Adaptive Energy Storage Management in Green Wireless Networks

Xinruo Zhang, Mohammad Reza Nakhai and Wan Nur Suryani Firuz Wan Ariffin

***Abstract*—Time-varying wireless channel as well as the variability of renewable energy supply and energy prices are practically unknown in advance. To address such dynamic statistics of wireless networks, this paper develops an adaptive strategy inspired by combinatorial multi-armed bandit model for energy storage management and cost-aware coordinated load control at the base stations. The proposed strategy makes online foresighted decisions on the amount of energy to be stored in storage to minimize the average energy cost over long time horizon. Simulation results validate the superiority of the proposed strategy over a recently proposed storage-free learning-based design.**

I. INTRODUCTION

Relying only on the fossil-fuel-based electric energy generation to power next generation dense wireless networks will significantly contribute to the global carbon footprint [1]. This coupled with the increasing operational expenditure of the network caused by the ever-increasing energy demand motivate the need to integrate the renewable energy generated from natural sources with the conventional electric grid to power base stations (BSs). However, renewable generation is naturally uncertain and the sudden ramp-up of a power plant to compensate for the real-time energy shortage of the BSs can be expensive or technically infeasible [2]. The dynamic nature of renewable generation not only introduces significant fluctuations on the electricity price, but can also destabilize the reliable and cost-efficient operation of the BSs supplied by hybrid grid/renewable energy generators. Deploying energy storage devices can address these concerns by providing the opportunity to achieve a cost efficient balance between the supply and demand at BSs. The authors in [2] formulate the system as a simplified two-level Stackelberg game and the authors in [3] conclude the demand-side power management solutions for a single BS in the smart-grid-powered green CoMP. The authors in [4] study energy allocation problem for renewable energy powered BSs using a noncooperative game. However, their designs require statistics of the system dynamics, which is not a realistic assumption in practice. Furthermore, none of these designs consider the impact of online learning in energy storage management without requiring upfront statistical knowledge and merely relying on learning the system dynamics during operation. Requiring no prior knowledge of traffic, the authors in [5] develop an adaptive resource management in vehicular access network. Assuming prior knowledge of statistical distribution of demand load, the authors in [6] propose an online learning algorithm for stochastic storage management in smart grid rather than cellular network based on Markov decision process. Using stochastic optimization rather than online learning over infinite time horizon, the authors in [7] propose a dynamic energy management scheme for the smart-grid-powered CoMP, where BSs are fully cooperated and governed by a central processor (CP). However, such full cooperation design requires high speed energy-hungry backhaul links to deliver user's data from the CP to all BSs. Authors in [8], [9] propose energy management designs based on sparse beamforming technique for partial BS cooperation to relax the backhaul link capacity without consideration of either the renewable energy dynamics or the deployment of energy storage devices. This paper develops an adaptive energy storage management strategy to integrate the intermittent nature of the renewable energy generators with the conventional electric grid to minimize the energy consumption cost at wireless networks in the long run. This is a challenging task since 1) the state of each storage is only known to the corresponding BS, 2) the actions of BSs are coupled in a complex way that affect the overall energy cost, 3) the storage charging decisions have strong temporal correlations, i.e., the current decisions affect the future energy consumption costs. We introduce a novel adaptive algorithm that iteratively alternates between two decision making layers by exchanging conjectured information. The first layer located at the CP designs the overall transmission strategy across the network of BSs using a convex semidefinite programming (SDP) and the second layer designs the pre-charging strategies for storages at distributed BSs via online learning, i.e., a combinatorial multi-armed bandit (CMAB) approach.

II. SYSTEM MODEL

This paper considers a downlink green wireless network, where a set of $\mathcal{L}_b = \{1, \cdots, N\}$ adjacent M-antenna BSs partially cooperated to serve a set of $\mathcal{L}_i = \{1, \cdots, K_i\}$ single-antenna user terminals (UTs) over a shared bandwidth as per their power budgets and backhaul link capacity restrictions. The CP designs all transmission strategies for the BSs based on perfect channel state information. Assume that the individual BSs are equipped with energy storage devices and are powered by local renewable energy generators, energy storage devices and the grid at various energy prices. The storage-deployed BSs not only prevent the shortage of energy, but also enable the optimization of time-average energy cost via charging the storage either from the grid in advance at cheaper price or from

The authors are with Centre for Telecommunications Research, Department of Informatics, King's College London, Strand, WC2R 2LS, UK. E-mail: [xinruo.zhang, reza.nakhai, k1206546]@kcl.ac.uk

the excessive renewable energy. Let the time horizon T be divided into discrete time slots, indexed as $\mathcal{T} = \{1, \cdots, T\}$, and assume the renewable energy supply varies across time slots but remains invariant within each time slot.

A. Downlink Transmission Model

Let $\mathbf{w}_{ni} \in \mathbb{C}^{M\times 1}$ and $\mathbf{h}_{ni} \in \mathbb{C}^{M\times 1}$ denote, respectively, the beamformer and the channel vector from the n-th BS to the i-th UT, $n \in \mathcal{L}_b, i \in \mathcal{L}_i$. Then, the signal-to-interference-plus-noise ratio (SINR) at the i-th UT, can be expressed as $\text{SINR}_i = \frac{|\sum\limits_{n\in\mathcal{L}_b}(\mathbf{h}_{ni}^H\mathbf{w}_{ni})|^2}{\sum\limits_{j\neq i, j\in\mathcal{L}_i}|\sum\limits_{n\in\mathcal{L}_b}(\mathbf{h}_{ni}^H\mathbf{w}_{nj})|^2+\sigma_i^2}$, where σ_i^2 is the additive white Gaussian noise variance. Let the binary function $\left\|\|\mathbf{w}_{ni}\|_2^2\right\|_0$ denote the scheduling choices between the n-th BS and the i-th UT, where $\|\mathbf{w}_{ni}\|_2^2 = 0$ indicates that the i-th UT is not served by the n-th BS. Then the backhaul capacity consumption of the n-th BS is given by $B_n^{[\text{backhaul}]} = \sum\limits_{i\in\mathcal{L}_i}\left\|\|\mathbf{w}_{ni}\|_2^2\right\|_0 R_i$, where $R_i = \log_2(1+\text{SINR}_i)$ is achievable data rate for the i-th UT.

B. Energy Storage Management Model

Assume that an amount of $G_n(t)$ units of renewable energy is generated at the n-th BS at the t-th time slot, $t \in \mathcal{T}$. Let the amounts of $E_n^{[\text{s}]}(t)$ and $E_n^{[\text{c}]}(t)$ denote the units of the initial energy contents of the storage in the beginning of the t-th time slot and the units of energy charged to the storage of the n-th BS prior to the actual time of energy demand at the t-th time slot, respectively. Notice that $E_n^{[\text{s}]}(t)+E_n^{[\text{c}]}(t) \in [0,\ E_n^{[\text{capacity}]}]$, where $E_n^{[\text{capacity}]}$ is the upper limit of the storage capacity at the n-th BS. Let an amount of $E_n^{[\text{r}]}(t)$ units of energy be the energy shortage to be real-time supplied by the grid to the n-th BS at the t-th time slot. From the supply and demand perspective, let $\pi^{[\text{r}]} \geq \pi^{[\text{c}]} \geq \pi^{[\text{g}]} \geq \pi^{[\text{s}]}$, where $\pi^{[\text{r}]}, \pi^{[\text{c}]}, \pi^{[\text{g}]}$ and $\pi^{[\text{s}]}$ be, respectively, the per unit energy prices for $E_n^{[\text{r}]}(t)$, $E_n^{[\text{c}]}(t)$, $G_n(t)$ and $E_n^{[\text{s}]}(t)$. Then, the total energy cost of the n-th BS at the t-th time slot, i.e., $C_n^{[\text{total}]}(t)$, is given by

$$C_n^{[\text{total}]}(t) = \pi^{[\text{r}]}E_n^{[\text{r}]}(t) + \pi^{[\text{c}]}E_n^{[\text{c}]}(t) + \pi^{[\text{g}]}G_n(t) + \pi^{[\text{s}]}E_n^{[\text{s}]}(t). \tag{1}$$

Let $P_n^{[\text{Tx}]}(t) = \sum\limits_{i\in\mathcal{L}_i}||\mathbf{w}_{ni}||_2^2$ and $P_n^{[\text{c}]}$ denote, respectively, the total transmit power from the n-th BS at the t-th time slot and the hardware circuit power consumption at the n-th BS. Then the total energy consumption of the n-th BS at the t-th time slot is upper-bounded by the energy budget at the n-th BS, as

$$P_n^{[\text{Tx}]}(t) + P_n^{[\text{c}]} \leq G_n(t) + E_n^{[\text{s}]}(t) + E_n^{[\text{c}]}(t) + E_n^{[\text{r}]}(t). \tag{2}$$

Thus, the initial energy storage of the n-th BS at the t-th time slot is constrained by the following expression:

$$\begin{aligned} E_n^{[\text{s}]}(t) = \min\{&E_n^{[\text{capacity}]}, \max\{G_n(t-1) + E_n^{[\text{s}]}(t-1) \\ &+E_n^{[\text{c}]}(t-1) + E_n^{[\text{r}]}(t-1) - P_n^{[\text{Tx}]}(t-1) - P_n^{[\text{c}]},\ 0\}\}. \end{aligned} \tag{3}$$

III. ADAPTIVE STORAGE MANAGEMENT STRATEGY

In this section, we will introduce an adaptive storage management algorithm used jointly by the CP to iteratively update the downlink beamforming vectors, i.e., $\mathbf{w}_{ni}, n \in \mathcal{L}_b, i \in \mathcal{L}_i$, at BSs as well as the amount of real-time energy purchase from the grid, i.e., $E_n^{[\text{r}]}(t), t \in \mathcal{T}$, and by the individual distributed BSs to update their strategies of charging their locally installed storages , i.e., $E_n^{[\text{c}]}(t), t \in \mathcal{T}$, in order to efficiently compensate for the randomness of the renewable generators. Individual BSs send their conjectured amount of required storage charges $E_n^{[\text{c}]}(t)$ to the CP and receive the corresponding instantaneous reward/regret from the CP. This process of iterative exchange of data allows the proposed adaptive algorithm to converge to optimal conjectured optimization variables, i.e., the beamforming vectors $\mathbf{w}_{ni}$, the amount of real-time energy provisioning from the grid $E_n^{[\text{r}]}(t)$ and the amount of charge to be deposited to the storage devices at a current time slot $E_n^{[\text{c}]}(t)$.

A. Problem Formulation

We formulate the problem of adaptive storage energy management as reinforcement learning problem based on CMAB model that insures the time-averaged cost efficiency of the BSs over a time horizon "T". The proposed adaptive algorithm is governed by a trade-off between exploring new sets of arms and exploiting the best set of arms that maximizes the accumulated reward at individual BSs, in a CMAB reinforcement learning model [10]. Let us consider a set of arms denoted as a super arm, where each arm corresponds to an energy size to be stored in the storage of a BS in advance of the actual time that the shortage of energy may occur. A super arm is comprised of N arms chosen for N BSs out of P possible arms, i.e., $N \subset P$. Let us define the reward of the arm chosen for the n-th BS at time slot t, as

$$\mathcal{R}_n(t) = C_n^{[\text{total}]}(0) - C_n^{[\text{total}]}(t),\ \forall n \in \mathcal{L}_b, t \in \mathcal{T}, \tag{4}$$

where $C_n^{[\text{total}]}(0)$ and $C_n^{[\text{total}]}(t)$ are the total energy cost of the n-th BS at the initial time slot and at the t-th time slot, respectively. The proposed CMAB based adaptive algorithm maximizes the time-averaged accumulated reward over the online decisions on the amount of electricity to be stored in the storage devices of individual BSs, as

$$\max_{E_n^{[\text{c}]}(t)}\left\{\lim_{T\to\infty}\frac{1}{T}\sum_{t=0}^{T-1}\mathcal{R}_n(t)\right\}. \tag{5}$$

The energy consumption at individual BSs is governed by the following optimization problem for resource allocation.

$$\begin{aligned} \min_{\mathbf{w}_{ni}, E_n^{[\text{r}]}(t)} \quad & \sum_{n\in\mathcal{L}_b} P_n^{[\text{Tx}]}(t) + \max_{n\in\mathcal{L}_b}\left\{E_n^{[\text{r}]}(t)\right\} \\ \text{s.t.} \quad & \text{C1}: \text{SINR}_i(t) \geq \gamma_i, \quad \forall i \in \mathcal{L}_i, \\ & \text{C2}: B_n^{[\text{backhaul}]}(t) \leq B_n^{[\text{limit}]}, \quad \forall n \in \mathcal{L}_b, \\ & \text{C3}: P_n^{[\text{Tx}]}(t) + P_n^{[\text{c}]} - G_n(t) - E_n^{[\text{s}]}(t) - E_n^{[\text{c}]}(t) \\ & \qquad\qquad \leq E_n^{[\text{r}]}(t), \quad \forall n \in \mathcal{L}_b, \\ & \text{C4}: P_n^{[\text{Tx}]}(t) \leq P_n^{[\text{Tmax}]},\ \text{C5}: E_n^{[\text{r}]}(t) \geq 0,\ \forall n \in \mathcal{L}_b, \end{aligned} \tag{6}$$

where C1 denotes the SINR requirements for the UTs and C2 is the backhaul link capacity limitations for each BS. C3 indicates that the energy shortage of the n-th BS will be provisioned by the grid as per (2), whilst $E_n^{[\mathrm{s}]}(t)$ is updated as per (3) in the beginning of the t-th time slot. C4 enforces the maximum transmit power allowance $P_n^{[\mathrm{Tmax}]}$ at the n-th BS. The objective function in (6) seeks an optimal schedule for beamforming vectors $\{\mathbf{w}_{ni}(t)\}_{t,i}$ and the real-time energy shortage of the n-th BS $\{E_n^{[\mathrm{r}]}(t)\}_t$ in order to minimize the total network cost $\sum_{n\in\mathcal{L}_b} P_n^{[\mathrm{Tx}]}(t)$. The purpose of the online learning part of the proposed algorithm at individual BSs is to determine proactively the optimal conjectured amount of storage charging, i.e., $E_n^{[\mathrm{c}]}(t)$, ahead of time, before experiencing a possible energy shortage at the time slot "t", such that when the CP reacts based on that, the resulting transmission strategy, i.e., the beamforming vectors $\{\mathbf{w}_{ni}(t)\}_{t,i}$ and the supporting real-time amount of energy purchase from the grid $\{E_n^{[\mathrm{r}]}(t)\}_t$, minimizes the overall energy cost of the network.

B. Reweighted ℓ_1-norm and SDP

The intractable ℓ_0-norm in constraint C2 of problem (6), is handled with *reweighted ℓ_1-norm method* [8], as $B_n^{[\mathrm{backhaul}]} \approx \sum_{i\in\mathcal{L}_i} \left\|[\xi_{ni}\|\mathbf{w}_{ni}\|_2^2]\right\|_1 R_i = \sum_{i\in\mathcal{L}_i} \xi_{ni}\mathrm{tr}(\mathbf{w}_{ni}\mathbf{w}_{ni}^H)R_i$, where the cooperative links between BSs and UTs will be gradually removed as per backhaul capacity constraints, via alternating between computing the optimal beamformers $\mathbf{w}_{ni}^*$ of problem (7) for a given weighting factors ξ_{ni}, and adjusting $\xi_{ni} = \frac{1}{\mathrm{tr}(\mathbf{w}_{ni}^*\mathbf{w}_{ni}^{*H})+\mu}$ and R_i based on $\mathbf{w}_{ni}^*$. Let us define the rank-one semidefinite matrices $\mathbf{W}_{ni} = \mathbf{w}_{ni}\mathbf{w}_{ni}^H$ and $\mathbf{H}_{ni} = \mathbf{h}_{ni}\mathbf{h}_{ni}^H$. The problem in (6) can be transformed as SDP after relaxing the rank-one constraints of $\mathrm{rank}(\mathbf{W}_{ni}) = 1$, as

$$\min_{\mathbf{W}_{ni}\succeq 0,\chi} \quad \sum_{n\in\mathcal{L}_b}\sum_{i\in\mathcal{L}_i} \mathrm{tr}(\mathbf{W}_{ni}) + \chi \tag{7}$$

$$\text{s.t.} \quad \mathrm{C1}: \gamma_i^{-1}\mathrm{tr}\Big(\sum_{n\in\mathcal{L}_b}\mathbf{H}_{ni}\mathbf{W}_{ni}\Big) \geq \sum_{j\in\mathcal{L}_i, j\neq i}\mathrm{tr}\Big(\sum_{n\in\mathcal{L}_b}\mathbf{H}_{ni}\mathbf{W}_{nj}\Big) + \sigma_i^2, \forall i\in\mathcal{L}_i,$$

$$\mathrm{C2}: \sum_{i\in\mathcal{L}_i}\xi_{ni}\mathrm{tr}(\mathbf{W}_{ni})R_i \leq B_n^{[\mathrm{limit}]}, \quad \forall n\in\mathcal{L}_b,$$

$$\mathrm{C3}: \sum_{i\in\mathcal{L}_i}\mathrm{tr}(\mathbf{W}_{ni}) + P_n^{[\mathrm{c}]} - E_n^{[\mathrm{s}]}(t) - E_n^{[\mathrm{c}]}(t) - G_n(t) \leq E_n^{[\mathrm{r}]}(t),\ \forall n\in\mathcal{L}_b,$$

$$\mathrm{C4}: \sum_{i\in\mathcal{L}_i}\mathrm{tr}(\mathbf{W}_{ni}) \leq P_n^{[\mathrm{Tmax}]}, \quad \forall n\in\mathcal{L}_b,$$

$$\mathrm{C5}: E_n^{[\mathrm{r}]}(t) \geq 0, \quad \mathrm{C6}: E_n^{[\mathrm{r}]}(t) \leq \chi,\ \forall n\in\mathcal{L}_b.$$

Lemma 1: The optimal solutions to the problems (7) satisfy $\mathrm{rank}(\mathbf{W}_{ni}^*) = 1$ with probability one.

Proof: Please refer to a similar proof in [8].

C. Proposed Online Learning Algorithm

Let $\mathcal{K} = \{1,\cdots,K\}$ denote the set of indexes used to identify the learning (exploration) iterations during a time slot, $\mathcal{P} = \{1,\cdots,P\}$ be the set of indexes associated to P arms, i.e., P discrete energy sizes (energy packets) $\{\mathcal{E}^1,\cdots,\mathcal{E}^P\}$, such that $\mathcal{E}^p = \mathcal{E}^{p-1} + \Delta\mathcal{E}, p\in\mathcal{P}$. In the k-th learning iteration of an exploration time slot, $k\in\mathcal{K}$, let the chosen super arm that consists of N energy packets to be stored at N BSs' storage devices, be denoted by the set $\mathcal{S}^{[\mathrm{set}]}(k) = \{E_1^{[\mathrm{c}]}(k),\cdots,E_N^{[\mathrm{c}]}(k)\}$. Let us define the reward of the arm selected for the n-th BS at the t-th time slot as

$$\mathcal{R}_t(E_n^{[\mathrm{c}]}(k)) = C_n^{[\mathrm{total}]}(0) - C_n^{[\mathrm{total}]}(k),\ \forall n\in\mathcal{L}_b, t\in\mathcal{T}, \tag{8}$$

where $C_n^{[\mathrm{total}]}(0)$ and $C_n^{[\mathrm{total}]}(k)$ are the total energy cost of the n-th BS in the first learning iteration of the initial time slot, t_0, and in the k-th learning iteration of current time slot, t, respectively, as per (1). The steps of the proposed online learning procedure during the time slots allocated for the exploration are detailed in Algorithm 1. Algorithm 1 explores a new super arm and assigns a set of energy packages to the BSs' storage devices for the next learning iteration based on the rewards acquired from the current and the previous learning iterations. Once a predefined number of K learning iterations are accomplished, the mean rewards for individual arms assigned to the n-th BS's storage device during the t-th time slot, i.e., $\hat{\mathbf{r}}_{\mathbf{n}}^{[\mathbf{t}]}$, are estimated and adjusted as per the steps 8 and 9 of Algorithm 2, respectively. The adjustment step 9 in Algorithm 2 implements the trade-off between exploiting the set of arms resulted in the highest accumulated reward so far and exploring new sets of arms that are not frequently selected and may result in a better accumulated reward during the future time slots. The proposed algorithm by design is not sensitive to the time scale due to the fact that the exploration cycle of Algorithm 2 responds to the variation in the environment by making adaptive decisions of $E_n^{[\mathrm{c}]}(t)$ for the upcoming exploitation cycles based on long-term time averaged accumulated rewards with a discount factor of $\mathcal{D}$, as detailed in step 13 in Algorithm 2.

Algorithm 1 *Super arm exploration at the t-th time-slot*

1: **For** $k = 1:K$
2: Solve problem in (7),
3: Compute $C_n^{[\mathrm{total}]}(k)$ as per (1) and $\mathcal{R}_t(E_n^{[\mathrm{c}]}(k))$ as per (8),
4: **if** $k=1$ (initial iteration) and $E_n^{[\mathrm{c}]}(k) \neq \mathcal{E}^P$
5: **then** $E_n^{[\mathrm{c}]}(k+1) = E_n^{[\mathrm{c}]}(k) + \Delta\mathcal{E},\ n\in\mathcal{L}_b,$
6: **else if** $\mathcal{R}_t(E_n^{[\mathrm{c}]}(k)) < \mathcal{R}_t(E_n^{[\mathrm{c}]}(k-1))$,
7: **then** Backward search as $E_n^{[\mathrm{c}]}(k+1) = E_n^{[\mathrm{c}]}(k) - \Delta\mathcal{E}$,
8: **else if** $\mathcal{R}_t(E_n^{[\mathrm{c}]}(k)) > \mathcal{R}_t(E_n^{[\mathrm{c}]}(k-1))$
9: **then** Forward search as $E_n^{[\mathrm{c}]}(k+1) = E_n^{[\mathrm{c}]}(k) + \Delta\mathcal{E}$,
10: **else** $E_n^{[\mathrm{c}]}(k+1) = E_n^{[\mathrm{c}]}(k)$,
11: **end if**
12: Compute the arm index p as $p = \frac{E_n^{[\mathrm{c}]}(k)}{\Delta\mathcal{E}}, n\in\mathcal{L}_b$,
13: Update the reward vector of the n-th BS in the k-th iteration, i.e., $\mathbf{r}_{\mathbf{n}}^{[\mathbf{k},\mathbf{t}]} = (r_{n,1}^{[k,t]}, r_{n,2}^{[k,t]},\cdots,r_{n,P}^{[k,t]})$, as
$r_{n,p}^{[k,t]} = \mathcal{R}_t(E_n^{[\mathrm{c}]}(k)),\ \forall p\in\mathcal{P}, n\in\mathcal{L}_b,$
14: Update super arm for next iteration as
$\mathcal{S}^{[\mathrm{set}]}(k+1) = \{E_1^{[\mathrm{c}]}(k+1),\cdots,E_N^{[\mathrm{c}]}(k+1)\}.$
15: **End for**

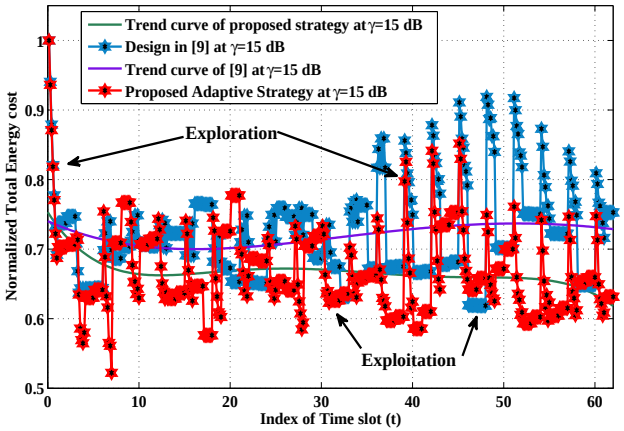
Algorithm 2 *Adaptive storage management algorithm*

1: **For** $t = 1 : T$
2: **if** $t = 1$ (initial time slot)
3: **then** Initialize the super arm for the first iteration ($k = 1$) as $\mathcal{S}^{[\text{set}]}(1) = \{0_1, \cdots, 0_N\}$ and $E_n^{[\text{s}]}(1) = 0$,
4: **else** $\mathcal{S}^{[\text{set}]*}(1) = \Delta\mathcal{E}[p_1^*, p_2^*, \cdots, p_N^*]$,
5: **end if**
6: **if** t is **Exploration**
7: **then** Run Algorithm 1,
8: **Estimation Stage :**
Compute the estimated mean reward vector, i.e., $\hat{\mathbf{r}}_{\mathbf{n}}^{[\mathbf{t}]} = (\hat{r}_{n,1}^{[t]}, \hat{r}_{n,2}^{[t]}, \ldots, \hat{r}_{n,P}^{[t]})$, as $\hat{r}_{n,p}^{[t]} = \frac{\sum_{k=1}^{K} r_{n,p}^{[k,t]}}{K}$, $\forall p \in \mathcal{P}$,
9: **Adjustment Stage :**
Update adjusted reward $\bar{\mathbf{r}}_{\mathbf{n}}^{[\mathbf{t}]} = (\bar{r}_{n,1}^{[t]}, \bar{r}_{n,2}^{[t]}, \ldots, \bar{r}_{n,P}^{[t]})$, as $\bar{r}_{n,p}^{[t]} = \hat{r}_{n,p}^{[t]} + \sqrt{\frac{3\ln t}{2N_p(t)}}$, where $N_p(t)$ is number of times the p-th arm has been played by the t-th time slot,
10: **else if** t is **Exploitation**
11: Solve problem in (7),
12: **end if**
13: Average $\bar{\mathbf{r}}_{\mathbf{n}}^{[\mathbf{t}]}$ over accumulated number of time slots, as $\bar{\mathbf{r}}_{\mathbf{n}} = \frac{\sum_{t'=1}^{t} \bar{\mathbf{r}}_{\mathbf{n}}^{[\mathbf{t}']} \mathcal{D}^{(t-t')}}{t} = [\bar{r}_{n,1}, \bar{r}_{n,2}, \cdots, \bar{r}_{n,P}], n \in \mathcal{L}_b$.
14: For the next time slot: find N optimum arm indexes as $p_n^* = \underset{p}{\text{argmax}}(\bar{r}_{n,p}), p \in \mathcal{P}, \forall n \in \mathcal{L}_b$.
15: **End for**

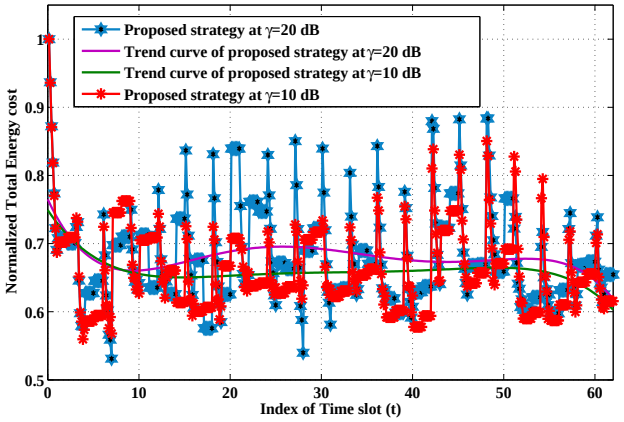
IV. SIMULATION RESULTS

Consider a downlink network comprised of 3 adjacent 8-antenna BSs that transmit towards 6 single-antenna UTs. The renewable energy supply at BSs at each time slot varies as $G_1 \in [0.5\ 1.0]$ W, $G_2 \in [0.1\ 0.5]$ W and $G_3 \in [0.03\ 0.1]$ W, respectively, at $\pi^{[\text{g}]} = £0.05/\text{W}$. Other simulation parameters are $\pi^{[\text{c}]} = £0.07/\text{W}$, $\pi^{[\text{r}]} = £0.15/\text{W}$, $\pi^{[\text{s}]} = £0.01/\text{W}$, $\mathcal{D} = 0.95$, $E_n^{[\text{capacity}]} = P_n^{[\text{c}]} = 30$ dBm, $P_n^{[\text{Tmax}]} = 46$ dBm, $B_n^{[\text{limit}]} = 35$ bits/s/Hz, $P = 20$ with $\Delta\mathcal{E} = 100$ mW and the exploration-exploitation trade-off is 1:3. The proposed algorithm is simulated with $K = 7$ iterations averaging over $F = 20$ independent channel realizations for $T = 62$ time slots. The simulation results are obtained via CVX [11] using Intel i7-3770 CPU@3.4GHz with 8GB RAM and the running time for per learning iteration is 7 seconds.

The normalized total energy cost of the proposed strategy at $\gamma = 15$ dB is compared in Fig. 1(a) against a simplified CMAB based storage-free design in [9]. For fair comparison, the overall energy cost is normalized to the energy cost at the initial iteration of the proposed algorithm. The burst at the start of an exploration cycle is due to the uncertain renewable energy generation and the perturbation in step 9 of Algorithm 2 to give priority to explore the less-explored arms. As shown in Fig. 1(a), the polynomial trend curves, fitted onto the actual experimental points, approximately indicate that the averaged performance of the proposed strategy achieves, respectively, 34 percent and 10 percent improvements over its initial learning state and the design in [9]. Furthermore, as the time-slot index increases, the design in [9] indicates larger variations in total energy cost and worse average performance than the proposed strategy. This is due to the single-directional search and the storage-free nature of the design in [9], which provides poorer adaptation to the wireless channel dynamics and variations in renewable generation. The proposed algorithm is evaluated at two more different targets of SINR in Fig. 1(b). It is shown that the average performance of the proposed algorithm slightly degrades as the target SINR increases within a substantial dynamic range, i.e., from $\gamma = 10$ dB to $\gamma = 20$ dB.



(a) Proposed strategy versus design in [9] at $\gamma = 15$ dB



(b) Proposed strategy at $\gamma = 10$ dB and $\gamma = 20$ dB

Fig. 1. Normalized total energy cost at individual time slots

V. CONCLUSION

This paper studies the problem of adaptive energy storage management at BSs in the presence of uncertain renewable energy generation and dynamic wireless channel environment. We adopt a CMAB model to formulate the problem as a combination of online learning and optimal cost-aware energy coordination amongst the BSs to minimize the network cost over an infinite time horizon. We introduce a storage management algorithm to address the uncertain variations in energy supply and energy prices via adaptive power balancing at BSs. Simulation results confirm a significant performance gain over a recent learning-based design.

REFERENCES

[1] A. Fehske, G. Fettweis, J. Malmodin, and G. Biczok, "The global footprint of mobile communications: The ecological and economic perspective," *IEEE Communications Magazine*, vol. 49, no. 8, pp. 55–62, Aug. 2011.

[2] S. Bu, F. Yu, Y. Cai, and X. Liu, "When the smart grid meets energy-efficient communications: Green wireless cellular networks powered by the smart grid," *IEEE Transactions on Wireless Communications*, vol. 11, no. 8, pp. 3014–3024, Aug. 2012.

[3] D. Niyato, X. Lu, and P. Wang, "Adaptive power management for wireless base stations in a smart grid environment," *IEEE Wireless Communications*, vol. 19, no. 6, pp. 44–51, Dec. 2012.

[4] D. Li, W. Saad, I. Guvenc, A. Mehbodniya, and F. Adachi, "Decentralized energy allocation for wireless networks with renewable energy powered base stations," *IEEE Transactions on Communications*, vol. 63, no. 6, pp. 2126–2142, Jun. 2015.

[5] N. Cordeschi, D. Amendola, M. Shojafar, and E. Baccarelli, "Performance evaluation of primary-secondary reliable resource-management in vehicular networks," *IEEE PIMRC*, Sep. 2014.

[6] Y. Zhang and M. Schaar, "Structure-aware stochastic storage management in smart grids," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 6, pp. 1098–1110, Dec. 2014.

[7] X. Wang, Y. Zhang, T. Chen, and G. Giannakis, "Dynamic energy management for smart-grid-powered coordinated multipoint systems," *IEEE Journal of Selected Areas in Communications*, vol. 34, no. 5, pp. 1348–1359, May 2016.

[8] W. N. S. F. W. Ariffin, X. Zhang, and M. R. Nakhai, "Sparse beamforming for real-time resource management and energy trading in green c-ran," *IEEE Transactions on Smart Grid*, Sep. 2016.

[9] W. N. S. F. W. Ariffin, X. Zhang, and M. R. Nakhai, "Combinatorial multi-armed bandit algorithms for real-time energy trading in green C-RAN," *IEEE ICC*, May 2016.

[10] W. Chen, Y. Wang, and Y. Yuan, "Combinatorial multi-armed bandit: General framework, results and applications," *International Conference on Machine Learning*, Jun. 2013.

[11] M. Grant and S. Boyd, *CVX: Matlab Software for Disciplined Convex Programming, Version 2.0 (Beta)*. [Online], 2013, availble:http://cvxr.com/cvx/doc/CVX.pdf.