Optimal Resource Management for Hierarchical Federated Learning over HetNets with Wireless Energy Transfer

Rami Hamdi, Member, IEEE, Ahmed Ben Said, Emna Baccour, Aiman Erbad, Senior Member, IEEE, Amr Mohamed, Senior Member, IEEE, Mounir Hamdi, Fellow, IEEE, and Mohsen Guizani, Fellow, IEEE

Abstract—Remote monitoring systems analyze the environment dynamics in different smart industrial applications, such as occupational health and safety, and environmental monitoring. Specifically, in industrial Internet of Things (IoT) systems, the huge number of devices and the expected performance put pressure on resources, such as computational, network, and device energy. Distributed training of Machine and Deep Learning (ML/DL) models for intelligent industrial IoT applications is very challenging for resource limited devices over heterogeneous wireless networks (HetNets). Hierarchical Federated Learning (HFL) performs training at multiple layers offloading the tasks to nearby Multi-Access Edge Computing (MEC) units. In this paper, we propose a novel energy-efficient HFL framework enabled by Wireless Energy Transfer (WET) and designed for heterogeneous networks with massive Multiple-Input Multiple-Output (MIMO) wireless backhaul. Our energy-efficiency approach is formulated as a Mixed-Integer Non-Linear Programming (MINLP) problem, where we optimize the HFL device association and manage the wireless transmitted energy. However due to its high complexity, we design a Heuristic Resource Management Algorithm, namely H2RMA, that respects energy, channel quality, and accuracy constraints, while presenting a low computational complexity. We also improve the energy consumption of the network using an efficient device scheduling scheme. Finally, we investigate device mobility and its impact on the HFL performance. Our extensive experiments confirm the high performance of the proposed resource management approach in HFL over HetNets, in terms of training loss and grid energy costs.

Index Terms—HetNets, Hierarchical federated learning, device association, wireless energy transfer, energy efficiency.

I. INTRODUCTION

Occupational health and safety (OHS) has recently attracted the attention of governmental agencies in order to provide workers' protection and minimize health risks [2]– [4]. More generally, remote monitoring systems that capture the environment dynamics and monitor individuals are used in different applications to help senior citizens, patients, and factory workers. Distributed Artificial Intelligence (AI) [5] and communication technologies allow continuous monitoring for Key Health Indicators (KHIs) of a large number of workers using relevant industrial IoT devices (smart phones, mobile devices, wearables, smart cameras, and smart watches). The collected KHIs are then transmitted to a central or a nearby edge Central Unit (CU) to be processed by machine and deep learning algorithms and help in the decision making tasks.

With an increasing number of connected devices, centralized Machine and Deep Learning (ML/DL) algorithms require more and more memory resources and huge processing (i.e. CPU) [6]. In a wireless network setting with real-time and low-latency requirements, a centralized learning scheme becomes inefficient as it induces delay and may compromise the privacy of exchanged data between devices. Hence, decentralized learning approaches have emerged, particularly Federated Learning (FL) [6], [7]. FL enables to train a shared model among multiple devices, where each device uses only its private local data [7]. Indeed, the devices are training and updating the weights of the global model they received using their local data samples. Then, they send the new model weights to a CU for aggregation into a shared learning model. Finally, the CU broadcasts the aggregated model weights to the FL participants in order to update their models. This routine is iterated until reaching the desired model performance. FL addresses the challenge of data privacy, while enhancing the bandwidth usage as the amount of transferred data to the CU decreases. However, with the nexus of thousands of devices, several challenges need to be addressed including communication bottlenecks, increase in energy consumption, congestion, and and data skewness i.e. non-i.i.d data [10]. To cope with these issues, Hierarchical Federated Learning (HFL) has been presented [8], [9]. HFL enables offloading the computation task to the Multi-Access Edge Computing (MEC) in device vicinity, hence increasing the energy efficiency and enhancing the response time of the IoT nodes. In this new configuration, the devices send their local updates to their corresponding MEC for learning and model aggregation. Then, the MECs transfer their aggregated models to the CU in order to compute the global aggregation. The most distinctive feature of HFL is the ability to address and manage diverse data distribution, e.g., non-i.i.d property of the data. This is achieved by carefully assigning devices to MECs, hence speeding up convergence and attaining higher performance levels [10]. We highlight that if the data is i.i.d, the HFL may

A preliminary version of this work [1] has appeared in the proceedings of IEEE Global Communications Conference in 2021.

R. Hamdi is with the AFG College with the University of Aberdeen, Doha, Qatar (email: rami.hamdi@afg-aberdeen.edu.qa).

A. B. Said and A. Mohamed are with the College of Engineering, Qatar University, Doha, Qatar (email: abensaid@qu.edu.qa; amrm@qu.edu.qa).

E. Baccour, A. Erbad, and M. Hamdi are with the Division of Information and Computing Technology, College of Science and Engineering, Hamad Bin Khalifa University, Qatar Foundation, Doha, Qatar (email: ebaccourepbesaid@hbku.edu.qa; aerbad@ieee.org; mhamdi@hbku.edu.qa).

M. Guizani is with the Machine Learning Department, Mohamed Bin Zayed University of Artificial Intelligence (MBZUAI), Abu Dhabi, UAE (email: mguizani@ieee.org).

not have a big contribution to the system performance.

In a wireless communication environment with constant inter-device interference, latency constraint, increase in energy consumption and device heterogeneity, the implementation of HFL becomes challenging. Therefore, many research efforts have recently focused on making HFL practical for wireless communication networks. More precisely, recent works have tackled resources allocation problems, including wireless resource allocation, edge-device assignment, device selection and energy management with the objective of optimizing the performance of such system fully exploiting its potential. On the other hand, Wireless Energy Transfer (WET) represents a promising technology whose objective is to address the problem of energy bottlenecks in wireless networks [11]. The system is based on using electromagnetic radiation waves transmitted from a power source to deliver energy to a wireless device. Delivering power using wireless technology satisfies the frequent charging for mobile devices. Hence, it represents a key technology for the design of energy-efficient IoT wireless networks [12].

One of the challenges facing the ML/DL distribution is the high-energy consumption. Thus, to enhance the energy efficiency of HFL processed over HetNets with wirless backhaul, we propose to incorporate energy harvesting [13], where each MEC transmits the harvested energy to its related devices via WET. Compared to the existing schemes in [8]–[10], [14]–[26], our work is distinctive as it focuses on designing an energy-efficient HFL architecture assisted by WET. As per our knowledge, we are the first to consider the implementation of FL over HetNets with massive MIMO wireless backhaul enabled by WET. Furthermore, a practical mobility scenario that represents the workers in a construction site is considered by developing an efficient dynamic resource allocation scheme. The contributions of our work include the following:

- We propose a novel energy-efficient HFL framework, designed for HetNets with massive MIMO wireless backhaul and enabled by WET.
- We formulate our energy-efficiency approach as a Mixed-Integer Non-Linear Programming, where we optimize the HFL device association and manage the wireless transmitted energy.
- We derive the optimal device association analytically and we use Brute-Force search to obtain the energy management decisions.
- Due to the high complexity of the optimal solution, we propose a device association and energy management algorithm that aims to enhance the energy-efficiency of the proposed HFL framework.
- Some devices with high energy requirements may be deactivated for some frames. Hence, we propose a device scheduling strategy to decrease the energy consumption without deteriorating the learning accuracy.
- We develop an efficient dynamic device association scheme, where we consider a mobility scenario that represents the workers in a construction site.
- Finally, we conduct an extensive evaluation of the proposed system to prove the energy efficiency and high accuracy performance of our resource management, de-

2

vice association, and device scheduling schemes under different network configurations.

We organize the remainder of this article as follows: In section II, we explore the related works that study one-layer and hierarchical FL. In section III, we present the system model. The optimization problem is formulated in Section IV. We investigate the resource management in Section V. The impact of mobility in FL performance is investigated in Section VI. Evaluation results are provided in Section VII. Lastly, conclusions are drawn in Section VIII.

II. RELATED WORK

A. Federated Learning

A number of studies have explored resource allocation for FL [27]-[37] over wireless networks. In [27], an energy harvesting FL framework was developed by proposing efficient user association and scheduling schemes. Minimizing the energy costs of FL wireless systems was investigated, while taking into consideration a limited energy budget by developing an efficient resource management scheme. In [29], a data augmentation and adaptive user scheduling scheme was developed to overcome the challenges related to the imbalanced distribution of data in FL. The convergence time was theoretically investigated and optimized in [30] by designing an efficient control scheme. The FL performance was improved in [31] by proposing an efficient user scheduling policy. Moreover, the authors of [32] designed an optimal user selection algorithm that minimizes the FL convergence rate. In [33], a user selection scheme was proposed to decrease the effect of unreliable users and enhance the accuracy of federated learning training. To address the issue of battery limitation for FL users, initial works [34]-[37] incorporated WET systems when performing FL tasks. In [34], the authors assumed that the unmanned aerial vehicles (UAVs) have WET capability and proposed a suitable FL solution in terms of energy efficiency. In [35], a heterogeneous mobile architecture is implemented and an optimal resource allocation solution is proposed that minimizes the energy consumption of FL devices. In [36], a low-complexity beamforming strategy was proposed to minimize the mean-square error in FL based systems enabled by WET. The authors of [37] developed an analytic FL framework that establishes a trade-off between the model convergence and the power consumption which represents a guideline to ensure efficient learning tasks.

B. Hierarchical Federated Learning

Others works have tackled the optimization of resource allocation for HFL [8]–[10], [14]–[26]. More specifically, the authors in [9] used HFL to design a reliable anomaly detection strategy for IoT systems. Moreover, to face the challenges caused by non-i.i.d. data, the authors in [10] designed an edgedevice assignment scheme to enhance the federated averaging performance. Considering the same constraint of non-i.i.d. data among users, the authors in [19] proposed an efficient resource allocation and user assignment scheme in HFL for heterogeneous IoT systems. Meanwhile, the authors of [14] proposed a novel HFL architecture aiming to minimize the consumed

TABLE I: List of key Notations.

Symbol	Description	Symbol	Description
N _{cu}	Number of antennas at the CU	\mathbf{H}_m	Channel matrix between the CU and MEC m
N _{mec}	Number of antennas at each MEC	${f g}_{m,k}$	Channel vector between device k and MEC m
K	Number of devices	$\beta_{m,k}$	Pathloss
L	Number of frames	σ^2	Noise variance
M	Number of MECs	\mathbf{W}_m	Detection matrix at MEC m
Т	Duration of each frame	$E^{\rm dev}_{m,k}(l)$	Energy required by device k assigned to MEC m at frame l
Λ_m	Set of devices associated with MEC <i>m</i>	E ^{cir}	Fixed circuit energy consumption
K_m	Cardinal of Λ_m	$E^{\mathrm{ac}}_{m,k}(l)$	Required transmit energy of device k at frame l
С	Number of classes	Q	Size of the local FL model
\mathcal{D}_k	Dataset of device k	B _{max}	Maximal battery capacity
$egin{array}{c} x_k \end{array}$	Input of device k	$A_{m,k}(l)$	Amount of harvested energy by device k assigned to MEC m at frame l
$oldsymbol{y}_k$	Output of device k	$B_k(l)$	Battery level of device k at frame l
S	Data matrix of the devices	$E_m^{\rm wet}(l)$	Transferred energy from MEC m at frame l
$S_{c,k}$	Number of training data samples of the device k for the class c	$\alpha_m(l)$	Grid's weights
$oldsymbol{w}_k(l)$	Weight vector of the local ML model of device k at frame l	Δ	Total grid energy consumption cost
$\cup_{j\in\Lambda_m}\mathcal{D}_j$	Virtual dataset available at MEC m	$\chi_{m,k}$	Device association index
$p^{(m)}(i)$	Probability of the class i per MEC m 's distribution	$p(\cdot)$	Global classes distribution

energy and the total training latency under challenging data distribution. Energy-efficiency was also the focus of the work in [17], where authors have introduced a power control and client association solution for HFL. The work in [8] targeted the minimization of the global cost using a novel resource allocation and device association algorithm. Another extremely important metric in FL systems is the latency, that was the objective of the works in [21] and [15]. While the former optimized the association between multiple servers and devices, the latter designed a HFL framework across heterogeneous cellular networks. The authors in [18] proposed a flexible decentralized control over the training process in order to enhance the privacy of HFL, while the authors of [22] suggested using blockchain to meet the privacy requirements of internet of vehicles composing the HFL system. In [23], a novel HFL framework based on clustering is proposed to enhance the test accuracy. The authors of [24] investigated HFL by proposing a hybrid data partitioning algorithm. In [25], an efficient HFL algorithm is presented with an objective to determine the optimal cluster structure. In [26], a resource allocation heuristic was designed to guarantee the HFL convergence rate subject to fairness constraint. Finally, the game theory was

used in [16] and [20] to design an efficient resource allocation strategy for HFL.

Unlike the works proposed in the literature, we present in this paper a novel HFL architecture considering massive MIMO wireless backhaul, i.e. the links between the MECs and the devices and the links between MECs are wireless. Moreover, we consider a practical mobility scenario that represents the workers in a construction site. Additionally, the use of energy harvesting for HFL was not considered in most of the related works. Hence, we consider in this work that the MECs have WET capability and the devices are equipped with energy harvesting batteries, which allows to enhance the energy efficiency. Taking into account these challenging constraints, we formulate and investigate a novel resource optimization problem.

III. SYSTEM MODEL AND ENERGY MODEL

A. Hierarchical Federated Learning Model

Massive MIMO is based on multiplexing a large number of antennas simultaneously to satisfy the demand of multiple users in the same time-frequency resource, which contributes to improving energy efficiency, spectral efficiency, and latency [38]. Furthermore, HetNets, which are based on network densification in multiple dimensions with different types of cells, allow to efficiently exploit the wireless resources and improve the QoS. Incorporating massive MIMO and HetNets in IoT wireless networks is suitable to optimize the FL performance. Hence, we opt for a HetNet as presented in Fig. 1. Our system is composed of a massive MIMO Central Unit (CU) equipped with N_{cu} antennas, and M MECs, each one is endowed with N_{mec} MIMO antennas ($N_{\text{cu}} > MN_{\text{mec}}$). Different MECs communicate with the CU through a wireless backhaul link [39]. Furthermore, each MEC m serves multiple single-antenna devices. Let Λ_m be the set of K_m devices served by the MEC m. We note that the total number of devices in the network is equal to K. We study our system over an interval of time that we divide into L frames of duration T. Table I presents different key notations introduced in our paper. The device association index is defined as:

$$\chi_{m,k} = \begin{cases} 1, & \text{if device } k \text{ is associated to MEC } m, \\ 0, & \text{otherwise.} \end{cases}$$
(1)

Each device k owns a dataset $\mathcal{D}_k = \{x_k, y_k\}$, where x_k and y_k denote the input and output of the device k. These devices train their FL models locally and send the obtained weights to their related MECs. Then, each MEC aggregates the weights of its devices and forwards the results to the CU. Finally, the CU aggregates the weights received from all MECs and shares the updated FL model with all devices. This process is repeated until reaching a desirable accuracy. Let the Matrix $S \in \mathbb{N}^{C \times K}$ describe the splitting of data among all devices, where C denotes the total number of classes in the learning problem. Particularly, each element $S_{c,k}$ represents the number of data samples owned by the device k and belonging to the class c.

Let $w_k(l)$ represent the vector of FL weights of the local model trained within the device k at a frame l. After aggregating all weights, first incoming from the devices, then collected from the MECs, the CU is considered globally fitting the resultant weights by minimizing a particular loss function. In our work, we opt for the cross-entropy loss function that can be presented as follows:

$$\mathcal{L}(\boldsymbol{w}) = \sum_{c=1}^{C} -p(c)\mathbb{E}_{\boldsymbol{x} \sim q(\boldsymbol{x}|y=c)} \left[\log d_c(\boldsymbol{x}; \boldsymbol{w})\right], \quad (2)$$

where $p(\cdot)$ defines the global distribution of classes, $q(\cdot | \cdot)$ denotes the likelihood function, and $d_c(\boldsymbol{x}; \boldsymbol{w})$ represents the probability that the input \boldsymbol{x} belongs to the class c, under the parameters \boldsymbol{w} .

After receiving the weights of the local FL models from the related devices, each MEC computes the aggregation following the equation in [10]:

$$\boldsymbol{w}_{m}^{\text{mec}}(l) = \sum_{k \in \Lambda_{m}} \frac{|\mathcal{D}_{k}|}{|\cup_{j \in \Lambda_{m}} \mathcal{D}_{j}|} \boldsymbol{w}_{k}(l), \qquad (3)$$

where $\bigcup_{j \in \Lambda_m} D_j$ is the grouping of local datasets owned by different devices related to the MEC *m*. Accordingly, it can be considered as a virtual dataset trained at the MEC *m*. The next

step is to aggregate the weights received from all the MECs at the central unit, as done in [10]:

$$\boldsymbol{w}^{\mathrm{cu}}(l) = \sum_{m=1}^{M} \frac{|\cup_{k \in \Lambda_m} \mathcal{D}_k|}{|\cup_{j=1:K} \mathcal{D}_j|} \boldsymbol{w}_m^{\mathrm{mec}}(l)$$
(4)

Finally, the aggregated weights are broadcasted to all devices and this process is iterated until reaching a targeted convergence rate.

Because of the distribution and federation of the learning, a deviation between the FL and aggregated weights may occur. This divergence is studied in [10] and quantified as follows:

$$\theta = \sum_{m=1}^{M} \frac{\left| \bigcup_{k \in \Lambda_m} \mathcal{D}_k \right|}{\left| \bigcup_{j=1:K} \mathcal{D}_j \right|} \times \left\| D^{(m)} \right\|_1, \tag{5}$$

where $D^{(m)} = \{ \|p(i) - p^{(m)}(i)\| \}_{i=1}^{C}$, and $p^{(m)}(i)$ is the probability of the class *i* per MEC *m*'s distribution.

B. Communication Model

We assume that the channel between the CU and the MEC *m* undergoes a Rician fading represented by $\mathbf{H}_m \in \mathbb{C}^{M_s \times N}$. On the other hand, we consider a Gaussian i.i.d. small-scale fading channel between the MEC *m* and each device $k \in \Lambda_m$, denoted as $\mathbf{f}_{m,k}$. The access link between the MEC *m* and each device $k \in \Lambda_m$ is denoted by a vector $\mathbf{g}_{m,k} = \beta_{m,k}^{\frac{1}{2}} \mathbf{f}_{m,k} \in \mathbb{C}^{N_{\text{mec}} \times 1}$, where $\beta_{m,k}$ represents the pathloss. To this end, the channel matrix is defined as $\mathbf{G}_m = [\mathbf{g}_{m,k}]_{k=1:K_m} \in \mathbb{C}^{N_{\text{mec}} \times K_m}$.

In our system, we assume a perfect Channel State Information (CSI) of all devices at each MEC. Thus, all MECs can apply a cooperative Zero Forcing (ZF) scheme to suppress the effect of inter-cell and intra-cell interferences as done in [40]. Let $\mathbf{Z}_m = \mathbf{G}_m (\mathbf{G}_m^H \mathbf{G}_m)^{-1}$ be the decoder matrix of MEC *m*. Accordingly, the rate of a device *k* connected to the MEC *m* can be formulated as:

$$r_{m,k}^{\text{dev}}(l) = \lambda \cdot \log_2\left(1 + \frac{p_{m,k}^{\text{dev}}(l)}{\sigma^2 \left[(\mathbf{G}_m^H \mathbf{G}_m)^{-1}\right]_{k,k}}\right), \quad (6)$$

where λ is the bandwidth, and σ^2 represents the variance of an Additive White Gaussian Noise (AWGN).

When the MECs receive the FL weights from their related devices, they proceed the aggregation and forward their aggregated weights to the central unit. At the central unit side, in case of multiple antennas receivers, the Block Diagonalization (BD) is used to mitigate the multi-device interference. This approach is known by its high performance in MIMO systems [41]. Therefore, we assign for each MEC m a detection matrix \mathbf{W}_m , in order to distinguish different signals received from different MECs. Following [41], \mathbf{W}_m lies in the null space of matrix $\overline{\mathbf{H}}_m$, which is given by $\overline{\mathbf{H}}_m = [\mathbf{H}_1^T ... \mathbf{H}_{m-1}^T \mathbf{H}_m^T + ... \mathbf{H}_M^T]^T$. The decoder matrix for MEC m is obtained as $\mathbf{W}_m = \mathbf{V}_m^{(0)} \mathbf{V}_m^{(1)}$, where $\mathbf{V}_m^{(0)}$ is composed of the last $N_{cu} - \overline{J}_m$ right singular vectors of $\overline{\mathbf{H}}_m, \mathbf{V}_m^{(0)}$, $\overline{J}_m = \operatorname{rank}(\overline{\mathbf{H}}_m)$, and $J_m = \operatorname{rank}(\overline{\mathbf{H}}_m \mathbf{V}_m^{(0)})$. Hence, the



Fig. 1: WET-enabled HFL over HetNets in battery powered end-devices.

uplink rate of MEC m is given as [42]:

$$r_m^{\text{mec}}(l) = \lambda \cdot \log_2\left(1 + \frac{p_m^{\text{mec}}(l) \left\|\mathbf{H}_m \mathbf{W}_m\right\|_F^2}{\sigma^2}\right).$$
(7)

C. Energy Model

In this section, we evaluate the energy needed by the MECs and their related devices to compute different FL operations. First, the energy required by a device k connected to an MEC m to complete the FL tasks in the frame l is expressed by:

$$E_{m,k}^{\text{dev}}(l) = E^{\text{cir}} + E_k^{\text{cmp}} + E_{m,k}^{\text{ac}}(l),$$
 (8)

where E^{cir} is a fixed circuit energy consumed by filters, mixers, and digital to analog converters. E_k^{cmp} is the energy required by the device k to train its local data, while $E_{m,k}^{\text{ac}}(l)$ is the energy required to transmit the resulted weights at frame l.

The energy needed by the device k to train its data can be calculated as done in [43]:

$$E_k^{\rm cmp} = \varsigma \omega \vartheta^2 Q,\tag{9}$$

Where ς is an energy coefficient that depends on the device chip, ω represents the number of CPU cycles, ϑ is the CPU clock frequency, and Q denotes the size of the local model that highly depends on the deployed FL algorithm. It is worth mentioning that all local and global models in all devices, MECs, and CU present the same weights' size.

Next, to send the weights of the local trained model to its assigned MEC m at a frame l, the device k requires an energy equal to:

$$E_{m,k}^{\rm ac}(l) = \frac{p_{m,k}^{\rm dev}(l)Q}{r_{m\,k}^{\rm dev}(l)}.$$
(10)

where $p_{m,k}^{\text{dev}}(l)$ is the required transmit power of the device k connected to the MEC m at a frame t. Based on equation (6), $p_{m,k}^{\text{dev}}(l)$ can be formulated as:

$$p_{m,k}^{\text{dev}}(l) = \sigma^2 \left[(\mathbf{G}_m^H \mathbf{G}_m)^{-1} \right]_{k,k} \left(2^{\frac{r^{\text{dev}}(k)}{\lambda}} - 1 \right).$$
(11)

Hence, by replacing $p_{m,k}^{\text{dev}}(l)$ by its expression in (10), we obtain:

$$E_{m,k}^{\mathrm{ac}}(l) = \frac{\sigma^2 Q \left[(\mathbf{G}_m^H \mathbf{G}_m)^{-1} \right]_{k,k} \left(2^{\frac{r_{m,k}^{\mathrm{dev}}(l)}{\lambda}} - 1 \right)}{r_{m,k}^{\mathrm{dev}}(l)}.$$
 (12)

Each device in our FL system has a battery with a maximum capacity equal to B_{max} . These on-device batteries are also used to store the wireless transferred energy. Let $A_{m,k}(l)$ represent the amount of energy transmitted to a device k connected to the MEC m at a frame of time l. Hence, we define the battery level update of any device k related to the MEC m as follows:

$$B_k(l+1) = \min\left(B_{\max}, B_k(l) - E_{m,k}^{\text{dev}}(l) + A_{m,k}(l)\right).$$
(13)

Furthermore, we define the energy consumed by any MEC m using WET to supply its associated devices at a frame l as follows:

$$E_m^{\rm mec}(l) = E^{\rm cir} + E_m^{\rm wet}(l) + E_m^{\rm bh}(l),$$
 (14)

where $E_m^{wet}(l)$ denotes the energy transferred by the MEC m to its related devices using the WET system, at a frame l. Meanwhile, $E_m^{bh}(l)$ represents the energy consumed by the MEC m to transmit the aggregated FL weights to the CU. We remind that E^{cir} is a fixed circuit energy consumption. Next, we adopt an optimal beamformer for a MIMO WET system to calculate the energy supplied to each device k related to MEC m at a frame l. This energy is expressed as:

$$A_{m,k}(l) = E_m^{\text{wet}}(l)\beta_{m,k}\xi_{m,k}N_{\text{mec}},\qquad(15)$$

where $\xi_{m,k}$ is the beamforming coefficient for device k. The energy required by the MEC m to transmit the aggregated FL weights to the central unit at a frame l is given by:

$$E_m^{\rm bh}(l) = \frac{p_m^{\rm mec}(l)Q}{r_m^{\rm mec}(l)}.$$
(16)

The transmit power $p_m^{\text{mec}}(l)$ can be derived from (7) as:

$$p_m^{\text{mec}}(l) = \frac{\sigma^2}{\left\|\mathbf{H}_m \mathbf{W}_m\right\|_F^2} \left(2^{\frac{r_m^{\text{mec}}(l)}{\lambda}} - 1\right).$$
(17)

Hence, by replacing $p_m^{\text{mec}}(l)$ by its expression in (16), we obtain:

$$E_m^{\rm bh}(l) = \frac{\sigma^2 Q\left(2^{\frac{r_m(U)}{\lambda}} - 1\right)}{\left\|\mathbf{H}_m \mathbf{W}_m\right\|_F^2 r_m^{\rm mec}(l)}.$$
(18)

We assume that the energy consumed by each MEC m at a frame of time l is weighted by a parameter $\alpha_m(l)$ [44]. Thus, the total cost of the energy consumed by all MECs is calculated as follows:

$$\Delta = \sum_{l=1}^{L} \sum_{m=1}^{M} \alpha_m(l) E_m^{\text{mec}}(l).$$
(19)

IV. PROBLEM FORMULATION

Our objective in this work is to design an energy-efficient HFL framework for heterogeneous networks, where devices are supplied by wireless energy transfer. More specifically, we aim at minimizing the cost of the total grid energy consumed by different MECs by optimizing the device association and the wireless energy, constrained by the divergence between the FL's weights and central weights. Since the deviation between the FL and central weights should not exceed a given threshold, the local data distributions at different devices should be considered when deriving the device association. On these bases, the optimization problem is formulated as (20a). Constraint (20a) ensures that the deviation between the central weights and the FL's weights does not exceed a threshold $\theta_{\rm max}$. Constraint (20b) guarantees that each device in the network is assigned to only one MEC. Equation (20c) ensures that the consumed energy by the devices at each frame is constrained by the available energy at the batteries. Equation (20d) ensures that we cannot store energy more than the maximum battery capacity. Finally, the last two constraints (20e) and (20f) guarantee that the transferred amounts of energy via wireless links are positive, and the device association decisions are binaries.

$$\min_{\substack{\{\chi_{m,k}, E_m^{\text{wet}}(l)\}\\ n=1,\dots,M, k=1,\dots,K, l=1,\dots,L}} \Delta = \sum_{l=1}^{L} \sum_{m=1}^{M} \alpha_m(l) E_m^{\text{mec}}(l) \quad (20)$$

s. t.
$$\theta \le \theta_{\max}$$
, (20a)

$$\sum_{m=1}^{M} \chi_{m,k} = 1, \ \forall k = 1, \dots, K,$$
(20b)

$$\chi_{m,k} \sum_{i=1}^{M} E_{m,k}^{m,k}(i) \leq B_k(0) + \sum_{i=1}^{M} A_{m,k}(i),$$

$$\forall m = 1, \dots, M, k = 1, \dots, K, l = 1, \dots, L, \qquad (20c)$$

$$\sum_{i=1}^{l} A_{m,k}(l) - \chi_{m,k} \sum_{i=1}^{l-1} E_{m,k}^{\text{dev}}(i) \le B_{\max},$$

$$\forall m = 1, \dots, M, k = 1, \dots, K, l = 2, \dots, L,$$
 (20d)

$$E_m^{\text{wet}}(l) \ge 0, \ \forall m = 1, \dots, M, l = 1, \dots, L,$$
 (20e)

$$\chi_{m,k} \in \{0,1\}, \ \forall m = 1, \dots, M, k = 1, \dots, K.$$
 (20f)

Furthermore, we investigate an extended version of the formulated problem (20) by incorporating device scheduling. Some devices with high amount of wireless transferred energy may be unscheduled for certain communication rounds in order to improve the energy-efficiency of the network without deteriorating the FL training loss. The optimization problem becomes more challenging. Moreover, we extend the formulated problem (20) by incorporating a mobility scenario and assuming dynamic device association where the association index became variable over time as $\chi_{m,k}(l)$.

V. RESOURCE MANAGEMENT

A. Optimal Solution

In this section, we derive the optimal solution of the problem formulated in (20). First of all, we can see that our optimization is a Mixed-Integer Non-Linear Programming (MINLP) due to its combinatorial nature and the non-linearity of the constraint (20a). Hence, we propose to use the following theorem to find the optimal energy transferred by each MEC at each time frame, and consequently simplify the problem.

Theorem 1. For a MEC m and the set Λ_m of its associated

devices, the optimal transferred energy $E_m^{\text{wet}*}(l)$ from MEC m to its associated devices is given by:

$$E_m^{\text{wet}*}(l) = \max_{k \in \Lambda_m} \frac{E_{m,k}^{\text{dev}}(l) - B_k(l)}{\beta_{m,k} \xi_{m,k} N_{\text{mec}}}.$$
 (21)

Proof. At a frame l, each device k related to a MEC m is supplied by the harvested energy via WET, in addition to its own battery. The total energy of k is expressed as follows:

$$E_{m,k}^{\text{dev}}(l) = B_k(l) + A_{m,k}(l).$$
(22)

Thus, by replacing $A_{m,k}(l)$ by its expression presented in (15), we can derive the energy $E_m^{\text{wet}}(l)$ transferred by the MEC *m* to device *k*:

$$E_m^{\text{wet}}(l) = \frac{E_{m,k}^{\text{dev}}(l) - B_k(l)}{\beta_{m,k}\xi_{m,k}N_{\text{mec}}}.$$
(23)

Each MEC serves its related devices based on the highest required energy. More precisely, the MEC finds the device that requires the highest amount of energy and serves it along with the others with this same amount. $\hfill \Box$

Using the expression of the transferred energy $E_m^{wet}(l)$ presented in equation (21), our optimization problem (20) can be simplified as it becomes an integer non-linear problem. This simplified problem that aims now to find the optimal device association is combinatorial and is known to be NP-hard [45]. However, it can be solved using Brute-Force Search (BFS). Although deriving the optimal solution is still highly complex, the results can be used as a benchmark to our heuristic approach that will be proposed in what follows. It is also worth mentioning that by neglecting the constraint (20a), the optimization becomes an integer linear problem that can be solved using numerical tools such as CVX.

B. Heuristic Approach

Due to the combinatorial complexity of the optimal solution that may prohibit its online deployment, we propose a heuristic algorithm that solves the device association and energy management problem with low complexity. Similarly to the optimization, our algorithm aims to minimize the cost of the consumed grid energy, while respecting the FL divergence constraints.

Let R_m denote the range of the MEC m, which is the pathloss of the farthest associated device. Based on (16), we can see that the device with the higher pathloss receives the lowest amount of energy via WET. In other words, any device assigned to the MEC m with a lower pathloss than R_m , i.e., $\beta_{m,k} \leq R_m$, may benefit from a higher transferred energy. Accordingly, each device k tries to find the set of MECs having a higher range than its pathloss. Next, it selects the one with minimum divergence θ . This device association process is done one by one. If one of the devices fails to find a MEC with a suitable range, it will be associated to the closest one. Our proposed device association heuristic reduces not only the energy consumption but also the divergence between FL and central weights, which contributes to minimize the training loss. Finally, for each frame l, we can calculate the amount of energy transmitted by each MEC via WET using equation (21). Algorithm 1 presents different steps of our proposed HFL Heuristic Resource Management Algorithm (H2RMA). Moreover, the proposed algorithm is described in Fig. 2 and we provide an example of the device association procedure in Fig. 3. We can see in Fig. 3 that Dev 3 is within the range of MEC 1 and MEC 2. Hence, it chooses to be associated with the MEC that minimizes the divergence. For Dev 4, it is outside the range of all the MECs. Hence, it chooses to be associated with the MEC under minimal energy requirement. Then, the range of MEC 4 is updated following Dev 4.



Fig. 2: Diagram of the proposed algorithm H2RMA.



Fig. 3: An example of the device association procedure.

We investigate the complexity of H2RMA. The device association process is presented in lines (3-20) using two loops for with a complexity equal to O(MK). In the **for** loop in lines (21-29), the power control is conducted with a complexity in the order O(L(2K+M)). Hence, the complexity of H2RMA can be expressed as:

$$C^{\text{H2RMA}} = O(MK + L(2K + M)).$$
 (24)

Hence, H2RMA can be executed in polynomial time.

Algorithm 1 HFL Heuristic Resource Management Algorithm (H2RMA)

- 1: Initialization:
- 2: $\chi_{i_k,k} \leftarrow 0, \ m = 1..M, \ k = 1..K,$
- 3: $R_m \leftarrow 0, \ m = 1..M$
- 4: for k = 1 : K do
- Initialize $i_k \leftarrow 0$ 5:
- Initialize $\theta_{\min} \leftarrow \theta_{\max}$ 6:
- for m = 1 : M do 7:
- if $\beta_{m,k} \leq R_m$ then, verify if device k is within 8: the range of MEC m.
- compute θ 9:
- if $\theta \leq \theta_{\min}$ then 10
- 11: $i_k \leftarrow m$, select the MEC with the minimum divergence.
 - $\theta_{\min} \leftarrow \theta$
- end if 13:
- end if 14:
- end for 15:

12:

- if $i_k = 0$ then, if device k is not within the range of 16: any MEC.
- $i_k \leftarrow \operatorname{argmin}_{m=1..M} \beta_{m,k}$, associate with the 17: closest MEC m*.
- $R_{m*} \leftarrow \beta_{m*,k}$, update the range of the MEC m*. 18: end if
- 19:
- 20: $\chi_{i_k,k} \leftarrow 1$ 21: **end for**
- 22: for l = 1 : L do
- compute the energy required by each device. 23:
- 24.
- compute $E_m^{\text{wet}}(l), m = 1..M$ using (21).
- send the weights $w_k(l)$ to the related MECs and 25: aggregation using (4).
- $B_k(l+1) \leftarrow \min\left(B_{\max}, B_k(l) E_{m,k}^{\text{dev}}(l) + A_{m,k}(l)\right),$ 26: battery update
- send the weights $\boldsymbol{w}_m^{\mathrm{mec}}(l)$ to the central unit to compute 27: the aggregation using (5).
- send the aggregated weights to all devices. 28:
- different devices update their FL training models. 29:
- 30: end for

C. Device Scheduling

We now investigate the device scheduling for our proposed HFL framework. In fact, at a given frame l, some devices may face bad channel conditions and require high amount of wireless transferred energy, without contributing much on the HFL procedure. This costs the system unnecessary energy consumption that can be avoided. In this scenario, a practical idea to save the grid energy is to deactivate some devices for some frames based on the available energy and time-varying channel. Moreover, the scheduled devices at each frame should respect the maximal divergence constraint so as not to degrade of FL performance.

Including device scheduling to the optimization problem (20) which is MINLP increases the complexity of finding the optimal solution. This optimal device scheduling can be obtained with BFS. However, its high computational complexity prevents it from being conceivable. Hence, we propose an efficient heuristic device scheduling scheme to save unnecessary grid energy cost. We start by considering that all devices are selected. The devices to be deactivated at the current frame are those that require high amount of energy exceeding a certain threshold $E_{\rm th}$ given by the set:

$$\Omega = \{k \in \{1, \dots, K\} | E_{m,k}^{ac}(l) > E_{th}\}.$$
(25)

We deactivate the devices in Ω one by one iteratively and we start by the ones that require higher transmit energy. At each iteration, the device is deactivated only if the maximal divergence constraint is satisfied so as not to degrade FL performance.

VI. IMPACT OF MOBILITY

In this section, we investigate an applied case of oc pational health and safety by applying our proposed H framework in order to minimize the health risks of the work in a construction site. In consequence, we choose a mobi scenario [46] that reflects the workers equipped with smart vices in a construction site. The dynamic interactions betw workers on the site makes the construction job one of most risky activities. The mobility of workers can be mode by Hidden Markov Models (HMMs) [46], which are powe statistical tools that represent the probability distributions of observation. We can categorize the workers into three sta i.e., static, normal, and risky, depending on their speed turning angle. Hence, the hidden states of the HMM may defined depending on the values of these two parameters, speed and turning angle. These values are provided in Table Moreover, the step length of the workers is modeled by Gamma distribution and the turning angle is modeled by Von Mises distribution. Let us divide the network into mic cells with length μ . The average number of frames that normal worker moves from a micro-cell to another is gi by:

$$\psi_{\rm nml} = \lceil \frac{\mu}{TV_{\rm nml}} \rceil, \tag{26}$$

where $\lceil \cdot \rceil$ is the ceiling function, and V_{nml} is the average speed of a normal worker. The average number of frames that a risky worker moves from a micro-cell to another is given by:

$$\psi_{\rm rsk} = \lceil \frac{\mu}{TV_{\rm rsk}} \rceil,\tag{27}$$

where V_{rsk} is the average speed of a risky worker. It is clear that the risky workers change their micro-cells more frequently than the normal workers. Also, some workers when they pass by certain places in the network (like office, meeting room, coffee break ...), they remain static for a certain period.

Considering this mobility scenario, the mobile devices move around and become far from their associated MECs. Hence,

TABLE II: Values of Hidden States.

State	Speed (steps/mn)	Turning angle (radians)
Static	0	angle < $\pi/2$
Normal	$0 < steps \le 84$	$\pi/2 \leq \text{angle} < \pi$
Risky	steps > 84	angle $\geq \pi$

they require higher amounts of wireless transferred energy. Indeed, performing fixed device association scheme when the devices are moving involves an increase in the energy consumption cost. Therefore, we need to design an adaptive device association scheme taking into account devices mobility.

1:	for $l = 1 : L$ do		
2:	if $l \mod \psi_{rsk} = 0$ or $l \mod \psi_{nrm} = 0$ then,		
3:	Update Γ the set of mobile devices that change		
	their micro-cell during the current frame.		
4:	for $k\in \Gamma$ do		
5:	$i_k \leftarrow 0$, initialization		
6:	$\theta_{\min} \leftarrow \theta_{\max}$, initialization		
7:	for $m = 1: M$ do		
8:	if $\beta_{m,k} \leq R_m$ then, verify if the mot		
	device k is within the range of MEC m .		
9:	compute θ		
10:	if $ heta \leq heta_{\min}$ then		
11:	$i_k \leftarrow m$, select the MEC that 1		
	the minimum divergence.		
12:	$ heta_{\min} \leftarrow heta$		
13:	end if		
14:	end if		
15:	end for		
16:	if $i_k = 0$ then, if the mobile device k is		
	within the range of any MEC.		
17:	$i_k \leftarrow \operatorname{argmin}_{m=1M} \beta_{m,k}$, associate w		
	the closest MEC $m*$.		
18:	$R_{m*} \leftarrow eta_{m*,k}$, update the range of		
	MEC $m*$.		
19:	end if		
20:	$\chi_{i_k,k} \leftarrow 1$		
21:	end for		
22:	end if		
23:	end for		

We propose a heuristic dynamic device association strategy for HFL over HetNets enabled by WET. We start by executing the original device association algorithm H2RM presented in section V. Then, we change the associated MECs of the normal workers after each ψ_{nml} frames and for the risky workers after each ψ_{rsk} frames. The mobile device k updates its new associated MEC as follows. It finds first the new set of MECs that have a range higher than its pathloss. We already defined the range R_m of a MEC m as the pathloss of the farthest associated device. Hence, being associated with one of these MECs will not induce any supplementary energy cost. The mobile device will choose the MEC having the minimum divergence θ . If this device fails to find a MEC with suitable range, the closest one will be selected. The proposed Dynamic HFL Device Association Algorithm (DHDA) is illustrated in Algorithm 2.

To find the complexity of DHDA, we need to multiply the computational complexity of the first **for** loop of H2RMA by the number of times that we repeat association of mobile device $\left\lceil \frac{L}{\psi_{rsk}} \right\rceil$. Hence, the complexity of DHDA is give by:

$$C^{\text{DHDA}} = O(\lceil \frac{L}{\psi_{\text{rsk}}} \rceil MK + L(2K + M)).$$
(28)

VII. NUMERICAL RESULTS

In this section, we investigate the performance of our proposed HFL framework under different network configurations. The simulation settings are as follows: we positioned the CU at the center of a circular macro-cell and uniformly distributed the MECs and devices within the considered coverage. First, we consider that devices are required to recognize digits between 0 and 9, which can be simulated by training a classification on MNIST dataset. In our simulation, we used a deep learning model to train on the MNIST dataset. This model is composed of two convolutional layers of 3x3 filters and we chose Rectified Linear Unit (ReLU) as an activation function. To downsize the resulting feature maps, we added a max pooling layer, which output is flattened to be fed into the subsequent fully connected layer. Finally, the last layer is designed to generate the probabilities of the 10 classes. We note that we used the dropout technique [47] to reduce the overfitting and regularize the model. Furthermore, we incorporated the information about the class weight in the training process to mitigate the problem of imbalanced classes, and we used simple Stochastic Gradient Descent (SGD) to train the model and optimize the cross-entropy loss, throughout 20 training epochs. Next, we investigate CIFAR-10 (Canadian Institute For Advanced Research) which is a well-known image dataset that is commonly used to validate the performance of machine learning and computer vision algorithms. The data consist of 60000 color images of size 32x32 and 10 classes; each class has 6000 images. The class labels are: horse, frog, dog, cat, deer, bird, automobile, airplane, truck, and ship. The data does not have overlap between classes, i.e. the classes are mutually exclusive. We summarize the simulation parameters used in this evaluation section in Table III.

A. Fixed Device Association without Scheduling

We start by investigating the performance of the proposed HFL framework considering a static network and fixed device association over time. In Fig. 4, we show the grid energy cost, when varying the number of devices. Deriving the optimal device association suffers from a high computation complexity. Thus, we run our simulation using only two MECs in order to establish the optimal solution as a baseline. Compared to the optimal solution, we can see that H2RMA achieves a good performance in terms of energy consumption cost, while offering a much lower computation complexity. In Fig. 5, we illustrate the grid energy cost incurred by our heuristic compared to the random device association scheme, when changing the number of devices K. Using a large number

TABLE III: Configuration of the HFL over HetNet environment in our experiments.

Symbol	Description	Value
	number of antennas at the	
N _{cu}	CU	128
M	number of MECs	8
	number of antennas at each	
N _{mec}	MEC	16
K	number of devices	20
L	number of frames	50
C	number of classes	10
B _{max}	max battery capacity	1000 J
$B_k(0)$	initial battery level	200 J
ν	path loss exponent	3.7
λ	bandwidth	20 MHz
ϑ	frequency of the CPU clock	10^{9}
ω	number of CPU cycles	40
ς	consumed energy coefficient	10^{-27}
$\theta_{\rm max}$	maximal divergence	0.5
	cell radius	200 m
	circuit power per RF chain	30 dBm
	noise power spectral density	-174 dBm/Hz



Fig. 4: Grid energy cost while varying the number of devices K: comparison between optimal device association, random association, and H2RMA ($M = 2, B_k(0) = 10$ J).

of devices, H2RMA presents a highly lower grid energy cost, due to its adequate device association method. Fig. 6 depicts the performance of our heuristic algorithm when varying the number of MECs. When the number of MECs increases, the cost of grid energy decreases, as different devices in the system are surrounded by a larger set of MECs; hence, they have a higher probability to be associated to closer MECs. Moreover, we can see clearly, that our H2RMA algorithm outperforms the random association scheme, for different numbers of MECs.

Fig. 7 presents the accuracy performance of H2RMA com-



Fig. 5: Grid energy cost while varying the number of devices K: comparison between H2RMA and random device association scheme ($M = 8, B_k(0) = 200$ J).



Fig. 6: Grid energy cost while varying the number of MECs M: comparison between H2RMA and random device association ($K = 20, B_k(0) = 200$ J).

pared to the random association scheme, when tested using MNIST dataset. We can see that our algorithm converges faster, owing to its efficient device association process that minimizes the divergence. In Fig. 8, the accuracy performance of H2RMA is evaluated, while considering CIFAR-10 dataset. Similar to the MNIST dataset, our proposed device association scheme performs well and outperforms the random device association scheme in terms of test accuracy.

B. Fixed Device Association with Scheduling

We investigate now in Fig. 9 the impact of the proposed device scheduling on the grid energy consumed by our system. It is clear that we can save a significant amount of energy while deactivating some devices that require a high amount of wireless transferred energy for a certain number of frames. Moreover, the proposed device scheduling scheme performs very well compared to the random scheme, because only far



Fig. 7: Test accuracy: comparison between H2RMA and random device association scheme considering MNIST (K = 15, M = 3, C = 3).



Fig. 8: Test accuracy: comparison between H2RMA and random device association scheme considering CIFAR-10 (K = 15, M = 3, C = 2).

devices are being deactivated, while respecting the maximal divergence constraint.

Fig. 10 depicts the performance of our device scheduling heuristic when varying the number of MECs. We can save a significant amount of grid energy specifically when the number of MECs increases thanks to the efficient device scheduling scheme.

Fig. 11 depicts the performance of our heuristic device scheduling algorithm in terms of test accuracy considering MNIST. It can be seen that when deactivating some devices randomly, the test accuracy decreases. However, the proposed heuristic device scheduling scheme allows to enhance the convergence rate since only a subset of devices that respect the divergence constraint is scheduled.



Fig. 9: Grid energy cost while varying the number of devices K: comparison between the proposed scheduling heuristic, random device association and system scheduling all devices $(M = 8, B_k(0) = 200 \text{ J}).$



Fig. 10: Grid energy cost while varying the number of MECs M: comparison between the proposed scheduling heuristic, random device association and system scheduling all devices $(K = 20, B_k(0) = 200 \text{ J}).$

C. Dynamic Device Association

Finally, we implement a mobility scenario for the devices which is described in Section V. The mobility parameters used in this simulation are illustrated in Table II. In Fig. 12, we evaluate the cost of grid energy consumed by our HFL framework, when considering the mobility scenario. Indeed, the proposed dynamic device association algorithm allows saving significant amount of energy because it allows us to efficiently adapt the association of the mobile devices. Furthermore, the performance gap between dynamic and fixed device association increases when the number of mobile devices in the network becomes higher.

Fig. 13 shows the performance of the proposed HFL framework in terms of convergence rate considering a mobility scenario. The proposed dynamic device association scheme ensures that mobile devices respect the maximal divergence



Fig. 11: Test accuracy considering MNIST: comparison between the proposed scheduling heuristic, random device association and system scheduling all devices (K = 15, M = 3, C = 3).



Fig. 12: Grid energy cost while varying the number of devices K considering a mobility scenario ($M = 8, B_k(0) = 200$ J).

constraints and hence achieve better convergence rate than the fixed device association.

To summarize, the performance of our HFL framework is evaluated by showing the energy-efficiency of the proposed resource management approaches, in addition to the high accuracy achieved under different network configurations and using two well-known datasets, namely MNIST and CIFAR-10. Moreover, we showed the efficiency of the proposed dynamic device association considering a mobility scenario in the network.

VIII. CONCLUSION

An energy-efficient HFL over HetNets with massive MIMO wireless backhaul enabled by WET has been investigated in this work. An optimization problem involving a grid energy consumption cost constrained by energy availability and maximal divergence, have been developed. In addition, an



Fig. 13: Test accuracy considering a mobility scenario and MNIST dataset (K = 15, M = 3, C = 3).

optimal energy management has been derived. To achieve this, we have formulated the main objective as a device association optimization problem that we solved using BFS. Because of the high complexity to get the optimal solution, we have proposed an efficient device association heuristic, with low computational complexity, namely H2RMA. Moreover, a device scheduling strategy has been designed to save the grid energy without deteriorating the learning accuracy. The impact of mobility on FL performance has also been investigated by developing an efficient dynamic device association scheme. Our extensive experiments confirm the efficiency of our HFL framework in terms of accuracy, convergence rate and grid energy consumption, when evaluated under different network configurations.

As a future work, we plan to develop an efficient clustering strategy based on reinforcement learning for HFL over HetNets enabled by WET.

ACKNOWLEDGMENT

This work was supported by National Priorities Research Program-Standard (NPRP-S) Thirteen (13th) Cycle grant # NPRP13S-0205-200265 from the Qatar National Research Fund (a member of Qatar Foundation). The findings achieved herein are solely the responsibility of the authors.

REFERENCES

- R. Hamdi, A. B. Said, A. Erbad, A. Mohamed, M. Hamdi, and M. Guizani, "Hierarchical federated learning over HetNets enabled by wireless energy transfer," *in Proc. IEEE Global Commun Conf. (GLOBE-COM)*, Madrid, Spain, Dec. 2021, pp. 1-6.
- [2] D. Roberts, "Risk management and electrical safety: Implementation of an occupational health and safety management system," *IEEE Ind. Appl. Mag.*, vol. 21, no. 3, pp. 67-74, Feb. 2015.
- [3] A. Ranjan, Y. Zhao, H. B. Sahu, and P. Misra, "Opportunities and challenges in health sensing for extreme industrial environment: Perspectives from underground mines," *IEEE Access*, vol. 7, pp. 139181-139195, Sep. 2019.
- [4] J. A. Palazon, J. Gozalvez, J. L. Maestre, and J. R. Gisbert, "Wireless solutions for improving health and safety working conditions in industrial environments," *in Proc. IEEE Int. Conf. e-Health Netw., Appl. Services (Healthcom)*, Lisbon, Portugal, Oct. 2013, pp. 544-548.
 [5] E. Baccour, N. Mhaisen, A. A. Abdellatif, A. Erbad, A. Mohamed, M.
- [5] E. Baccour, N. Mhaisen, A. A. Abdellatif, A. Erbad, A. Mohamed, M. Hamdi, and M. Guizani, "Pervasive AI for IoT applications: Resource-efficient distributed artificial intelligence," *IEEE Commun. Surveys Tuts.*, vol. 24, no. 4, pp. 2366-2418, Aug. 2022.

- [6] M. Chen, D. Gündüz, K. Huang, W. Saad, M. Bennis, A. V. Feljan, and H. V. Poor, "Distributed learning in wireless networks: Recent progress and future challenges," *IEEE J. Sel. Areas Commun.*, vol.39, no. 12, pp. 3579-3605, Dec. 2021.
- [7] W. Y. B Lim, N. C. Luong, D. T. Hoang, Y. Jiao, Y. C. Liang, Q. Yang, D. Niyato, and C. Miao, "Federated learning in mobile edge networks: A comprehensive survey," *IEEE Commun. Surveys Tuts.*, vol. 22, no. 3, pp. 2031-2063, Apr. 2020.
- [8] S. Luo, X. Chen, Q. Wu, Z. Zhou, and S. Yu, "HFEL: Joint edge association and resource allocation for cost-efficient hierarchical federated edge learning," *IEEE Trans. Wireless Commun.*, vol. 19, no. 10, pp. 6535-6548, Jun. 2020.
- [9] X. Wang, S. Garg, H. Lin, J. Hu, G. Kaddoum, M. J. Piran, and M. S. Hossain, "Towards accurate anomaly detection in industrial internetof-things using hierarchical federated learning," *IEEE Internet Things J.*, vol.9, no. 10, pp. 7110-7119, May 2022.
- [10] N. Mhaisen, A. Awad, A. Mohamed, A. Erbad, and M. Guizani, "Optimal user-edge assignment in hierarchical federated learning based on statistical properties and network topology constraints," *IEEE Trans. Netw. Sci. Eng.*, vol. 9, no. 1, pp. 55-66, Feb. 2022.
- [11] G. Yang, C. K. Ho, R. Zhang, and Y. L. Guan, "Throughput optimization for massive MIMO systems powered by wireless energy transfer," *IEEE J. Sel. Areas Commun.*, vol. 33, no. 8, pp. 1640-1650, Aug. 2015.
- [12] J. Huang, Y. Zhou, Z. Ning, and H. Gharavi, "Wireless power transfer and energy harvesting: Current status and future prospects," *IEEE Wireless Commun.*, vol. 26, no. 4, pp. 163-169, Aug. 2019.
- [13] R. Hamdi, E. Driouch, and W. Ajib, "Energy management in hybrid energy large-scale MIMO systems," *IEEE Trans. Veh. Technol.*, vol. 66, no. 11, pp. 10183-10193, Nov. 2017.
- [14] L. Liu, J. Zhang, S. H. Song, and K. B. Letaief, "Client-edge-cloud hierarchical federated learning," in Proc. IEEE Int. Conf. Commun. (ICC), Virtual Conference, Jun. 2020, pp. 1-6.
- [15] M. S. H. Abad, E. Ozfatura, D. Gunduz, and O. Ercetin, "Hierarchical federated learning across heterogeneous cellular networks," *in Proc. IEEE Int. Conf. Acoustics, Speech, Signal Process. (ICASSP)*, Barcelona, Spain, May 2020, pp. 1-5.
- [16] W. Y. B. Lim, J. S. Ng, Z. Xiong, D. Niyato, C. Miao, and D. I. Kim, "Dynamic edge association and resource allocation in self-organizing hierarchical federated learning networks," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 12, pp. 3640-3653, Dec. 2021.
- [17] B. Xu, W. Xia, W. Wen, P. Liu, H. Zhao, and H. Zhu, "Adaptive hierarchical federated learning over wireless networks," *IEEE Trans. Veh. Technol.*, vol. 71, no. 2, pp. 2070-2083, Feb. 2022.
- [18] A. Wainakh, A. S. Guinea, T. Grube, and M. Mühlhäuser, "Enhancing privacy via hierarchical federated learning," *in Proc. IEEE Euro. Symp. Security Privacy Workshops (EuroS&PW)*, Genoa, Italy, Sep. 2020, pp. 1-4.
- [19] A. A. Abdellatif, N. Mhaisen, A. Mohamed, A. Erbad, M. Guizani, Z. Dawy, and W. Nasreddine, "Communication-efficient hierarchical federated learning for IoT heterogeneous systems with imbalanced data," *Future Generation Computer Syst.*, vol. 128, pp. 406-419, Mar. 2022.
- [20] W. Y. B. Lim, J. S. Ng, Z. Xiong, J. Jin, Y. Zhang, D. Niyato, C. Leung, and C. Miao, "Decentralized edge intelligence: A dynamic resource allocation framework for hierarchical federated learning," *IEEE Trans. Parallel Distrib. Syst.*, vol.33, no.3, pp. 536-550, Mar. 2022.
- [21] S. R. Pandey, M. N. Nguyen, T. N. Dang, N. H. Tran, K. Thar, Z. Han, and C. S. Hong, "Edge-assisted democratized learning towards federated analytics," *IEEE Internet Things J.*, vol. 9, no.1, pp. 572-588, Jan. 2022.
- [22] H. Chai, S. Leng, Y. Chen, and K. Zhang, "A hierarchical blockchainenabled federated learning algorithm for knowledge sharing in internet of vehicles," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 7, pp. 3975-3986, Jul. 2021.
- [23] C. Briggs, Z. Fan, and P. Andras, "Federated learning with hierarchical clustering of local updates to improve training on non-IID data," *in Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Glasgow, UK, Jul. 2020, pp. 1-9.
- [24] L. Su and V. K. N. Lau, "Hierarchical federated learning for hybrid data partitioning across multitype sensors," *IEEE Internet Things J.*, vol. 8, no. 13, pp. 10922-10939, Jul. 2021.
- [25] Z. Wang, H. Xu, J. Liu, H. Huang, C. Qiao, and Y. Zhao, "Resourceefficient federated learning with hierarchical aggregation in edge computing," *in Proc. IEEE Conf. Comput. Commun. (INFOCOM)*, Virtual Conference, May 2021, pp. 1-9.
- [26] J. Feng, L. Liu, Q. Pei, and K. Li, "Min-max cost optimization for efficient hierarchical federated learning in wireless edge networks," *IEEE Trans. Parallel Distrib. Syst.*, vol.33, no.11, pp. 2687-2700, Nov. 2022.

- [27] R. Hamdi, M. Chen, A. B. Said, M. Qaraqe, and H. V. Poor, "Federated learning over energy harvesting wireless networks," *IEEE Internet Things J.*, vol. 9, no. 1, pp. 663-675, Jan. 2022.
- [28] Z. Yang, M. Chen, W. Saad, C. S. Hong, and M. Shikh-Bahaei, "Energy efficient federated learning over wireless communication networks," *IEEE Trans. Wireless Commun.*, vol. 20, no. 3, pp. 1935-1949, Mar. 2021.
- [29] M. Duan, D. Liu, X. Chen, R. Liu, Y. Tan, and L. Liang, "Self-balancing federated learning with global imbalanced data in mobile systems," *IEEE Trans. Parallel Distrib. Syst.*, vol. 32, no. 1, pp. 59-71, Jan. 2021.
- [30] S. Wang, T. Tuor, T. Salonidis, K. K. Leung, C. Makaya, T. He, and K. Chan, "Adaptive federated learning in resource constrained edge computing systems," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 6, pp. 1205-1221, Jun. 2019.
- [31] H. H. Yang, Z. Liu, T. Q. Quek, and H. V. Poor, "Scheduling policies for federated learning in wireless networks," *IEEE Trans. Commun.*, vol. 68, no. 1, pp. 317-333, Jan. 2020.
- [32] M. Chen, Z. Yang, W. Saad, C. Yin, H. V. Poor, and S. Cui, "A joint learning and communications framework for federated learning over wireless networks," *IEEE Trans. Wireless Commun.*, vol. 20, no. 1, pp. 269-283, Jan. 2021.
- [33] J. Kang, Z. Xiong, D. Niyato, Y. Zou, Y. Zhang, and M. Guizani, "Reliable federated learning for mobile networks," *IEEE Wireless Commun.*, vol. 27, no. 2, pp. 72-80, Apr. 2020.
- [34] Q. V. Pham, M. Le, T. Huynh-The, Z. Han, and W. J. Hwang, "Energy-efficient federated learning over UAV-enabled wireless powered communications," *IEEE Trans. Veh. Technol.*, vol. 71, no. 5, pp. 4977-4990, Feb. 2022.
- [35] J. Feng, W. Zhang, Q. Pei, J. Wu, and X. Lin, "Heterogeneous computation and resource allocation for wireless powered federated edge learning systems," *IEEE Trans. Commun.*, vol. 70, no. 5, pp. 3220-3233, Mar. 2022.
- [36] M. Arslan, C. Cruz, and D. Ginhac, "Battery-less massive access for simultaneous information transmission and federated learning in WPT networks," in Proc. IEEE Conf. Comput. Commun. Workshops (INFOCOM WKSHPS), New York, USA, May 2022, pp. 1-2.

- [37] Q. Zeng, Y. Du, and K. Huang, "Wirelessly powered federated edge learning: Optimal tradeoffs between convergence and power transfer," *IEEE Trans. Wireless Commun.*, vol. 21, no. 1, pp. 680-695, Jan. 2022.
- [38] T. E. Bogale, and L. B. Le, "Massive MIMO and mmWave for 5G wireless HetNet," *IEEE Veh. Technol. Mag.*, vol. 11, no. 1, pp. 64-75, Mar. 2016.
- [39] R. Hamdi, E. Driouch, and W. Ajib, "New efficient transmission technique for HetNets with massive MIMO wireless backhaul," *IEEE Trans. Veh. Technol.*, vol. 69, no. 1, pp. 663-675, Jan. 2020.
- [40] O. Somekh, O. Simeone, Y. Bar-Ness, A. M. Haimovich, and S. Shamai, "Cooperative multicell zero-forcing beamforming in cellular downlink channels," *IEEE Trans. Inf. Theory*, vol. 55, no. 7, pp. 3206-3219, Jul. 2009.
- [41] W. Ni and X. Dong, "Hybrid block diagonalization for massive multiuser MIMO systems," *IEEE Trans. Commun.*, vol. 64, no. 1, pp. 201-211, Jan. 2016.
- [42] T. X. Tran and K. C. Teh, "Performance analysis of massive multiuser multiple-input multiple-output systems with block diagonalisation," *IET Commun.*, vol. 10, no. 7, pp. 832–838, May 2016.
- [43] M. Chen, Z. Yang, W. Saad, C. Yin, H. V. Poor, and S. Cui, "A joint learning and communications framework for federated learning over wireless networks," *IEEE Trans. Wireless Commun.*, vol. 20, no. 1, pp. 269-283, Jan. 2021.
- [44] Y. Che, L. Duan, and R. Zhang, "Dynamic base station operation in large-scale green cellular networks," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 12, pp. 3127-3141, Dec. 2016.
- [45] Z. Q. Luo and S. Zhang, "Dynamic spectrum management: Complexity and duality," *IEEE J. Sel. Topics Signal Process.*, vol. 2, no. 1, pp. 57-73, Feb. 2008.
- [46] M. Arslan, C. Cruz, and D. Ginhac, "Understanding worker mobility within the stay locations using HMMs on semantic trajectories," *in Proc. IEEE Int. Conf. Emerg. Technol. (ICET)*, Islamabad, Pakistan, Nov. 2018, pp. 1-6.
- [47] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Research*, vol. 15, no. 1, pp. 1929-1958, Jun. 2014.