Quantization-Aware Federated Learning with Coarsely Quantized Measurements

Alireza Danaee*, Rodrigo C. de Lamare*,†, and Vítor H. Nascimento‡
* Centre for Telecommunications Studies, Pontifical Catholic University of Rio de Janeiro, Brazil
† Department of Electronic Engineering, University of York, United Kingdom
‡ Department of Electronic Systems Engineering, University of São Paulo, Brazil
alireza@cetuc.puc-rio.br, delamare@cetuc.puc-rio.br, vitor@lps.usp.br

***Abstract*—In this work, we present an energy-efficient federated learning framework using coarsely quantized measured data for Internet of Things (IoT) networks. In particular, we develop a quantization-aware Federated Averaging Least Mean Square (QA-FedAvg-LMS) algorithm that can learn parameters in an energy-efficient fashion using measurements quantized with few bits and develop a bias compensation strategy to further improve the performance of the QA-FedAvg-LMS algorithm. We carry out a statistical analysis of the proposed QA-FedAvg-LMS algorithm. Numerical results assess the QA-FedAvg-LMS algorithm against existing techniques for a parameter estimation task in a scenario where IoT devices operate in federated learning mode.**

***Index Terms*—Federated learning, energy-efficient signal processing, mean-square error, coarse quantization**

I. Introduction

The goal of federated learning [1], [2] is to learn a global statistical model from data stored at tens to millions of devices subject to storing the data locally at devices and only communicating the intermediate updates generated by devices to the server. In this context, Internet of Things (IoT) networks include smart devices such as mobile phones, smart watches, and autonomous vehicles which are generating new data every day [3]. Federated learning offers IoT networks local data storage at devices and transfers network computation to the devices due to the growing computational capability of these devices. This can mitigate concerns over transmitting private information. Moreover, IoT devices contain many sensors that allow them to interact with the physical world, collecting and processing streaming data in real time [4], [5]. They integrate various sensors such as temperature, humidity, accelerometer, gyroscope, magnetometer, altimeter, heart rate, light, microphone, camera, battery monitor, infrared proximity, gas, ultraviolet, capacitive sensors. The type of sensor determines the accuracy of the analog interface and the resolution of the analog-to-digital converter (ADC). The ADC resolution requirement varies greatly with the sensing application, ranging from 6 to 16 bits (see [6] Table 1), and has a trade-off between sensing performance and energy consumption since the energy consumption of ADCs strongly depends on the number of bits used to represent digital samples [7].

Prior work on energy efficiency has reported many contributions in signal processing for communications and electronic systems that operate with coarsely quantized signals [8]–[10]. Even though there have been many studies on federated learning that evaluated the need for communication-efficiency to reduce the cost of communication between devices and the server [2], [11], prior work on energy-efficient techniques that reduce the cost of ADCs deployed at sensors is rather limited. In adaptive IoT networks, a distributed quantization-aware least-mean square (DQA-LMS) algorithm was proposed in [12], [13] to reduce the power consumption of ADCs in sensors that measure the input data in an energy-efficient framework.

In this work, we propose an energy-efficient federated learning framework using coarsely quantized measured data for IoT networks. In particular, we present a quantization-aware Federated Averaging Least Mean Square (QA-FedAvg-LMS) algorithm that can learn parameters in an energy-efficient fashion using measurements quantized with few bits, and devise a bias compensation strategy to further improve the performance of the proposed QA-FedAvg-LMS algorithm. We carry out a statistical analysis of the proposed QA-FedAvg-LMS algorithm. Simulations assess the QA-FedAvg-LMS algorithm against existing techniques for a parameter estimation task in a scenario where IoT devices operate with federated learning.

Throughout this paper, we denote scalars, vectors and matrices with lowercase, boldface lowercase and boldface uppercase letters, respectively. $\overline{a}$ and $\widehat{a}$ denote the estimated and coarsely quantized versions of a, respectively. $\mathbf{I}_M$ is the $M \times M$ identity matrix. $(.)^T$ and $(.)^*$ denotes transposition and complex conjugate (Hermitian) transposition, respectively.

II. System Model and Problem Statement

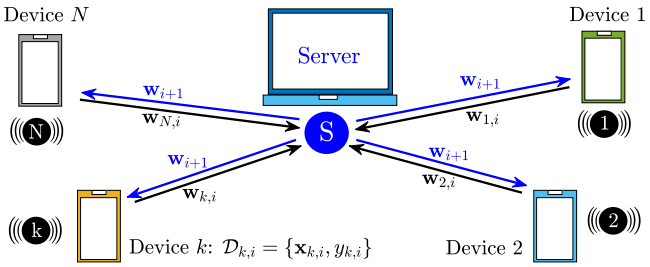

Fig. 1. A federated IoT network

Fig. 1 shows the architecture of an IoT network consisting of N IoT edge devices orchestrated by a server under federated learning strategy. Each device k, $k = 1, \ldots, N$, has access to local training data $\mathcal{D}_k$ including $n_k = |\mathcal{D}_k|$ data samples.

This work was supported in part by CNPq, CAPES, FAPERJ and by the ELIOT Project (FAPESP 2018/12579-7, ANR-18-CE40-0030).

Data sample i is represented by $\mathcal{D}_{k,i} = \{\mathbf{x}_{k,i}, y_{k,i}\}$ where $\mathbf{x}_{k,i} \in \mathbb{C}^M$ and $y_{k,i}$ are the ith input data vector and associated output response at device k, respectively. The goal of the learning task of the network is typically determined by:

$$\mathbf{w}_{opt} \triangleq \min_{\mathbf{w}\in\mathbb{R}^M} j(\mathbf{w}) = \sum_{k=1}^{N} a_k j_k(\mathbf{w}), \tag{1}$$

where $j(\mathbf{w})$ is the global objective function, $j_k(\mathbf{w})$ is the local objective function, $a_k \geq 0$ denote weights and $\sum_k a_k = 1$. It is common to set $a_k = \frac{1}{N}$ in homogeneous networks to give an equal weight to every device k. The local objective function is given by:

$$j_k(\mathbf{w}) = \frac{1}{n_k} \sum_{i=1}^{n_k} \ell(\mathbf{w}; \mathbf{x}_{k,i}, y_{k,i}), \tag{2}$$

where $\ell(\mathbf{w}; \mathbf{x}_{k,i}, y_{k,i})$ quantifies the loss of the model parameterization $\mathbf{w}$ on data sample $\{\mathbf{x}_{k,i}, y_{k,i}\}$. One solution of (1) can be obtained by applying iterative methods such as stochastic gradient descent to (2) as follows:

$$\mathbf{w}_{k,i} = \mathbf{w}_{i-1} - \mu \nabla j_k(\mathbf{w}_{i-1}), \tag{3}$$

where μ is the step size, and the server receives the updated parameter $\mathbf{w}_{k,i}$ from devices and sends the following updated global parameter to N devices:

$$\mathbf{w}_i = \frac{1}{N} \sum_{k=1}^{N} \mathbf{w}_{k,i}. \tag{4}$$

The server update (4) and local update (3) are key steps of FedAvg [1]. The data at IoT devices are described by the model:

$$y_{k,i} = \mathbf{w}_{opt}^* \mathbf{x}_{k,i} + v_{k,i},\ k = 1, \ldots, N, \text{and}\, i = 1, \ldots, n_k, \tag{5}$$

where $v_{k,i}$ represents Gaussian noise with zero mean and variance $\sigma_{v_k}^2$ at each device k that is uncorrelated with $\mathbf{x}_{k,i}$. We consider the mean square error (MSE) as the local objective function (2) to estimate $\mathbf{w}_i$ as defined by:

$$\begin{aligned} j_k(\mathbf{w}_{i-1}) &= \mathbb{E}\Big[\big\|e_{k,i}\big\|^2\Big] \triangleq \mathbb{E}\Big[\big\|y_{k,i} - d_{k,i}\big\|^2\Big] \\ &= \mathbb{E}\Big[\big\|y_{k,i} - \mathbf{w}_{i-1}^* \mathbf{x}_{k,i}\big\|^2\Big], \end{aligned} \tag{6}$$

where $d_{k,i} = \mathbf{w}_{i-1}^* \mathbf{x}_{k,i}$ is the desired estimation of $y_{k,i}$ in (5) and $e_{k,i} = y_{k,i} - \mathbf{w}_{i-1}^* \mathbf{x}_{k,i}$ is the estimation error. The gradient of (6) with respect to $\mathbf{w}_{i-1}^*$ is $\nabla j_k(\mathbf{w}_{i-1}) = -\mathbf{x}_{k,i} e_{k,i}^*$. Replacing it into (3), we arrive at the distinctive adaptive algorithm, Least Mean Square (LMS) at device k as follows:

$$\mathbf{w}_{k,i} = \mathbf{w}_{i-1} + \mu \mathbf{x}_{k,i} (y_{k,i} - \mathbf{w}_{i-1}^* \mathbf{x}_{k,i})^*. \tag{7}$$

Combining (7) with (4) results in the Federated Averaging-LMS (FedAvg-LMS) algorithm that has been used in adaptive federated learning tasks [14], [15].

As shown in Fig. 1, each IoT device uses sensors and hence the measurement data are analog and should be converted to digital data for processing. The average ADC resolution for different types of sensors in IoT devices varies from 5 to 16 bits (see Table. 1 in [6]). Let $\{\mathbf{x}_k(t), y_k(t)\}$ denote the analog measurement data that are converted to digital samples $\{\mathbf{x}_{k,i}, y_{k,i}\}$ with high resolution ADCs. Note that we denote the continuous time (analog) data with (t) whereas we use subscript i for discrete-time data. One concern is that the cost and power consumption of ADCs increase exponentially with the number of quantization bits [7] for each device. This motivates us to quantize the measurement data with few bits to support the low-cost and low-power consumption features of IoT sensors.

A. Signal Decomposition with Coarse Quantization

Let $\{\widehat{\mathbf{x}}_{k,i}, \widehat{y}_{k,i}\}$ denote the coarsely quantized version of high precision data samples, i.e., $\widehat{\mathbf{x}}_{k,i} = Q_b(\mathbf{x}_{k,i})$ and $\widehat{y}_{k,i} = Q_b(y_{k,i})$, where Q_b is the b-bit quantization operation. Since the quantization results in a biased estimation of model parameter $\mathbf{w}$, we propose a bias-compensation method for Gaussian measurement data using the Bussgang decomposition theorem [16]. We use the following assumption that is very common in parameter estimation [17]–[19] and adaptive signal processing [20].

Assumption 1: The input data regressors $\mathbf{x}_{k,i}$ are zero-mean with covariance matrices $\mathbf{R}_{x_k} = \mathbb{E}[\mathbf{x}_{k,i}\mathbf{x}_{k,i}^*]$ and temporally independent. This assumption also applies to the additive noise sequences $v_{k,i}$ with variance $\sigma_{v_k}^2$ and the quantized regressors $\widehat{\mathbf{x}}_{k,i}$ with covariance matrices $\widehat{\mathbf{R}}_{x_k} = \mathbb{E}[\widehat{\mathbf{x}}_{k,i}\widehat{\mathbf{x}}_{k,i}^*]$. Moreover, covariance matrices are time-invariant and all data are assumed spatially independent.

Let $\widehat{\mathbf{x}}_k = Q_b(\mathbf{x}_k)$ denote the b-bit quantized output of an ADC at device k, described by a set of $2^b + 1$ thresholds $\mathcal{T}_b = \{\tau_0, \tau_1, ..., \tau_{2^b}\}$, such that $-\infty = \tau_0 < \tau_1 < ... < \tau_{2^b} = \infty$, and the set of 2^b labels $\mathcal{L}_b = \{l_0, l_1, ..., l_{2^b-1}\}$ where $l_p \in (\tau_p, \tau_{p+1}]$, for $p \in [0, 2^b - 1]$. Let us assume that $\mathbf{x}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_{x_k})$, where $\mathbf{R}_{x_k} \in \mathbb{R}^{M\times M}$ is the covariance matrix of $\mathbf{x}_k$. We now use Bussgang's theorem [16] to derive a model for the quantized vector $\widehat{\mathbf{x}}_k$, which we will use later to derive our QA-FedAvg-LMS algorithm. Employing Bussgang's theorem, $\widehat{\mathbf{x}}_k$ can be decomposed as

$$\widehat{\mathbf{x}}_k = \mathbf{G}_{x_k} \mathbf{x}_k + \mathbf{q}_{x_k}, \tag{8}$$

where the distortion $\mathbf{q}_{x_k}$ is uncorrelated with $\mathbf{x}_k$, and $\mathbf{G}_{x_k} \in \mathbb{R}^{M\times M}$ is a diagonal matrix described by

$$\begin{aligned} \mathbf{G}_{x_k} = \operatorname{diag}(\mathbf{R}_{x_k})^{-\frac{1}{2}} \sum_{j=0}^{2^b-1} \frac{l_j}{\sqrt{\pi}} \big[&\exp(-\tau_j^2 \operatorname{diag}(\mathbf{R}_{x_k})^{-1}) \\ &- \exp(-\tau_{j+1}^2 \operatorname{diag}(\mathbf{R}_{x_k})^{-1})\big]. \end{aligned} \tag{9}$$

For the particular case that $\mathbf{R}_{x_k} = \sigma_{x_k}^2 \mathbf{I}_M$, we have $\mathbf{G}_{x_k} = g_{x_k} \mathbf{I}_M$, where

$$g_{x_k} = \frac{1}{\sqrt{\sigma_{x_k}^2}} \sum_{j=0}^{2^b-1} \frac{l_j}{\sqrt{\pi}} \big(e^{-\frac{\tau_j^2}{\sigma_{x_k}^2}} - e^{-\frac{\tau_{j+1}^2}{\sigma_{x_k}^2}}\big). \tag{10}$$

Note that this signal decomposition is also applied to the quantized output response, $\widehat{y}_k$.

B. Design of ADCs

We assume that the analog data samples arriving at the ADCs have been adjusted (e.g., by using automatic gain control (AGC)) to have approximately unit power and design the ADCs for the unit variance data samples. We show in the numerical results that even imperfect gain control at the analog inputs

will not degrade the performance of the proposed algorithm. To minimize the mean square error (MSE) between $\mathbf{x}_k$ and $\widehat{\mathbf{x}}_k$, we need to characterize the probability density function (PDF) of $\mathbf{x}_k$ to find the optimal quantization labels. Because choosing these labels based on such PDF is ineffective in practice, we assume the data input $\mathbf{x}_{k,i}$ is Gaussian, and compute the thresholds and labels as follows:

1) We generate an auxiliary Gaussian random variable x_{aux} with unit variance and then use the Lloyd-Max algorithm [21], [22] to find a set of thresholds $\widetilde{\mathcal{T}}_b = \{\tau_1, \ldots, \tau_{2^b-1}\}$ and labels $\mathcal{L}_b = \{l_0, \ldots, l_{2^b-1}\}$ that minimize the MSE between the unquantized and the quantized data samples.
2) We wrap up the set of thresholds $\mathcal{T}_b$ by adding $\tau_0 = -\infty$ and $\tau_{2^b} = \infty$ to $\widetilde{\mathcal{T}}_b$.
3) We quantize x_{aux} using $\mathcal{T}_b$ and $\mathcal{L}_b$, generate the quantized signal $\widehat{x}_{\text{aux}}$, and estimate the variance of the distortion, $\sigma^2_{q_k}$ with the subtraction of the variance of the quantized auxiliary signal from the variance of the auxiliary signal as follows:
$$\overline{\sigma}^2_{q_k} = \sigma^2_{x_{\text{aux}}} - \sigma^2_{\widehat{x}_{\text{aux}}}. \tag{11}$$

Note that step 1 designs the ADC thresholds and labels, step 2 completes the thresholds needed for (9), and steps 3 is useful to estimate $\mathbf{R}_{x_k}$ later in (27).

III. Proposed QA-FedAvg-LMS Algorithm

Let $\beta_{k,i}$ be a bias compensation coefficient to be chosen, define the desired estimation $d_{k,i} = \beta_{k,i}\mathbf{w}^*_{i-1}\widehat{\mathbf{x}}_{k,i}$ and construct an MSE cost function as described by

$$\begin{aligned} j_k(\mathbf{w}_{k,i-1}) &= \mathbb{E}[\|\widehat{y}_{k,i} - d_{k,i}\|^2] \\ &= \mathbb{E}[\|\widehat{y}_{k,i} - \beta_{k,i}\mathbf{w}^*_{i-1}\widehat{\mathbf{x}}_{k,i}\|^2], \end{aligned} \tag{12}$$

which is defined based on the quantized data samples $\widehat{y}_{k,i}$ and $\widehat{\mathbf{x}}_{k,i}$, and a bias correction term $\beta_{k,i}$. The gradient of (12) with respect to $\mathbf{w}^*_{k,i-1}$ is given by:

$$\nabla j_k(\mathbf{w}_{i-1}) = -\beta_{k,i}\widehat{\mathbf{x}}_{k,i}(\widehat{y}_{k,i} - \beta_{k,i}\mathbf{w}^*_{i-1}\widehat{\mathbf{x}}_{k,i})^*. \tag{13}$$

Replacing (13) into (3) and using (4), we obtain the QA-FedAvg-LMS algorithm as follows:

$$\mathbf{w}_{k,i} = \mathbf{w}_{i-1} + \mu\beta_{k,i}\widehat{\mathbf{x}}_{k,i}(\widehat{y}_{k,i} - \beta_{k,i}\mathbf{w}^*_{i-1}\widehat{\mathbf{x}}_{k,i})^* \tag{14a}$$

$$\mathbf{w}_i = \frac{1}{N}\sum_{k=1}^{N}\mathbf{w}_{k,i}. \tag{14b}$$

We call (14a) and (14b) the adaptation and averaging steps which are performed on the devices and the server, respectively. The next section shows how the bias compensation $\beta_{k,i}$ should be chosen such that (14) is asymptotically unbiased in the mean.

A. Convergence Analysis

To analyze the performance of QA-FedAvg-LMS, we use the weight-error vector defined as:

$$\widetilde{\mathbf{w}}_{k,i} = \mathbf{w}_{opt} - \mathbf{w}_{k,i}, \quad \text{and} \quad \widetilde{\mathbf{w}}_i = \mathbf{w}_{opt} - \mathbf{w}_i. \tag{15}$$

Under Assumption 1, if the entries in the input regressors $\mathbf{x}_{k,i}$ are uncorrelated and with equal variance, we have $\mathbf{R}_{x_k} = \mathbb{E}[\mathbf{x}_{k,i}\mathbf{x}^*_{k,i}] \approx \sigma^2_{x_k}\mathbf{I}_M$ and the matrix $\mathbf{G}_{x_k}$ reduces to $g_{x_k}\mathbf{I}_M$. Let us denote $\mathbf{R}_{q_k} = E[\mathbf{q}_{x_k,i}\mathbf{q}^*_{x_k,i}] \approx \overline{\sigma}^2_{q_k}\mathbf{I}_M$ where $\overline{\sigma}^2_{q_k}$ is given by (11).

Using (8), we can decompose $\widehat{\mathbf{x}}_{k,i}$ and $\widehat{y}_{k,i}$ as follows:

$$\widehat{\mathbf{x}}_{k,i} = g_{x_k}\mathbf{x}_{k,i} + \mathbf{q}_{x_k,i}, \tag{16}$$

$$\widehat{y}_{k,i} = g_{d_k}y_{k,i} + q_{d_k,i} = g_{d_k}\mathbf{w}^*_{opt}\mathbf{x}_{k,i} + p_{k,i}, \tag{17}$$

where $p_{k,i} = g_{d_k}v_{k,i} + q_{d_k,i}$. Using this decomposition, we write the error $e_{k,i}$ as follows:

$$\begin{aligned} e_{k,i} &= \widehat{y}_{k,i} - \beta_{k,i}\mathbf{w}^*_{i-1}\widehat{\mathbf{x}}_{k,i} \\ &= g_{d_k}\mathbf{w}^*_{opt}\mathbf{x}_{k,i} + p_{k,i} - \beta_{k,i}\mathbf{w}^*_{i-1}\big(g_{x_k}\mathbf{x}_{k,i} + \mathbf{q}_{x_k,i}\big). \end{aligned} \tag{18}$$

Replacing (18) into (14a) and subtracting from $\mathbf{w}_{opt}$ yields

$$\begin{aligned} \widetilde{\mathbf{w}}_{k,i} &= \widetilde{\mathbf{w}}_{i-1} - \mu\widehat{\mathbf{x}}_{k,i}e^*_{k,i} \\ &= \widetilde{\mathbf{w}}_{i-1} - \mu(g_{x_k}\mathbf{x}_{k,i} + \mathbf{q}_{x_k,i})e^*_{k,i} \\ &= \widetilde{\mathbf{w}}_{i-1} - \mu\Big(g_{x_k}g_{d_k}\mathbf{x}_{k,i}\mathbf{x}^*_{k,i}\mathbf{w}_{opt} + g_{x_k}\mathbf{x}_{k,i}p_{k,i} \\ &\quad - g^2_{x_k}\beta_{k,i}\mathbf{x}_{k,i}\mathbf{x}^*_{k,i}\mathbf{w}_{i-1} - g_{x_k}\beta_{k,i}\mathbf{x}_{k,i}\mathbf{q}^*_{x_k,i}\mathbf{w}_{i-1} \\ &\quad + g_{d_k}\mathbf{q}_{x_k,i}\mathbf{x}^*_{k,i}\mathbf{w}_{opt} + \mathbf{q}_{x_k,i}p_{k,i} \\ &\quad - g_{x_k}\beta_{k,i}\mathbf{q}_{x_k,i}\mathbf{x}^*_{k,i}\mathbf{w}_{i-1} - \beta_{k,i}\mathbf{q}_{x_k,i}\mathbf{q}^*_{x_k,i}\mathbf{w}_{i-1}\Big) \\ &= \widetilde{\mathbf{w}}_{i-1} - \mu\Big(\big(g_{x_k}g_{d_k}\mathbf{x}_{k,i}\mathbf{x}^*_{k,i} + g_{d_k}\mathbf{q}_{x_k,i}\mathbf{x}^*_{k,i}\big)\mathbf{w}_{opt} \\ &\quad - \big(g^2_{x_k}\beta_{k,i}\mathbf{x}_{k,i}\mathbf{x}^*_{k,i} + g_{x_k}\beta_{k,i}\mathbf{q}_{x_k,i}\mathbf{x}^*_{k,i} \\ &\qquad + g_{x_k}\beta_{k,i}\mathbf{x}_{k,i}\mathbf{q}^*_{x_k,i} + \beta_{k,i}\mathbf{q}_{x_k,i}\mathbf{q}^*_{x_k,i}\big)\mathbf{w}_{i-1} \\ &\quad + g_{x_k}\mathbf{x}_{k,i}p_{k,i} + \mathbf{q}_{x_k,i}p_{k,i}\Big). \end{aligned} \tag{19}$$

We take the expectation from both sides of (19). Since $\mathbf{x}_{k,i}$, $\mathbf{q}_{x_k,i}$, and $p_{k,i}$ are uncorrelated pairwise, the expectations of these cross terms vanish. Considering this, we obtain

$$\begin{aligned} \mathbb{E}[\widetilde{\mathbf{w}}_{k,i}] &= \mathbb{E}[\widetilde{\mathbf{w}}_{i-1}] - \mu\Big(\mathbb{E}\big[g_{x_k}g_{d_k}\mathbf{x}_{k,i}\mathbf{x}^*_{k,i}\big]\mathbf{w}_{opt} \\ &\quad - \mathbb{E}\big[\big(g^2_{x_k}\beta_{k,i}\mathbf{x}_{k,i}\mathbf{x}^*_{k,i} + \beta_{k,i}\mathbf{q}_{x_k,i}\mathbf{q}^*_{x_k,i}\big)\mathbf{w}_{i-1}\big]\Big) \\ &= \mathbb{E}[\widetilde{\mathbf{w}}_{i-1}] - \mu\Big(g_{x_k}g_{d_k}\mathbf{R}_{x_k}\mathbf{w}_{opt} \\ &\quad - \big(g^2_{x_k}\beta_{k,i}\mathbf{R}_{x_k} + \beta_{k,i}\mathbf{R}_{q_k}\big)\mathbb{E}\big[\mathbf{w}_{i-1}\big]\Big). \end{aligned} \tag{20}$$

In the last line of (20), we use a common assumption that states that $\mathbf{x}_{k,i}$ varies slowly in relation to $\widetilde{\mathbf{w}}_{i-1}$ [20]. Thus, when they appear inside the expectations we decouple their expected values. This also applies to $\mathbf{q}_{x_k,i}$ in relation to $\widetilde{\mathbf{w}}_{i-1}$.

We show next that a necessary but not sufficient condition to have an asymptotically unbiased solution in the mean is that

$$g_{x_k}g_{d_k}\mathbf{R}_{x_k} = g^2_{x_k}\beta_{k,i}\mathbf{R}_{x_k} + \beta_{k,i}\mathbf{R}_{q_k}, \tag{21}$$

and we show in the next section that this condition is possible by appropriately choosing $\beta_{k,i}$. Assuming (21) and using (15), we can write (20) as follows:

$$\mathbb{E}[\widetilde{\mathbf{w}}_{k,i}] = \big(\mathbf{I}_M - \mu g_{x_k}g_{d_k}\mathbf{R}_{x_k}\big)\mathbb{E}[\widetilde{\mathbf{w}}_{i-1}]. \tag{22}$$

Subtracting $\mathbf{w}_{opt}$ from both sides of (14b) we observe that adding (22) results in the recursion

$$\mathbb{E}[\widetilde{\mathbf{w}}_i] = \Big(\mathbf{I}_M - \frac{\mu}{N}\sum_{k=1}^{N}g_{x_k}g_{d_k}\mathbf{R}_{x_k}\Big)\mathbb{E}[\widetilde{\mathbf{w}}_{i-1}]. \tag{23}$$

Define $\mathbf{R}_x = \frac{1}{N}\sum_{k=1}^{N} g_{x_k} g_{d_k} \mathbf{R}_{x_k}$. To ensure stability of the recursion (23), we must have $|\mathrm{eig}(\mathbf{I}_M - \mu\mathbf{R}_x)| < 1$. Using the eigenvalue decomposition $\mathbf{R}_x = \mathbf{\Phi}_x \mathbf{\Lambda}_x \mathbf{\Phi}_x^*$, where $\mathbf{\Lambda}_x$ is an $M \times M$ diagonal matrix consisting of the eigenvalues of $\mathbf{R}_x$, and the matrix $\mathbf{\Phi}_x$ is an $M \times M$ square matrix whose columns are the eigenvectors of $\mathbf{R}_x$ associated with these eigenvalues, the condition on the step size can be written as $\left\|\mathbf{I}_M - \mu\mathbf{\Lambda}_x\right\|_\infty < 1$. Therefore, the stability condition for QA-FedAvg-LMS is given by:

$$0 < \mu < \frac{2}{\lambda_{\max}(\mathbf{R}_x)}, \tag{24}$$

where $\lambda_{\max}$ is the largest eigenvalue of $\mathbf{R}_x$.

B. Bias Compensation

From (21), we must have

$$\beta_{k,i}\mathbf{I}_M = g_{x_k} g_{d_k} \mathbf{R}_{x_k} \left(g_{x_k}^2 \mathbf{R}_{x_k} + \mathbf{R}_{q_k}\right)^{-1}. \tag{25}$$

Therefore, the bias compensation term is expressed by

$$\beta_{k,i} = \frac{g_{x_k} g_{d_k} \sigma_{x_k}^2}{g_{x_k}^2 \sigma_{x_k}^2 + \overline{\sigma}_{q_k}^2}. \tag{26}$$

Remark 1: *(One ADC for each sensor).* To reduce the cost and energy consumption of sensors, we consider one ADC to quantize the measurement data $\{\mathbf{x}_{k,i}, y_{k,i}\}$. Then g_{x_k} and g_{d_k} can be considered equal and this reduces the complexity of our algorithm as well.

Remark 2: *(Approximation of data variance).* Since the devices receive quantized data and have access to the covariance of the quantized data, $\mathbf{R}_{\widehat{x}_k} = \mathbb{E}[\widehat{\mathbf{x}}_{k,i}\widehat{\mathbf{x}}_{k,i}^*] \approx \sigma_{\widehat{x}_k}^2 \mathbf{I}_M$, we approximate the variance of high precision data as follows:

$$\overline{\sigma}_{x_k}^2 = \overline{\sigma}_{\widehat{x}_k}^2 + \overline{\sigma}_{q_k}^2, \tag{27}$$

where $\overline{\sigma}_{\widehat{x}_k}^2 = \frac{1}{M}\sum_{l=1}^{M}(\widehat{x}_k(l) - \mathrm{mean}\{\widehat{x}_k\})^2$, and $\widehat{x}_k(l)$ are the lth entry of vector $\widehat{x}_k$. Therefore, at each data sample i, the bias correction term is given by:

$$\beta_{k,i} = \frac{g_{x_k}^2 \overline{\sigma}_{x_k}^2}{g_{x_k}^2 \overline{\sigma}_{x_k}^2 + \overline{\sigma}_{q_k}^2}. \tag{28}$$

The QA-FedAvg-LMS algorithm is summarized in table I.

TABLE I
PSEUDO CODE FOR THE QA-FEDAVG-LMS ALGORITHM

Initialization: $\mathbf{w}_0 = 0$

Design ADC with $\mathcal{T}_b$ and $\mathcal{L}_b$ and **Compute** $\overline{\sigma}_{q_k}^2$ from (11)

At each time instant i

At each device k

For a quantized data sample: $\{\widehat{\mathbf{x}}_{k,i}, \widehat{y}_{k,i}\}$ **do**

$\overline{\sigma}_{x_k}^2 = \overline{\sigma}_{\widehat{x}_k}^2 + \overline{\sigma}_{q_k}^2$

$g_{x_k} = \frac{1}{\sqrt{\overline{\sigma}_{x_k}^2}} \sum_{j=0}^{2^b-1} \frac{l_j}{\sqrt{\pi}} \left(e^{-\frac{\tau_j^2}{\overline{\sigma}_{x_k}^2}} - e^{-\frac{\tau_{j+1}^2}{\overline{\sigma}_{x_k}^2}} \right)$

$\beta_{k,i} = \frac{g_{x_k}^2 \overline{\sigma}_{x_k}^2}{g_{x_k}^2 \overline{\sigma}_{x_k}^2 + \overline{\sigma}_{q_k}^2}$

$\mathbf{w}_{k,i} = \mathbf{w}_{i-1} + \mu\beta_{k,i}\widehat{\mathbf{x}}_{k,i}(\widehat{y}_{k,i} - \beta_{k,i}\mathbf{w}_{i-1}^*\widehat{\mathbf{x}}_{k,i})^*$

At server: Receive $\mathbf{w}_{k,i}$ from IoT devices

Send $\mathbf{w}_i = \frac{1}{N}\sum_{k=1}^{N} \mathbf{w}_{k,i}$ **to IoT devices**

IV. Simulation Results

In this section, we assess the performance of the QA-FedAvg-LMS algorithm for a parameter estimation problem in an IoT network with $N = 100$ devices. The unknown parameter vector has a length of $M = 32$, is generated randomly and normalized to unit norm. We generated 10^5 $M \times 1$ vectors with multivariate Gaussian distribution as the input data samples $\mathbf{x}_{k,i}$ for 100 devices (1000 data samples for each device) with the covariance matrix $\mathbf{R}_{x_k} = \sigma_{x_k}^2 \mathbf{I}_M$ where $\sigma_{x_k}^2 \in (0.5, 1)$. The noise samples of each device are drawn from a zero mean white Gaussian process with variance $\sigma_{v_k}^2 \in (0.01, 0.05)$. The data samples are quantized with $\widetilde{\mathcal{T}}_b$ and $\mathcal{L}_b$ to generate $\widehat{\mathbf{x}}_{k,i}$ and $\widehat{y}_{k,i}$. We choose $\mu = 0.05$ as the step size of QA-FedAvg-LMS and FedAvg-LMS.

We use the mean-square deviation (MSD) to investigate the performance of the network and use the excess mean square error (EMSE) to compare the performance of each device k as given by:

$$\begin{aligned} \mathrm{MSD} &\triangleq \lim_{i \to +\infty} \mathbb{E}[\|\mathbf{w}_{opt} - \mathbf{w}_i\|^2], \\ \mathrm{EMSE}_k &\triangleq \lim_{i \to +\infty} \mathbb{E}[\|(\mathbf{w}_{opt} - \mathbf{w}_{k,i})^* \mathbf{x}_{k,i}\|^2]. \end{aligned} \tag{29}$$

The simulated learning curves are obtained by ensemble averaging over 200 independent trials and the steady-state values are averaged over the last 10% data samples. We have compared QA-FedAvg-LMS (14) with FedAvg-LMS (7) with generated data quantized with different numbers of bits. Full resolution FedAvg-LMS refers to the case where the data $\{\mathbf{x}_{k,i}, y_{k,i}\}$ is not quantized.

Fig. 2 shows the evaluation of the global MSD (29) for 1000 communication rounds between server and devices. Fig. 3 compares the steady-state MSD values for the different signal-to-noise ratios (SNR) (keeping $\sigma_{x_k}^2 \in (0.5, 1)$ and changing $\sigma_{v_k}^2$) where the SNR value is averaged over devices. Fig. 4 compares the steady-state EMSE (29) performance of 10 randomly chosen devices. As it can be seen in the numerical results, the network MSD and device-wise EMSE performance of the proposed QA-FedAvg-LMS algorithm are closer to the full resolution FedAvg-LMS while it substantially reduces the power consumption related to the ADCs in the input sensors.

V. Conclusion

In this paper, we have proposed an energy-efficient framework for federated learning and developed the QA-FedAvg-LMS algorithm along with bias compensation strategies for IoT networks. The QA-FedAvg-LMS algorithm has comparable computational complexity to the standard FedAvg-LMS algorithm while it substantially reduces the power consumption of the ADCs in the network. Simulations have shown excellent performance of QA-FedAvg-LMS as compared to FedAvg-LMS for coarsely quantized signals.

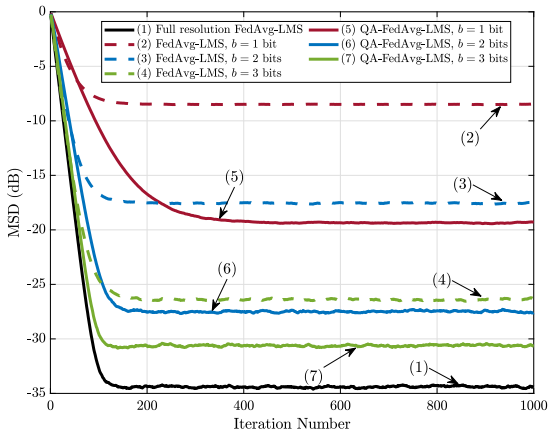

Fig. 2. MSD curves for the FedAvg-LMS (7) and QA-FedAvg-LMS (14) algorithms.

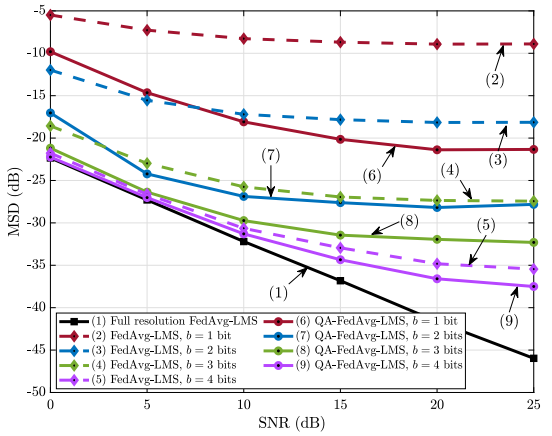

Fig. 3. Steady-state MSD versus SNR for the FedAvg-LMS (7) and QA-FedAvg-LMS (14) algorithms.

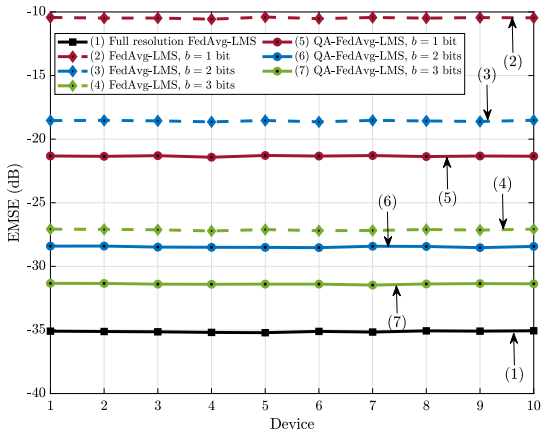

Fig. 4. Steady-state EMSE curves for the FedAvg-LMS (7) and QA-FedAvg-LMS (14) algorithms.

REFERENCES

[1] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y. Arcas, "Communication-efficient learning of deep networks from decentralized data," in *Artificial intelligence and statistics*. PMLR, 2017, pp. 1273–1282.

[2] J. Konečnỳ, H. B. McMahan, F. X. Yu, P. Richtárik, A. T. Suresh, and D. Bacon, "Federated learning: Strategies for improving communication efficiency," *arXiv preprint arXiv:1610.05492*, 2016.

[3] T. Li, A. K. Sahu, A. Talwalkar, and V. Smith, "Federated learning: Challenges, methods, and future directions," *IEEE Signal Processing Magazine*, vol. 37, no. 3, pp. 50–60, 2020.

[4] L. U. Khan, W. Saad, Z. Han, E. Hossain, and C. S. Hong, "Federated learning for internet of things: Recent advances, taxonomy, and open challenges," *IEEE Communications Surveys & Tutorials*, 2021.

[5] D. C. Nguyen, M. Ding, P. N. Pathirana, A. Seneviratne, J. Li, and H. V. Poor, "Federated learning for internet of things: A comprehensive survey," *IEEE Communications Surveys & Tutorials*, 2021.

[6] M. Alioto and M. Shahghasemi, "The internet of things on its edge: Trends toward its tipping point," *IEEE Consumer Electronics Magazine*, vol. 7, no. 1, pp. 77–87, 2017.

[7] R. H. Walden, "Analog-to-digital converter survey and analysis," *IEEE Journal on selected areas in communications*, vol. 17, no. 4, pp. 539–550, 1999.

[8] S. Jacobsson, G. Durisi, M. Coldrey, U. Gustavsson, and C. Studer, "Throughput analysis of massive MIMO uplink with low-resolution ADCs," *IEEE Transactions on Wireless Communications*, vol. 16, no. 6, pp. 4038–4051, 2017.

[9] A. Mezghani, M.-S. Khoufi, and J. A. Nossek, "A modified MMSE receiver for quantized MIMO systems," *Proc. ITG/IEEE WSA, Vienna, Austria*, pp. 1–5, 2007.

[10] Z. Shao, L. Landau, and R. C. de Lamare, "Adaptive RLS channel estimation and sic for large-scale antenna systems with 1-bit adcs," in *WSA 2018; 22nd International ITG Workshop on Smart Antennas*. VDE, 2018, pp. 1–4.

[11] J. Park, S. Samarakoon, A. Elgabli, J. Kim, M. Bennis, S. Kim, and M. Debbah, "Communication-efficient and distributed learning over wireless networks: Principles and applications," *Proceedings of the IEEE*, vol. 109, no. 5, pp. 796–819, 2021.

[12] A. Danaee, R. C. de Lamare, and V. H. Nascimento, "Energy-efficient distributed learning with coarsely quantized signals," *IEEE Signal Processing Letters*, vol. 28, pp. 329–333, 2021.

[13] A. Danaee, R. C. de Lamare, and V. H. Nascimento, "Energy-efficient distributed learning with adaptive bias compensation for coarsely quantized signals," in *2021 IEEE Statistical Signal Processing Workshop (SSP)*. IEEE, 2021, pp. 61–65.

[14] P. Di Lorenzo, C. Battiloro, M. Merluzzi, and S. Barbarossa, "Dynamic resource optimization for adaptive federated learning at the wireless network edge," in *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2021, pp. 4910–4914.

[15] V. C. Gogineni, S. Werner, Y. Huang, and A. Kuh, "Communication-efficient online federated learning framework for nonlinear regression," *arXiv preprint arXiv:2110.06556*, 2021.

[16] J. J. Bussgang, "Crosscorrelation functions of amplitude-distorted gaussian signals," Tech. Rep. 216, Research Laboratory of Electronics, Massachusetts Institute of Technology, 1952.

[17] P. Koukoulas and N. Kalouptsidis, "Nonlinear system identification using gaussian inputs," *IEEE Transactions on signal Processing*, vol. 43, no. 8, pp. 1831–1841, 1995.

[18] T. Koh and E. Powers, "Second-order volterra filtering and its application to nonlinear system identification," *IEEE Transactions on acoustics, speech, and signal processing*, vol. 33, no. 6, pp. 1445–1455, 1985.

[19] R. van der Merwe and E. A. Wan, "The square-root unscented kalman filter for state and parameter-estimation," in *2001 IEEE international conference on acoustics, speech, and signal processing. Proceedings (Cat. No. 01CH37221)*. IEEE, 2001, vol. 6, pp. 3461–3464.

[20] Ali H Sayed, *Fundamentals of adaptive filtering*, John Wiley & Sons, 2003.

[21] S. Lloyd, "Least squares quantization in PCM," *IEEE transactions on information theory*, vol. 28, no. 2, pp. 129–137, 1982.

[22] J. Max, "Quantizing for minimum distortion," *IRE Transactions on Information Theory*, vol. 6, no. 1, pp. 7–12, 1960.