Hindawi
Wireless Communications and Mobile Computing
Volume 2023, Article ID 8559701, 8 pages
https://doi.org/10.1155/2023/8559701

Research Article

Over-the-Air Computation with Quantized CSI and Discrete Power Control Levels

Christos Tsinos,[1] Sotirios Spantideas,[2] Anastasios Giannopoulos,[2] and Panagiotis Trakadas[2]

[1]*Department of Digital Industry Technologies, National and Kapodistrian University of Athens, Evia 34400, Greece*
[2]*Department of Ports Management and Shipping, National and Kapodistrian University of Athens, Evia 34400, Greece*

Correspondence should be addressed to Anastasios Giannopoulos; angianno_8@hotmail.com

Received 20 September 2022; Revised 10 May 2023; Accepted 28 October 2023; Published 13 November 2023

Academic Editor: Ding Xu

In this paper, an Over-the-Air Computation (AirComp) scheme for fast data aggregation is considered. Multisource data are simultaneously transmitted by single-antenna mobile devices to a single-antenna fusion center (FC) through a wireless multiple-access channel. The optimal power levels at the devices and a postprocessing scaling function at the FC are jointly derived such that mean square error of the computation is minimized. Different than the existing approaches that rely on perfect channel state information (CSI) at the FC and assume that the devices' optimal power levels can be selected from an infinite solution set, in the present paper, it is assumed that only quantized CSI is available at the FC and that the aforementioned optimal power levels lie in a finite discrete set of solutions. To derive the optimal power levels and FC's scaling factor, a difficult nonconvex constrained optimization problem is formulated. An efficient and robust solution to quantization errors is developed via the deep reinforcement learning framework. Numerical results verify the good performance of the proposed approach while it exhibits a significant reduction in the required feedback.

1. Introduction

The sixth generation (6G) of wireless communications is foreseen to accommodate a huge number of mobile devices within the context of the so-called internet-of-things (IoT) for enabling novel and demanding applications such as smart cities, interconnected autonomous vehicles, and so forth [1]. These devices require the aggregation of massive data distributed to them, to support their functions. To that end, a promising technology called Over-the-Air Computation (AirComp) has recently emerged for fast wireless data aggregation [2, 3].

AirComp exploits the signal superposition property of the multiple-access channel (MAC) between the devices and a fusion center (FC) for averaging their simultaneously transmitted data over the wireless medium. By properly applying processing at both the devices and the FC ends, AirComp can be used to also calculate different data functions from their average that belong to the class of the so-called nomographic functions, for example, geometric mean and polynomial expressions. Recent works in the field of AirComp have expanded the original ideas in Nazer and Gastpar [2] and Soundararajan and Vishwanath [3] under different system models [4–7].

To deal with the fading characteristics of the wireless medium, the work in Cao et al. [8] and later works in the federated learning domain [9, 10] presented optimized power control schemes for AirComp systems by minimizing the computation error at the FC. This presented significant performance gains since it avoided the suboptimal approach of channel inversion power control, used in the previous works [4–7].

On the other side, the approaches in Cao et al. [8–10] they require perfect channel state information (CSI) at the FC side. This requirement can be very restrictive, especially when the CSI is typically estimated at the devices from the downlink training symbols and then has to be fed back to the FC. If highly accurate CSI is fed back to the FC, the required overhead could be extremely high. In the literature of

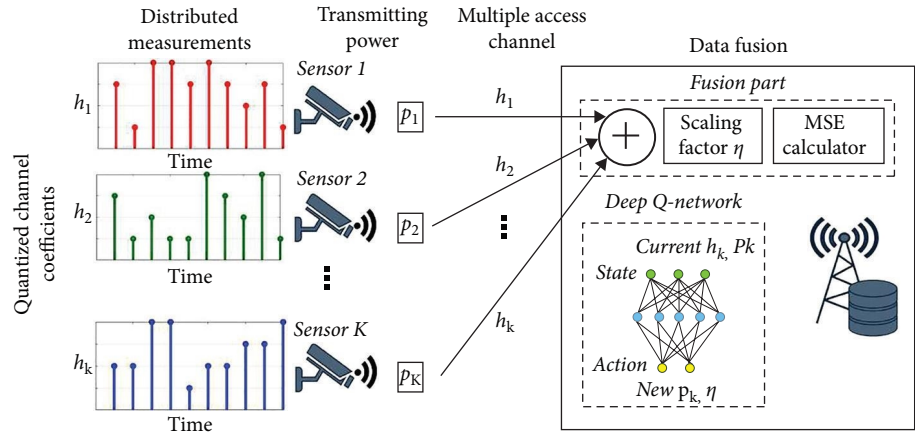

FIGURE 1: System model. The IoT devices are measuring time-varying parameters that are transmitted with device-specific power p_k over channel h_k. The FC aggregates the received signals and applies a denoising factor η.

communication systems, feedback mechanisms have been developed based on quantized CSI (QCSI) [11, 12]. In such approaches, the estimated CSI is quantized to one of the known states and then only the index of the detected state is fed back, thus reducing the required overhead. The latter comes at the cost of inferior performance, if no robust methods to quantization errors are used. AirComp schemes with QCSI have yet to be developed and this is the first objective of this paper.

Furthermore in Cao et al. [8–10], the optimal power levels at the devices are selected from an infinite space of values. Given that these devices are mostly of low hardware complexity/capabilities, for example, sensors, it is highly probable that they can support only a limited finite number of power levels. This perplexes the derivation of the optimal setup for the considered AirComp regime since it requires the solution to difficult discrete optimization problems. Moreover, it is very challenging to achieve satisfactory performance, due to the reduced solution set compared to the original approaches. Contrariwise, such an approach results in reduced overhead for feeding back the optimal power levels to the mobile devices. Such solutions are not yet available for AirComp systems and their development is the second objective of the present paper. Analytically, the contributions of this work are as follows.

An IoT network applying AirComp over a fading MAC is assumed based on a single antenna FC and single antenna devices. The devices simultaneously transmit their sensing data to the FC to calculate their average value. The objective is to determine the optimal transmit power levels for the devices and the postprocessing scaling factor applied at the FC given that the FC has only QCSI knowledge and the devices can set their transmit power levels from a finite discrete set. The optimal transmit power levels and the FC center's scaling factor are jointly derived such that the mean square error (MSE) of the computation is minimized. To that end, first a difficult nonconvex constrained optimization problem is defined with the view to minimize the MSE given the QCSI knowledge and the discrete set of transmit power levels. Then, a solution to the defined problem is developed based on the deep reinforcement learning (DRL) framework [13]. The DRL-based solution is able to exploit the computational power of the deep neural network (NN) is order to efficiently solve the difficult optimization problem while being robust to the CSI imperfections due to the quantization. Overall, this paper aims to show the efficacy of the DRL in providing better suboptimal power level suggestions compared with the typical scheme (i.e., direct inference of the optimal AirComp power control policy with the QCSI values as input) in terms of MSE. The comparisons are conducted with realistic system assumptions, including coarse CSI knowledge and discrete power levels. Numerical results show that the performance of the proposed approach is very satisfactory under coarse QCSI knowledge when compared to the perfect CSI approaches and their extensions to the QCSI case.

The rest of the paper is organized as follows: Section 2 describes the considered system model and formulates the problem to be solved. Section 3 derives the DRL-based algorithmic solution to the defined optimization problem. Section 4 presents the numerical results and Section 5 concludes this work.

2. System Model and Problem Formulation

An AirComp scheme is considered over a MAC based on which K single-antenna mobile devices are sending information to a single-antenna FC (Figure 1). Let us assume that the devices are measuring a set of time-varying parameters of the environment they are deployed to. The FC aggregates the received information with the view to calculate the average of the measured data from the mobile devices. That is, in timeslot t, the FC calculates the function

$$f(t) = (1/K)\sum_{k=1}^{K} s_k(t), \tag{1}$$

where $s_k = g(d_k(t))$, $d_k(t)$ is the measurement at node k, $1 \leq k \leq K$, at timeslot t and $g(\cdot)$ is a scaling function applied for power control. The function $g(\cdot)$ is actually a linear and

uniform to all devices normalization operator such that the variables $\{s_k(t)\}$ are of zero mean and unit variance. Then, the FC is able to recover the desired mean value by simply applying a de-normalization operation in Equation (1), that is,

$$\widetilde{f}(t) = g^{-1}(f(t)), \tag{2}$$

where $g^{-1}(\cdot)$ is the inverse function of $g(\cdot)$. The received signal at FC, at timeslot t, is given by

$$y(t) = \sum_{k=1}^{K} h_k(t) w_k(t) s_k(t) + z(t), \tag{3}$$

where $w_k(t) = \frac{\sqrt{p_k(t)} h_k^*(t)}{|h_k(t)|}$, $p_k(t) \in \mathbb{R}_+$ is the transmit power of device k, $1 \le k \le K$, $(\cdot)^*$, $h_k(t)$ is the channel coefficient of channel k (also called perfect CSI) and $|\cdot|$ denote the conjugate and the absolute value of a complex number, respectively and $z(t)$ is a complex additive white Gaussian noise variable at the FC of zero mean and variance σ^2. Upon receiving $y(t)$, the FC applies a denoising factor $\eta(t)$ for recovering the average measurement by the devices. Thus, the signal at the FC after postprocessing is given by,

$$\widehat{f}(t) = \sum_{k=1}^{K} \frac{\sqrt{p_k(t)}}{K\sqrt{\eta(t)}} |h_k(t)| s_k(t) + \frac{z(t)}{K\sqrt{\eta(t)}}, \tag{4}$$

where the scaling factor K is introduced for averaging. As it is evident, the values of the power allocation $p_k(t)$, $1 \le k \le K$ and denoising $\eta(t)$ variables have to be determined in order to apply the AirComp scheme. A common approach for deriving the values of the required variables is by the minimization of the MSE between the calculated average of the transmitted data $\widehat{f}(t)$ in Equation (4) and the actual one $f(t)$. Under the assumption of statistically independent observations $\{s_k(t)\}$ among the users, the instantaneous MSE can be shown to be given by,

$$\begin{aligned} \mathrm{MSE}(t) &= \mathbb{E}\left\{ \left(\widehat{f}(t) - f(t) \right)^2 \right\} \\ &= \frac{1}{K^2} \left[\sum_{k=1}^{K} \left(\frac{\sqrt{p_k(t)} |h_k(t)|}{\sqrt{\eta(t)}} - 1 \right)^2 + \frac{\sigma^2}{\eta(t)} \right], \end{aligned} \tag{5}$$

where $\mathbb{E}\{\cdot\}$ is the expectation operator.

By dropping the time index t for simplicity and based on Equation (5), the values of the power allocation p_k, $1 \le k \le K$ and denoising η variables can be derived as the solution to the following minimization problem:

$$\begin{aligned} (P_1): &\min_{p_k \ge 0, \eta > 0} \sum_{k=1}^{K} \left(\frac{\sqrt{p_k} |h_k|}{\sqrt{\eta}} - 1 \right)^2 + \frac{\sigma^2}{\eta} \\ &\text{s.t. } p_k \le \bar{P}_k, \forall k \in K, \end{aligned} \tag{6}$$

where $\bar{P}_k$ is the power level of each sensor.

This problem is nonconvex since the sensor power vector $\{p_k\}$ and the denoising factor η are coupled in the objective function. On to top of this, problem (P_1) requires CSI knowledge at the transmitter's side. The required CSI is estimated at each device via training symbols transmitted from the FC at the downlink. Thus, by exploiting the uplink–downlink channel reciprocity, the devices estimate the required CSI and then, they feed it back to the FC for solving P_1. If perfect (or highly accurate to be more practical) CSI is assumed at the FC, the feedback phase results in huge communication overhead. To that end, in this paper, we assume a QCSI feedback estimation scheme based on a predetermined QCSI codebook.

Let us now assume that the FC and the devices have knowledge of this predetermined QCSI codebook. Based on the estimated CSI via the previously described procedure, each device locates the closest representative entry in the codebook and feeds back to the FC only the index associated with the detected QCSI state requiring reduced communication overhead.

By straightforwardly applying the closed form of the solution in Cao et al. [8] for P_1 under QCSI information, the performance exhibits severe degradation, especially for very coarse CSI quantization. Moreover, the situation is further perplexed if it is assumed that the devices can set their power levels through a discrete and finite codebook. This results to a feasible solution set for P_1 that is discrete and thus, in a very difficult optimization problem with no known efficient solution that has, in general, exponential complexity for its solution. The defined problem to be addressed under the QCSI and finite power levels set is defined as:

$$\begin{aligned} (P_2): &\min_{p_k \ge 0, \eta > 0} \sum_{k=1}^{K} \left(\frac{\sqrt{p_k} \left| \widehat{h}_k \right|}{\sqrt{\eta}} - 1 \right)^2 + \frac{\sigma^2}{\eta} \\ &\text{s.t. } p_k \in \mathscr{P}_k \, \forall k \in K, \end{aligned} \tag{7}$$

where $\mathscr{P}_k = \{P_{k,1}, \ldots, P_{k,M}\}$ is the set of discrete power levels for the kth device, $\widehat{h}_k$ is the QCSI feedback of channel k, M is the number of power levels assumed to be the same for all the devices without generality loss and $P_{k,m} \in [0, \bar{P}_k]$, for $1 \le m \le M$.

In the following, a solution will be developed for solving P_2 based on the DRL framework that effectively deals with the QCSI errors and the discrete levels of the devices' power.

3. DRL-Based Solution

In this section, the solution to P_2 is derived via a deep Q-learning (DQL) method. DQL is a DRL method that utilizes a NN as a quality function estimator (Q-value) [14]. In principle, the deep Q-network (DQN) agent observes the wireless environment in the form of a state $s \in S$ and performs an action $a \in A$, where S and A correspond to the state and action spaces, respectively [15]. Then, depending on the quality of the performed action, the agent receives a reward r. The DQL method involves the Bellman equation, given by

$$Q_t(s_t, a_t) = (1-\alpha)Q_{t-1}(s_t, a_t) + \alpha(r(s_t, a_t) + \gamma \max_{a'} \{Q(s_{t+1}, a')\}, \quad (8)$$

where s_t is the state of the environment at time t and a_t is the action performed by the agent. The hyperparameters $\alpha \in [0, 1]$ and $\gamma \in [0, 1]$ correspond to the learning rate and discount factor and are used as a trade-off between previous Q-values $(Q_{t-1}(s_t, a_t))$, immediate rewards $(r(s_t, a_t))$ and the optimal future rewards $\gamma \max_{a'} \{Q(s_{t+1}, a')\}$.

The Bellman equation in practice quantifies the quality of being in state s_t and performing the action a_t, designating the learning strategy framework. The DQN agent interacts with the environment in a trial-and-error process, ideally performing all possible actions A from all possible states S. Therefore, the agent gains experience regarding the favorable and disadvantageous actions from any current state through the reward function during the training phase of the DQL algorithm [16]. Regarding the deep learning context, two identical (in dimensions) NNs are involved: (i) the Q-network which is used to estimate the current best action (considered to contain the input features) and (ii) the target Q-network which is used to estimate the next action (or action policy) that will return the maximum long-term reward (considered to contain the output labels). Once the training phase of the DQL algorithm has been finalized and the hyperparameters γ and α of the Bellman equation have been stabilized, the pretrained agent may be utilized for inference purposes in order to determine the action selection policy.

A solution for P_2, is derived based on a DQN agent, located at the FC which interacts with the wireless environment. The design parameters of the DQN agent may be described:

State space: The state space describes the wireless environment from the communications' perspective. In the proposed solution, the state space includes the combined QCSI and power information of each wireless sensor. At a given time t, the system space can be expressed as $s_t = [s_1, s_2, \ldots, s_K]$ with s_k being related to the power $p_k \in \mathcal{P}_k$ and channel coefficient h_k of sensor k. The value of the sensor k QCSI is represented by $\widehat{h}_k = W(|h_k|^2)$, where $W(\cdot)$ is a quantization function that depends on the number of quantization bits J. The set of QCSI values is also defined as $\widehat{H} = \{0, 1, \ldots, 2^J - 1\}$. The state values of sensor k can be then derived by $s_k \in \mathcal{P}_k \times \widehat{H}, \forall k$ (all possible states are 2^{JK}). In this context, the DQN agent at the FC receives the combined information related to the QCSI and power values for all sensors (Figure 1).

Action space: Upon observing the system state, the DQN agent selects an action during a specific training episode. Specifically, at a given time step t, a discretized power level is selected by the agent and assigned to each sensor $p_{k,m}$, along with a discretized denoising factor η. Formally, the DQN agent action is described as $a_t = [(p_{1,m}, p_{2,m} \ldots, p_{K,m}, \eta_\phi)]$ and its dimensionality is $K + 1$. As aforementioned, the power values that are assigned to the wireless sensors depend lie in $\mathcal{P}_k\}$. Similarly, the values selected by the agent for the denoising factor η from Φ levels lie in set $\mathcal{H} = \{0, \eta_1, \ldots, \eta_\Phi\}$ (all possible actions are $M^K \times \Phi$). The selected action is then implemented on the wireless environment and the state space is updated at the next time step, since it encompasses the updated power vector of all wireless sensors.

Require: $K, M, \Phi, J, \tilde{P}_k$
Ensure: $\arg\min_{p_k \geq 0, \eta > 0} F$
  Initialize a, γ, and $\epsilon = 1$
  Initialize a Replay Memory D
  Initialize action-value function Q with random weights θ
  Load 2^J-level Channel Quantizer $W(\cdot)$
  for *episode* $\leftarrow 1, \ldots, T$ **do**
    Draw Channel Coefs h_k
    Assign Power Levels p_k randomly
    $S_0 = W\left(p_k \left|\widehat{h}_k\right|^2\right), \forall k$ ▷ Initial State Quantization
    while $r_t > 0$ **do**
      Select a random number *Choice* $\in [0, 1]$
      if *Choice* $> \epsilon$ **then** ▷ Exploration
        Select a random action a_t
      else ▷ Exploitation
        $a_t \leftarrow \arg\max_a Q^*(S_t, a|\theta)$
      end if
      Take action a_t, Observe reward r_t and state S_{t+1}
      Store transition (S_t, a_t, r_t, S_{t+1}) in D ▷ Experience Replay
      Select random minibatch of transitions (S_t, a_t, r_t, S_{t+1}) from D
      Set $y_j = \begin{cases} r_j, & \text{if } S_{j+1} terminal \\ r_j + \gamma \max_{a'} Q(S_{j+1}, a'), & \text{otherwise} \end{cases}$
      Perform Gradient Descent on $(y_j - Q(S_j, a_j|\theta))^2$
      $S_t \leftarrow S_{t+1}$
    end while
    $\epsilon \leftarrow \epsilon \times \frac{T - 2episode}{T}$ ▷ ϵ-greedy decaying
  end for

Algorithm 1: AirComp-DRL Training for MSE Minimization.

Reward function: The performed action a_t from a state s_t results in a new state s_{t+1} and a positive or zero reward, depending on whether this action was beneficial toward the optimization goal. At a given time t the reward function is defined as:

$$r_t(s_{t-1}, a_{t-1}) = \begin{cases} F_{t-1} - F_t, & \text{If } F_t < F_{t-1} \\ 0, & \text{Otherwise} \end{cases}, \quad (9)$$

where the objective function F can be expressed by:

$$F = \sum_{k \in K} \left(\frac{\sqrt{p_{k,m}} |h_k|}{\sqrt{\eta}} - 1 \right)^2 + \frac{\sigma^2}{\eta}. \quad (10)$$

Evidently, the reward function leads the DRL agent during the training process to gradually favor a sequence of

actions that minimize the F function and thus, also the MSE in Equation (5).

Note that the training procedure is based on DRL principles with experience replay [14]. The complete procedure for training the AirComp-DRL model is summarized in Algorithm 1. Following the initialization of the learning hyperparameters (α and γ), a replay memory D (filled with experience/transition tuples corresponding to random actions), the two Q-function approximators (Q- and target Q-neural networks with random weights) and the 2^J-level channel quantizer $W(\cdot)$ are also initialized. In each training episode, the wireless environment is initialized before the agent begins to perform actions by randomly selecting the power levels p_k to the sensors and the η value which constitute the initial system state S_0. Depending on the phase of the training process, an action a_t is selected, that is, the power levels of the sensors are randomly selected in the exploration mode, whereas the power vector is estimated by the Q-network during the exploitation mode of the algorithm. The performed action leads the environment in a new system state S_{t+1} and a reward r_t is returned to the DRL agent, according to the objective of the reward function. The transition tuple (S_t, a_t, r_t, S_{t+1}) is stored in the replay memory D, while a minibatch of experience tuples is randomly selected from D and, based on the Bellman equation, the Q-network is used to estimate the quality of immediate and future actions (in case that S_{j+1} is not a terminal system state). Thereafter, the gradient descent method is utilized to update the weights of the Q-network neurons (backpropagation), using the target Q-network estimations as output labels (every N_c steps of the algorithm, the weights of the Q-network are inherited to the target Q-network neurons). Finally, to gradually transit from exploration to exploitation, an ϵ-greedy method with linear decaying is adopted. Noteworthy, the DRL training efficiency is highly influenced by the degree of exploration completeness, which in turn depends on the state/action space dimensionality. A sufficient number of training episodes T should ensure that the agent visits as many as possible state/action pairs.

Regarding the inference procedure of a pretrained model, the DRL agent performs actions only in exploitation mode ($\epsilon = 0$), while the storing of experience tuples in the replay memory and the backpropagation processes are simply omitted.

4. Simulation Results

In this section, numerical results are demonstrated both for the DRL training phase and MSE comparison between different schemes. The presented simulations were conducted in Python 3.8, whereas the libraries TensorFlow (version 2.3), Keras, and Scikit-Learn were used for constructing and training the AI/ML models. Coding scripts ran on a personal PC (CPU i7-8700; 3.2 GHz; RAM 8 GB; no GPU usage).

4.1. DRL Training. A QCSI-based AirComp system with 20 sensors is considered during the DRL hyperparameter stabilization (see Figure 2). In addition, a time-varying channel

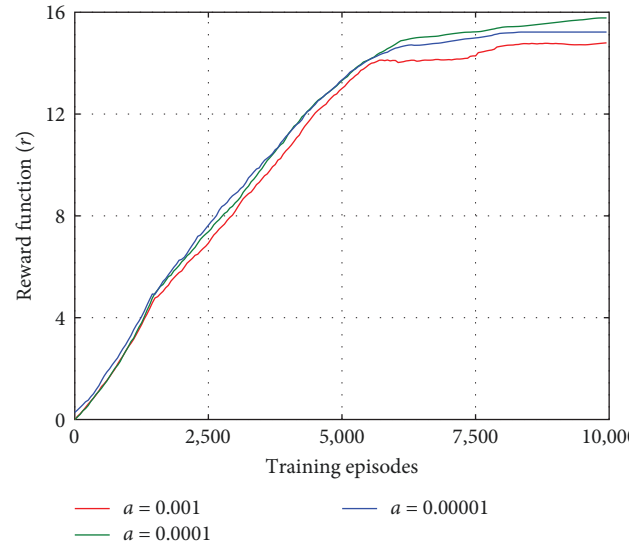

FIGURE 2: Learning curves for different values of learning rate α as a function of the training episodes.

model, composed by a dominant pathloss component and a Rayleigh fading component with variance $\sigma_c^2 = 0.1$ is adopted for the rest of the simulations to represent dynamic channel conditions. The channel quantizer $W(\cdot)$ is implemented via a k-means clustering algorithm trained over 10,000 channel samples (where k represents the quantization levels $2^J = 4$). In this sense, time-varying QCSI values are represented by the k-means centroids based on a minimum Euclidean distance criterion. Without loss of generality, it is assumed that the number of power and discrete FC scaling factor levels involved in the DRL solution are $M = \Phi = 10$. The values for sets $\mathscr{P}_k$, $1 \leq k \leq K$, and $\mathscr{H}$ are derived by uniformly discretizing the continuous sets $(0, P_{\max}]$ and $(0, \eta_{\max}]$, respectively, where $P_{\max} = 1$W and $\eta_{\max} = 1$. The noise variation at the receiver (FC) is set to $\sigma^2 = 0.01$. Upon testing multiple NN setups, we concluded to a NN with three fully connected hidden layers with sizes $3\times$, $2\times$, $1\times(MK + \Phi)$, while the update frequency of the Q-target network is set to $N_c = 100$ steps. The activation function of all neurons included in the hidden layers was the rectified linear (ReLU) one, whereas the neurons of the output layer employed the linear activation one.

Notably, the DRL reward convergence defines the extent to which the resulted policy can significantly optimize the objective function. Initially, two of the most critical hyperparameters involved in the DRL training process, namely the learning rate α (monitors the update ratio between new and previous Q-values) and discount factor γ (balances the degree of which immediate or future-expected rewards are preferred), were fine-tuned to ensure optimal reward convergence. To that end, hyperparameter stabilization was obtained by inspecting the training/learning curve for varying values α (see Figure 2) and γ (see Figure 3). As shown in Figures 2 and 3, the reward time course gradually transits from the exploration to the exploitation stage, reaching the highest values for $\alpha = 0.0001$ and $\gamma = 0.9$ (reward function

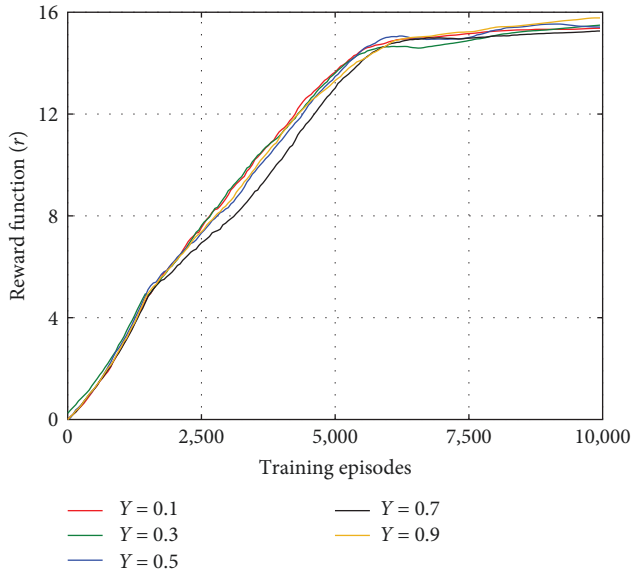

Figure 3: Learning curves for different values of discount factor γ as a function of the training episodes.

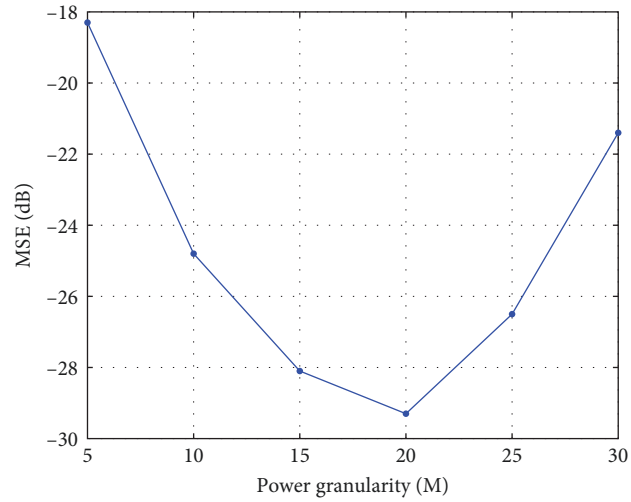

Figure 4: MSE performance derived by DRL as a function of the number of power levels M.

was relatively insensitive to γ parameters). Both parameters were set to their optimal values for the rest of the simulations.

4.2. Impact of Power Granularity. This section includes further simulations related to the impact of the power granularity (number of available power levels M that can be selected by the DRL agent). In general, every MSE minimization model in AirComp systems presents a total estimation error (in MSE solutions) which is usual to the sum of four individual error terms: (i) channel quantization error (introduced by the bit-based representation of CSI), (ii) power discretization error (derived by the realistic and discrete power level configuration of the sensors), (iii) eta discretization error (scaling factor of data fusion takes practically discrete values), and (iv) model fitting error (resulted by the model itself in attempting to ensure a good trade-off between over- and under-fitting, also called generalization error). Power granularity comprises a crucial parameter for the DRL performance and MSE optimization, since it defines the extent to which the agent can precisely tune the transmitting power of the sensors for a given power range. The target is to investigate whether the increasing M actually improves the DRL performance, given stable CSI quantization (here 4-level quantization), number of sensors ($K = 30$) and power range (here 0.1–1 W). Variations in the CSI quantization and/or power range do not result in loss of generality of the conclusions.

In specific, the higher the number of power levels for a specific available power range (e.g., from 0.1 to 1 W), the better the training performance. This is attributed to the fact that the power granularity is increased with increasing M, therefore making the action information (i.e., the outputs of the DRL agent) to better approximate the perfect (selected by the optimal solution) power level. Ideally, one could expect that when the number of available power levels is infinite (i.e., $M \to \infty$), then MSE of DRL better approaches the MSE of the optimal solution, given that the power level suggestions are almost continuous values, and not discretized levels. Noteworthy, the optimal solution outcome depends drastically on the inputs of the DRL agents, which are the QCSI values. Intuitively, when both perfect CSI and continuous power levels are considered, then the DRL solution is closer to the optimal one. However, in realistic conditions, as the power granularity increases, the DRL becomes more demanding in the DRL network dimensionality, given that the number of the output layer neurons is increased with power levels. Thus, there is an upper bound of the power granularity, above which MSE starts to degrade due to the concurrent increment of model dimensionality and complexity. The higher the dimensionality of the output layer, the more demanding the training phase due to the fact that the number of available actions is increased (i.e., more Q-values have to be estimated in the output neurons of the DRL agent).

As shown in Figure 4, the MSE performance follows a U-shaped form as a function of the number of power levels M. This means that, for a given power range, number of sensors, and quantization level, MSE is improved (i.e., lower MSE values) until M reaches a threshold (here critical $M = 20$). Beyond this critical value of M, MSE performance starts to degrade (i.e., higher MSE values) because of the higher complexity and dimensionality of the DRL model. Specifically, complexity is proportional to the NN density, which is also increased with the number of available power levels. We also note that, as the number of available actions that can be selected by the DRL agent is increased, NN dimensionality should be increased to accurately estimate the large number of available power configurations. This U-shaped function of MSE versus M implies that the DRL performance in AirComp MSE minimization comes with a power granularity limitation, with the latter requiring no more than $M = 20$ power levels. In conclusion, power granularity has to be thoroughly selected in MSE minimization problems, so as

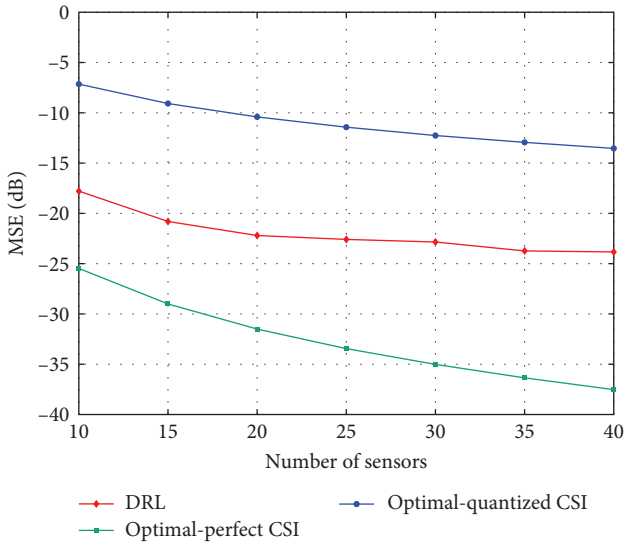

Figure 5: Comparison of MSE as a function of the number of sensors between the DRL (red), and optimal schemes with perfect (green) and quantized (blue) CSI. Two-level quantization is considered.

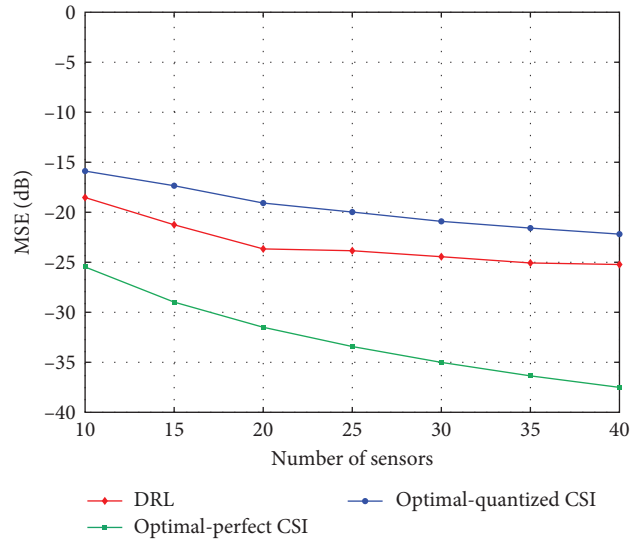

Figure 6: Comparison of MSE as a function of the number of sensors between the DRL (red), and optimal schemes with perfect (green) and quantized (blue) CSI. Four-level quantization is considered.

to ensure the minima of the U-shaped function between MSE and M.

4.3. Comparative Results. For comparison purposes, the MSE at the FC is computed for the proposed approach and compared to the one of two baseline schemes: (i) the optimal power allocation and FC denoising factor scheme using perfect CSI knowledge in Cao et al. [8] and (ii) the optimal power and η allocation strategy computed in Cao et al. [8] under QCSI knowledge.

The MSE can be calculated as F/K^2 for the three aforementioned schemes, taking into account the sensors' assigned power levels and the assigned denoising factor that emerge from each allocation framework. Toward this direction, a comparison of the MSE computed at the FC receiver for the three solutions is shown in Figure 5 for $J = 1$ and varying number of wireless sensors. It is observed that, the DRL-assisted solution reaches lower MSE values (∼10 dB) compared to the optimal solution under coarse QCSI knowledge, regardless of the number of sensors that participate in the AirComp system.

Similar results can be observed in Figure 6, where the MSE value at the FC is compared amongst the three solutions for $J = 2$ with respect to the number of IoT sensors. Notably, the MSE gap between the optimal solution with perfect CSI and with QCSI knowledge is reduced due to the increased number of quantization levels. Nevertheless, the power vector and denoising factor allocation strategy provided by the DRL solution accomplish decreased MSE values (∼3–4 dB) in contrast to the optimal QCSI solution.

As indicated from the results, the potency of the deployed DRL framework on an AirComp system becomes apparent in cases that quantization error significantly degrades the AirComp MSE performance. To this end, the communication overhead between the IoT sensors and the FC can be effectively reduced (low number of transmitted quantization bits), without significant degradation of the MSE value and the overall performance of the AirComp system.

4.4. DRL Versus Optimal Solution under Coarse CSI Knowledge. The main drawback of the optimal solution is that it requires perfect CSI knowledge, as well as it assumes perfect precision in the power configuration. In realistic conditions, the perfect CSI is not known and the precise power level proposition requires a high number of bits to be transmitted. The assumption of discrete power levels is adopted in this work primarily to reduce the information required to be exchanged by low-capacity and low-memory devices. Thus, here we aimed to demonstrate the efficacy of applying DRL under coarse QCSI knowledge.

Assuming that the optimal solution is inferred with the QCSI values as input, the resulting solution deviates from the perfect power allocation and MSE minimization, primarily due to the quantization error introduced by the QCSI inputs. These quantization errors cannot be corrected by the optimal solution itself, since the latter is basically a closed-form equation requiring only the perfect CSI values as inputs. The reason for which DRL outperforms the optimal solution inferred with QCSI as input may be attributed to the rewarding function that is used to train the agent. Specifically, the rewards received by the agent (only during the training) take into account the perfect CSI to better estimate the Q-values of all power actions. Note that, the training is performed offline with simulated data, whereas during the inference phase, the agent is directly compared with the optimal QCSI method using only the QCSI. The inference output (i.e., power suggestions) is used to calculate the resulting MSE, which was proven to outperform the optimal with QCSI method.

5. Conclusion

In this work, a wireless sensor AirComp system with QCSI feedback and discrete power control levels is studied. To mitigate the quantization error introduced in the AirComp's MSE calculation, a DRL-assisted framework for power level and denoising factor selection is thoroughly described and implemented, jointly exploiting quantized channel and power information. The proposed DRL model is compared against the analytical (optimal) MSE optimization solution, assuming both perfect and QCSI knowledge and power level configuration. Numerical results confirm the potency of the centrally placed DRL agent in reducing the performance gap between the optimal solution with versus without ideal CSI. Overall, this study demonstrates the dominance of the DRL-assisted solution under coarse QCSI conditions, highlighting the effective communication overhead reduction without considerable degradation of the AirComp system's performance.

Data Availability

The simulation data used to support the findings of this study can be made available upon request to the corresponding author.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work has been partially supported by the European project entitled "Trustworthy and Resilient Decentralized Intelligence for Edge Systems (TaRDIS)" funded by EU HORIZON, under grant agreement no 101093006.

References

[1] M. Gupta, R. K. Jha, and S. Jain, "Tactile based intelligence touch technology in IoT configured WCN in B5G/6G-a survey," *IEEE Access*, vol. 11, pp. 30639–30689, 2022.

[2] B. Nazer and M. Gastpar, "Computation over multiple-access channels," *IEEE Transactions on Information Theory*, vol. 53, no. 10, pp. 3498–3516, 2007.

[3] R. Soundararajan and S. Vishwanath, "Communicating linear functions of correlated gaussian sources over a MAC," *IEEE Transactions on Information Theory*, vol. 58, no. 3, pp. 1853–1860, 2012.

[4] G. Zhu and K. Huang, "MIMO over-the-air computation for high-mobility multimodal sensing," *IEEE Internet of Things Journal*, vol. 6, no. 4, pp. 6089–6103, 2019.

[5] X. Li, G. Zhu, Y. Gong, and K. Huang, "Wirelessly powered data aggregation for IoT via over-the-air function computation: beamforming and power control," *IEEE Transactions on Wireless Communications*, vol. 18, no. 7, pp. 3437–3452, 2019.

[6] D. Yu, S. Park, O. Simeone, and S. S. Shitz, "Optimizing over-the-air computation in IRS-aided C-RAN systems," in *2020 IEEE 21st International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, pp. 1–5, IEEE, 2020.

[7] G. Zhu, Y. Wang, and K. Huang, "Broadband analog aggregation for low-latency federated edge learning," *IEEE Transactions on Wireless Communications*, vol. 19, no. 1, pp. 491–506, 2020.

[8] X. Cao, G. Zhu, J. Xu, and K. Huang, "Optimized power control for over-the-air computation in fading channels," *IEEE Transactions on Wireless Communications*, vol. 19, no. 11, pp. 7498–7513, 2020.

[9] X. Cao, G. Zhu, J. Xu, Z. Wang, and S. Cui, "Optimized power control design for over-the-air federated edge learning," *IEEE Journal on Selected Areas in Communications*, vol. 40, no. 1, pp. 342–358, 2022.

[10] X. Cao, G. Zhu, J. Xu, and S. Cui, "Transmission power control for over-the-air federated averaging at network edge," *IEEE Journal on Selected Areas in Communications*, vol. 40, no. 5, pp. 1571–1586, 2022.

[11] A. D. Dabbagh and D. J. Love, "Multiple antenna MMSE based downlink precoding with quantized feedback or channel mismatch," *IEEE Transactions on Communications*, vol. 56, no. 11, pp. 1859–1868, 2008.

[12] C. Tsinos, A. Galanopoulos, and F. Foukalas, "Low-complexity and low-feedback-rate channel allocation in CA MIMO systems with heterogeneous channel feedback," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 5, pp. 4396–4409, 2017.

[13] A. Giannopoulos, S. Spantideas, N. Kapsalis, P. Karkazis, and P. Trakadas, "Deep reinforcement learning for energy-efficient multi-channel transmissions in 5G cognitive HetNets: centralized, decentralized and transfer learning based solutions," *IEEE Access*, vol. 9, pp. 129358–129374, 2021.

[14] V. Mnih, K. Kavukcuoglu, D. Silver et al., "Human-level control through deep reinforcement learning," *Nature*, vol. 518, no. 7540, pp. 529–533, 2015.

[15] A. Giannopoulos, S. Spantideas, C. Tsinos, and P. Trakadas, "Power control in 5G heterogeneous cells considering user demands using deep reinforcement learning," in *IFIP International Conference on Artificial Intelligence Applications and Innovations*, pp. 95–105, Springer, 2021.

[16] A. Giannopoulos, S. Spantideas, N. Kapsalis et al., "Supporting intelligence in disaggregated open radio access networks: architectural principles, AI/ML workflow, and use cases," *IEEE Access*, vol. 10, pp. 39580–39595, 2022.