IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received October 19, 2018, accepted November 5, 2018, date of publication November 13, 2018,
date of current version December 18, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2881073

Energy Saving and Interference Coordination in HetNets Using Dynamic Programming and CEC

JOSE A. AYALA-ROMERO, JUAN J. ALCARAZ, AND JAVIER VALES-ALONSO
Department of Information and Communications Technologies, Technical University of Cartagena, 30202 Cartagena, Spain

Corresponding author: Jose A. Ayala-Romero (josea.ayala@upct.es)

This work was supported by Project Grant AEI/FEDER TEC2016-76465-C2-1-R (AIM). The work of Jose A. Ayala-Romero was supported by personal Grant FPU14/03701.

ABSTRACT Energy efficiency in cellular networks has gained great relevance due to the increasing power supply demands in new generation heterogeneous network (HetNet). On the other hand, interference coordination is a key aspect in HetNets resource management which directly affects the performance and energy consumption. Although these two aspect are intertwined, they have been studied separately so far. In this paper, we address the joint problem of energy saving and interference coordination in HetNets. We formulate the problem as a finite horizon Markov decision process (MDP) leveraging two facts: the user traffic demands usually follow periodic patterns, and the knowledge and prediction of network load is crucial in order to select efficient network configurations. Quality of Service (QoS) in the network is defined as the ratio of users meeting a requirement specified by the operator, and allows us to account for a minimum QoS requirement by including a constraint in the formulation of the MDP. To address this MDP, we propose an approximate dynamic programming (ADP) algorithm which selects energy efficient farsighted configurations with QoS guarantees achieving near optimal performance. This ADP algorithm is built upon: 1) the certainty equivalent control principle, which simplifies the complexity of the MDP, and 2) machine learning techniques: a neural network and a polynomial regressor which allow us to predict the QoS and the consumption of the network in advance. We evaluate our proposal in a LTE-A network simulator following the 3GPP guidelines, and the results obtained show that a joint control of the energy saving and interference coordination mechanisms results in a notably performance improvement compared to a disjoint control, in terms of both energy savings and QoS guarantees. Moreover, our proposal has the advantage of being adaptable to the operator QoS requirements.

INDEX TERMS Dynamic programming, energy saving, green networking, heterogeneous networks, interference coordination.

I. INTRODUCTION

One of the most prominent features of 5G networks is expected to be a dense deployment of small cells to increase the spatial spectrum reuse [1]. However, small cell densification entails two problems: a higher energy consumption in the access network infrastructure, and an increased inter-cell interference [2]. These two issues have been addressed separately before, but they should not be decoupled because of two reasons: first, the energy saving mechanisms can affect the interference profile (particularly when these mechanisms switch base stations on or off). And second, because the interference coordination mechanism affects the traffic load and the available spectral resources at each station, which are aspects that the energy saving policies should consider and exploit.

This paper addresses energy saving at the base stations and interference coordination as a joint problem, using a stochastic control approach. We show that, addressing both problems simultaneously, a considerable improvement in energy saving is achieved.

Micallef *et al.* [2] explain that over 80% of the energy consumed in a mobile network is due to the base stations. There are, however, two aspects enabling energy saving strategies: the overlapping coverage areas of many cells and the variable traffic demands. By exploiting these aspects, some underutilized cells can be switched off without perceptible degradation on the user QoS. This strategy outperforms other approaches based on transmission power optimization, which incurs in load-independent energy consumption [3].

We consider a heterogeneous network (HetNet) composed of pico eNBs overlapping the macro eNB coverage area in each sector. The pico eNBs can be in active or sleeping mode (on or off), according to the energy saving mechanism, while the macro eNB is always active. The energy consumption model in this work is load dependent and considers an activation cost when a sleeping pico eNB is switched on. We apply the enhanced Inter Cell Interference Coordination (eICIC) mechanism defined by the 3GPP for LTE-A Networks [4]. The eICIC parameters are Cell Range Expansion (CRE) bias and Almost Blank Subframe (ABS) ratio.

We formulate the problem as a finite horizon Markov Decision Process (MDP) encompassing the daily pattern of traffic demands [5], [6]. This allows us to capture the variable traffic conditions during each day. However, the use of this formulation and inserting the eICIC parameters in the control space increases notably the complexity of the problem, compared to action spaces restricted to energy decisions [7], [8].

To address this complex problem, we propose an Approximate Dynamic Programming (ADP) algorithm that combines stochastic control and machine learning techniques. It comprises an offline phase where the cost-to-go function can be efficiently estimated from stored network data, and an online phase, where the algorithm selects, at each decision stage, efficient farsighted controls (considering their impact in future decision stages) with QoS guarantees.

The remainder of the paper is organized as follows. In section II the related work and contribution are discussed. In section III we describe the power consumption model and the interference coordination mechanism. The system model and the formulation of the problem are given in Section IV. In Section V we describe the proposed mechanism. The simulation framework and the numerical results are given in Section VI. Finally, the conclusions are summarized in Section VII.

II. RELATED WORK AND CONTRIBUTION

A. RELATED WORK

The optimization of energy efficiency decisions in HetNet has been previously addessed as an stochastic control problem. The most usual approach is to formulate the problem as a Markov Decision Process (MDP) including restrictions that account for the QoS objectives [7], [8], [21] as in our proposal. The inherent computational complexity of the MDPs implies the use of ADP approaches like Reinforcement Learning (RL) [21]. Other works [24] combine game theoretic techniques with RL algorithms.

Similarly, the stochastic control approach has also been used to address interference coordination in heterogeneous networks [25]–[27]. However, as far as we know, our work is the first addressing the joint control of energy saving and interference coordination in HetNets as a stochastic control problem. The improvement associated to the inclusion of interference coordination control can be seen in our numerical results in Section VI-D. Previously, Virdis *et al.* [18] addressed the energy saving problem in HetNets exploiting the ABS configuration, which strongly indicates that it has a significant impact in the power consumption, as our hypothesis. However, user association and eNB on-off switching are not considered in this work.

A common simplification (e.g. [7], [8], [21]) is to model the traffic intensity as a homogeneous Markov process. Nevertheless, real network traces exhibit a daily pattern [5], [6], [28]. Thus, we aim at exploiting this fact by formulating a finite horizon MDP with one-day horizon. This formulation allows us, in contrast to previous works, to apply the certainty equivalent control (CEC) which is an approximate dynamic programing technique to approximate the cost-to-go function with a deterministic model [29].

In order to reduce the problem dimensionality some works make a compact representation of the state space using a function approximation [21]. In our case, the control space is specially large because it comprises not only energy saving but also interference coordination decisions. For this reason, we propose to transform the state and control spaces into reduced dimensional ones achieving a notable reduction of the computational cost of both offline and online phases.

Other works [9]–[16], [18], [30] address the eNB on-off switching problem as an optimization problem. Since the computational complexity of these problems is NP-Hard, they are addressed using iterative algorithms or decompositing the problem in relaxed ones aimed at finding operative (but suboptimal) solutions [31]. An additional issue is that these solutions have to be recomputed whenever the state of the network changes (e.g. traffic intensity). In contrast, our solution operates for any network state. Some of the aforementioned works also consider the user association [10]–[13], but none of them take into account the interference management, which is an essential issue in HetNets.

The eNB switching problem has been also addressed using Stochastic Geometry [17], [19], [20], [32], [33]. However, this approach implies some simplifications in the network model (e.g. path loss as a channel model, eNBs deployed following a Poisson Point Process). In contrast, our proposal is capable of using real data from the network in both offline and online phases and does not require any simplification, neither does assumptions in the network model.

Regarding the power consumption model, some works [5], [28], [34] consider two values of power consumption (cell on or off). Other works [7], [21], [32] consider a more detailed model where the power consumption of a cell depends on its traffic load. In our work, we consider the load dependent consumption model proposed by the 3GPP [35] with additional consumption spikes associated to the activation of sleeping eNBs [6].

B. CONTRIBUTION

The main contributions of this paper are:

- To consider the joint control of energy saving and interference coordination mechanisms in HetNets leading to

TABLE 1. Comparison of energy saving related works.

	[5], [9]	[10]	[11]–[16]	[17]	[18]	[19]	[20]	[21]	[7]	[22], [23]	This work
eNB on-off Switching	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	No	**Yes**
User Association	No	Yes	Yes	No	No	Yes	Yes	Yes	No	Yes	**Yes**
Power Control	No	Yes	No	No	Yes	No	Yes	No	No	Yes	**Yes**
Interference Management	No	No	No	No	Yes	No	No	No	No	Yes	**Yes**
Approach	O	O	O	SG	O	SG	SG	RL	MDP	O	**DP**

(O): Optimization, (SG): Stochastic Geometry, (RL): Reinforcement Learning, (MDP): Markov Decision Process, (DP): Dynamic Programming

remarkable performance improvement compared to a disjoint control.

- To formulate the problem as a constrained finite horizon MDP leveraging two facts: the traffic demands usually follow periodic patterns, and the knowledge and prediction of network load is crucial in order to select efficient network configurations.
- To propose an ADP algorithm achieving near optimal performance in terms of energy consumption and QoS of the users, reducing energy consumption up to 24% with respect to a fixed default configuration.
- To build the ADP algorithm using a novel combination of stochastic control and machine learning techniques. In particular, we use Certainty Equivalent Control (CEC) in combination with Neural Networks and polynomial regression.

III. INTERFERENCE MANAGEMENT AND POWER CONSUMPTION MODEL

A. INTERFERENCE MANAGEMENT IN LTE-A: eICIC

The eICIC is an interference management mechanism for heterogeneous networks defined by the 3GPP from Release 10 (LTE-A) [4]. It schedules the radio resources for macro and pico eNBs in different time periods (called subframes) to avoid inter-cell interference. The eICIC comprises the following features: *Cell Range Expansion (CRE)* and *Almost Blank Subframe (ABS)*.

Pico eNBs are low transmission power cells aimed at increasing the spatial reuse of radio resources. However, LTE cell association is based on the received signal reference power (RSRP), that is, each UE associates with the highest reference signal received from each cell. Therefore, UEs located close to a pico could end up associating to the macro eNB, precluding spatial reuse. For this reason, CRE allows pico eNBs to extend their footprint by adding a bias to their RSRP.

An associated issue is that the UEs located at pico eNB extended region (CRE region) will, in general, have a low channel quality due to the high interference received from macro eNB. In order to improve that quality eICIC allows to mute all data symbol in certain subframes (ABS subframes). Then, the UEs located at CRE region can be scheduled in ABS subframes obtaining a higher signal-to-interference-plus-noise ratio (SINR) due to the absence of interfering signal from the macro eNB. The ABS pattern has a periodicity of 8 subframes. We define the ABS ratio as the portion of muted subframes over conventional ones.

B. ENB POWER CONSUMPTION MODEL

In this section, we describe the eNB consumption model used in this work. It is based on the 3GPP guidelines [35] and also includes the consumption associated to the activation of a sleeping eNB [6].

The eNB transceivers (TRXs) are composed of an AC-DC unit for main supply, a power amplifier, a cooling system, a baseband interface (a transmitter for downlink and a receiver for uplink), a radio frequency small-signal transceiver section and an antenna interface. The power consumption of some of these components depends on the load of the eNB. Therefore, a common assumption is to approximate the relation between RF output power and power consumption of TRXs using a linear model [35]. Then, the power consumption model of a pico eNB j is given by:

$$C_{\mathrm{p}}^{j} = e^{j} \cdot N_{\mathrm{TRX}} \cdot (P_0 + R^{j} \cdot P_{\max}) + (1 - e^{j}) \cdot N_{\mathrm{TRX}} \cdot P_{\mathrm{sleep}} + \Delta \quad (1)$$

$$\Delta = \begin{cases} \beta \cdot P_0 & \text{when eNB } j \text{ is switched on from} \\ & \text{sleeping mode} \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

where:

- $e^{j} = 1$ when the pico eNB j is active and $e^{j} = 0$ otherwise.
- N_{TRX} is the number of TRXs.
- $P_{\max}$ is the maximum RF output.
- P_0 represents the power consumption at zero RF output power.
- $R^{j} \in [0, 1]$ is the load factor of the pico eNB j and depends on the ABS ratio, the CRE bias, the traffic intensity and the location of UEs.
- P_{sleep} is the power consumption of TRX components in sleep mode.
- β is the portion of P_0 needed to switch on the pico eNB TRXs.

Note that Δ captures the consumption associated to switching on a sleeping pico eNB. The power consumption of the macro eNB i is given by

$$C_{\mathrm{m}}^{i} = N_{\mathrm{TRX}} \cdot (P_0^{\mathrm{m}} + R^{i} \cdot P_{\max}^{\mathrm{m}}) \cdot (1 - \gamma) + N_{\mathrm{TRX}} \cdot P_0^{\mathrm{m}} \cdot \gamma \quad (3)$$

where γ denotes the ABS ratio, $P^{\mathrm{m}}_{\max}$ is the maximum power output of the macro eNB and P^{m}_{0} is the power consumption at zero RF output power of the macro eNB. Note that the ABS ratio, the CRE bias and the activation controls of the eNBs can affect the load factor R^j of an eNB j, making the global power consumption in the network an unknown non-convex function. This motivates the inclusion of these parameters in the control of energy consumption, as explained in next Section.

IV. SYSTEM MODEL AND PROBLEM FORMULATION

A. SYSTEM MODEL

We consider a sector of a LTE-A access network comprising a macro eNB and P pico eNBs. Let $\mathcal{P} = \{1, 2, \ldots, P\}$ be the set of pico eNBs overlapping the macro sector.

The ABS ratio and the CRE bias are denoted by $\gamma \in \Gamma$ and $\phi \in \Phi$, where Γ and Φ are the finite sets of all possible configurations of these parameters. Time is divided into discrete time stages $k \in \{0, 1, \ldots\}$. At each stage, the decision maker must select the values for γ, ϕ and the activation state of the pico eNBs in $\mathcal{P}$. This sequential decision procedure is modeled by a Markov Decision Processes characterized by the following elements:

1) States: Let e^j_k be the state of the pico eNB $j \in \mathcal{P}$ at stage k, where $e^j_k = 1$ when the pico eNB j is switched on and $e^j_k = 0$ otherwise. The vector $p_k = (e^1_k, \ldots, e^P_k) \in \mathcal{E}$ ($\mathcal{E} = \{0, 1\}^P$) denotes the activation state of all picos in $\mathcal{P}$. Let $\lambda_k \in \Lambda$ refer to the mean number of UEs in the sector (traffic load) at stage k, where $\Lambda = [0, \lambda^{\max}]$. The state of the system at stage k is given by $x_k = (\lambda_k, p_{k-1}) \in \mathcal{X}$, where $\mathcal{X} = \Lambda \times \mathcal{E}$ is the state space. Thus, the state of the system is defined by the traffic load at current stage and the activation state of the pico eNBs in $\mathcal{P}$ established in the previous stage.

2) Controls: At the beginning of each stage, the decision maker observes x_k and makes a decision regarding which stations are switched on and off, as well as the values of γ and ϕ to be used during stage k. These decisions define the control $u_k = (p_k, \gamma_k, \phi_k) \in \mathcal{U}$, where $\mathcal{U} = \mathcal{E} \times \Gamma \times \Phi$ is the control space.

3) State transitions: The state of the system at next stage is given by $x_{k+1} = f_k(x_k, u_k)$, where f_k is a function defining the stochastic state transition. In our case, the state transition can be written as follows: $x_{k+1} = (\lambda_{k+1}, p_k)$, where the random variable λ_{k+1} entails the stochastic nature of state transition. Fig. 1 depicts the system state transition.

4) Payoff functions: We define two global payoff functions: $C : \mathcal{X} \times \mathcal{U} \to \mathbb{R}^+$ providing the aggregated power consumption in the network and $Q : \mathcal{X} \times \mathcal{U} \to [0, 1]$ which gives us the ratio of UEs in the network satisfying the QoS requirement specified by operator. In particular, the QoS requirement considered is the UE throughput being greater than a given (specified) threshold. Therefore, the QoS function indicates the ratio of UEs whose throughput is greater than the specified threshold. Note that C and Q are random since they depend on the random UE locations and traffic demands.

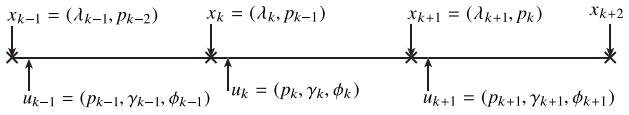

FIGURE 1. Scheme of system state transition where crosses indicate the beginning of each stage.

B. FINITE-HORIZON DP PROBLEM FORMULATION

It has been widely reported that data traffic in a cellular network presents a periodic pattern [5], [6], [28], according to which the average traffic load can be described by a periodic function with a one-day period length. We leverage this fact by formulating the problem as a finite-horizon MDP where the length of the horizon is one day. We divide the day into N discrete time stages $k = 0, 1, \ldots, N-1$. An action u_k is selected for each state x_k for each $k = 1, \ldots, N-1$. Then, at the end of the N stages, we use the *boundary control* $u_N = (p_N, \gamma_N, \phi_N)$ for resetting the process to an initial state $x_0 = (\lambda_0, p_N)$.

The mapping $\pi_k : \mathcal{X} \to \mathcal{U}$, determines the action u_k at stage k given the current state x_k. A policy for the finite-horizon MDP is given by $\pi = \{\pi_0, \ldots, \pi_{N-1}\}$. The objective is to find the optimal policy π^* minimizing the expected accumulated consumption over the N stages and satisfying the QoS constraint at every stage:

$$\pi^* = \underset{\{\pi_0,\ldots,\pi_{N-1}\}}{\operatorname{argmin}} \; E\left\{C(x_N, u_N) + \sum_{k=0}^{N-1} C(\lambda_k, p_{k-1}, \pi_k(x_k))\right\}$$
$$\text{s.t. } E\{Q(\lambda_k, \pi_k(x_k))\} > Q_{\min} \text{ for } k = 0, \ldots, N-1 \tag{4}$$

where $Q_{\min}$ defines the QoS requirement and $C(x_N, u_N)$ is the consumption associated to resetting the system for the next day using the boundary control u_N.

Let $J_k(x_k)$ be the cost-to-go which indicates the minimal consumption of the system from stage k to the final stage N, provided that, at stage k, the system is in state x_k and considering that the QoS requirement is met at every stage.

$$J_N(x_N) = E\{C(x_N, u_N)\}$$
$$J_k(x_k) = \min_{u_k \in \mathcal{U}} E\{C(x_k, u_k) + J_{k+1}(f_k(x_k, u_k))\}$$
$$\text{s.t. } E\{Q(x_k, u_k)\} > Q_{\min} \text{ for } k = 0, \ldots, N-1 \tag{5}$$

The computational overhead associated to the computation of the cost-to-go may be prohibitively large. The reason is twofold: first, the dimensions of $\mathcal{X}$ and $\mathcal{U}$ grow exponentially with the size of $\mathcal{P}$, which also limits the scalability of the problem. Second, the stochastic nature of state transition makes it necessary to compute, for each stage k, the probability of all possible traffic intensities λ_{k+1} at next stage. Specifically, considering Λ as a finite set of traffic intensities, the cost-to-go function consists of $N \cdot |\Lambda| \cdot 2^P$ values. To obtain each of these values, we need to compute the expected values of the consumption and the QoS for each of the $2^P \cdot |\Gamma| \cdot |\Phi|$

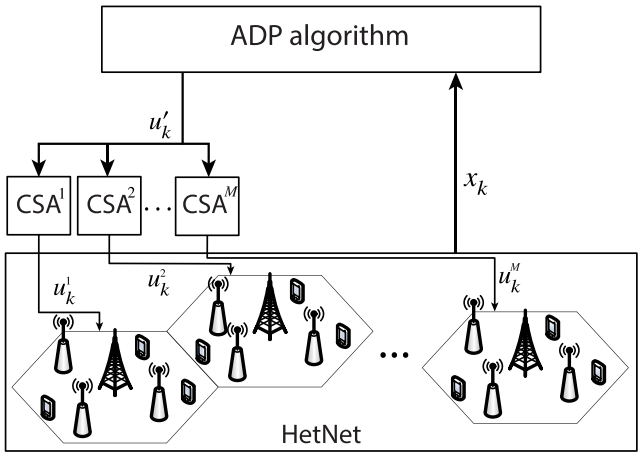

FIGURE 2. Scheme of the online operation of our proposal.

controls in $\mathcal{U}$. Note that Λ has been defined as a continuous set in Sec. IV-A. In this case, the cost-to-go is be a continuous function for each value of k. In the next section we present our proposal aimed at avoiding these problems.

V. PROPOSED CONTROL SCHEME

Our proposal to address problem (4) in a computationally feasible manner, comprises the combination of the following techniques: first, we make use of the Certainty Equivalent Control (CEC) principle to reduce the complexity associated to the stochastic nature of the state transitions. Second, we transform the state and control spaces into a reduced dimensional ones. Third, we use a machine learning approach to infer instant power consumption and to evaluate QoS constraint satisfaction.

In the offline phase we obtain an estimator for the consumption, $\tilde{C}$, and an estimator of the QoS, $\tilde{Q}$. These estimators should be trained offline with a data set comprising user performance samples and network energy consumption for diverse states and configurations of the system. This data set can be readily available to networks operators from the real network, testbeds and/or simulation. Using the estimators $\tilde{C}$ and $\tilde{Q}$ and the CEC principle (Sec. V-A), the estimation of the Cost-to-go function $\tilde{J}$ is computed (Sec. V-D).

Fig. 2 shows the operation of the online phase where the ADP algorithm entity controls a cluster comprising M sectors. The ADP algorithm receives the system state x_k at each stage and, using $\tilde{C}$, $\tilde{Q}$ and $\tilde{J}$, computes the dimensionally-reduced global control u'_k which is sent to the Control Space Augmentors (CSAs) (Sec. V-B). The i-th CSA (corresponding to the i-th sector) computes the full-dimensional control u^i_k from u'_k using the specific information about the network topology of the i-th sector. Each sector i operates during the stage k using the control u^i_k computed by its corresponding CSA. This semi-distributed approach reduces the dimensionality of state/control spaces, simplifying notably the computational complexity of the problem. For the sake of simplicity in the notation, we formulate the problem for only one sector. Note that, while the offline phase is performed once, the online phase is continuously executed in a stage by stage basis. The following subsections describe and discuss the elements of our proposal in detail.

A. CERTAINTY EQUIVALENT CONTROL (CEC)

The Certainty Equivalent Control (CEC) is an approximated control scheme that applies at each stage the control that would be optimal if the uncertain quantities were fixed to their expected values [29]. CEC differs from exact dynamic programming in the following aspects: the cost-to-go $J_{k+1}(x_{k+1})$ is replaced by its estimation $\tilde{J}_{k+1}(x_{k+1})$, and the uncertain quantities are replaced with their expected values. Specifically, we replace the traffic intensity at each stage k, λ_k, with its expected value $\bar{\lambda}_k \in \bar{\Lambda}$, where $\bar{\Lambda}$ is a finite set whose cardinality is $|\bar{\Lambda}| = N$ (it contains one expected value for each decision stage).

Therefore, considering $\bar{\lambda}_k$ at each stage k, the state transitions are now deterministic $x_{k+1} = (\bar{\lambda}_{k+1}, p_k)$, since $\bar{\lambda}_{k+1}$ is known in advance and p_k is provided by the control u_k.

The objective is to find the control that minimizes the sum of the expected power consumption at current stage and the cost-to-go estimation while satisfying the QoS requirement at each stage:

$$u_k = \underset{u\in\mathcal{U}}{\operatorname{argmin}}\, E[C(x_k,u)] + \tilde{J}_{k+1}(\bar{\lambda}_{k+1},u) \quad \text{s.t. } E[Q(x_k,u)] > Q_{\min} \text{ for } k = 0,1,\ldots,N-1. \quad (6)$$

Note that this minimization is performed at each stage, while the network operates (this is referred to as online operation). Then, given the observed current state x_k, we need to predict the expectation of the energy consumption for all feasible controls: $E[C(x_k,u)] \quad \forall u \in \{u : E[Q(x_k,u)] > Q_{\min}\}$, which is addressed later in this Section. Note at this point that the computational cost associated to solve (6) depends on $|\mathcal{U}|$. For this reason, we propose a state/control space transformation to make the problem computationally feasible.

B. STATE/CONTROL SPACE TRANSFORMATION

The possible values of the vector p (which indicates the activation state of each pico in $\mathcal{P}$) increase exponentially with P: $|\mathcal{E}| = 2^P$. This section describes a strategy to reduce the relationship between P and state/control spaces from exponential to linear. We define $p' = \Pi(p)$ as the linear transformation of the parameter p where $\Pi(x) = \sum_i x_i$. Thus, $p' \in \mathcal{E}'$ where $|\mathcal{E}'| = P + 1$. Let $u' = (p', \gamma, \phi) \in \mathcal{U}'$ denote a transformed control, where $\mathcal{U}' = \mathcal{E}' \times \Gamma \times \Phi$. This dimensionally-reduced control will use in next sections to define the estimation of the cost-to-go function and the online operation of our proposal. Although the transformation $\mathcal{U} \rightarrow \mathcal{U}'$ is straightforward, it must be undone in the last step of our algorithm (Fig. 2). CSA entities are in charge of reversing the transformation ($\mathcal{U}' \rightarrow \mathcal{U}$). In the reminder of this section we detail the procedure implemented by CSAs.

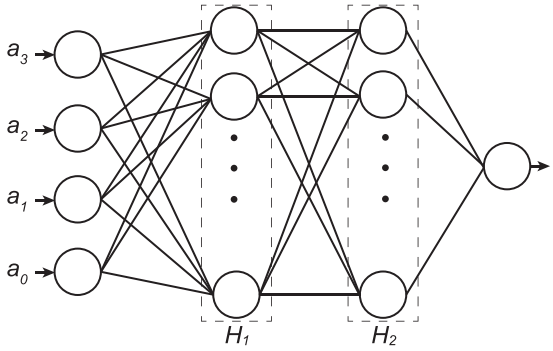

FIGURE 3. Neural network Architecture where the inputs are the traffic load λ and a control u', i.e., $a = (\lambda, u')$.

We define the *cell adjacency value* d^j associated to a pico eNB $j \in \mathcal{P}$ as a weighted average of the distances to the remaining eNBs in the sector:

$$d^j = w \cdot d^j_{\mathrm{m}} + (1 - w) \cdot d^j_{\mathrm{p}} \tag{7}$$

where d^j_{m} is the distance to the macro eNB, d^j_{p} is the average distance to the rest of the pico eNBs in the sector and $w \in [0, 1]$ is a weighting factor. Let $\mathcal{D} = \{d^i\}_{i\in\mathcal{P}}$ be the set containing the cell adjacency values of all pico eNBs in $\mathcal{P}$. The higher the value of the cell adjacency, the higher the priority of a pico eNB to be switched off. Therefore, the transformation can reverted ($p' \to p$), by computing the values of each element e^j of p as follows:

$$e^i = \begin{cases} 1 & \text{if } i \in \mathcal{I} \text{where } \mathcal{I} = M_{p'}(\mathcal{D}) \\ 0 & \text{otherwise} \end{cases} \tag{8}$$

where the operator $M_{p'}(\mathcal{D})$ returns the indexes of the p' greatest values in $\mathcal{D}$, i.e., the indexes of the p' pico eNBs with the greatest values of cell adjacency.

The idea behind this approach is the following. The cell adjacency value of a pico eNB indicates its closeness to the rest of the eNBs in the sector. Thus, our strategy consists on switching off firstly the eNBs with lower values of cell adjacency, that is, the eNBs with more eNBs in their surroundings because they can offload their traffic to near eNBs (macro or picos) incurring lower losses in the channel quality. We evaluate this strategy in Section VI.

C. MACHINE LEARNING FOR ENERGY CONSUMPTION AND QoS ESTIMATION

We define $\tilde{C} : \Lambda \times \mathcal{U}' \to \mathbb{R}$ as an estimator of the expected aggregated power consumption. The estimator $\tilde{C}$ is based on a polynomial regression model of the relation between the input variables (the system state and the selected control) and output variable (the observed power consumption). We evaluated polynomials of several degrees ranging from 1 to 9, using two data sets, one for training the model and other for testing its estimation accuracy. In these experiments, degree 6 polynomials obtained the minimum Root Mean Squared Error (RSME) in the test data set.

Let $\tilde{Q} : \Lambda \times \mathcal{U}' \to \{0, 1\}$ be a classifier estimator indicating whether the current expected QoS is above the defined threshold $Q_{\min}$ ($\tilde{Q}(\lambda, u') = 1$) or not ($\tilde{Q}(\lambda, u') = 0$). The function $\tilde{Q}$ is implemented using a neural network (NN) whose architecture is shown in Figure 3. The number of nodes in the hidden layers are $H_1 = H_2 = 50$. A rectified linear unit (ReLU) is used as a nonlinear activation function in both hidden layers and we use a sigmoid activation function in the output layer. This neural network is trained using the gradient-based algorithm in [36] to minimize the cross-entropy cost function. The NN was trained during 400 epochs with a learning-rate equal to 0.001, using data previously gathered from the network. The data set used to train both models consists of 23000 samples of power consumption and QoS obtained by simulation.

D. COST-TO-GO ESTIMATION

In this section, we address the cost-to-go estimation presented in equations (6). First, we need to introduce some notation. Let $\bar{x}_k = (\bar{\lambda}_k, p'_{k-1}) \in \bar{\mathcal{X}}$ be a transformed CEC state value, where $\bar{\mathcal{X}} = \bar{\Lambda} \times \mathcal{E}'$. We define $\Delta' : \bar{\mathcal{X}} \times \mathcal{U}' \to \mathbb{R}$ as the function providing the power consumption associated to the activation of the pico eNBs. According to (2):

$$\Delta'(\bar{x}_k, u') = \beta \cdot P_0 \cdot \max(0, p'_k - p'_{k-1}) \tag{9}$$

The estimation of the cost-to-go at stage k is defined as follows:

$$\begin{aligned} \tilde{J}_N(\bar{x}_N) &= \Delta'(\bar{x}_N, u'_N) \\ \tilde{J}_k(\bar{x}_k) &= \min_{u'_k \in U'} \tilde{C}(\bar{\lambda}_k, u'_k) + \Delta'(\bar{x}_k, u'_k) + \tilde{J}_{k+1}(\bar{f}(\bar{x}_k, u'_k)) \\ &\quad \text{s.t. } \tilde{Q}(\bar{\lambda}_k, u'_k) = 1 \text{ for } k = 0, 1, \ldots, N-1. \end{aligned} \tag{10}$$

where $\bar{f}_k$ is the function defining the state transitions which are now deterministic and $\tilde{J}_N(\bar{x}_N, u_N)$ correspond to the power consumption associated to the activation of the pico eNBs at the initial stage of the next day.

We summarize the offline phase as follows:

1) Retrieve the data set from the network whose j-th sample is $\big(x_j, u_j, C(x_j, u_j), Q(x_j, u_j)\big)$.
2) Train the estimators $\tilde{C}$ and $\tilde{Q}$ according to the procedure detailed in Sec. V-C.
3) Compute $\tilde{J}$ using (10).

E. ONLINE OPERATION

In the online phase, we select the control u'_k for each stage k minimizing the power consumption afterwards and satisfying the QoS constraint, given the current state x_k observed from the network, i.e.,

$$\begin{aligned} u'_k \in &\operatorname*{argmin}_{u' \in \mathcal{U}'} \tilde{C}(x_k, u') + \Delta'(x_k, u') + \tilde{J}_{k+1}(\bar{\lambda}_{k+1}, u') \\ &\text{s.t. } \tilde{Q}(\lambda_k, u') = 1 \text{ for } k = 0, 1, \ldots, N-1. \end{aligned} \tag{11}$$

The online phase is summarized in Algorithm 1. It shows the operation of our proposal from a high-level point of view during one day. Steps 3, 4 and 5 in Alg. 1 are executed

TABLE 2. Notation table.

Notation	Definition
$\mathcal{P}$	Set of pico eNBs in the sector
$\gamma \in \Gamma, \phi \in \Phi$	ABS ratio, CRE bias
$p \in \mathcal{E}$	Activation state of pico eNBs in the sector
$\lambda \in \Lambda$	Traffic intensity
$x = (\lambda, p) \in \mathcal{X}$	Network state, network state space
$u = (p, \gamma, \phi) \in \mathcal{U}$	Control, control space
C, Q	Power consumption and QoS pay-off functions
$p' \in \mathcal{E}'$	Activation state of pico eNBs in the sector
$u' = (p', \gamma, \phi) \in \mathcal{A}$	Transformed (dimensionally-reduced) control
$\bar{\lambda} \in \bar{\Lambda}$	Expected traffic intensity
$\bar{x} = (\bar{\lambda}, p') \in \bar{\mathcal{X}}$	CEC state, CEC state space
$\tilde{C}, \tilde{Q}$	Power consumption and QoS estimators
Δ'	Power consumption activation function

once per stage. Note that the computational complexity of solving (11) only depends on the size of $\mathcal{U}'$, independently of the size of the network. This means that the computational complexity per stage in the online phase is $O(1)$. That is, our proposal can be applied to an arbitrarily large network without increasing the computational complexity, only by replicating the instances of the ADP algorithm.

We propose to group the sectors in clusters characterized by a similar traffic profile (e.g. city center, business area). Thus, one ADP controller can be assigned to each cluster to cover the whole network. In other words, we need as many instances of the ADP algorithm as clusters of sectors in the network. Table 2 summarizes the notation in this paper.

Algorithm 1 Online Phase (ADP Algorithm Operation)

1: **for** $k = 0, 1, \ldots, N-1$ **do**
2: Receive the network state x_k
3: Compute u'_k using (11)
4: Send u'_k to CSAs
5: **end for**

VI. NUMERICAL RESULTS

A. DESCRIPTION OF THE SIMULATION FRAMEWORK

The network layout is based on the 3GPP guidelines for evaluation of LTE networks [37]. It consists of 5 sectorial macro eNBs (120 degrees) and several pico eNBs overlapping each macro eNB coverage area (Figure 4). We simulate the central sector (shadowed in Figure 4) and the rest of macro eNBs emulate the aggregated interference of a larger network. The wireless channel comprises two components: the deterministic pathloss and the stochastic shadow fading.

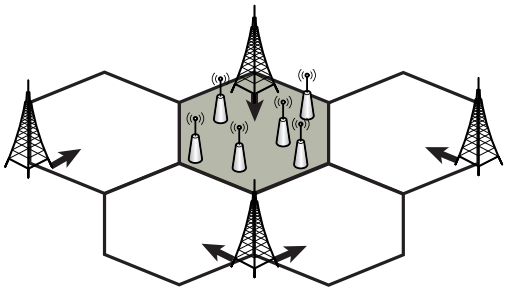

FIGURE 4. Illustration of the simulated scenario with the simulated (shadowed) sector and four sectorial macro eNBs used to compute the aggregated interference, emulating the effect of a larger network.

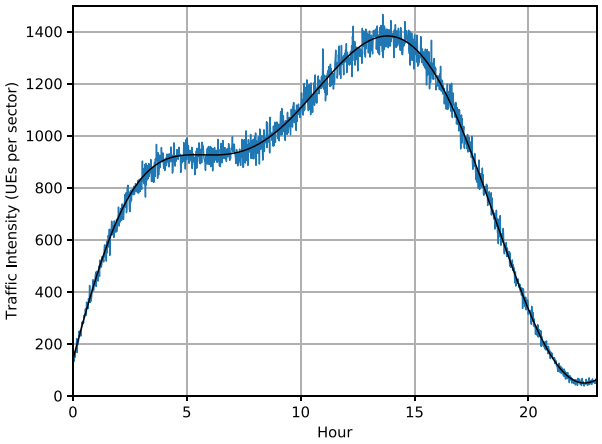

FIGURE 5. One day traffic pattern. The black line corresponds to the average traffic values and the blue line is a random traffic obtained from the mean values.

The aggregated interference at each UE receiver consists of the power received from all interfering eNBs in the sector (picos and macro) plus the interference from the macro eNBs from other sectors. Fig. 5 shows, for one simulation run, the average traffic load following a daily periodic pattern, and the per-stage traffic load randomly generated from this pattern.

A throughput sample is defined, according to the 3GPP guidelines [37], as the quotient between the size of a downloaded file and the time required to download it. Each UE has to download one file, and therefore generates one throughput sample. We consider that a UE fulfills the QoS requirement if its throughput is over $T_{\min} = 100$ kbps. In our simulations, the QoS function Q provides the expected ratio of UEs satisfying $T_{\min}$. The power consumption model is defined in Section III-B and its parameter configuration [35] is shown in Table 3.

We consider an horizon of $N = 24$ stages, corresponding to the hours of the day. For each stage, we perform 60 evaluations of 100 LTE frames with varying traffic corresponding to each minute within the hour. At each minute, we generate the traffic load according to a Poisson process whose mean is given by the average traffic profile shown in Fig. 5. Each UE is dropped uniformly over the macro eNB coverage area

TABLE 3. Simulation parameters.

Network layout	5 sectorial macro eNBs, 500 m ISD, $P = 6$ pico eNBs per sector
System bandwidth	10 MHz
LTE frame duration	Subframe 1 ms, Protected-subframe pattern 8 ms, Frame 10 ms
Transmit power	Macro eNB 46 dBm, pico eNB 30 dBm
Macro sector antenna pattern	$A_H(\phi) = -min[12(\frac{\phi}{\phi_{3dB}})^2, A_m]$, $A_m = 70$ degrees $A_m = 25$ dB
Pico antenna pattern	Omnidirectional
Antenna gains	macro: 14 dBi; pico: 5 dBi; UE: 0 dBi
Macro to UE path loss	$128.1 + 37.6 \cdot \log_{10}(R[\text{Km}])$ where R is the macro eNB to UE distance
Pico to UE path loss	$149.7 + 36.7 \cdot \log10(R[\text{Km}])$ where R is the pico eNB to UE distance
Shadow fading	Lognormal distribution with 10 dB standard deviation
Thermal noise	-176 dBm
Scheduling algorithm	Proportional Fair (PF)
Traffic model	File Transfer Protocol (FTP)
File size	0.5 Mbytes
Minimum distances	Macro - pico: 70 m; Macro - UE: 35 m; Pico - pico: 40 m; Pico - UE: 10 m
Macro consumption parameters	$N_{\text{TRX}} = 6$, $P_0^{\text{m}} = 130$ W, $P_{\text{max}}^{\text{m}} = 20$ W
Pico consumption parameters	$N_{\text{TRX}} = 2$, $P_0 = 56$, $P_{\text{max}} = 6.3$, $P_{\text{sleep}} = 39$ W, $\beta = 0.5$

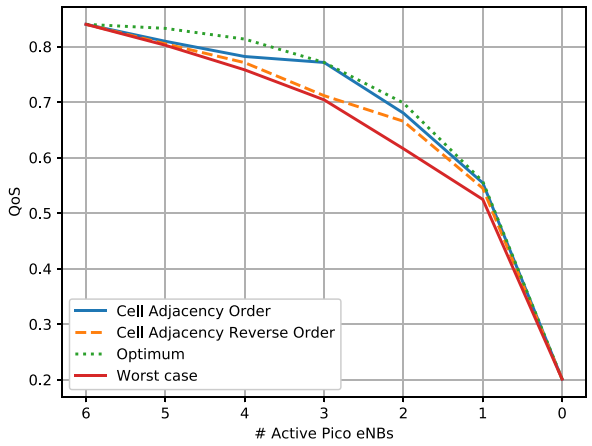

FIGURE 6. Performance of the switching on-off policies in terms of QoS for different numbers of active pico eNBs.

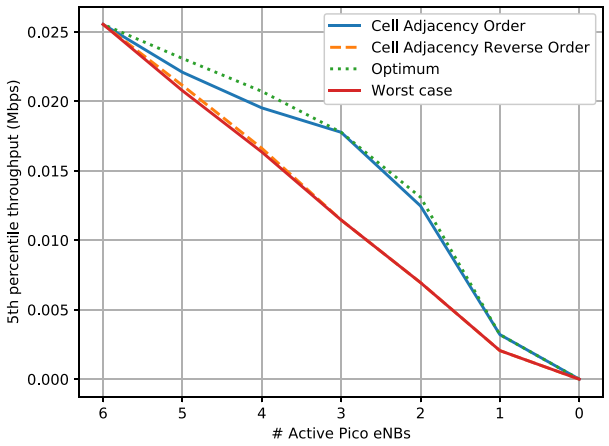

FIGURE 7. Performance of the switching on-off policies in terms of 5th percentile throughput for different numbers of active pico eNBs.

with probability $\frac{1}{3}$, or over a pico eNB coverage area with probability $\frac{2}{3}$.

Regarding the interference coordination controls, we consider $\Gamma = \{0, \frac{1}{8}, \frac{2}{8}, \ldots, \frac{7}{8}\}$ and $\Phi = \{0, 6, 9, 12, 18\}$. The number of pico eNBs per sector is $P = 6$. The rest of the simulation parameters are shown in Table 3.

B. EVALUATION OF CSA STRATEGY

In this subsection we evaluate the losses associated with the state/control space transformation. In order to revert the transformation, the eNBs with lower adjacency values are switched off firstly, as explained in Sec. V-B. We compare the proposed strategy with other alternative strategies, including the optimal one. We consider for this evaluation $P = 6$, $\gamma = 6/8$ and $\phi = 6$ dBs.

Figures 6 and 7 show the performance in terms of QoS and 5th percentile throughput [37] of the following policies:

- Cell adjacency order: eNBs with lower values of cell adjacency are deactivated firstly.
- Cell adjacency reverse order: eNBs with higher values of cell adjacency are deactivated firstly.
- Optimum case: We select the best of the $P!$ possible orderings of P pico eNBs in the sector.
- Worst case: We select the worst of the $P!$ possible orderings of P pico eNBs in the sector.

Both figures show that Cell Adjacency Order performs similarly to the optimal policy. We performed the simulations of our proposed policy for different values of the weighting parameter w ranging from 0 to 1. The best performing value was 0.4, which is the one used in the following results.

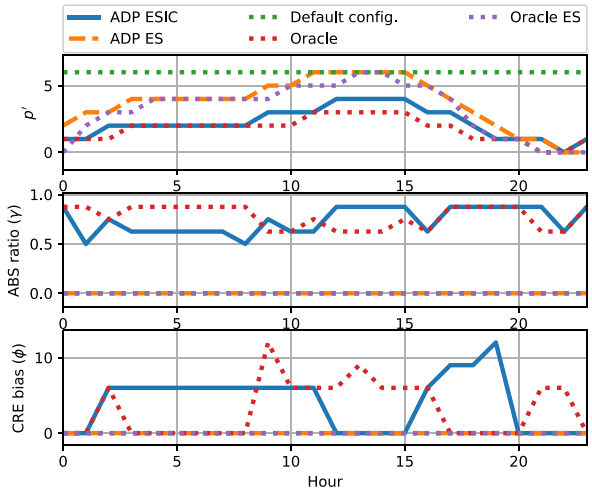

FIGURE 8. Selected control u' at each hour for $Q_{\min} = 0.6$ under the different policies.

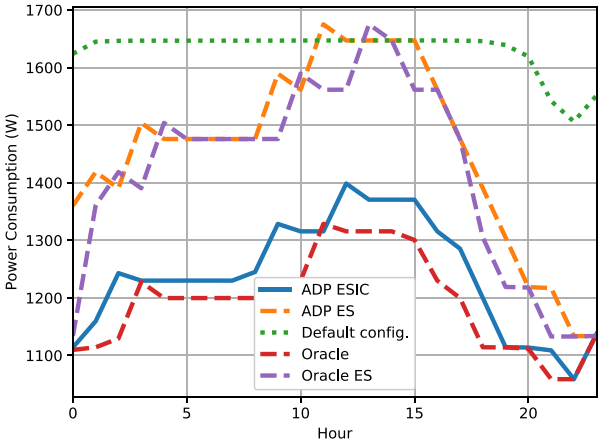

FIGURE 9. Comparison of the aggregated instantaneous power consumption at each hour for $Q_{\min} = 0.6$ under different policies.

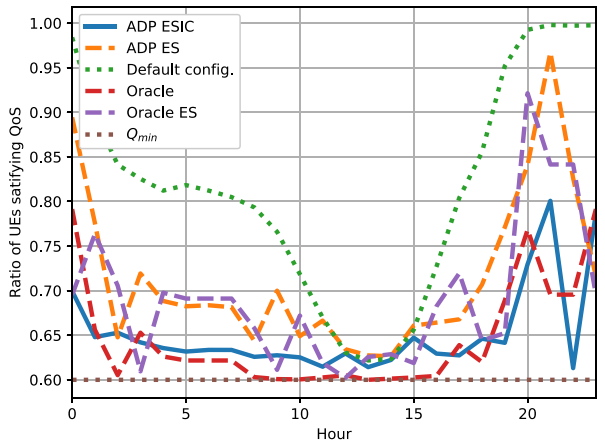

FIGURE 10. Ratio of UEs satisfying the minimum throughput $T_{\min}$ at each hour for $Q_{\min} = 0.6$.

C. EVALUATION OF ESTIMATORS PERFORMANCE

The estimators $\tilde{C}$ and $\tilde{Q}$ are obtained in the offline phase. These estimators are used firstly in the computation of the cost-to-go function, and then in the control selection during the online operation (11). The error incurred by these estimators may affect the final performance of our ADP algorithm. For example, if a control u' and a traffic λ are misclassified as meeting the QoS constraint ($Q(\lambda_k, u') = 1$), the control u' could be selected under traffic λ, resulting in a non-fulfillment of the QoS requirement in this stage.

We evaluate the errors of the estimators against the available data in the offline phase. That is, we have computed the estimators $\tilde{C}$ and $\tilde{Q}$ for *training sets* of different sizes, ranging from 50 to 23000 samples. For all the obtained estimators, their errors are estimated using the same *test set*. This way, we obtain better quality estimators as the size of the training set increases. For the polynomial regressor $\tilde{C}$, we evaluate the RMS error, and for the classifier $\tilde{Q}$, we evaluate the accuracy (ratio of correct classifications), and the ratio of type I and type II errors. Table 4 summarizes the results of these evaluations and the performance of the proposed ADP algorithm in terms of daily energy consumption and QoS satisfaction when using these estimators.

D. COMPARATIVE EVALUATION OF THE ADP ALGORITHM

In our numerical evaluation we compare the following policies:

- Our proposed ADP algorithm controlling the Energy Saving (ES) and the Interference Coordination (IC) mechanisms (ADP ESIC).
- Our proposal, but only controlling the ES mechanism (ADP ES).
- A default network configuration where ES and IC mechanisms are not operative.
- An Oracle policy which minimizes the overall daily consumption satisfying the QoS requirement at every stage. This policy has been computed by means of exhaustive search over the possible controls at each stage.
- An Oracle policy (Oracle ES) which, as ADP ES, only controls the ES mechanism. This policy is an upper bound of previous works that only consider the ES scheme.

Fig. 8 shows the controls selected by each policy at each hour (decision stage) for a threshold $Q_{\min} = 0.6$. Fig. 9 shows the aggregated power consumption at each stage, i.e. the consumption of all the eNBs within the sector. Fig. 10 shows the QoS observed at each step, i.e. the portion of total UEs in the sector satisfying $T_{\min}$.

ADP ESIC uses, in most stages, fewer pico eNBs than ADP ES, as Fig. 8 shows, implying significant energy savings. The use of IC makes it possible to satisfy the QoS threshold using fewer pico eNBs. Specifically, the Oracle, ADP ESIC and ADP ES policies attain an energy saving with respect to the default configuration of 26.79%, 24.07% and 10.96%, respectively.

Note that the QoS threshold balances power consumption and QoS. If we impose a higher QoS threshold, more active

TABLE 4. Performance of estimators and ADP algorithm as a function of the number of samples in the training set.

Number of samples	50	300	1000	7000	18000	23000
RMSE in $\tilde{C}$	263.80 ± 59.10	13.08 ± 1.06	3.95 ± 0.14	3.11 ± 0.01	3.07 ± 0.02	3.08 ± 0.03
Accuracy in $\tilde{Q}$	0.83 ± 0.03	0.88 ± 0.01	0.92 ± 0.01	0.95 ± 0.01	0.95 ± 0.01	0.96 ± 0.01
Type I error in $\tilde{Q}$	0.13 ± 0.04	0.09 ± 0.03	0.06 ± 0.02	0.04 ± 0.01	0.04 ± 0.01	0.03 ± 0.01
Type II error in $\tilde{Q}$	0.25 ± 0.09	0.16 ± 0.05	0.13 ± 0.05	0.06 ± 0.02	0.05 ± 0.02	0.07 ± 0.02
Daily cons. (kWh)	29.95 ± 0.30	28.70 ± 0.03	28.64 ± 0.00	28.64 ± 0.00	28.64 ± 0.00	28.64 ± 0.00
QoS fulfillment prob.	0.31 ± 0.11	0.54 ± 0.12	0.59 ± 0.11	0.78 ± 0.05	0.85 ± 0.06	0.90 ± 0.05

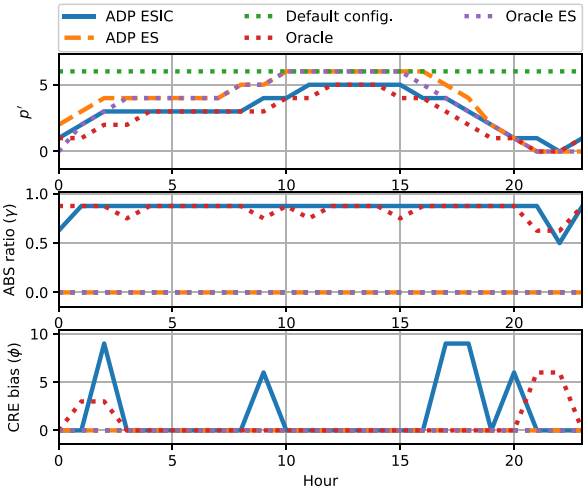

FIGURE 11. Selected control u' at each hour for $Q_{\min} = 0.68$ under the different policies.

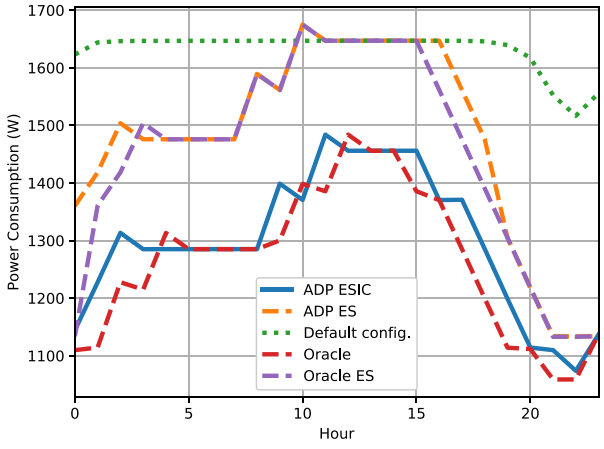

FIGURE 12. Comparison of the aggregated instantaneous power consumption at each hour for $Q_{\min} = 0.68$ under different policies.

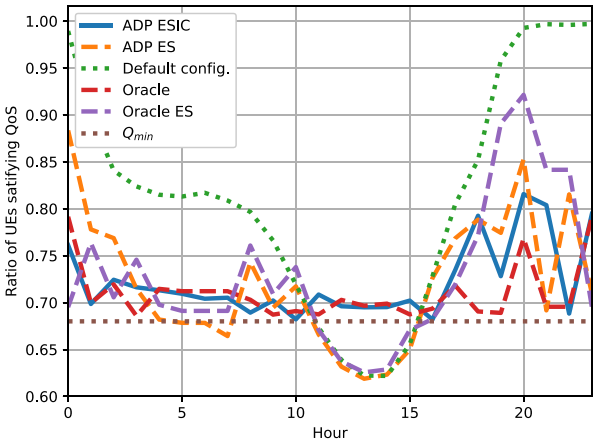

FIGURE 13. Ratio of UEs satisfying the minimum throughput $T_{\min}$ at each hour for $Q_{\min} = 0.68$.

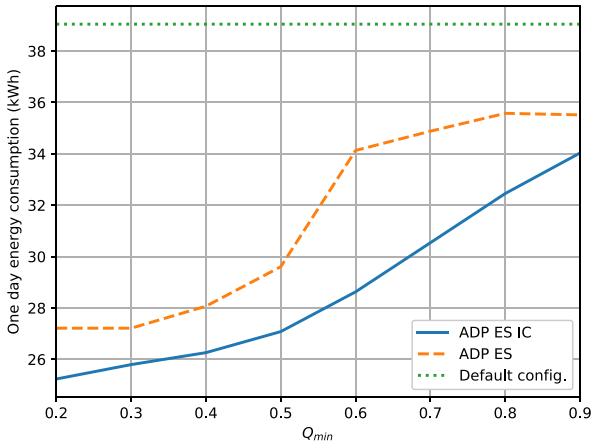

FIGURE 14. Accumulated energy consumption in one day for different values of $Q_{\min}$.

pico eNBs are needed and therefore the power consumption increases. We illustrate that in Figures 12, 13 and 11 where $Q_{\min} = 0.68$. In this case, the energy savings of Oracle, ADP ESIC and ADP ES with respect to the default configuration are 22.54%, 20.43% and 9.84%, respectively. It is worth mentioning that in Fig. 12 the instantaneous power consumption of the Oracle policy exceeds that of our proposal in three stages due to the activation power consumption spikes. However, the Oracle policy attains the minimal overall daily consumption, which is the objective according to Eq. (4).

For both $Q_{\min}$ values our proposal performs very closely to the Oracle policy. This indicates that the the potential efficiency losses associated to the approximation strategies of our proposal are not significant, and the performance of our ADP policies are very close to the optimal ones. It is worth mentioning that when IC is not considered (ADP ES and

Default configuration), the QoS threshold cannot be satisfied during the traffic peak between stages 11^{th} and 15^{th}.

Finally, Fig. 14 illustrates the relation between the QoS threshold, $Q_{\min}$, and the best achievable daily energy consumption (fulfilling this threshold). The higher the value of $Q_{\min}$, the more energy is needed to satisfy the QoS requirement. This figure also shows that combining the energy saving and interference coordination mechanisms attains the lowest energy consumption under any QoS requirement.

VII. CONCLUSION

This paper presented a novel stochastic control approach addressing the energy saving and the interference coordination problems jointly. It is applied to an LTE-A HetNet to minimize the energy consumption assuring QoS guarantees. Considering the periodic pattern of daily traffic demands, the problem is formulated as a finite horizon MDP where we apply the CEC principle for the sake of computational tractability. Specifically, we propose an ADP algorithm combining the CEC principle and machine learning techniques. We have shown the considerable improvement associated to addressing energy saving and interference coordination problems jointly. In addition, our proposal allows a configurable QoS requirement.

There are, however, two aspects left for future research: The first one is clustering of sectors explained in Sec. V. The fewer the sectors per cluster, the more similar the traffic profile among clusters, which can increase the performance of our proposal, at the cost of requiring more instances of the ADP controller. On the other hand, larger clusters lead to more variance in the traffic intensity of the sectors. This trade-off has to be further investigated. Second, we assume a daily periodic pattern in the traffic demands which can change depending on the day. For example, the traffic demands can change on weekends or during special events like sport games or concerts. In the case of regular weekends or vacation days the prediction is straightforward. However, for special event the traffic demands can change depending on the amount of people, the profile of these people, the type of the event, etc. In order to address this issue, we can use external information (e.g. event program in the city) or detect the changes in the traffic patterns in an online fashion and adapt the algorithm to them.

REFERENCES

[1] N. Bhushan *et al.*, "Network densification: The dominant theme for wireless evolution into 5G," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 82–89, Feb. 2014.

[2] G. Micallef, P. Mogensen, and H.-O. Scheck, "Cell size breathing and possibilities to introduce cell sleep mode," in *Proc. IEEE Eur. Wireless Conf. (EW)*, Apr. 2010, pp. 111–115.

[3] L. Saker, S. E. Elayoubi, and H. O. Scheck, "System selection and sleep mode for energy saving in cooperative 2G/3G networks," in *Proc. IEEE 70th Veh. Technol. Conf. Fall (VTC-Fall)*, Sep. 2009, pp. 1–5.

[4] *Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN); Overall Description (Release 10)*, document 3GPP TS 36.300 version 10.5.0, 3rd Generation Partnership Project (3GPP), 2011.

[5] E. Oh, K. Son, and B. Krishnamachari, "Dynamic base station switching-on/off strategies for green cellular networks," *IEEE Trans. Wireless Commun.*, vol. 12, no. 5, pp. 2126–2136, May 2013.

[6] M. Ismail and W. Zhuang, "Network cooperation for energy saving in green radio communications," *IEEE Wireless Commun.*, vol. 18, no. 5, pp. 76–81, Oct. 2011.

[7] L. Saker, S. E. Elayoubi, R. Combes, and T. Chahed, "Optimal control of wake up mechanisms of femtocells in heterogeneous networks," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 3, pp. 664–672, Apr. 2012.

[8] Y.-H. Chiang and W. Liao, "Genie: An optimal green policy for energy saving and traffic offloading in heterogeneous cellular networks," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Jun. 2013, pp. 6230–6234.

[9] S. Cai, Y. Che, L. Duan, J. Wang, S. Zhou, and R. Zhang, "Green 5G heterogeneous networks through dynamic small-cell operation," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 5, pp. 1103–1115, May 2016.

[10] M. Feng, S. Mao, and T. Jiang, "BOOST: Base station on-off switching strategy for green massive MIMO HetNets," *IEEE Trans. Wireless Commun.*, vol. 16, no. 11, pp. 7319–7332, Nov. 2017.

[11] J. He *et al.*, "Energy efficient BSs switching in heterogeneous networks: An operator's perspective," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Apr. 2016, pp. 1–6.

[12] J. Kim, W. S. Jeon, and D. G. Jeong, "Base-station sleep management in open-access femtocell networks," *IEEE Trans. Veh. Technol.*, vol. 65, no. 5, pp. 3786–3791, May 2016.

[13] L. Tang, W. Wang, Y. Wang, and Q. Chen, "An energy-saving algorithm with joint user association, clustering, and on/off strategies in dense heterogeneous networks," *IEEE Access*, vol. 5, pp. 12988–13000, 2017.

[14] T. Zhou, N. Jiang, Z. Liu, and C. Li, "Joint cell activation and selection for green communications in ultra-dense heterogeneous networks," *IEEE Access*, vol. 6, pp. 1894–1904, 2018.

[15] B. Zhuang, D. Guo, and M. L. Honig, "Energy-efficient cell activation, user association, and spectrum allocation in heterogeneous networks," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 4, pp. 823–831, Apr. 2016.

[16] Q. Kuang and W. Utschick, "Energy management in heterogeneous networks with cell activation, user association, and interference coordination," *IEEE Trans. Wireless Commun.*, vol. 15, no. 6, pp. 3868–3879, Jun. 2016.

[17] C. Liu, B. Natarajan, and H. Xia, "Small cell base station sleep strategies for energy efficiency," *IEEE Trans. Veh. Technol.*, vol. 65, no. 3, pp. 1652–1661, Mar. 2016.

[18] A. Virdis, G. Stea, D. Sabella, and M. Caretti, "A distributed power-saving framework for LTE HetNets exploiting Almost Blank Subframes," *IEEE Trans. Green Commun. Netw.*, vol. 1, no. 3, pp. 235–252, Sep. 2017.

[19] X. Xu, C. Yuan, W. Chen, X. Tao, and Y. Sun, "Adaptive cell zooming and sleeping for green heterogeneous ultradense networks," *IEEE Trans. Veh. Technol.*, vol. 67, no. 2, pp. 1612–1621, Feb. 2017.

[20] S. Zhang, N. Zhang, S. Zhou, J. Gong, Z. Niu, and X. Shen, "Energy-aware traffic offloading for green heterogeneous networks," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 5, pp. 1116–1129, May 2016.

[21] X. Chen, J. Wu, Y. Cai, H. Zhang, and T. Chen, "Energy-efficiency oriented traffic offloading in wireless networks: A brief survey and a learning approach for heterogeneous cellular networks," *IEEE J. Sel. Areas Commun.*, vol. 33, no. 4, pp. 627–640, Apr. 2015.

[22] U. Siddique, H. Tabassum, E. Hossain, and D. I. Kim, "Channel-access-aware user association with interference coordination in two-tier downlink cellular networks," *IEEE Trans. Veh. Technol.*, vol. 65, no. 7, pp. 5579–5594, Jul. 2016.

[23] J. Zheng, L. Gao, H. Wang, J. Niu, X. Li, and J. Ren, "EE-eICIC: Energy-efficient optimization of joint user association and ABS for eICIC in heterogeneous cellular networks," *Wireless Commun. Mobile Comput.*, vol. 2017, Sep. 2017, Art. no. 6768415. [Online]. Available: https://www.hindawi.com/journals/wcmc/2017/6768415/abs/

[24] X. Chen, H. Zhang, T. Chen, and M. Lasanen, "Improving energy efficiency in green femtocell networks: A hierarchical reinforcement learning framework," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Jun. 2013, pp. 2241–2245.

[25] O.-C. Iacoboaiea, B. Sayrac, S. Ben Jemaa, and P. Bianchi, "SON coordination in heterogeneous networks: A reinforcement learning framework," *IEEE Trans. Wireless Commun.*, vol. 15, no. 9, pp. 5835–5847, Sep. 2016.

[26] M. Simsek, M. Bennis, and I. Güvenç, "Learning based frequency- and time-domain inter-cell interference coordination in HetNets," *IEEE Trans. Veh. Technol.*, vol. 64, no. 10, pp. 4589–4602, Oct. 2015.
[27] Q. Li, H. Xia, Z. Zeng, and T. Zhang, "Dynamic enhanced inter-cell interference coordination using reinforcement learning approach in heterogeneous network," in *Proc. 15th IEEE Int. Conf. Commun. Technol. (ICCT)*, Nov. 2013, pp. 239–243.
[28] M. A. Marsan, L. Chiaraviglio, D. Ciullo, and M. Meo, "Optimal energy savings in cellular access networks," in *Proc. IEEE Int. Conf. Commun. Workshops (ICC Workshops)*, Jun. 2009, pp. 1–5.
[29] D. P. Bertsekas, *Dynamic Programming and Optimal Control*, vol. 1, no. 2. Belmont, MA, USA: Athenea Scientific 1995.
[30] B. Shen, Z. Lei, X. Huang, and Q. Chen, "An interference contribution rate based small cells on/off switching algorithm for 5G dense heterogeneous networks," *IEEE Access*, vol. 6, pp. 29757–29769, 2018.
[31] E. Baccarelli, N. Cordeschi, and V. Polli, "Optimal self-adaptive QoS resource management in interference-affected multicast wireless networks," *IEEE/ACM Trans. Netw.*, vol. 21, no. 6, pp. 1750–1759, Dec. 2013.
[32] Y. S. Soh, T. Q. S. Quek, M. Kountouris, and H. Shin, "Energy efficient heterogeneous cellular networks," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 5, pp. 840–850, May 2013.
[33] J. Chen, X. Ge, X. Song, and Y. Zhong. (Jun. 2018). "Base station switch-off with mutual repulsion in 5G massive MIMO networks." [Online]. Available: https://arxiv.org/abs/1806.03183
[34] Z. Niu, Y. Wu, J. Gong, and Z. Yang, "Cell zooming for cost-efficient green cellular networks," *IEEE Commun. Mag.*, vol. 48, no. 11, pp. 74–79, Nov. 2010.
[35] *Evolved Universal Terrestrial Radio Access Network (E-UTRAN); Study on Energy Saving Enhancement for E-UTRAN (Release 12)*, document 3GPP TR 36.887, 3rd Generation Partnership Project (3GPP), 2014.
[36] D. Kingma and J. Ba. (Dec. 2014). "Adam: A method for stochastic optimization." [Online]. Available: https://arxiv.org/abs/1412.6980
[37] *Evolved Universal Terrestrial Radio Access (EUTRA); Further Advancements for E-UTRA Physical Layer Aspects*, document 3GPP TR 36.814, 3rd Generation Partnership Project (3GPP), 2010.

JOSE A. AYALA-ROMERO received the B.Sc. degree in telematics engineering and the M.Sc. degree in telecommunication engineering from the Technical University of Cartagena, Spain, in 2012 and 2014, respectively, where he is currently pursuing the Ph.D. degree with the Department of Information and Communications Technologies. His research interests are cellular networks and learning algorithms.

JUAN J. ALCARAZ received the Engineering degree from the Polytechnical University of Valencia in 1999 and the Ph.D. degree from the Technical University of Cartagena (UPCT) in 2007. After working for several telecommunication companies, he joined UPCT in 2004. He was a Fulbright Visiting Scholar at the Electrical Engineering Department, UCLA, in 2013, and a Visiting Professor with the Department of Information Engineering, University of Padova, in 2017. He is currently an Associate Professor with the Department of Information and Communication Technologies, UPCT. His current research focuses on learning algorithms for wireless network management.

JAVIER VALES-ALONSO received the M.Sc. degree in telecommunication engineering from the University of Vigo, Vigo, Spain, in 2000, and the Ph.D. degree in telecommunication engineering from the Technical University of Cartagena, Cartagena, Spain, in 2005, and the M.Sc. degree in mathematics from the National University of Distance Education, Madrid, Spain, in 2015. He is involved in different research topics related to modeling and optimization.