-interference-plus-noise ratio (SINR) at the i-th IU, $i \in \mathcal{L}_i$, is formulated as

$$\mathrm{SINR}_i^{[\mathrm{IU}]} = \frac{|\mathbf{h}_i^H \mathbf{w}_i|^2}{\sum\limits_{j\in\mathcal{L}_i, j\neq i} |\mathbf{h}_i^H \mathbf{w}_j|^2 + \sum\limits_{e\in\mathcal{L}_e} |\mathbf{h}_i^H \mathbf{v}_e|^2 + \sigma_i^2}, \tag{4}$$

where $|\mathbf{h}_i^H \mathbf{w}_i|^2$ indicates the desired power received at the i-th IU and $|\mathbf{w}_i|^2$ is the required transmit power at the RRHs. Let us define the scheduling arrangements between the i-th IU and the n-th RRH for partial cooperation [18], i.e., $\left\| \|\mathbf{w}_{ni}\|_2^2 \right\|_0$, as

$$\left\| \|\mathbf{w}_{ni}\|_2^2 \right\|_0 = \begin{cases} 0, & \text{if } \|\mathbf{w}_{ni}\|_2^2 = 0, \\ 1, & \text{if } \|\mathbf{w}_{ni}\|_2^2 \neq 0, \end{cases} \tag{5}$$

where $\|\mathbf{w}_{ni}\|_2^2 = 0$ betokens that the i-th IU is not selected to be supported by the n-th RRH and, hence, the front-haul link between the CU and the n-th RRH is not employed for joint data transmission to the i-th IU. Hence, the front-haul link capacity consumption of the n-th RRH is expressed as

$$C_n^{[\mathrm{front}]} = \sum_{i\in\mathcal{L}_i} \left\| \|\mathbf{w}_{ni}\|_2^2 \right\|_0 R_i, \quad \forall n \in \mathcal{L}_b, \tag{6}$$

where $R_i = \log_2(1 + \mathrm{SINR}_i^{[\mathrm{IU}]})$ is the achievable data-flow rate (bit/s/Hz) for the i-th IU and directly depends on the transmit power and the wireless channel fading condition. The total energy received by the e-th active EU, $e \in \mathcal{L}_e$, is defined as

$$\mathcal{G}_e^{[\mathrm{EU}]} = \eta \left(|\mathbf{g}_e^H \mathbf{v}_e|^2 + \sum_{j\in\mathcal{L}_e, j\neq e} |\mathbf{g}_e^H \mathbf{v}_j|^2 + \sum_{i\in\mathcal{L}_i} |\mathbf{g}_e^H \mathbf{w}_i|^2 \right), \tag{7}$$

where the terms on the right-hand side of (7) represent the intended energy-carrying signal for the e-th active EU, the inter-user interference caused by all other non-desired energy beams, and the inter-user interference caused by information beams, respectively. Let $0 \leq \eta \leq 1$ denote the conversion efficiency to convert the harvested RF energy into the functional electrical energy form, and $\mathbf{g}_e = [\mathbf{g}_{1e}^H, \cdots, \mathbf{g}_{Ne}^H]^H \in \mathbb{C}^{MN\times 1}$ indicates the set of indexes of the channel vector between all the RRHs and the e-th active EU. The collective energy that can be harvested from the ambiances and atmospheres by the z-th idle EU, $z \in \mathcal{L}_e^{[\mathrm{idle}]}$, is presented as

$$\mathcal{G}_z^{[\text{ET-idle}]} = \eta \Big(\sum_{i\in\mathcal{L}_i} |\mathbf{f}_z^H \mathbf{w}_i|^2 + \sum_{e\in\mathcal{L}_e} |\mathbf{f}_z^H \mathbf{v}_e|^2 \Big), \tag{8}$$

where $\mathbf{f}_z = [\mathbf{f}_{1z}^H, \cdots, \mathbf{f}_{Nz}^H]^H \in \mathbb{C}^{MN\times 1}$ represents the set of indexes of the channel vector between all the RRHs and the z-th idle EU.

3. Real-Time Energy Trading on an Individual Time Frame

This paper relies on foresighted optimization based on CMAB learning to minimize the long-term average energy cost. In accordance with (2), the total energy cost at a given trial of a frame within a time slot is determined by four parameters, i.e., $B_n^{[\text{spot}]}$, S_n, E_n, and $B_n^{[\text{ahead}]}$. It is assumed that the amount of renewable energy supply E_n is given at the beginning of each time slot, whereas $B_n^{[\text{ahead}]}, \forall n \in \mathcal{L}_b$ is determined in advance at the end of the previous time slot via the proposed online learning algorithm, i.e., Algorithms 1 and 2 in Section 4, to prepare for future demands.

3.1. Problem Formulation

Let us define $P_n^{[\text{Tx}]} = \sum_{i \in \mathcal{L}_i} ||\mathbf{w}_{ni}||_2^2 + \sum_{e \in \mathcal{L}_e} ||\mathbf{v}_{ne}||_2^2$ as the total power transmitted by the n-th RRH to its scheduled users and the degree of partial cooperation among RRHs as

$$\begin{aligned} \mathcal{P}^{[\text{coop}]} &= \Big(\sum_{i \in \mathcal{L}_i} \Big\| \|\mathbf{w}_{1i}\|_2^2 \Big\|_0 + \cdots + \sum_{i \in \mathcal{L}_i} \Big\| \|\mathbf{w}_{Ni}\|_2^2 \Big\|_0\Big) \\ &+ \Big(\sum_{e \in \mathcal{L}_e} \Big\| \|\mathbf{v}_{1e}\|_2^2 \Big\|_0 + \cdots + \sum_{e \in \mathcal{L}_e} \Big\| \|\mathbf{v}_{Ne}\|_2^2 \Big\|_0\Big). \end{aligned}$$

To minimize the average energy cost, let us consider the following cooperative energy trading model in each trial of a frame within a time slot for the given $B_n^{[\text{ahead}]}$ as

$$\begin{aligned} \min_{\substack{\mathbf{w}_{ni}, \mathbf{v}_{ne}, \\ B_n^{[\text{spot}]}, S_n}} \quad & \alpha \mathcal{P}^{[\text{coop}]} + \sum_{n \in \mathcal{L}_b} P_n^{[\text{Tx}]} + \sum_{n \in \mathcal{L}_b} \Big\{ B_n^{[\text{spot}]} \Big\} \qquad (9) \\ \text{s.t.} \quad & \text{C1}: \text{SINR}_i^{[\text{IU}]} \geq \gamma_i, \quad \forall i \in \mathcal{L}_i, \\ & \text{C2}: \mathcal{G}_e^{[\text{EU}]} \geq P_e^{[\text{min}]}, \quad \forall e \in \mathcal{L}_e, \\ & \text{C3}: \mathcal{G}_z^{[\text{EU-idle}]} \geq P_z^{[\text{idle}]} \quad \forall z \in \mathcal{L}_e^{[\text{idle}]}, \\ & \text{C4}: P_n^{[\text{Tx}]} \leq E_n + B_n^{[\text{ahead}]} + B_n^{[\text{spot}]} - S_n - P_n^{[\text{circ}]} \\ & \text{C5}: P_n^{[\text{Tx}]} \leq P_n^{[\text{Tmax}]}, \quad \forall n \in \mathcal{L}_b, \\ & \text{C6}: C_n^{[\text{front}]} \leq C_n^{[\text{limit}]}, \quad \forall n \in \mathcal{L}_b, \\ & \text{C7}: \sum_{n \in \mathcal{L}_b} B_n^{[\text{ahead}]} + \sum_{n \in \mathcal{L}_b} B_n^{[\text{spot}]} \leq P_{\text{CU}}^{[\text{max}]} - P_{\text{CU}}^{[\text{circ}]} \\ & \text{C8}: B_n^{[\text{spot}]} \geq 0,, \quad \forall n \in \mathcal{L}_b, \\ & \text{C9}: S_n \geq 0, \quad \forall n \in \mathcal{L}_b, \end{aligned}$$

where $\alpha \geq 0$ is the maximal energy cost in the front-haul link for the degree of partial cooperation among RRHs. C1 guarantees the minimum SINR requirements γ_i for the i-th IUs. $P_e^{[\text{min}]}$ in C2 indicates the minimal energy demanded by the active EUs, whereas $P_z^{[\text{idle}]}$ in C3 is the minimal conditions of energy harvested from the ambiance and atmospheres by the idle EUs. C4 betokens that the individual RRHs' power budget restrains the total transmit power as per (1), while C5 emphasizes that the total transmit power is upper-limited by the maximum transmit power permitted, i.e., $P_n^{[\text{Tmax}]}$, at the n-th RRH. C6 expresses the front-haul link capacity limitations for the individual RRHs. C7 implies the restriction for the total power provided by the grid to the RRHs, where $P_{\text{CU}}^{[\text{circ}]}$ is the hardware circuit power consumption and $P_{\text{CU}}^{[\text{max}]}$ is the maximum power generated by the grid at the CU [19]. C8 and C9 are the non-negative constraints set for the optimization variables.

3.2. Re-Weighted ℓ_1-Norm and Semidefinite Programming

The optimization problem in (9) is an NP-hard (nondeterministic polynomial time) problem due to the non-convexity of the ℓ_0-norm term in the objective function and the

constraints C1 and C6. These non-convexity terms can be reformulated by using one of the powerful convex optimization techniques, i.e., semidefinite programming (SDP). Note that the ℓ_1-norm approximation is commonly adopted in compressed sensing to handle ℓ_0-norm optimization problems [20]. Then, let us consider the following property:

$$\mathbf{x}^H\mathbf{A}\mathbf{x} = \mathrm{tr}(\mathbf{x}\mathbf{x}^H\mathbf{A}) = \mathrm{tr}(\mathbf{A}\mathbf{x}\mathbf{x}^H). \tag{10}$$

From a mathematical point of view, the property in (10) can be interpreted as the inner vector product being equal to the trace of the outer product. If $\mathbf{A} = \mathbf{I}$, then

$$\mathbf{x}^H\mathbf{x} = \mathrm{tr}(\mathbf{x}\mathbf{x}^H). \tag{11}$$

By adding this property, denoting $\mathbf{H}_i = \mathbf{h}_i\mathbf{h}_i^H$, $\mathbf{G}_e = \mathbf{g}_e\mathbf{g}_e^H$, and $\mathbf{F}_z = \mathbf{f}_z\mathbf{f}_z^H$, and specifying the rank-one semidefinite matrices as $\mathbf{W}_i = \mathbf{w}_i\mathbf{w}_i^H$ and $\mathbf{V}_e = \mathbf{v}_e\mathbf{v}_e^H$ in the optimization problem, the constraint C1 can be reformulated as

$$\mathrm{C1}: \frac{|\mathbf{h}_i^H\mathbf{w}_i|^2}{\sum\limits_{j\in\mathcal{L}_i, j\neq i}|\mathbf{h}_i^H\mathbf{w}_j|^2 + \sum\limits_{e\in\mathcal{L}_e}|\mathbf{h}_i^H\mathbf{v}_e|^2 + \sigma_i^2} \geq \gamma_i, \ \ \forall i \in \mathcal{L}_i, \tag{12}$$

$$\mathrm{C1}: \mathrm{tr}(\mathbf{H}_i\mathbf{W}_i) \geq \gamma_i \sum_{j\in\mathcal{L}_i, j\neq i}\mathrm{tr}(\mathbf{H}_i\mathbf{W}_j) + \gamma_i \sum_{e\in\mathcal{L}_e}\mathrm{tr}(\mathbf{H}_i\mathbf{V}_e) + \gamma_i\sigma_i^2, \ \ \forall i \in \mathcal{L}_i. \tag{13}$$

Following a procedure similar to that in [13], the intractability of the ℓ_0-norm term in the objective function and constraint C6 is overcome by approximating by their respective re-weighted ℓ_1-norms [20], as follows:

$$\begin{aligned}\mathcal{P}^{[\mathrm{coop}]} \approx & \sum_{i\in\mathcal{L}_i}\left\|[\xi_{1i}\|\mathbf{w}_{1i}\|_2^2]\right\|_1 + \cdots + \sum_{i\in\mathcal{L}_i}\left\|[\xi_{Ni}\|\mathbf{w}_{Ni}\|_2^2]\right\|_1 \\ & + \sum_{e\in\mathcal{L}_e}\left\|[\kappa_{1e}\|\mathbf{v}_{1e}\|_2^2]\right\|_1 + \cdots + \sum_{e\in\mathcal{L}_e}\left\|[\kappa_{Ne}\|\mathbf{v}_{Ne}\|_2^2]\right\|_1 \\ = & \sum_{n\in\mathcal{L}_b}\left(\sum_{i\in\mathcal{L}_i}\xi_{ni}\mathrm{tr}(\mathbf{w}_i\mathbf{w}_i^H\mathbf{D}_n) + \sum_{e\in\mathcal{L}_e}\kappa_{ne}\mathrm{tr}(\mathbf{v}_e\mathbf{v}_e^H\mathbf{D}_n)\right),\end{aligned} \tag{14}$$

$$C_n^{[\mathrm{front}]} \approx \sum_{i\in\mathcal{L}_i}\left\|[\xi_{ni}\|\mathbf{w}_{ni}\|_2^2]\right\|_1 R_i = \sum_{i\in\mathcal{L}_i}\xi_{ni}\mathrm{tr}(\mathbf{w}_i\mathbf{w}_i^H\mathbf{D}_n)R_i, \tag{15}$$

where $\mathbf{D}_n \triangleq \mathrm{diag}(\overbrace{0,...,0}^{(n-1)M}, \overbrace{1,...,1}^{M}, \overbrace{0,...,0}^{(N-n)M}) \succeq 0, \forall n \in \mathcal{L}_b$, is used for extracting the corresponding beamformer $\mathbf{w}_{ni}$. ξ_{ni} and κ_{ne}, respectively, are the weighting factors associated with the n-th RRH and the i-th IU/the e-th active EU, which will be updated as per the re-weighted ℓ_1-norm method algorithm in [13] to iteratively remove the collaborative links between the RRHs and the IUs/active EUs in the circumstances of front-haul link capacity limitations at the individual RRHs. Hence, the non-convex problem formulation introduced in (9) can be modified to a convex optimization problem with significantly reduced complexity [21] after relaxing the rank-one constraints of $\mathrm{rank}(\mathbf{W}_i) = 1$ and $\mathrm{rank}(\mathbf{V}_e) \leq 1$, as (16).

Lemma 1. *The optimal solutions to the problems* (16) *satisfy* $\mathrm{rank}(\mathbf{W}_i^*) = 1$ *and* $\mathrm{rank}(\mathbf{V}_e^*) \leq 1$ *with a probability of one.*

Proof. The proof is straightforward by following similar steps to those in [13]. □

As per [22], the interior point methods for solving SDP have polynomial (quadratic) worst-case complexity and are superb for medium- and large-scale problems, e.g., those

bounded by $O(\log n)$, where n is the problem size. Furthermore, as the size of the optimization problem grows large, the computational complexity tends to grow more slowly and even remains almost constant according to [23]. Hence, with increasing size of the problem, e.g., an increasing total number of RRHs, users, and per-RRH antennas, the number of iterations needed to solve the optimization problem grows sub-linearly with the size of the problem, and even tends to remain almost constant.

$$
\begin{aligned}
\min_{\substack{\mathbf{W}_i,\\ \mathbf{V}_e,S_n,\\ B_n^{[\text{spot}]}}} \quad & \sum_{n\in\mathcal{L}_b}\left(\sum_{i\in\mathcal{L}_i}\xi_{ni}\text{tr}(\mathbf{W}_i\mathbf{D}_n)+\sum_{e\in\mathcal{L}_e}\kappa_{ne}\text{tr}(\mathbf{V}_e\mathbf{D}_n)\right) \\
& +\left(\sum_{i\in\mathcal{L}_i}\text{tr}(\mathbf{W}_i)+\sum_{e\in\mathcal{L}_e}\text{tr}(\mathbf{V}_e)\right)+\sum_{n\in\mathcal{L}_b}\left\{B_n^{[\text{spot}]}\right\} \\
\text{s.t.} \quad & \text{C1}: \text{tr}(\mathbf{H}_i\mathbf{W}_i)\geq\gamma_i\sum_{j\in\mathcal{L}_i,j\neq i}\text{tr}(\mathbf{H}_i\mathbf{W}_j)+ \\
& \qquad \gamma_i\sum_{e\in\mathcal{L}_e}\text{tr}(\mathbf{H}_i\mathbf{V}_e)+\gamma_i\sigma_i^2,\ \forall i\in\mathcal{L}_i, \\
& \text{C2}: \text{tr}(\mathbf{G}_e\mathbf{V}_e)+\sum_{\substack{j\in\mathcal{L}_e,\\ j\neq e}}\text{tr}(\mathbf{G}_e\mathbf{V}_j)+\sum_{i\in\mathcal{L}_i}\text{tr}(\mathbf{G}_e\mathbf{W}_i) \\
& \qquad \geq P_e^{[\min]}\eta^{-1},\ \forall e\in\mathcal{L}_e, \\
& \text{C3}: \sum_{i\in\mathcal{L}_i}\text{tr}(\mathbf{F}_z\mathbf{W}_i)+\sum_{e\in\mathcal{L}_e}\text{tr}(\mathbf{F}_z\mathbf{V}_e)\geq P_z^{[\text{idle}]}\eta^{-1}, \\
& \qquad \forall z\in\mathcal{L}_e^{[\text{idle}]}, \\
& \text{C4}: \sum_{i\in\mathcal{L}_i}\text{tr}(\mathbf{W}_i\mathbf{D}_n)+\sum_{e\in\mathcal{L}_e}\text{tr}(\mathbf{V}_e\mathbf{D}_n)\leq[E_n-S_n \\
& \qquad +B_n^{[\text{ahead}]}+B_n^{[\text{spot}]}-P_n^{[\text{circ}]}],\ \forall n\in\mathcal{L}_b, \\
& \text{C5}: \sum_{i\in\mathcal{L}_i}\text{tr}(\mathbf{W}_i\mathbf{D}_n)+\sum_{e\in\mathcal{L}_e}\text{tr}(\mathbf{V}_e\mathbf{D}_n)\leq P_n^{[\text{Tmax}]}, \\
& \text{C6}: \sum_{i\in\mathcal{L}_i}\xi_{ni}\text{tr}(\mathbf{W}_i\mathbf{D}_n)R_i\leq C_n^{[\text{limit}]},\ \forall n\in\mathcal{L}_b, \\
& \text{C7}: \sum_{n\in\mathcal{L}_b}B_n^{[\text{ahead}]}+\sum_{n\in\mathcal{L}_b}B_n^{[\text{spot}]}\leq P_{\text{CU}}^{[\max]}-P_{\text{CU}}^{[\text{circ}]} \\
& \text{C8}: B_n^{[\text{spot}]}\geq 0, \qquad \text{C9}: S_n\geq 0, \quad \forall n\in\mathcal{L}_b, \\
& \text{C10}: \mathbf{W}_i\succeq 0,\ \forall i\in\mathcal{L}_i,\ \text{C11}: \mathbf{V}_e\succeq 0,\ \forall e\in\mathcal{L}_e.
\end{aligned}
\tag{16}
$$

4. Predictive Energy Trading Strategy

The multi-armed bandit (MAB) problem is expressed as a J-arm system, with each being associated with independent and identically distributed (i.i.d.) stochastic rewards. The objective is to maximize the accumulated profits by observing the associated reward of new arms during the exploration stage while simultaneously optimizing the decisions among a set of arms based on existing knowledge at the exploitation stage in multiple trials [24]. Let consider a combinatorial generalization of the classical MAB problem, where a super arm consisting of a set of N base arms, $N\subset J$, is played, and the rewards of its relevant base arms are observed individually in each trial [25].

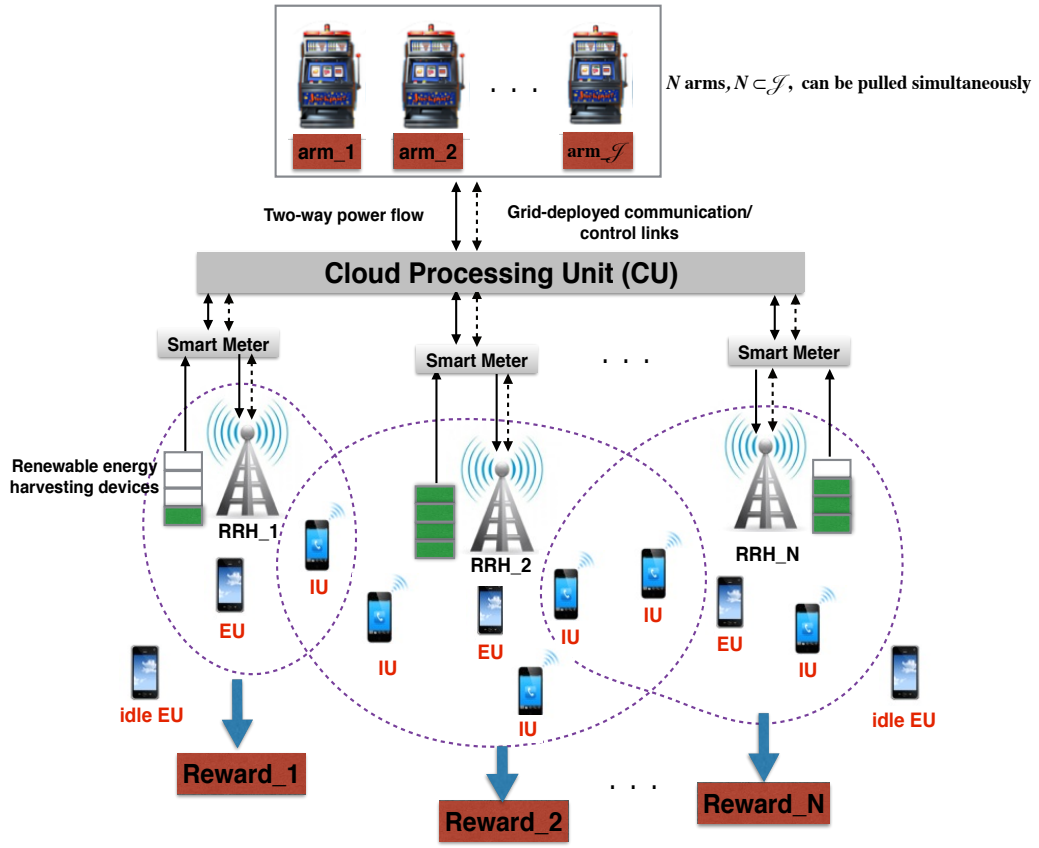


Figure 1. Combinatorial multi-armed bandit (CMAB) problem for predictive real-time energy trading in a cloud radio access network (Cloud-RAN) with the sparse beamforming technique.

As illustrated in Figure 1, the problem scrutinized in this paper is categorized as a combinatorial MAB problem, where a super arm is composed of N base arms and each base arm corresponds to an energy package purchased for an RRH from the day-/hour-ahead market at each trial $k, k \in \mathcal{K}$, before the real-time energy demand. The CU adapts its cooperative energy trading strategies to the intermittent environment in the Cloud-RAN by dynamically forming super arms to maximize the averaged rewards accumulated over the period T, which is equivalent to lessening the averaged energy expense in the long run. Let $\mathcal{J} = \{1, \cdots, J\}$ be defined as a set of indexes for possible energy packages offered in the day-/hour-ahead market by the grid, and let $\mathcal{E}^{[\text{total}]} = \{\mathcal{E}^1, \cdots, \mathcal{E}^J\}$ denote all energy packages offered by the grid in the day-/hour-ahead market, where $\mathcal{E}^p = \mathcal{E}^{p-1} + \Delta\mathcal{E}, p \in \mathcal{J}$. Furthermore, let $\mathcal{A}_k^{[\text{set}]} = \{B_1^{[\text{ahead}]}(k), \cdots, B_N^{[\text{ahead}]}(k)\}$ represent a super arm, i.e., a set of N energy packages purchased in advance for N RRHs from the day-/hour-ahead market, at the k-th trial. Let us further define the reward for the individual arms at the n-th RRH and the reward for the super arm at the k-th trial as $\mathcal{R}(B_n^{[\text{ahead}]}(k))$ and $\mathcal{R}(\mathcal{A}_k^{[\text{set}]})$, respectively, as

$$\mathcal{R}(B_n^{[\text{ahead}]}(k)) = B_n^{[\text{total}]}(1) - B_n^{[\text{total}]}(k), \tag{17}$$

$$\mathcal{R}(\mathcal{A}_k^{[\text{set}]}) = \sum_{n \in \mathcal{L}_b} \mathcal{R}(B_n^{[\text{ahead}]}(k)), \tag{18}$$

where $B_n^{[\text{total}]}(1)$ and $B_n^{[\text{total}]}(k)$ in (17) are the total energy cost incurred by the n-th RRH at the initial trial and the k-th trial of a frame, respectively. Furthermore, let $\boldsymbol{\mu}_n^{[k,f,t]} = (\mu_{n,1}^{[k,f,t]}, \mu_{n,2}^{[k,f,t]}, \cdots, \mu_{n,J}^{[k,f,t]})$ be defined as the reward vector for the n-th RRH, where $\mu_{n,p}^{[k,f,t]} = \mathcal{R}(B_n^{[\text{ahead}]}(k)), p \in \mathcal{J}$, is the reward associated to the p-th energy package in the k-th trial of the f-th frame at the t-th time slot.

In the following, we propose CUCB-based [25] predictive energy trading strategy, which is shown in Figure 2 and detailed in Algorithms 1 and 2, to find the best possible combination of energy packages to be purchased from the day-/hour-ahead market for N RRHs for the next time slot by exploring the rewards of new combinations of energy packages within a limited number of trials at the current time slot and exploiting the past captured information on rewards of super arms from the previous time slots so that the long-term averaged rewards, i.e., the total energy cost in the long run, can be optimized.

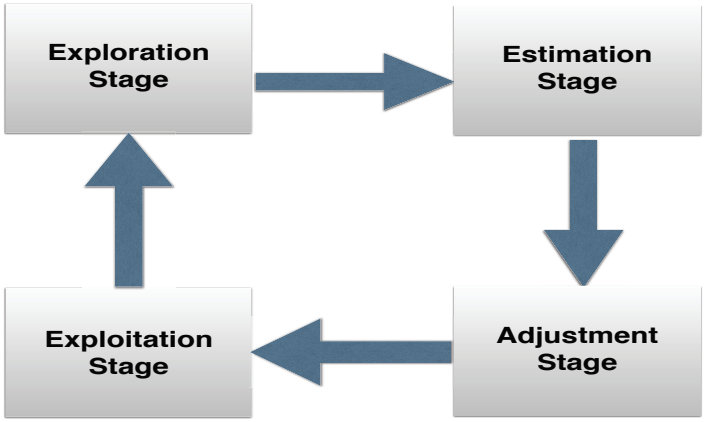


Figure 2. Proposed predictive energy trading strategy in a Cloud-RAN.

Algorithm 1 Super Arm Exploration

Initialize: Total number of trials K
for $k = 1 : K$
  Solve problem (16) for a given $B_n^{[\mathbf{ahead}]}(k)$,
  CU calculates $B_n^{[\mathbf{total}]}(k)$ **as per (2),** $\mathcal{R}(B_n^{[\mathbf{ahead}]}(k))$ **as per (17), and** $\mathcal{R}(\mathcal{A}_k^{[\mathbf{set}]})$ **as per (18).**
    If $k = 1$
    then
      $B_n^{[\mathbf{ahead}]}(k+1) = B_n^{[\mathbf{ahead}]}(k) + \Delta\mathcal{E}, \qquad n \in \mathcal{L}_b.$
    else if the super arm reward of all the RRHs
      $\mathcal{R}(\mathcal{A}_k^{[set]}) \leq \mathcal{R}(\mathcal{A}_{k-1}^{[set]})$,
    then
      $B_n^{[\mathbf{ahead}]}(k+1) = B_n^{[\mathbf{ahead}]}(k-1), \quad \forall n \in \mathcal{L}_b,$
    else if the individual reward for the n**-th RRH,** $n \in N$
      $\mathcal{R}(B_n^{[\mathbf{ahead}]}(k)) \geq \mathcal{R}(B_n^{[\mathbf{ahead}]}(k-1))$
      and
      $B_n^{[\mathbf{ahead}]}(k) \neq \mathcal{E}^J$,
    then $B_n^{[\mathbf{ahead}]}(k+1) = B_n^{[\mathbf{ahead}]}(k) + \Delta\mathcal{E}$,
    else
      $B_n^{[\mathbf{ahead}]}(k+1) = B_n^{[\mathbf{ahead}]}(k)$.
    end If
  Calculate the total energy cost of all the RRHs, $\beta^{[k,f,t]}$ **as**
    $\beta^{[k,f,t]} = \sum_{n \in \mathcal{L}_b} B_n^{[\mathbf{total}]}(k)$.
  Calculate the energy package index p **at all RRHs from**
    $p = \frac{B_n^{[\mathbf{ahead}]}(k)}{\Delta\mathcal{E}}, \qquad n \in \mathcal{L}_b.$
  Update
    $\mu_{n,p}^{[k,f,t]} = \mathcal{R}(B_n^{[\mathbf{ahead}]}(k)), \ \forall p \in \mathcal{J}, \qquad n \in \mathcal{L}_b;$
  Update
    $\mathcal{A}_{k+1}^{[\mathbf{set}]} = \{B_1^{[\mathbf{ahead}]}(k+1), \cdots, B_N^{[\mathbf{ahead}]}(k+1)\};$
end for
    Estimated mean reward for K trials
      $\hat{\mu}_{n,p}^{[f,t]} = \frac{\sum_{k=1}^{K} \mu_{n,p}^{[k,f,t]}}{K}, \ \forall p \in \mathcal{J}, n \in \mathcal{L}_b.$

Algorithm 2 Main Online Learning Algorithm

1: **Initialize**: Time slot count: $t = 0$;
2: **while** $t \neq T$ **do**
3: **Increment the iteration index** $t = t + 1$**;**
4: **for** $f = 1 : F$
5: **if** $t = 1$ **(initial time slot)**
6: **then Initialize the super arm for the first trial (**$k = 1$**) as**
$\mathcal{A}_1^{[\mathbf{set}]} = \{0_1, \cdots, 0_N\}$**,**
7: **else**
$\mathcal{A}_1^{[\mathbf{set}]} = S^*$**,**
8: **end if**
9: **Exploration Stage: Run Algorithm 1**
10: **Estimation Stage:**
11: **Calculate the mean reward vector for the frame**
$\hat{\boldsymbol{\mu}}_{\boldsymbol{n}}^{[\boldsymbol{f},\boldsymbol{t}]} = (\hat{\mu}_{n,1}^{[f,t]}, \hat{\mu}_{n,2}^{[f,t]}, \ldots, \hat{\mu}_{n,\mathcal{J}}^{[f,t]})$**, where**
$\hat{\mu}_{n,p}^{[f,t]} = \frac{\sum_{k=1}^{K} \mu_{n,p}^{[k,f,t]}}{K}$, $\forall p \in \mathcal{J}, n \in \mathcal{L}_b$**.**
12: **Adjustment Stage :**
13: **if** Ψ_p **(number of times the** p**-th arm is played)** $\neq 0$
14: **then**
adjust $\bar{\mu}_{n,p}^{[f,t]} = \hat{\mu}_{n,p}^{[f,t]} + \sqrt{\frac{3\ln K}{2\Psi_p}}$**,**
15: **else**
$\bar{\mu}_{n,p}^{[f,t]} = \hat{\mu}_{n,p}^{[f,t]}, \qquad \forall p \in \mathcal{J}, n \in \mathcal{L}_b$**.**
16: **end if**
17: **end for**
18: **Average adjusted mean reward vector over all frames**
$$\bar{\boldsymbol{\mu}}_{\boldsymbol{n}}^{[\boldsymbol{t}]} = \left(\frac{\sum_{f \in \mathcal{F}} \bar{\mu}_{n,1}^{[f,t]}}{F}, \frac{\sum_{f \in \mathcal{F}} \bar{\mu}_{n,2}^{[f,t]}}{F}, \ldots, \frac{\sum_{f \in \mathcal{F}} \bar{\mu}_{n,J}^{[f,t]}}{F} \right), \qquad n \in \mathcal{L}_b.$$
19: **Exploitation Stage :**
20: **Average** $\bar{\boldsymbol{\mu}}_{\boldsymbol{n}}^{[\boldsymbol{t}]}$ **over accumulated number of time slots, as**
$$\bar{\boldsymbol{\mu}}_{\boldsymbol{n}} = \frac{\sum_{t'=1}^{t} \bar{\mu}_n^{[t']}}{t} = [\bar{\mu}_{n,1}, \bar{\mu}_{n,2}, \cdots, \bar{\mu}_{n,J}], \qquad n \in \mathcal{L}_b.$$
21: **For the next time slot: find** N **optimum arm indexes as**
$$p_n^* = \underset{p}{\operatorname{argmax}}(\bar{\mu}_{n,p}), p \in \mathcal{J}, \qquad \forall n \in \mathcal{L}_b,$$
and the updated super arm as
$$S^* = \Delta\mathcal{E}[p_1^*, p_2^*, \cdots, p_N^*].$$
22: **end while**

Let $\hat{\boldsymbol{\mu}}_{\boldsymbol{n}}^{[\boldsymbol{f},\boldsymbol{t}]} = (\hat{\mu}_{n,1}^{[f,t]}, \hat{\mu}_{n,2}^{[f,t]}, \ldots, \hat{\mu}_{n,J}^{[f,t]})$ and $\bar{\boldsymbol{\mu}}_{\boldsymbol{n}}^{[\boldsymbol{f},\boldsymbol{t}]} = (\bar{\mu}_{n,1}^{[f,t]}, \bar{\mu}_{n,2}^{[f,t]}, \ldots, \bar{\mu}_{n,J}^{[f,t]}), \forall n \in \mathcal{L}_b, f \in \mathcal{F}, t \in \mathcal{T}$ denote the estimated mean reward vector and the adjusted reward vector of individual energy packages, respectively. In the exploration stage within each frame, Algorithm 1 explores new combinations of energy packages (super arms) for the next trial based on the rewards obtained at the current and the previous trials. Once a given number of K trials are completed, the mean rewards for individual energy packages, i.e., $\hat{\boldsymbol{\mu}}_{\boldsymbol{n}}^{[\boldsymbol{f},\boldsymbol{t}]}$, in each frame are estimated. The estimated mean rewards are, first, adjusted and averaged over a total number of F frames of a time slot as per step 18, then averaged again over the total number of past time slots as per step 20 [26], and, finally, used to update the super arm S^*, i.e., the optimal set of energy packages purchased from the day-ahead market, to be exploited in the next time slot, as detailed in Algorithm 2.

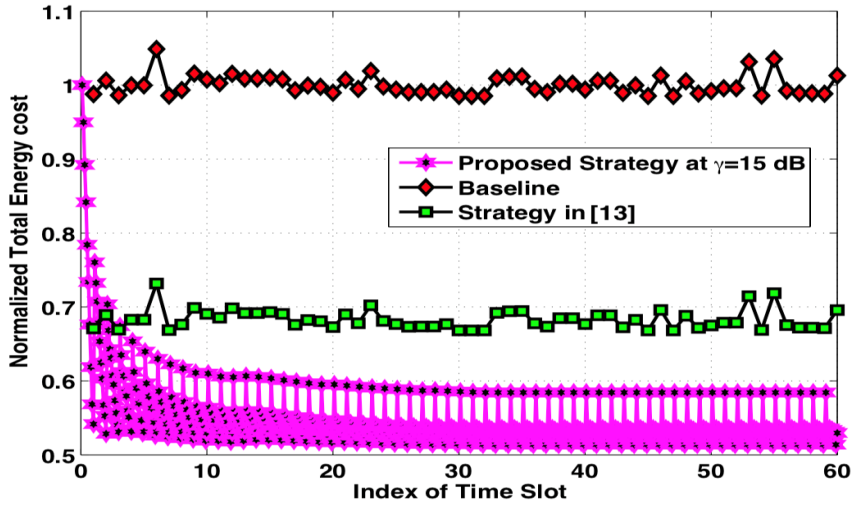
The proposed learning-based algorithm can be considered as a mixed online learning and convex optimization problem with linear matrix inequality constraints. The optimization problem is solved once per learning trial. Therefore, the complexity of the resulting algorithm is mainly due to the number of iterations required for solving a convex optimization problem that has polynomial worst-case complexity [22] and whose total number of learning trials depends on the dynamic range of variations in the environment.

5. Simulation Results

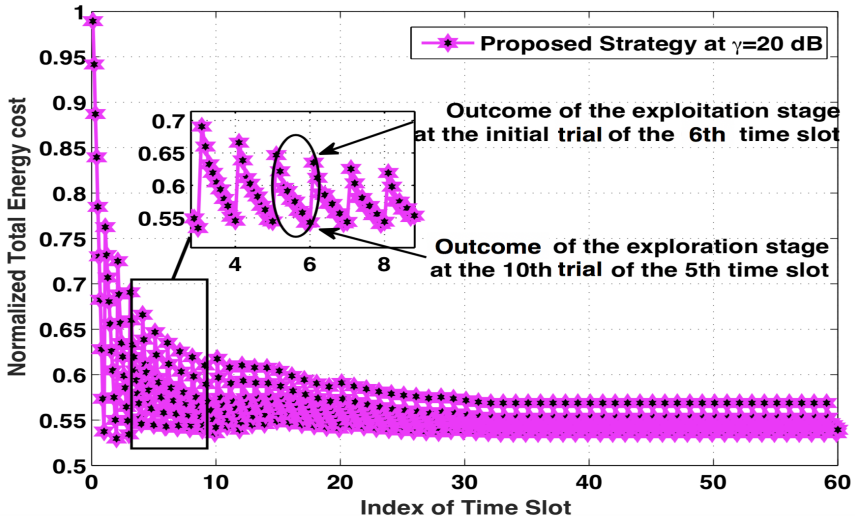
A downlink Cloud-RAN consisting of three adjacent RRHs with SWIPT towards six single-antenna IUs and six single-antenna EUs was considered in this paper. The proposed Cloud-RAN operated under the channel bandwidth of 20 MHz. All of the RRHs were installed with eight antennas and placed 500 m away from each other. The performance of the proposed scheme was assessed with $K = 10$ trials per frame, $F = 10$ frames per time slot, $T = 60$ time slots, and a total number of $J = 20$ energy packages with $\Delta\mathcal{E} = 100$ mW, i.e., $\mathcal{E}_t^{[\text{total}]} = \{100, 200, \cdots, 2000\}$ mW. The renewable energy generation values at the individual RRHs were $E_1 = 1.5$, $E_2 = 0.2$, and $E_3 = 0.05$ W, respectively, at a price of $\pi^{[\text{renew}]} = 0.02$ GBP/W. It was assumed that $\pi^{[\text{ahead}]} = 0.07$, $\pi^{[\text{spot}]} = 0.15$, and $\pi^{[\text{sell}]} = 0.05$ GBP/W. A correlated channel model, $\mathbf{h}_{ni} = \mathbf{R}^{1/2}\mathbf{h}_w$, was adopted [17,27], where $\mathbf{h}_w \in \mathbb{C}^{M\times 1}$ are zero-mean circularly symmetric complex Gaussian random variables with unit variance, $\mathbf{R} \in \mathbb{C}^{M\times M}$ is the spatial covariance matrix, and its (m, n)-th element is given by $G_a L_p \sigma_F^2 e^{-0.5\frac{(\sigma_s \ln 10)^2}{100}} e^{j\frac{2\pi\delta}{\lambda}[(n-m)sin\theta]} e^{-2\left[\frac{\pi\delta\sigma}{\lambda}(n-m)cos\theta\right]^2}$, where $G_a = 15$ dBi denotes the antenna gain, $L_p(\text{dB}) = 125.2 + 36.3\log_{10}(d)$ represents the path loss model over a distance of d km, σ_F^2 is the variance of the complex Gaussian fading coefficient, $\sigma_s = 8$ dB is the log-normal shadowing standard deviation, $\sigma = 2^\circ$ is the angular offset standard deviation, and θ is the estimated angle of departure. The simulation parameters were assumed, unless otherwise stated, to be $P_{\text{CU}}^{[\text{circ}]} = 40$ dBm, $P_{\text{CU}}^{[\text{max}]} = 50$ dBm, $P_n^{[\text{circ}]} = 30$ dBm, $P_n^{[\text{Tmax}]} = 46$ dBm, $C_n^{[\text{limit}]} = 30$ bits/s/Hz, $P_e^{[\text{min}]} = -60$ dBm, $P_z^{[\text{idle}]} = -90$ dBm [18], and $\eta = 0.5$, respectively. The simulation results were accomplished via CVX [28] using an Intel i7-3770 CPU at 3.4GHz with 8 GB RAM, and the running time for each learning trial was approximately seven seconds without use of parallelization. Our proposed online learning strategy was compared against a baseline design that had no ahead-of-time energy preparation and the non-learning based design in [13], which always assumes that a fixed set of energy packages is prepared from the day-/hour-ahead market, i.e., $\mathcal{A}^{[\text{set}]} = \{B_1^{[\text{ahead}]} = B_2^{[\text{ahead}]} = B_3^{[\text{ahead}]}\} = 700$ mW. For fair comparison, identical constraints were applied to all the strategies.

Note that the convergence speed of the proposed online learning strategy to achieve its steady-state is based on the total number of learning trials, which also depends on the dynamic range of variations in the environment. Due to the limitations of our simulation tool, we downsized the total number of learning trials and the other simulation parameters according to the scale of our problem size. In a practical scenario, with a large number of users, the resulting amount of look-ahead energy purchased from the day-/hour-ahead market will be increased proportionally, which may increase the number of arms or increase the difference between two adjacent arms, and may also increase the number of learning trials needed to speed up the convergence. Therefore, the practical enlarged scenario does not affect the scalability of the proposed algorithm, as it may only increase the computational burden.

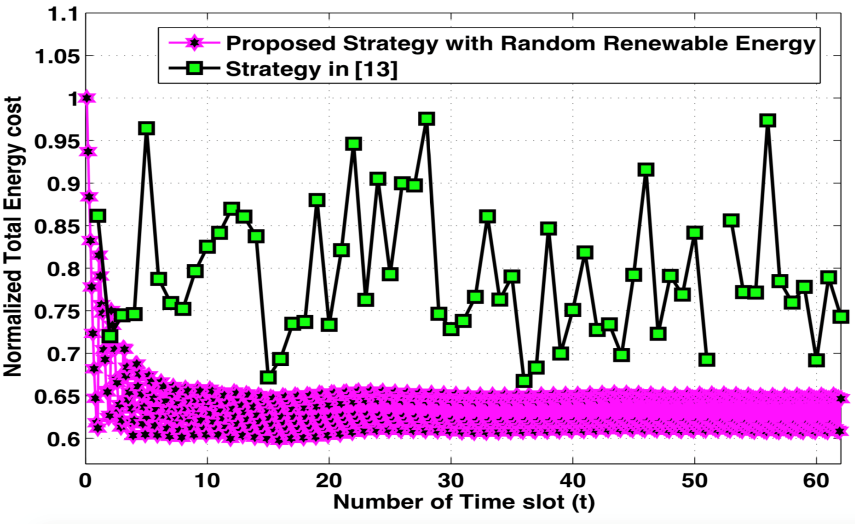
Figure 3a compares the normalized total energy cost over discrete time slots for different strategies at $\gamma = 15$ dB. It can be observed that, at its steady-state, the proposed strategy achieves performance gains of 43 percent and 11 percent, respectively, as compared with the baseline scheme and the design in [13], since their designs provide no adaption to the dynamic wireless channel conditions in Cloud-RANs. Figure 3b shows the normalized total energy cost of our proposed strategy at $\gamma = 20$ dB. One may observe that the performance of the proposed strategy slightly degrades with increasing target SINR, i.e., from $\gamma = 15$ dB to $\gamma = 20$ dB. Figure 3c represents the normalized total energy cost of our proposed strategy at $\gamma = 20$ dB in a more complex scenario, where it is assumed that the number of per-RRH antennas is six and the renewable energy generation at individual RRHs ranges from [0.5 2.5], [0.3 1.5], and [0.1 1.0] W, respectively.



(**a**)



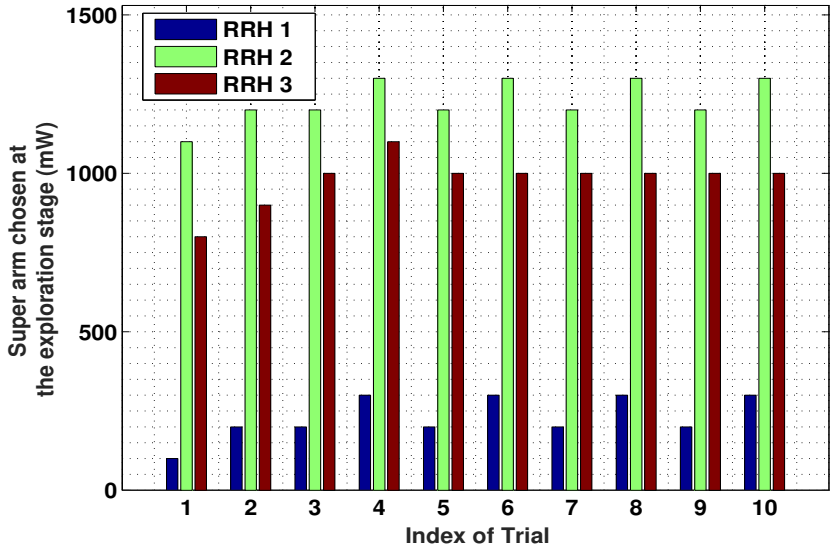
(**b**)



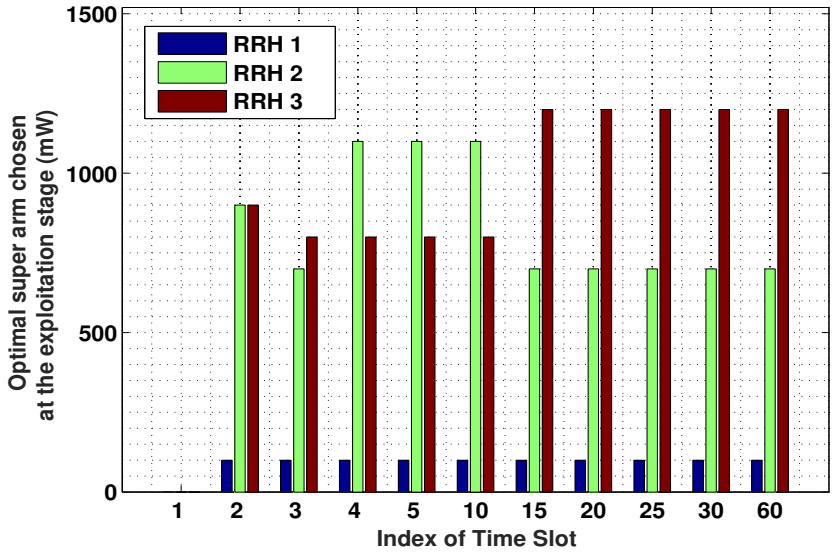
(**c**)

Figure 3. Normalized total energy cost at (**a**) $\gamma = 15$, (**b**) $\gamma = 20$, and (**c**) $\gamma = 20$ dB with random variations in renewable generation.

It is clear from the Figure 3c that the performance of the proposed strategy was slightly degraded compared to Figure 3b, which was simulated in a simpler scenario. However, as the time-slot index increases, the performance of our proposed strategy indicates considerable smaller variations in total energy cost and much better average performance compared to that of [13] under the same system setup. This validates the ability of our proposed algorithm to adapt to more realistic wireless networks.



(**a**)



(**b**)

Figure 4. Illustration of super arm decisions according to the proposed strategy: (**a**) super arms chosen in the individual trials at the fifth time slot; (**b**) look-ahead energy purchase decisions (i.e., final super arm) for individual time slots.

Figure 4 presents in detail the procedure of a super arm being selected in accordance with Algorithms 1 and 2. Figure 4a illustrates the procedure of a super arm, i.e., an optimal set of energy packages purchased for a set of RRHs from the day-/hour-ahead market, in different trials at the fifth time slot. In each trial, a new combination of energy packages is

explored on the basis of the individual and the averaged accumulated rewards obtained from the current and the previous trials, as per Algorithm 1. Figure 4b demonstrates the optimal super arm that was selected at the t-th time slot to be exploited as the starting point at the $(t+1)$-th time slot, as per Algorithm 2. It can be observed that from the 15th time slot onwards, nearly identical super arms that associate with the highest rewards for the RRHs are selected, which demonstrates the convergence of the proposed algorithm for the given simulation.

The normalized accumulated reward and regret at each time slot for different strategies are shown in Figure 5. The normalized accumulated reward at time slot t, denoted by $\mathcal{R}_t^{[\mathrm{acc}]}$, is calculated by averaging the difference of the total energy cost at the t-th time slot and the initial time slot over all frames, i.e., $\mathcal{R}_t^{[\mathrm{acc}]} = \frac{\sum_{f\in\mathcal{F}}(\beta^{[k,f,1]}-\beta^{[k,f,t]})}{F}$. In contrast, the regret of the strategies is defined as the difference in the accumulated reward between always playing the optimal super arm and playing the super arm according to the proposed strategy at the t-th time slot, i.e., $\mathcal{Q}_t = \mathcal{R}_{\mathrm{opt}}^{[\mathrm{acc}]} - \mathcal{R}_t^{[\mathrm{acc}]}$, where $\mathcal{R}_{\mathrm{opt}}^{[\mathrm{acc}]}$ is the accumulated reward after the convergence. Figure 5 confirms that a significant performance gap exists between the proposed strategy and the baseline scheme, as well as the design in [13]. One can conclude that, although the regret of the proposed strategy has the worst performance at the initial time slot, it declines rapidly with the continuous learning process until convergence due to the fact that the proposed strategy learns from the past captured behavior of cooperative energy trading and adapts to the dynamic wireless environment.

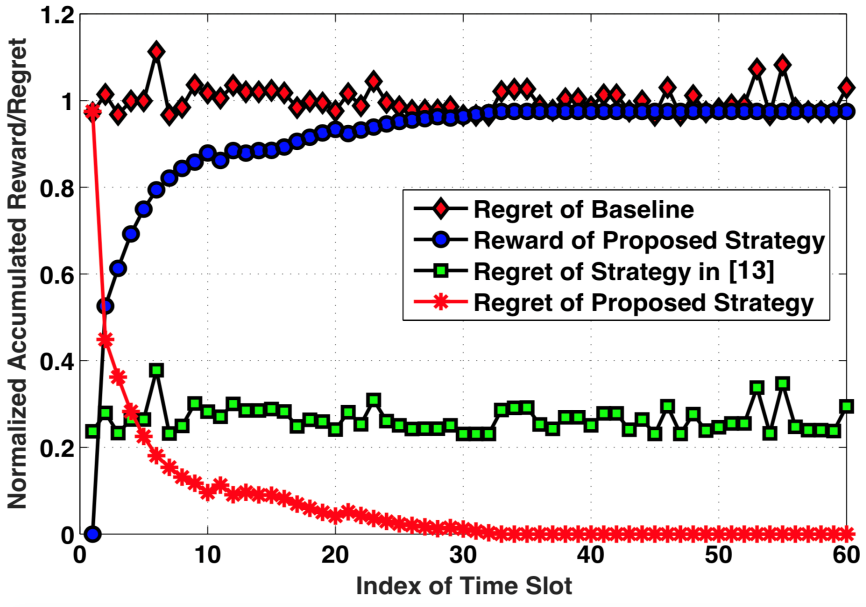


Figure 5. Normalized accumulated reward/regret for different strategies.

6. Conclusions

This paper proposes a predictive cooperative energy trading mechanism based on a CMAB model in a green Cloud-RAN with SWIPT, which adapts to the temporal variations of energy demands in a statistically unknown changing environment and improves its performance gain over time, with the objective of minimizing the time-averaged overall energy cost in the long run. The proposed strategy anticipates future energy demand and supplies the instantaneous energy demand at the current time slot with energy prepared in advance based on existing knowledge of uncertain wireless system dynamics at the previous time slots. The presented simulation results confirmed a reduction of the long-term running cost. Our proposed scheme outperforms a baseline scheme that purchases no ahead-of-time energy packages and a recently proposed non-learning-based design that assumes fixed energy purchases from the day-ahead market.

Author Contributions: Conceptualization, W.N.S.F.W.A.; Data curation, W.N.S.F.W.A.; Formal analysis, W.N.S.F.W.A. and M.R.N.; Funding acquisition, H.A.R. and R.B.A.; Investigation, W.N.S.F.W.A.; Methodology, W.N.S.F.W.A. and M.R.N.; Project administration, W.N.S.F.W.A. and M.R.N.; Resources, W.N.S.F.W.A. and X.Z.; Software, W.N.S.F.W.A.; Supervision, M.R.N.; Validation, W.N.S.F.W.A. and X.Z.; Visualization, W.N.S.F.W.A. and M.R.N.; Writing—original draft, W.N.S.F.W.A., X.Z. and M.R.N.; Writing—review & editing, W.N.S.F.W.A., X.Z., M.R.N. and H.A.R. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Fundamental Research Grant Scheme, FRGS/1/2018/TK10/UNIMAP/02/11, from the Ministry of Higher Education (MoHE), Malaysia and Universiti Malaysia Perlis (UniMAP), Malaysia.

Data Availability Statement: Not applicable.

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

Abbreviations

The following abbreviations are used in this manuscript:

$\mathcal{L}_b$	Index of the number N of BSs.
$\mathcal{L}_i$	Index of the number K_i of information users.
$\mathcal{L}_e$	Index of the number K_e of active energy users (EUs).
$\mathcal{L}_e^{[\text{idle}]}$	Index of the number $K_e^{[\text{idle}]}$ of idle EUs.
$P_n^{[\text{Tx}]}$	Total transmit power at the n-th RRH.
$P_n^{[\text{circ}]}$	Hardware circuit power consumption at the n-th RRH.
$P_{\text{CU}}^{[\text{circ}]}$	Hardware circuit power consumption at the CU.
$P_n^{[\text{Tmax}]}$	Maximum transmit power allowance of the n-th RRH.
$P_{\text{CU}}^{[\text{max}]}$	Maximum power provision by the grid at the CU.
$\mathcal{T}$	Index of the number T of time slots.
$\mathcal{F}$	Index of the number F of frames within a time slot.
$\mathcal{K}$	Index of the number K of trials within a frame.
$B_n^{[\text{ahead}]}$	Amount of energy purchased from the day-ahead market (Arm).
$B_n^{[\text{spot}]}$	Amount of energy to be purchased from the spot-market.
S_n	Amount of excessive energy to be sold back to the grid.
E_n	Amount of renewable energy generation at the n-th RRH.
$B_n^{[\text{total}]}(k)$	Total energy cost of the n-th RRH at the k-th trial.
$\mathcal{E}^{[\text{total}]} = \{\mathcal{E}^1, \cdots, \mathcal{E}^J\}$	All energy packages (arms) offered by the grid in the day-ahead market.
$\mu_{n,p}^{[k,f,t]} = \mathcal{R}(B_n^{[\text{ahead}]}(k))$	Reward associated with arm $B_n^{[\text{ahead}]}$ at the k-th trial of the f-th frame at the t-th time slot.
$\mathcal{A}_k^{[\text{set}]} = \{B_1^{[\text{ahead}]}(k), \cdots, B_N^{[\text{ahead}]}(k)\}$	N energy packages purchased a day ahead at the k-th trial (super arm).
$\mathcal{R}(\mathcal{A}_k^{[\text{set}]})$	Reward for the super arm $\mathcal{A}_k^{[\text{set}]}$ at the k-th trial.
$\boldsymbol{\mu_n^{[k,f,t]}} = (\mu_{n,1}^{[k,f,t]}, \mu_{n,2}^{[k,f,t]}, \cdots, \mu_{n,J}^{[k,f,t]})$	Reward vector for the n-th RRH
$\boldsymbol{\hat{\mu}_n^{[f,t]}} = (\hat{\mu}_{n,1}^{[f,t]}, \hat{\mu}_{n,2}^{[f,t]}, \ldots, \hat{\mu}_{n,J}^{[f,t]})$	Estimated mean reward vector
$\boldsymbol{\bar{\mu}_n^{[f,t]}} = (\bar{\mu}_{n,1}^{[f,t]}, \bar{\mu}_{n,2}^{[f,t]}, \ldots, \bar{\mu}_{n,J}^{[f,t]})$	Adjusted reward vector of individual arms
$\mathcal{R}_t^{[\text{acc}]}$	Accumulated reward at time slot t.
$\mathcal{Q}_t$	Regret at time slot t.

References

1. Peng, M.; Li, Y.; Jiang, J.; Li, J.; Wang, C. Heterogeneous Cloud Radio Access Networks: A New Perspective for Enhancing Spectral and Energy Efficiencies. *IEEE Wirel. Commun.* **2014**, *21*, 126–135. [CrossRef]
2. Tan, Z.; Yang, C.; Song, J.; Liu, Y.; Wang, Z. Energy consumption analysis of C-RAN architecture based on 10G EPON front-haul with daily user behaviour. In Proceedings of the 2015 14th International Conference on Optical Communications and Networks (ICOCN), Nanjing, China, 3–5 July 2015.

3. Habibi, M.A.; Nasimi, M.; Han, B.; Schotten, H.D. A Comprehensive Survey of RAN Architectures Toward 5G Mobile Communication System. *IEEE Access* **2019**, *7*, 70371–70421. [CrossRef]
4. Mahapatra, R.; Nijsure, Y.; Kaddoum, G.; Hassan, N.U.; Yuen, C. Energy Efficiency Tradeoff Mechanism towards Wireless Green Communication: A Survey. *IEEE Commun. Surv. Tutor.* **2016**, *18*, 686–705. [CrossRef]
5. China Mobile Research Institute. *C-RAN: The Road Towards Green RAN*, version 2.5; White Paper; 2011. Available online: labs.chinamobile.com/cran/ (accessed on 30 March 2016).
6. Fehske, A.; Fettweis, G.; Malmodin, J.; Biczok, G. The Global Footprint of Mobile Communications: The Ecological and Economic Perspective. *IEEE Commun. Mag.* **2011**, *49*, 55–62. [CrossRef]
7. Bu, S.; Yu, F.R.; Cai, Y.; Liu, X.P. When the smart grid meets energy efficient communications: Green wireless cellular networks powered by the smart grid. *IEEE Trans. Wirel. Commun.* **2012**, *11*, 3014–3024. [CrossRef]
8. Chen, S.; Shroff, N.B.; Sinha, P. Energy trading in the smart grid: From end-user's perspective. In Proceedings of the 2013 Asilomar Conference on Signals, Systems and Computers, Pacific Grove, CA, USA, 3–6 November 2013 ; pp. 327–331.
9. Xu, J.; Zhang, R. CoMP meets smart grid: A new communication and energy cooperation paradigm. *IEEE Trans. Veh. Technol.* **2015**, *64*, 2476–2488. [CrossRef]
10. Xu, J.; Zhang, R. Cooperative Energy Trading in CoMP Systems Powered by Smart Grids. In Proceedings of the 2014 IEEE Global Communications Conference, Austin, TX, USA, 8–12 December 2014; pp. 2697–2702. [CrossRef]
11. Wang, Y.; Saad, W.; Han, Z.; Poor, H.V.; Başar, T. A game-theoretic approach to energy trading in the smart grid. *IEEE Trans. Smart Grid* **2014**, *5*, 1439–1450. [CrossRef]
12. Leithon, J.; Lim, T.J.; Sun, S. Online energy management strategies for base stations powered by the smart grid. In Proceedings of the 2013 IEEE International Conference on Smart Grid Communications (SmartGridComm), Vancouver, BC, Canada, 21–24 October 2013; pp. 199–204. [CrossRef]
13. Ariffin, W.N.S.F.W.; Zhang, X.; Nakhai, M.R. Sparse Beamforming for Real-Time Resource Management and Energy Trading in Green C-RAN. *IEEE Trans. Smart Grid* **2016**, *8*, 2022–2031. [CrossRef]
14. Dall'Anese, E.; Zhu, H.; Giannakis, G.B. Distributed Optimal Power Flow for Smart Microgrids. *IEEE Trans. Smart Grid* **2013**, *4*, 1464–1475. [CrossRef]
15. Ariffin, W.N.S.F.W.; Zhang, X.; Nakhai, M.R. Combinatorial multi-armed bandit algorithms for real-time energy trading in green C-RAN. In Proceedings of the 2016 IEEE International Conference on Communications (ICC), Kuala Lumpur, Malaysia, 23–27 May 2016; pp. 1–6. [CrossRef]
16. Zhang, X.; Nakhai, M.R.; Ariffin, W.N.S.F.W. A Bandit Approach to Price-Aware Energy Management in Cellular Networks. *IEEE Commun. Lett.* **2017**, *21*, 1609–1612. [CrossRef]
17. Ariffin, W.N.S.F.W.; Zhang, X.; Nakhai, M.R. Sparse beamforming for real-time energy trading in CoMP-SWIPT networks. In Proceedings of the 2016 IEEE International Conference on Communications (ICC), Kuala Lumpur, Malaysia, 23–27 May 2016; pp. 1–6. [CrossRef]
18. Dai, B.; Yu, W. Sparse Beamforming and User-Centric Clustering for Downlink Cloud Radio Access Network. *IEEE Access Recent Adv. Cloud RAN* **2014**, *2*, 1326–1339.
19. Ng, D.W.K.; Schober, R. Resources Allocation for Coordinated Multipoint Networks with Wireless Information and Power Transfer. In Proceedings of the 2014 IEEE Global Communications Conference, Austin, TX, USA, 8–12 December 2014; pp. 4281–4287. [CrossRef]
20. Candes, E.; Wakin, M.; Boyd, S. Enhancing Sparsity by Reweighted ℓ_1 Minimization. *J. Fourier Anal. Appl.* **2008**, *14*, 877–905. [CrossRef]
21. Boyd, S.; Vandenberghe, L. *Convex Optimization*; Cambridge University Press: Cambridge, UK, 2004.
22. Vandenberghe, L.; Boyd, S. Semidefinite programming. *SIAM Rev.* **1996**, *38*, 4995. [CrossRef]
23. Nesterov, Y.; Nemirovsky, A. *Interior-Point Polynomial Methods in Convex Programming*; Studies in Applied Mathematics; Society for Industrial and Applied Mathematics: Philadelphia, PA, USA, 1994; Volume 13.
24. Blasco, P.; Gunduz, D. Learning-Based Optimization of Cache Content in a Small Cell Base Station. In Proceedings of the 2014 IEEE International Conference on Communications (ICC), Sydney, Australia, 10–14 June 2014; pp. 1897–1903. [CrossRef]
25. Chen, W.; Wang, Y.; Yuan, Y. Combinatorial multi-armed bandit: General framework, results and applications. In Proceedings of the International Conference on Machine Learning, Atlanta, GA, USA, 16–21 June 2013.
26. Yang, Y.; Zhu, D. Randomized Allocation with nonparametric estimation for a multi-armed bandit problem with covariates. *Ann. Statist.* **2002**, *30*, 100–121.
27. Ariffin, W.N.S.F.W.; Zhang, X.; Nakhai, M.R. Real-time power balancing in green CoMP network with wireless information and energy transfer. In Proceedings of the 2015 IEEE 26th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC), Hong Kong, 30 August–2 September 2015; pp. 1574–1578. [CrossRef]
28. Grant, M.; Boyd, S. CVX: Matlab Software for Disciplined Convex Programming, Version 2.0 (Beta). March 2013. Available online: http://cvxr.com/cvx/ (accessed on 30 January 2014).