

Received August 23, 2021, accepted September 14, 2021, date of publication September 16, 2021, date of current version September 28, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3113461

# Reinforcement Learning for Energy-Efficient 5G Massive MIMO: Intelligent Antenna Switching

MARCIN HOFFMANN<sup>1</sup>, (Graduate Student Member, IEEE),

AND PAWEŁ KRYSZKIEWICZ<sup>1</sup>, (Senior Member, IEEE)

Institute of Radiocommunications, Poznań University of Technology, Polanka, 61-131 Poznań, Poland

Corresponding author: Marcin Hoffmann (marcin.ro.hoffmann@doctorate.put.poznan.pl)

This work was supported by the Polish Ministry of Science and Higher Education subvention within the task “New methods of increasing energy and spectral efficiency and localization awareness for mobile systems”, in 2020, under Grant 0312/SBAD/8158.

**ABSTRACT** To provide users with high throughputs, the fifth generation (5G) and beyond networks are expected to utilize the Massive Multiple-Input Multiple-Output technology (MMIMO), i.e., large antenna arrays. However, additional antennas require the installation of dedicated hardware. As a result the power consumption of a 5G MMIMO network grows. This implies, e.g., higher operator costs. From this angle, the improvement of Energy Efficiency (EE) is identified as one of the key challenges for the 5G and beyond networks. EE can be improved through intelligent antenna switching, i.e., disabling some of the antennas installed at a 5G MMIMO Base Station (BS) when there are few User Equipments (UEs) within the cell area. To improve EE in this scenario we propose to utilize a sub-class of Machine Learning techniques named Reinforcement Learning (RL). Because 5G and beyond networks are expected to come with accurate UE localization, the proposed RL algorithm is based on UE location information stored in an intelligent database named a Radio Environment Map (REM). Two approaches are proposed: first EE is maximized independently for every set of UEs' positions. After that the process of learning is accelerated by exploiting similarities between data in REM, i.e., REM-Empowered Action Selection Algorithm (REASA) is proposed. The proposed RL algorithms are evaluated with the use of a realistic simulator of the 5G MMIMO network utilizing an accurate 3D-Ray-Tracing radio channel model. The utilization of RL provides about 18.5% EE gains over algorithms based on standard optimization methods. Moreover, when REASA is used the process of learning can be accomplished approximately two times faster.

**INDEX TERMS** Reinforcement learning, massive MIMO, energy efficiency, radio environment maps, 5G, machine learning, wireless communication.

## I. INTRODUCTION

Mobile network throughput can be significantly increased by the utilization of the Massive Multiple-Input Multiple-Output (MMIMO) technology [1]. The idea of MMIMO is to equip a Base Station (BS) with an antenna array of a high number of elements, i.e., much greater than 1 [2]. The large number of antenna elements allows the system to, e.g., increase the amount of the wanted signal's power being received by the User Equipment (UE). It can be achieved through proper weighting of the signal transmitted from each of the BS antennas, i.e., beamforming. However, the underlying cost of improved network throughput is reduced Energy

Efficiency (EE). This phenomenon is mainly related to the increased power consumption caused by additional hardware related to each of the BS antennas [3]. As high power consumption affects network operators' costs and contributes to the world's carbon footprint, EE optimization is identified as one of the key challenges for fifth generation (5G) and beyond networks [4], [5]. There are several possibilities to improve EE of the MMIMO network. The highest gains are expected to be obtained by switching off underutilized BSs. Several algorithms have already been reported in the literature [6]. A representative example of those techniques is an algorithm that switches off a BS if its traffic can be offloaded to the neighboring cells [7]. However, some research has shown that frequent BS on/off switching can reduce the lifetime of BSs, and cause high replacement costs [8]. On the other hand,

The associate editor coordinating the review of this manuscript and approving it for publication was Miguel López-Benítez<sup>1</sup>.

it has been shown that energy consumption scales nearly linearly with the number of active BS antennas, e.g., during the following studies [9], [10]. This number can be adjusted to the network state, e.g., the number of connected UEs. It has been shown that not all antennas contribute equally to the overall array gain. Therefore, a proper antenna selection algorithm should be used to obtain a subset of active antennas, to, e.g., improve EE [11]. While this problem is difficult to be fully specified and solved, because of its complexity, several heuristic metrics have been proposed to reduce its computational complexity. The simplest one is to select antennas characterized by the highest mean channel gains to all served UEs [12]–[14]. A more advanced solution additionally takes into the account spatial channel response correlations within the antenna array installed at the BS [15], [16]. Another approach is to use a channel gain-based antenna selection algorithm as an initial solution and then run a bio-inspired optimization algorithm [17]. On the other hand it has been shown that greedy algorithms can be successfully used for the purpose of antenna selection [18]. Finally antenna selection can be realized with the use of a machine learning techniques [19], [20]. However, machine learning-based antenna selection requires large training sets. Moreover learning of the models is very time-consuming when compared against e.g. channel gain-based algorithms.

The major drawback of the above mentioned papers is that they do not provide a good answer to the question: How many antennas do we need? Most of them only propose the order in which the antennas should be deactivated, i.e., an antenna selection algorithm. Though there are some papers where closed-form expressions are obtained for the number of active antennas that maximize MMIMO network EE [21]–[23]. However, the authors assume that every antenna contributes equally to the array gain, and simple Shannon formula-based throughput estimation is used. For real radio conditions, the network throughput is typically lower than a Shannon formula-based one. As a result of this inaccuracy, the obtained number of active antennas is misestimated, and the resultant EE is not optimal. One should also notice that, it is not a trivial task to use standard optimization methods to improve MMIMO network EE under real conditions. Mainly because the MMIMO system is too complex to obtain its accurate analytical model including: precoder, scheduler, intra-cell interference and realistic radio channel model.

Taking into account the limitations of the standard optimization methods, our proposal is to use machine learning for intelligent antenna switching in a complex MMIMO system. The solution is divided into two parts. First, antennas are sorted, e.g., based on their mean channel gains. Next, the optimal number of active antennas which maximize EE under a given pattern of UE positions is obtained using Reinforcement Learning (RL). The process of RL requires memory to store the results of learning, and necessary parameters. For this purpose, MMIMO BS is expected to be equipped with an intelligent database of location-dependent data known as a Radio Environment Map (REM) [24]. Although state of

the art REMs were designed to store value-location tuples, we propose to map a set of all connected UE positions onto the number of active antennas. A similar representation has been used in the context of base station switching in [25]. We propose to learn how many antennas are needed through interaction with the MMIMO network, i.e., using RL. The high-level idea of RL application is that for a given set of UE positions, various numbers of active antennas would be tested, and their resultant EE would be observed, in order to learn how to act in the future. In RL there is a problem of balancing gaining new knowledge (exploration), and exploitation of the current knowledge. At first, we utilize for this purpose a state-of-the-art algorithm named Upper-Confidence-Bound [26]. Then, we show that the utilization of similarities between UE positions saved in REM, can accelerate the learning phase. It shows that a combination of RL with REM produces a synergy effect. For the evaluation purpose, an advanced system-level simulator of the MMIMO network is used, together with an accurate 3D-Ray-Tracing radio channel model. While the paper focuses on downlink transmission only, being the most important factor to obtain high EE for a given network operator, the proposed method can easily be extended for uplink transmission.

The main contributions of this paper are as follows:

- Comparison between the optimal number of active antennas indicated by an algorithm based on standard optimization methods [21], [22], [27], and the optimal number of active antennas based on the observation of EE in a realistic simulator based on a 3D-Ray-Tracing radio channel model.
- Proposal of an algorithm combining REM, and RL schemes to map the network state (specified by UE position pattern) to the number of active antennas maximizing EE.
- Exploiting the similarities between sets of UE positions saved in REM in order to accelerate the learning phase.

The paper is organized as follows: Sec. II provides the reader with an overview of the system model. In Sec. III presents an intelligent antenna switching algorithm, including a description of the antenna selection algorithm and two approaches to active antenna number computation: state-of-the-art, and REM based. The proposed RL scheme aimed at providing REM with information about the optimal number of active antennas is described in Sec. IV. In Sec. V, the results of computer simulations are presented. The paper is concluded in Sec. VI. To improve readability all acronyms are listed in Tab. 1.

## II. SYSTEM MODEL

This paper aims at maximization of EE through Intelligent Antenna Switching, while considering downlink in a single MMIMO cell. In the cell there is one BS equipped with  $M$  antennas arranged in a rectangular array, and serving  $K$  single-antenna UEs. The BS total transmission power does not depend on the number of active antennas, and

TABLE 1. List of abbreviations.

Acronym	Full term(s)
AANCB	Active Antenna Number Computation Block
ASB	Antenna Selection Block
BS	Base Station
ETP	Effective Transmitted Power
EE	Energy Efficiency
5G	fifth generation
MMIMO	Massive Multiple-Input Multiple-Output
MRC	Maximum Ration Combining
OFDMA	Orthogonal Frequency-Division Multiple Access
REM	Radio Environment Map
RTK	Real Time Kinematics
RL	Reinforcement Learning
REASA	REM-Empowered Action Selection Algorithm
RB	Resource Block
TCP	Transceiver Chains Power
UCB	Upper Confidence Bound
UE	User Equipment
ZF	Zero-Frocing

has a constant value of  $P_T$  for fair comparison. A downlink in the considered MMIMO cell is based on Orthogonal Frequency-Division Multiple Access (OFDMA), where the available spectrum is divided into  $N_{rb}$  Resource Blocks (RBs). UEs are capable to report their positions with the use of standard cellular network mechanisms. They are claimed to utilize high accuracy satellite navigation available in 5G and beyond networks, e.g., Real Time Kinematics (RTK), providing cm-level accuracy [28], [29]. Less accurate location information is expected to decrease the performance of the proposed algorithms but can be an interesting subject of future studies similar to [30]. The process of intelligent antenna switching consists of the two stages, and requires two additional functional blocks to be installed at the BS: Active Antenna Number Computation Block (AANCB) and Antenna Selection Block (ASB), as depicted in Fig. 1. AANCB is responsible for providing information about the number of active antennas  $\hat{m}$  (out of all available  $M$ ) to the MMIMO BS, e.g., on the basis of current positions of UEs. Later on, ASB selects  $\hat{m}$  active antennas indicated by AANCB, according to a given selection metric. AANCB and ASB cooperate to improve EE of the considered MMIMO cell.

A. POWER CONSUMPTION MODEL

Although there are several definitions of EE in the literature, all of them consist of the term related to the throughput and the term related to the power consumption. Thus before the problem of the EE optimization can be formulated, first, a power consumption model proper for the MMIMO BS must be defined. There are already several power consumption models proposed, for a single-antenna BS, or a MIMO BS [31], [32]. These are not adequate for MMIMO, where BSs are equipped with possibly above hundred-element antenna arrays. A popular model of MMIMO BS power consumption can be found in [2]. Similarly as in our previous work, we will consider three components having major impact on power consumption [25]:

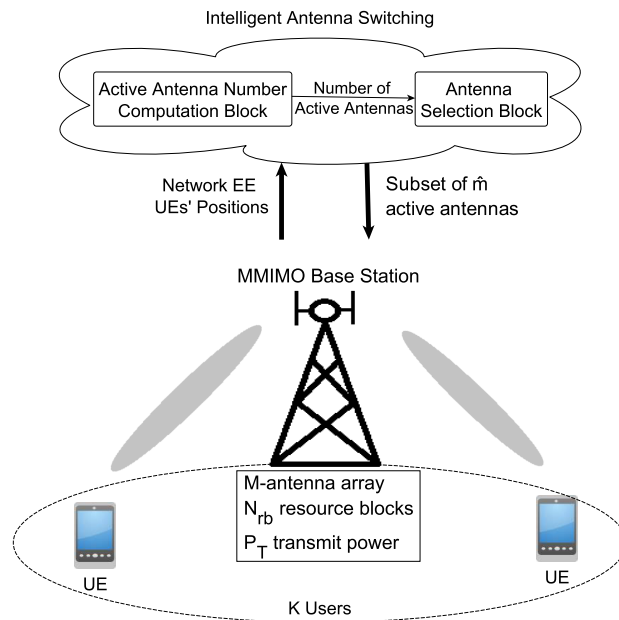


FIGURE 1. The concept of Intelligent Antenna Switching realized by extending the MMIMO BS by Active Antenna Number Computation Block, and Antenna Selection Block.

- **Effective Transmitted Power (ETP)**  $P_{ETP}$  is the total power  $P_T$  transmitted by the MMIMO BS, taking into the account the power amplifier efficiency  $\eta$ :

$$P_{ETP} = \frac{P_T}{\eta}. \tag{1}$$

- **Transceiver Chains Power (TCP)**  $P_{TCP}(m)$ , is the power consumed by the local oscillator  $P_{LO}$ , and  $m$  active transceiver chains (equivalent to the number of active antennas), each of power consumption  $P_{TCP}$ :

$$P_{TCP}(m) = m \cdot P_{TCP} + P_{LO}. \tag{2}$$

- **Fixed Power**  $P_{fix}$ , is the constant amount of power consumed by the BSs for, e.g., backbone communication or processing of signaling information.

The overall power consumption of the considered MMIMO BS is given by:

$$P_{tot}(m) = P_{ETP} + P_{TCP}(m) + P_{fix}. \tag{3}$$

B. PROBLEM FORMULATION

Having a power consumption model proper for MMIMO BS, a definition of EE can be introduced, and optimization problem can be formulated. Several definitions of EE can be found in the literature [33]. The most commonly used is the one, where EE is given by the ratio of average network throughput, and average network power consumption. We proposed a slight modification to this definition [25]:

$$EE(\mathbf{s}, m) = \frac{c_{50}(\mathbf{s}, m)}{P_{tot}(m)}, \tag{4}$$

where  $\mathbf{s}$  denotes the set of UE positions, i.e., state,  $m$  is the number of active antennas,  $P_{\text{tot}}(m)$  is the power consumption related to  $m$  active antennas, computed according to (3), and  $c_{50}(\mathbf{s}, m)$  stands for the median user bitrate related to the set of UEs at positions  $\mathbf{s}$ , and  $m$  active antennas. In the original definition of EE, users having high bitrates contribute much to the average network EE. Instead, our proposition improves the fairness by protecting users characterized by poor radio conditions. The optimization problem addressed in this paper is to maximize EE (4) independently for all visited states, i.e., sets of UE positions  $\mathbf{s}$ . In every state  $\mathbf{s}$ , the optimization is achieved by adapting the number of active antennas  $\hat{m}$ , so as to maximize EE (4):

$$\max_m EE(\mathbf{s}, m), \forall \mathbf{s} \in \mathcal{S}, \quad (5)$$

where  $\mathcal{S}$  denotes a set of all considered states.

### III. INTELLIGENT ANTENNA SWITCHING ALGORITHM

To obtain the subset of active antennas we propose to extend the MMIMO BS by two additional functional blocks creating together the Intelligent Antenna Switching Algorithm, i.e., AANCB and ASB. Our aim is to split the process between AANCB deciding on the number of active antennas, and ASB, responsible for obtaining the exact subset of active antennas. The motivation behind this split is to reduce the computational complexity. Without ASB, there would be  $2^M$  possible configurations of active antennas to be evaluated. With the introduction of ASB there are only  $M$  possible numbers of active antennas to be selected.

#### A. ANTENNA SELECTION BLOCK

The aim of ASB is to select a subset of  $\hat{m}$  active antennas from all  $M$  antennas installed at the MMIMO BS. This procedure is known as the antenna selection algorithm. In this paper we assumed to arbitrarily choose the antenna selection algorithm that will be used by ASB. To focus on learning REM with a proper number of active antennas, rather than the improvement of the antenna selection algorithm, a state-of-the-art solution based on average channel gains has been chosen [11]. The  $\hat{m}$  antennas having the highest channel gain averaged over  $K$  users, and  $N_{\text{rb}}$  RBs, have been chosen to be active. This algorithm is both intuitive and characterized by simple implementation.

#### B. ACTIVE ANTENNA NUMBER COMPUTATION BLOCK

The aim of AANCB is to compute the number of active antennas that will maximize the EE of a MMIMO cell under a given set of UE positions  $\mathbf{s}$ . The computed number of active antennas is passed to ASB to obtain the exact subset of active antennas. Two versions of AANCB will be described: a state-of-the-art solution based on analytical computations, and a REM-based idea.

##### 1) STATE-OF-THE-ART SOLUTION

The MMIMO network EE has been optimized under the Zero-Forcing (ZF) precoding scheme [21], [22], [27], and

the Maximum Ratio Combining (MRC) precoding scheme [34], [35]. The focus of this paper is on a more complex ZF precoder (see Sec. V). The number of active antennas therein is adapted to the number of users connected to MMIMO BS. While the UE position has a significant impact on wireless channel characteristics, it is not taken into account in the state-of-the-art papers. Although the authors obtained closed form expressions for the number of antennas, their simulation scenarios, EE definition and power models slightly differ from the ones presented in this paper. Thus, these expressions can't be directly implemented. Instead, we will adapt them to the considered definition of EE (4). The bitrate of the UE located at position  $s$  can be calculated as a function of the number of active antennas  $m$ , using an approximation proper for the ZF precoder [27]:

$$\hat{c}(s, m) = B \cdot \log \left( 1 + \frac{(m - K)P_{\text{T}}\beta(s)}{K\sigma^2} \right), \quad (6)$$

where  $\beta(s)$  is a mean channel gain between a user located at position  $s$  and the BS, averaged over all antennas and RBs,  $B$  is the system bandwidth, and  $\sigma^2$  is the power of the white Gaussian noise at the user's receiver. Using the above formula, a median user bitrate  $\hat{c}_{50}(\mathbf{s}, m)$  can be estimated in the function of active antennas  $m$ :

$$\hat{c}_{50}(\mathbf{s}, m) = \text{median} [\{\hat{c}(s, m)\} | s \in \mathbf{s}]. \quad (7)$$

Then, the number of active antennas  $\hat{m}$  related to the set of UE positions  $\mathbf{s}$  can be computed by solving:

$$\hat{m} = \arg \max_m \left\{ \frac{\hat{c}_{50}(\mathbf{s}, m)}{P_{\text{tot}}(m)} \right\}. \quad (8)$$

Because obtaining a closed form expression of this equation is out of this paper's scope, numerical methods will be used.

##### 2) REM-BASED SOLUTION

The state-of-the-art solution assumes that only the number of UEs connected to MMIMO BS and their path losses should be taken into account while computing the number of active antennas. However,  $K$  users can create various spatial patterns, resulting in different radio conditions, e.g., their channels can be more or less correlated or follow various non-Rayleigh distributions. Such properties of realistic MMIMO channels has been confirmed by measurements, e.g., [36]. Thus, we propose to optimize the number of active antennas separately within each set of UE positions. For this purpose an intelligent database called REM is employed at the MMIMO BS. The aim of REM is to store and process information related to the given set of UE positions, e.g., power consumption, bitrates, number of active antennas. As a long-term result of machine learning, REM provides MMIMO BS with information about the optimal number of active antennas, maximizing EE for a given set of UE positions.

The data stored in REM is organized in entries, as depicted in Fig. 2. Every new set of UE positions  $\mathbf{s}$  creates a unique REM entry, labeled with  $\mathbf{s}$ . Besides the set of UE positions,

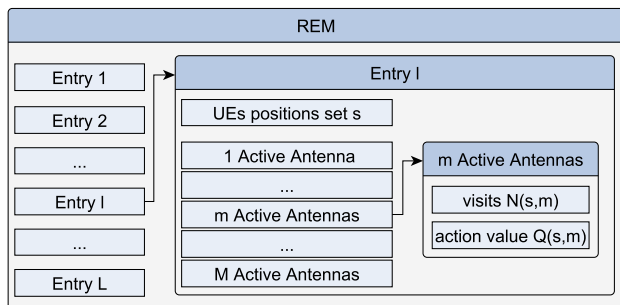


FIGURE 2. The structure of data stored in REM.

each REM entry contains, for each considered number of active antennas  $m$ , the so-called action value  $Q(s, m)$  used by RL, and computed on the basis of  $EE(s, m)$  observed previously. These variables will be described in detail in the next section. Finally, information about the number of times a particular number of active antennas has been tested  $N(s, m)$ , is stored in each REM entry as well.

#### IV. REINFORCEMENT LEARNING SCHEME

The solution proposed in this paper is to provide the network with an optimal number of active antennas, maximizing EE, separately for each set of UE positions. This information is to be stored in REM. However, REM must first get this knowledge. An effective approach to this task is to use RL, where the process of learning is based on interaction, i.e., a so called agent interacts with an environment in discrete time steps, by taking so-called actions, and observing their outcome-reward [26]. This procedure is depicted in Fig. 3 in the context of filling REM with information about the most energy efficient number of active antennas. The main elements of an RL scheme can be described as follows:

- **Environment** is a single cell in the MMIMO network equipped with  $M$  antennas and serving  $K$  UEs, as described in Sec. II.
- **A State  $s$**  is the set of UE positions. First, the currently reported set of UE positions is compared against REM entries. If the reported set of UE positions cannot be found in REM, a new entry is created.
- **An Action  $m$**  is the considered number of active antennas,  $m \in \{1, \dots, M\}$  being the output of AANCB. The exact configuration of active antennas is obtained by ASB, as described in Sec. III-A.
- **A Reward  $r(s, m)$**  in this case is EE computed according to (4). The reward is obtained after an observation period called *step*, when a given action  $m$  is under evaluation. The step duration has to be long enough for an average instantaneous EE metric variation caused by short-term changes in scheduling and channel coefficients.
- **An Agent** is a REM unit deployed at MMIMO BS as AANCB. It is responsible for taking actions  $m$  according to the environment state  $s$ , and updating action preferences on the basis of observed reward  $r(s, m)$ .

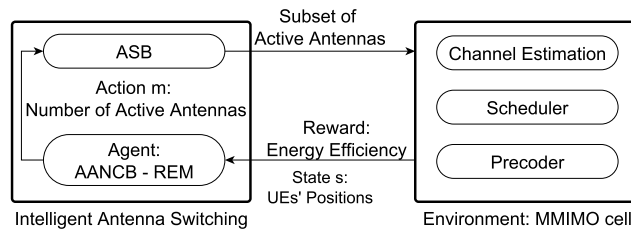


FIGURE 3. The RL scheme for training REM.

#### A. ACTION VALUES UPDATE RULE

In REM there is information about potential EE related to each possible action  $m$  (the number of active antennas) in state  $s$ , based on previous experience. In RL it is known as an action value  $Q(s, m)$ . Each time an agent receives a reward, the related action value is updated. In general, the action values depend both on the reward obtained after the current step and the action values of the next state. However, in the considered model the movement of UEs does not depend on the number of active antennas, i.e., the action does not affect the future state. This case is known in the literature as the problem of *Contextual Bandit* or *Associative Search* [26]. As a result it is enough to use the reward observed after performing the current action. Thus, the update rule is a simple exponential average of rewards observed after each visit in state  $s$  taking action  $m$ :

$$Q(s, m) \leftarrow (1 - \alpha) \cdot Q(s, m) + \alpha \cdot r(s, m), \quad (9)$$

where  $\alpha \in [0, 1]$  is a step size parameter.

#### B. UPPER CONFIDENCE BOUND SELECTION ALGORITHM

One of the challenges related to RL is the so-called exploration-exploitation problem, i.e., the problem of how much time should an agent spend on learning the environment by taking fewer explored actions in contrast to exploiting previously obtained knowledge by taking the best known action. There are several algorithms designed to balance exploration and exploitation, e.g.,  $\epsilon$ -greedy, soft-max distribution, Upper Confidence Bound (UCB) [37]. Due to the so-called channel hardening property of MMIMO systems, the radio signal after the procedure of precoding is almost immune to small-scale fading [38]. Thus, from the perspective of UE, achievable bitrates should be relatively stable, i.e., affected mainly by slow-varying large-scale fading. Due to that fact, we decided to utilize the UCB algorithm, because it is not utilizing randomness for action selection, i.e., the algorithm is not blindly exploring actions that are expected to result in low reward. Action selection, following UCB algorithm, is given as follows [26]:

$$\hat{m} = \arg \max_{1 \leq m < M} \left\{ Q(s, m) + c \cdot \sqrt{\frac{\ln \sum_{k=1}^M N(s, k)}{N(s, m)}} \right\}, \quad (10)$$

where  $N(s, m)$  stands for the number of times action  $m$  was selected in state  $s$ , term  $\sum_{k=1}^M N(s, k)$  is the total number of

visits in state  $s$ , and  $c$  is an arbitrary constant. The overall idea of UCB is to promote potentially promising actions that have been rarely selected. It is achieved through the introduction of the term  $c \cdot \sqrt{\frac{\ln \sum_{k=1}^M N(s, k)}{N(s, m)}}$  in (10). This term grows for uncertain  $Q(s, m)$  estimates, i.e., obtained with a low number of measurements. It is smaller for the actions that are selected often [26]. To control the balance between exploration and exploitation the constant  $c$  is introduced. For a high value of  $c$  algorithms tend to be more focused on exploration, while for low values of  $c$ , on exploitation. In the extreme case of  $c = 0$ , UCB becomes a greedy algorithm focused only on exploitation.

Additionally, in every state  $s$ , the UCB algorithm is forced to take each action at least once, even when  $c$  is set to zero. This behavior is called optimistic initialization [26]. It is obtained by initially setting  $Q(s, m)$  to very large values. After the first reward is observed the update rule is  $Q(s, m) \leftarrow r(s, m)$  instead of (9), that is used for next rewards.

### C. REM-EMPOWERED ACTION SELECTION ALGORITHM

In the case of the UCB algorithm, the EE optimization is strictly following the *Associative Search* rule, i.e., no information about other states is used. However, it can be expected that similar sets of UE positions would result in a similar number of antennas that should be active. Such a phenomenon was observed during our previous studies, exploiting the problem of BSs switching [25]. Therein, we have taken advantage of these similarities between states (REM entries) to speed up the process of learning. Thus, in this paper we propose an approach named REM-Empowered Action Selection Algorithm (REASA) for the purpose of active antenna selection. We expect that due to utilization of the similarities between REM entries REASA can achieve the same end level of EE as in the case of UCB, but requiring less time. REASA is designed as an extension of the UCB algorithm. In the case of UCB action selection is based on one action value related to the current state, while REASA uses action values  $Q(s, m)$  from all states averaged proportionally to the distance to the current state. Suppose that  $\{s_l\}_{l=1}^L$  denotes states saved in REM, and  $s_i$  is the currently reported set of UE positions. REASA is given by:

$$\hat{m} = \arg \max_{1 \leq m < M} \left\{ \hat{Q}(s_i, m) + c \cdot \sqrt{\frac{\ln \sum_{k=1}^M \hat{N}(s_i, k)}{\hat{N}(s_i, m)}} \right\}, \quad (11)$$

where:

$$\hat{Q}(s_i, m) = \frac{Q(s_i, m) + \sum_{l=1, l \neq i, Q(s_l, m) \neq 0}^L \frac{Q(s_l, m)}{d_h(s_i, s_l)^\gamma}}{1 + \sum_{l=1, l \neq i, Q(s_l, m) \neq 0}^L \frac{1}{d_h(s_i, s_l)^\gamma}}, \quad (12)$$

$$\hat{N}(s_i, m) = \frac{N(s_i, m) + \sum_{l=1, l \neq i}^L \frac{N(s_l, m)}{d_h(s_i, s_l)^\gamma}}{1 + \sum_{l=1, l \neq i}^L \frac{1}{d_h(s_i, s_l)^\gamma}}, \quad (13)$$

$\gamma$  is an arbitrary constant that scales the impact of more distant REM entries on the current action selection, and  $d_h(s_i, s_l)$  is the Hausdorff Distance between  $s_i$  and  $s_l$ . It is computed according to [39]:

$$d_h(s_i, s_l) = \max\{h(s_i, s_l), h(s_l, s_i)\}, \quad (14)$$

where:

$$h(s_a, s_b) = \max_{\mathbf{x}_k \in s_a} \{ \min_{\mathbf{x}_j \in s_b} \delta(\mathbf{x}_k, \mathbf{x}_j) \}, \quad (15)$$

and  $\delta(\cdot, \cdot)$  denotes a Euclidean Distance between two points on the Cartesian plane:  $\mathbf{x}_k$  and  $\mathbf{x}_j$ . Each of these points represents the coordinates of a single UE. One should note that the Hausdorff Distance is used, as it is a state-of-the-art measure of similarity between two sets of points of possibly different sizes, used e.g., for the purpose of image processing [39]. However, the performance of other distance metrics should be compared in the future, similarly to [30]. Comparing to the UCB, tuning parameters of REASA may be more challenging because there is an additional parameter:  $\gamma$ .

The procedure of RL-based Intelligent Antenna Switching using either UCB or REASA action selection is summarized as Algorithm 1.

---

### Algorithm 1 RL-Based Intelligent Antenna Selection

---

**Require:** Sequence of states  $\mathcal{S}$

**for each:**  $s \in \mathcal{S}$

- 1: **if**  $s$  not in REM **then** create new entry
  - 2: **end if**
  - 3: obtain action  $\hat{m}$  (number of active antennas) using UCB eq. (10) or REASA eq. (11)
  - 4: increment  $N(s, \hat{m})$
  - 5: select  $\hat{m}$  active antennas using ASB (as described in Sec. III-A)
  - 6: wait till new reward  $r(s, \hat{m})$  is obtained
  - 7: **if**  $N(s, \hat{m}) == 1$  **then**  $Q(s, \hat{m}) \leftarrow r(s, \hat{m})$
  - 8: **else** update action value  $Q(s, \hat{m})$  according to eq. (9)
  - 9: **end if**
- 

## V. SIMULATIONS RESULTS

The algorithms aimed at increasing the EE of a MMIMO network by switching off selected antennas, are evaluated in this section. For this purpose an advanced system-level simulator of a 5G-like, OFDMA-based MMIMO cell described in Sec. II is used. The cell covers mainly a park area surrounded by approximately 45 m tall buildings, according to the *Madrid Grid Model* [40]. The simulation parameters are presented in Tab. 2. The multi-stage system-level simulator used in this manuscript includes, e.g., channel estimation using Sounding Reference Signals, user scheduling with the proportional fair rule and ZF precoding, as presented in detail in [25]. We are considering a medium range BS, of transmit power equal to 38 dBm [41]. The BS is equipped with a rectangular antenna array of 128 elements placed in 8 rows and 16 columns. The available bandwidth equals 300 MHz

TABLE 2. Simulation Parameters.

Parameter	Value
Simulated Time	15 s
Time Slot Duration	0.5 ms
Number of UEs $N_{UE}$	40
UEs Speed $v$	$1.5 \frac{m}{s}$
Central Frequency	3.55 GHz
Bandwidth	300 MHz
Subcarrier Spacing	30 kHz
Number of RBs $N_{rb}$	272
Number of BS Antennas $M$	128 ( $16 \times 8$ )
Precoder Type	ZF
Transmitted Power $P_T$	38 dBm [41]
Amplifier Efficiency $\eta$	0.5 [2]
Transceiver Chain Power $P_{TC}$	0.4 W [2]
Local Oscillator Power $P_{LO}$	0.2 W [2]
BS Fixed Power $P_{fix}$	10 W [2]

around the center frequency of 3.55 GHz. Every 0.5 ms the scheduling algorithm allocates all 272 RBs to users connected to the MMIMO BS. Resource allocation is done independently for each RB following the proportional fair rule, i.e., the ratio between potential and past bitrate [42]. Beamforming is realized with the use of the so-called Zero-Forcing (ZF) precoding, aiming at the suppression of intra-cell interference [10]. There are 7 UEs in the cell moving with the speed of 1.5 m/s. Their initial positions are drawn from a uniform distribution. The location of the BS (larger dot) together with a single realization of UEs' initial positions (smaller dots) is depicted in Fig. 4. This is a low traffic scenario, that can be observed e.g., during night time. This scenario is expected to allow for a significant number of transmission chains to be deactivated, significantly increasing EE. The coefficients of radio channels are generated with the use of a realistic 3D-Ray-Tracing model [43]. The channel model individually tracks propagation paths between the UE antenna and each of the BS's antennas. Up to two reflections of each ray are considered (to limit computational complexity), but with many scattering rays possible. To produce accurate coefficients of the radio channel, 3D-Ray-Tracer takes into account the urban scenario, including reflections from the buildings and other obstacles, e.g., randomly distributed and moving pedestrians. To mimic the radio channel estimation error, a zero-mean Gaussian-distributed random variable is added to the real channel coefficients [44].

### A. TERMINOLOGY

The process of evaluation of the EE-improving algorithms consists of performing RL for several sets of initial UE positions. For each set of initial UE's positions, these users and their movement is observed over some time period. To clarify the design of the simulation experiment the following terminology is introduced:

- **A Step** is a time period of 60 ms when one cycle of RL is performed. First, initial 10 ms are discarded from statistics as a start-up phase, allowing, e.g., the scheduler to stabilize. After these 10 ms, the state is identified

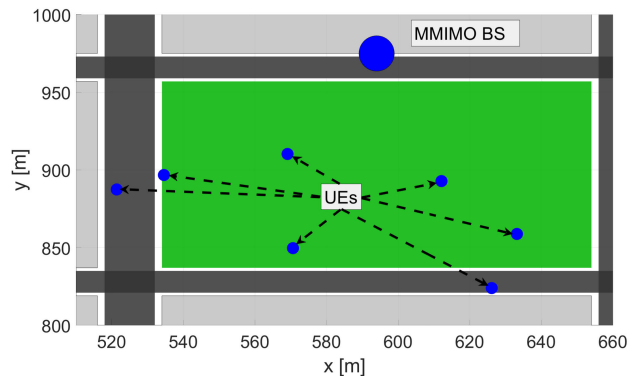


FIGURE 4. Deployment of MMIMO BS, together with a single realization of UEs' initial positions.

and the agent takes an action (decides on the number of active antennas) using values of  $Q(s, m)$  and  $N(s, m)$  stored in REM. After the remaining 50 ms the EE of the network is reported as a reward. While the computer simulator utilizing the Ray-Tracing-based channel is highly computationally complex, after these 60 ms, a time shift of 1 s is introduced and another step begins, starting from the recognition of a new state. This allows us to observe significant changes in UEs positions in a reasonable simulation time.

- **An Episode** is a sequence of several steps with the same UE path. During every episode, the same set of states is visited. From the network perspective the UE pattern and their path being the same as at some time in the past means that a given episode repeats. To assess algorithms in terms of learning speed, the average EE obtained during consecutive episodes would be observed. After many repetitions of a single episode, the UCB algorithm should focus on exploitation. In this case the mean EE over a single episode shows the average EE over UE paths (over all visited states).
- **An Experiment** consists of a constant number of episodes. Different experiments are characterized by different initial positions of the UEs. The presented results would often be averaged over several experiments, to average simulation results over different possible positions of UEs in the cell.
- **Ref** stands for the state-of-the art algorithm determining the number of active antennas, defined in Sec. III-B1.
- **No EE** denotes a scenario, where there is no EE optimization, i.e., all  $M$  antennas are active.
- **REM-UCB** denotes the algorithm of learning REM with the use of UCB described in Sec. IV-B
- **REASA** is the algorithm of learning REM, utilizing similarities between REM entries from Sec. IV-C

The dependencies between Experiment, Episode and Step, are summarized in Fig. 5.

### B. HOW MANY ANTENNAS ARE NEEDED?

The aim of the first simulations is to show that the optimal number of active antennas obtained for the full system model

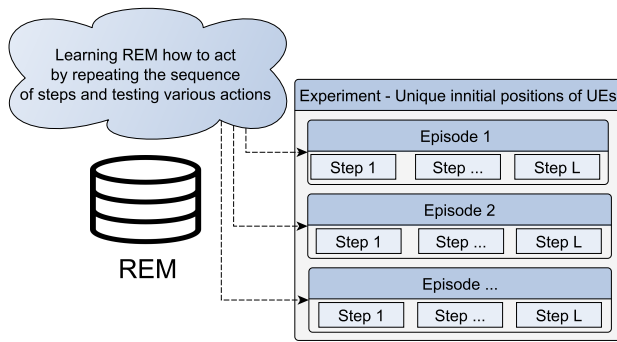


FIGURE 5. Diagram of dependencies between Experiments, Episodes and Steps.

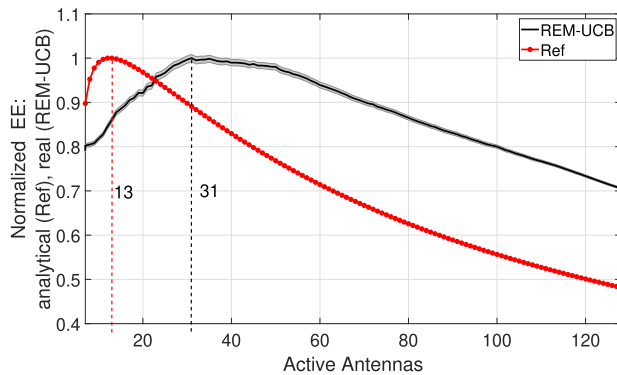


FIGURE 6. Comparison of normalized EE: analytical (ref), and real (REM-UCB), on the basis of 50 experiments. Shaded area around REM-UCB denotes a 95% confidence interval.

with the REM-UCB algorithm is different than obtained for simplified modeling with the Ref algorithm. For this purpose 50 experiments are conducted assuming perfect channel knowledge to make environment invariable. For such conditions, the number of episodes required to obtain knowledge about every action is equal to the number of antennas  $M = 128$ . Therefore, the number of episodes is set to the value of 128, making REM-UCB only to explore all possible actions, i.e., an exhaustive search is performed. Each episode consists of 1 step, i.e., 50 different random patterns of UE positions are tested. This allows us to estimate the average EE for 7 UEs randomly distributed in a cell. Fig. 6 presents the reward related to each action, averaged over all experiments, and normalized by the highest value. The red line refers to the Ref algorithm, while the black one to REM-UCB. The shaded area around REM-UCB marks the 95% confidence interval for a mean estimate. It can be seen that the Ref algorithm on average chooses a lower number of active antennas than REM-UCB, i.e., 13 against 31. It is because the Ref algorithm utilizes a simplified system model. It assumes that all antennas contribute equally to BS throughput, and that all users can be simultaneously served by the BS. In practice, due to channel correlations, some UEs cannot be served using the same time-frequency resources. Due to the antenna array geometry and correlations between antennas, they do not contribute equally to the throughput. That is why the

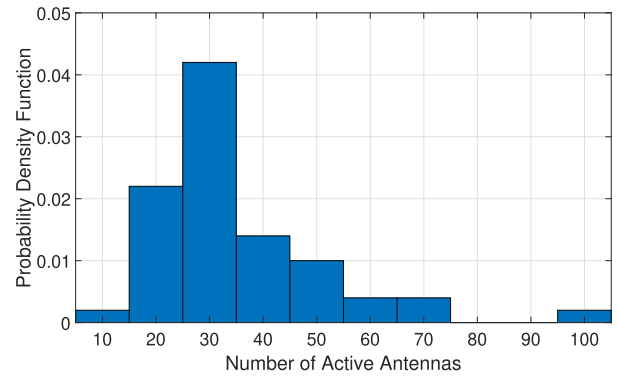


FIGURE 7. Probability Density Function of the optimal number of active antennas over 50 experiments obtained using the REM-UCB algorithm.

average optimal number of active antennas obtained through direct observations of EE via REM-UCB is larger. However, the results depicted in Fig. 6 are averaged over experiments related to different sets of UE positions. For each set of UE positions, a different number of active antennas can be optimal. In Fig. 7 a probability density function is shown for the optimal number of active antennas obtained for a full system model, over considered 50 experiments. It can be seen that, although the average optimal number of active antennas is around 30 (exactly 31 while considering Fig. 6), different patterns of UE positions can result in an optimal number of active antennas as low as 10, or as high as 100. This result justifies the deployment of REM, where the number of active antennas can be optimized independently for different patterns of UE positions.

### C. EVALUATION OF REM-UCB

The previous observations have shown that the number of active antennas indicated by the Ref algorithm is typically underestimated, and that the optimal number of active antennas should be adjusted to the particular pattern of UE positions. For this purpose, a REM-UCB learning algorithm is utilized. To evaluate the phase of learning, 10 experiments were conducted. Each experiment exploits 160 episodes. Because REM-UCB utilizes the so-called optimistic initialization, the Q-values have initially very big values to enforce checking each action at least once, as explained at the end of Sec. IV-B. Thus the 128 initial episodes are necessary to evaluate every possible action in each visited state once, while the remaining 32 episodes provide an evaluation of the algorithm performance under different values of  $c$  parameter, after initial “exploration” is finalized. Every episode consists of 15 steps. The resultant average reward-EE obtained after each episode and averaged over experiments is depicted in Fig. 8. We can see that during the initial 128 episodes, when each possible action is tested, the reward first increases and after about 40 episodes starts to decrease. The first reason is related to the implementation, i.e., actions are tested starting from one to all active antennas over initial 128 episodes. Maximum can be observed because both having too many and not enough active antennas causes poor EE. After every



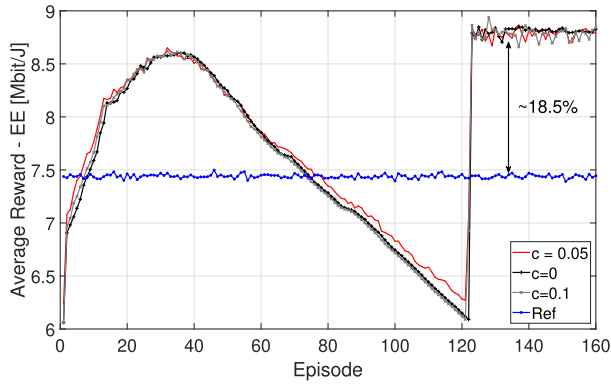


FIGURE 8. Average reward (EE) obtained with the use of REM-UCB, after each episode, and averaged over 10 experiments.

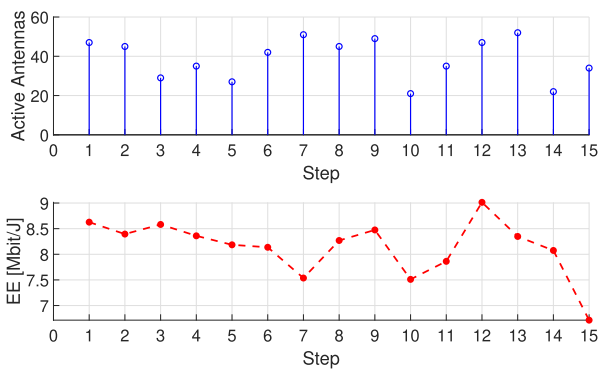


FIGURE 9. Optimal number of active antennas and related EE in the consecutive steps of Experiment 1.

action is taken at least once the algorithm is no longer forced to take unprofitable actions. As a result, after about 120-th episode reward rapidly grows, and remains relatively stable and provides about 18.5% EE improvement over the Ref algorithm. During the last 32 episodes, the least fluctuations are observed, when  $c = 0$ . In this case, the algorithm performs greedily, i.e., it exploits the current knowledge by always taking the best known action. This is caused by the fact that in the considered scenario all UEs have a line-of-sight connection to the BSs, and thus related radio channels are stable, even though the channel estimation error is introduced.

To highlight the fact that every set of UE positions can be potentially related to a different number of active antennas, we have plotted the optimal number of antennas and EE related to each step in Experiment 1. The results are depicted in Fig. 9. It can be seen that the optimal number of active antennas varies from about 20 to about 50 between steps. Similarly the related EE depends on the UE positions.

EE is defined as the ratio of the median-user bitrate and mean power consumption as shown in (4). Fig. 10 depicts separately power consumption and median-user bitrate obtained with the use of Ref algorithm, REM-UCB, and under No EE scenario. Results are averaged over the last 32 episodes, i.e., the period when reward becomes stable. REM-UCB results are obtained for  $c = 0$ . It can be seen that both power

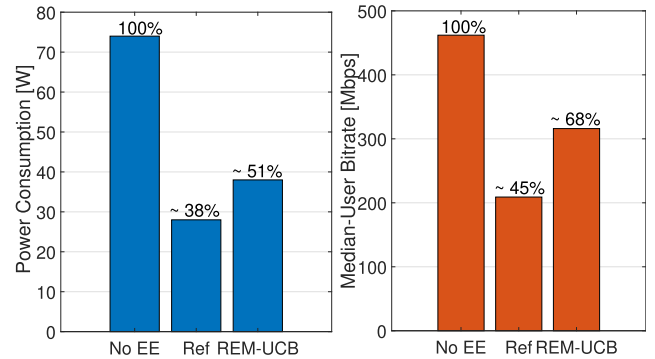
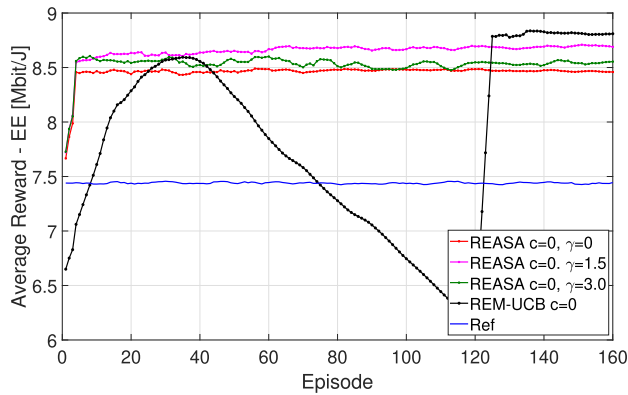


FIGURE 10. Power Consumption and Median-User Bitrate averaged over the last 32 episodes. REM-UCB results obtained for  $c = 0$ .

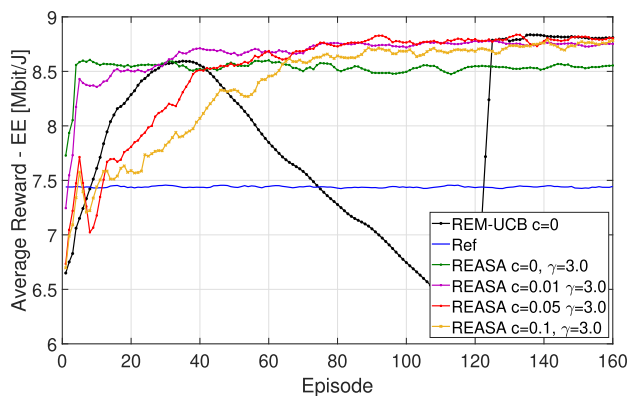
consumption and median-user bitrate are decreased as a result of the Intelligent Antenna Switching in comparison to the No EE scenario. In the case of the Ref algorithm, higher power savings are observed, comparing to the REM-UCB: about 38% of No EE scenario power consumption against about 51%. However, the Ref algorithm can achieve median-user bitrate only as high as about 45% of bitrate achieved under the No EE scenario, while REM-UCB achieves about 68%. That is why the resultant EE achieved by REM-UCB is better than the Ref algorithm.

#### D. EVALUATION OF REASA

Although the REM-UCB algorithm provides REM with the optimal number of active antennas for a given set of UE positions, we hope that the learning phase can be speed up, i.e., EE gains provided by REM-UCB after about 120 episodes can be reached faster. The speed-up can be achieved by the utilization of similarities between the patterns of UE positions, i.e., a similar pattern of UE positions would probably be related to a similar optimal number of active antennas. This idea is implemented in the REASA algorithm described in Sec. IV-C. We have evaluated the REASA algorithm under the same scenario as REM-UCB: 10 experiments, each consisting of 160 episodes exploiting 15 steps. In addition, the REM is from the beginning filled with 50 entries. These entries had been obtained during simulations reported in Sec. V-B. Most importantly, the experiments generated here are independent from those reported in Sec. V-B, i.e., UE positions and paths are not repeated. The resultant average reward-EE obtained after each episode and averaged over experiments is depicted in Fig. 11. As it can be observed, due to the utilization of REASA together with REM possessing initial knowledge about past experiments, EE can be improved almost instantly for a completely new UE positions pattern. The drawback of the utilization of  $c = 0$  in the case of REASA is that the algorithm tends to permanently exploit sub-optimal solutions. However, these solutions are much better than Ref. The quality of the sub-optimal solution depends on the  $\gamma$  constant. The higher the  $\gamma$ , the smaller the impact of similar REM entries on current action selection. The best sub-optimal solution is observed for  $\gamma = 1.5$ .



**FIGURE 11.** Average reward (EE) obtained with the use of REASA, after each episode, and averaged over 10 experiments. In order to smooth the plot, a moving average of reward over 10 consecutive episodes is applied.



**FIGURE 12.** The impact of the  $c$  constant on the REASA algorithm, for the  $\gamma$  parameter equal to 3.0. In order to smooth the plot, a moving average of the reward over 10 consecutive episodes is applied.

The same level of EE as for REM-UCB can be achieved by REASA by making the algorithm less greedy (increasing  $c$ ), and proper balancing of the impact of similarities on the current action, i.e., tuning the  $\gamma$  parameter. Although  $\gamma = 1.5$  gave the best sub-optimal result, the optimal solution couldn't be achieved for this value of  $\gamma$  by increasing the constant  $c$ . Thus, we decided to slightly reduce the impact of similar REM entries on the current action by setting  $\gamma = 3.0$ . The performance of REASA under  $\gamma = 3.0$ , and varying constant  $c$  is depicted in Fig. 12. It can be seen that constant  $c$  must be properly chosen. Too low value of  $c$  favors a sub-optimal solution because of excessive exploitation, i.e.,  $c = 0$  (green line),  $c = 0.01$  (purple line). On the other hand, too high value of the  $c$  constant results in too frequent exploration of the sub-optimal actions. As a result, the performance is slightly worse than the performance of REM-UCB. The best results were observed for  $c$  equal to 0.05. Most importantly, this value of  $c$  guarantees convergence to a solution as good as the one observed for REM-UCB algorithm. However, due to the utilization of similarities between REM entries, the procedure of learning REM can be reduced from about 120 episodes to only about 60. This shows that the proposed REASA can reduce the required learning phase at least twofold. These

results are promising from the perspective of potential practical implementations of a long time of network operation. In this case, similar states should occur very often. Then, REASA is expected to provide an even higher speed-up of the learning phase in relation to the REM-UCB.

## VI. CONCLUSION

The results have shown that, state of the art optimization of EE, in the case of MMIMO BS, based on analytical formulas, does not provide an optimal solution under realistic conditions. It is because of some simplified assumptions, e.g., that every antenna contributes equally to the resultant array gain. Instead we propose to extend the network with a REM unit, and utilize an RL scheme, to provide mapping between patterns of UE positions and the number of active antennas. The results have shown that this approach can increase EE by 18.5% compared to the reference algorithm. Moreover, by exploiting the similarities between REM entries, the learning phase can be significantly reduced, i.e., when the REASA algorithm is used instead of the REM-UCB method.

In the future more focus can be put on the selection and improvement of the antenna selection algorithm itself. More advanced metrics can be used to assess network and user bitrates, e.g., by taking into the account the correlations between radio channels. Finally studies similar to [30] can be performed to compare the performance of REASA under an alternative distance metric, e.g., the Sum of Minimums.

## ACKNOWLEDGMENT

The simulations were based on the QCM simulator from Huawei Technologies Sweden Research Center.

## REFERENCES

- [1] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 186–195, Feb. 2014.
- [2] E. Björnson, J. Hoydis, and L. Sanguinetti, *Massive MIMO Networks: Spectral, Energy, and Hardware Efficiency*, vol. 11, nos. 3–4. Foundations and Trends in Signal Processing, 2017, pp. 154–655, doi: 10.1561/20000000093.
- [3] K. N. R. S. V. Prasad, E. Hossain, and V. K. Bhargava, "Energy efficiency in massive MIMO-based 5G networks: Opportunities and challenges," *IEEE Wireless Commun.*, vol. 24, no. 3, pp. 86–94, Jun. 2017.
- [4] M. Masoudi et al., "Green mobile networks for 5G and beyond," *IEEE Access*, vol. 7, pp. 107270–107299, 2019.
- [5] S. Buzzi, C.-L. I, T. E. Klein, H. V. Poor, C. Yang, and A. Zappone, "A survey of energy-efficient techniques for 5G networks and challenges ahead," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 4, pp. 697–709, Apr. 2016.
- [6] F. Han, S. Zhao, L. Zhang, and J. Wu, "Survey of strategies for switching off base stations in heterogeneous networks for greener 5G systems," *IEEE Access*, vol. 4, pp. 4959–4973, 2016.
- [7] 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.
- [8] L. Chiaraviglio, P. Wiatr, P. Monti, J. Chen, J. Lorincz, F. Idzikowski, M. Listanti, and L. Wosinska, "Is green networking beneficial in terms of device lifetime?" *IEEE Commun. Mag.*, vol. 53, no. 5, pp. 232–240, May 2015.
- [9] S. Cui, A. J. Goldsmith, and A. Bahai, "Energy-efficiency of MIMO and cooperative MIMO techniques in sensor networks," *IEEE J. Sel. Areas Commun.*, vol. 22, no. 6, pp. 1089–1098, Aug. 2004.

- [10] E. Björnson, M. Bengtsson, and B. Ottersten, "Optimal multiuser transmit beamforming: A difficult problem with a simple solution structure [lecture notes]," *IEEE Signal Process. Mag.*, vol. 31, no. 4, pp. 142–148, Jul. 2014.
- [11] X. Gao, O. Edfors, F. Tufvesson, and E. G. Larsson, "Massive MIMO in real propagation environments: Do all antennas contribute equally?" *IEEE Trans. Commun.*, vol. 63, no. 11, pp. 3917–3928, Nov. 2015.
- [12] C. Jiang and L. J. Cimini, "Antenna selection for energy-efficient MIMO transmission," *IEEE Wireless Commun. Lett.*, vol. 1, no. 6, pp. 577–580, Dec. 2012.
- [13] Z. Chen, J. Yuan, and B. Vucetic, "Analysis of transmit antenna selection/maximal-ratio combining in Rayleigh fading channels," *IEEE Trans. Veh. Technol.*, vol. 54, no. 4, pp. 1312–1321, Jul. 2005.
- [14] E. Kurniawan, A. Madhukumar, and F. Chin, "Antenna selection technique for MIMO-OFDM systems over frequency selective Rayleigh fading channel," *IET Commun.*, vol. 1, pp. 458–463, Jul. 2007.
- [15] T.-H. Tai, H.-J. Chen, W.-H. Chung, and T.-S. Lee, "Energy efficient norm-and-correlation-based antenna selection algorithm in spatially correlated massive multi-user MIMO systems," in *Proc. IEEE Int. Workshop Signal Process. Syst. (SiPS)*, Oct. 2017, pp. 1–5.
- [16] A. F. Molisch, M. Z. Win, Y.-S. Choi, and J. H. Winters, "Capacity of MIMO systems with antenna selection," *IEEE Trans. Wireless Commun.*, vol. 4, no. 4, pp. 1759–1772, Jul. 2005.
- [17] J. C. Marinello, T. Abrão, A. Amiri, E. de Carvalho, and P. Popovski, "Antenna selection for improving energy efficiency in XL-MIMO systems," *IEEE Trans. Veh. Technol.*, vol. 69, no. 11, pp. 13305–13318, Nov. 2020.
- [18] M. O. K. Mendonca, P. S. R. Diniz, T. N. Ferreira, and L. Lovisolio, "Antenna selection in massive MIMO based on greedy algorithms," *IEEE Trans. Wireless Commun.*, vol. 19, no. 3, pp. 1868–1881, Mar. 2020.
- [19] X. Chai, H. Gao, J. Sun, X. Su, T. Lv, and J. Zeng, "Reinforcement learning based antenna selection in user-centric massive MIMO," in *Proc. IEEE 91st Veh. Technol. Conf. (VTC-Spring)*, May 2020, pp. 1–6.
- [20] W. Yu, T. Wang, and S. Wang, "Multi-label learning based antenna selection in massive MIMO systems," *IEEE Trans. Veh. Technol.*, vol. 70, no. 7, pp. 7255–7260, Jul. 2021.
- [21] Z. Zheng, J. Zhang, X. Wu, D. Liu, and L. Tian, "Zero-forcing precoding in the measured massive MIMO downlink: How many antennas are needed?" *Int. J. Antennas Propag.*, vol. 2019, pp. 1–10, Apr. 2019.
- [22] K. Senel, E. Björnson, and E. G. Larsson, "Joint transmit and circuit power minimization in massive MIMO with downlink SINR constraints: When to turn on massive MIMO?" *IEEE Trans. Wireless Commun.*, vol. 18, no. 3, pp. 1834–1846, Mar. 2019.
- [23] J. Hoydis, S. T. Brink, and M. Debbah, "Massive MIMO in the UL/DL of cellular networks: How many antennas do we need?" *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 160–171, Feb. 2013.
- [24] J. Van De Beek, T. Cai, S. Grimoud, B. Sayrac, P. Mahonen, J. Nasreddine, and J. Riihijarvi, "How a layered REM architecture brings cognition to today's mobile networks," *IEEE Wireless Commun.*, vol. 19, no. 4, pp. 17–24, Aug. 2012.
- [25] M. Hoffmann, P. Kryszkiewicz, and A. Kliks, "Increasing energy efficiency of massive-MIMO network via base stations switching using reinforcement learning and radio environment maps," *Comput. Commun.*, vol. 169, pp. 232–242, Mar. 2021. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0140366421000335>
- [26] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*. Cambridge, MA, USA: A Bradford Book, 2018.
- [27] E. Björnson, L. Sanguinetti, J. Hoydis, and M. Debbah, "Optimal design of energy-efficient multi-user MIMO systems: Is massive MIMO the answer?" *IEEE Trans. Wireless Commun.*, vol. 14, no. 6, pp. 3059–3075, Oct. 2015.
- [28] M. Hoffmann, P. Kryszkiewicz, and G. P. Koudouridis, "Modeling of real time kinematics localization error for use in 5G networks," *EURASIP J. Wireless Commun. Netw.*, vol. 2020, no. 1, pp. 1–19, Dec. 2020.
- [29] H. Viswanathan and P. E. Mogensen, "Communications in the 6G era," *IEEE Access*, vol. 8, pp. 57063–57074, 2020.
- [30] M. Hoffmann and P. Kryszkiewicz, "Similarity measures for location-dependent MMIMO, 5G base stations on/off switching using radio environment map," 2021, *arXiv:2105.01948*. [Online]. Available: <http://arxiv.org/abs/2105.01948>
- [31] H. Holtkamp, G. Auer, V. Giannini, and H. Haas, "A parameterized base station power model," *IEEE Commun. Lett.*, vol. 17, no. 11, pp. 2033–2035, Nov. 2013.
- [32] A. Capone, S. D'Elia, I. Filippini, A. E. C. Redondi, and M. Zangani, "Modeling energy consumption of mobile radio networks: An operator perspective," *IEEE Wireless Commun.*, vol. 24, no. 4, pp. 120–126, Aug. 2017.
- [33] S. Tombaz, K. W. Sung, and J. Zander, "On metrics and models for energy-efficient design of wireless access networks," *IEEE Wireless Commun. Lett.*, vol. 3, no. 6, pp. 649–652, Dec. 2014.
- [34] S. Mukherjee and S. K. Mohammed, "On the energy-spectral efficiency trade-off of the MRC receiver in massive MIMO systems with transceiver power consumption," 2014, *arXiv:1404.3010*. [Online]. Available: <http://arxiv.org/abs/1404.3010>
- [35] U. Amin, S. Anwar, W. Khan, and S. I. Khan, "Energy efficiency analysis of single-cell massive MIMO under pilot contamination," in *Proc. 13th Int. Bhurban Conf. Appl. Sci. Technol. (IBCAST)*, Jan. 2016, pp. 704–709.
- [36] S. Payami and F. Tufvesson, "Channel measurements and analysis for very large array systems at 2.6 GHz," in *Proc. 6th Eur. Conf. Antennas Propag. (EUCAP)*, Mar. 2012, pp. 433–437.
- [37] A. D. Tijmsma, M. M. Drugan, and M. A. Wiering, "Comparing exploration strategies for Q-learning in random stochastic mazes," in *Proc. IEEE Symp. Ser. Comput. Intell. (SSCI)*, Dec. 2016, pp. 1–8.
- [38] A. O. Martínez, J. Ø. Nielsen, E. De Carvalho, and P. Popovski, "An experimental study of massive MIMO properties in 5G scenarios," *IEEE Trans. Antennas Propag.*, vol. 66, no. 12, pp. 7206–7215, Dec. 2018.
- [39] D. P. Huttenlocher, G. A. Klanderma, and W. J. Rucklidge, "Comparing images using the Hausdorff distance," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 15, no. 9, pp. 850–863, Sep. 1993.
- [40] *Deliverable D6.1, Simulation Guidelines, Version: V1.0 Ed. Mobile and Wireless Communications Enablers for the Twenty-Two Information, METIS*, Sweden, 2013.
- [41] *Base Station (BS) Radio Transmission and Reception*, document TR 38.104, Version 15.9.0, 3GPP, Apr. 2020.
- [42] T. Yoo and A. Goldsmith, "On the optimality of multi-antenna broadcast scheduling using zero-forcing beamforming," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 3, pp. 528–541, Mar. 2006.
- [43] Qamcom. *Qamcom/QCM*. Accessed: Mar. 10, 2021. [Online]. Available: <https://github.com/qamcom/QCM>
- [44] A. S. Pagès, J. Olmos, and M. Lema, "Modelling channel estimation error in LTE link level simulations," *Eur. Cooperat. Sci. Technol., COST IC1004 Cooperat. Radio Commun. Green Smart Environ., 3rd Sci. Meeting*, Barcelona, Spain, Feb. 2012.



of network management.

**MARCIN HOFFMANN** (Graduate Student Member, IEEE) received the M.Sc. degree (Hons.) in electronics and telecommunication from Poznań University of Technology, in 2019, where he is currently pursuing the Ph.D. degree with the Institute of Radiocommunications. He is gaining scientific experience by being involved in both national and international research projects. His research interests include utilization of machine learning and location-dependent information for the purpose



**PAWEŁ KRYSZKIEWICZ** (Senior Member, IEEE) received the M.Sc. and Ph.D. degrees (Hons.) in telecommunications from Poznań University of Technology (PUT), Poland, in 2010 and 2015, respectively. He is currently an Assistant Professor with the Institute of Radiocommunications, PUT. He was involved in a number of national and international research projects. His research interests include the problems concerning the physical layer of the dynamic spectrum access systems, multi-carrier signal design for green communications, and problems related to practical implementation of massive MIMO systems.

...