Bandwidth Expansion for Over-the-Air Computation with One-Sided CSI

Nitish Mital*, and Deniz Gündüz*†
*Department of Electrical & Electronics Engineering, Imperial College London
†'Enzo Ferrari' Department of Engineering, University of Modena and Reggio Emilia, Italy
Email: {n.mital,d.gunduz}@imperial.ac.uk

***Abstract*—We consider a distributed computation problem over a multiple access channel (MAC), with N devices. It is known that *over-the-air computation (OAC)* can provide significant gains for this problem, but existing works are limited to the scenario with matched source and channel bandwidths. We propose OAC schemes for block-fading MACs that modulate the source to fit the available channel bandwidth in a wideband channel, while having channel state information (CSI) only at the transmitter or the receiver. Our results show that the proposed OAC scheme outperforms even ideal capacity-achieving digital schemes when the CSI is available only at the transmitter, and the distortion does not scale with the number of participating devices. We demonstrate the effectiveness of our proposed scheme in federated edge learning (FEEL), where OAC is used to aggregate model updates from the participating devices.**

I. Introduction

In emerging edge intelligence applications, the goal is often the computation of some pre-defined function of data at different devices, such as the arithmetic mean, the maximum or minimum value, or different polynomials, rather than reconstruction of the individual data sequences. This can be formulated as a distributed computation problem over a multiple-access channel (MAC). A standard approach to this problem is to let each device transmit a quantized version of its data samples to the central receiver, where an approximate value of the desired function is computed. However, the sub-optimality of this approach is well-known, and an alternative approach, called *over-the-air computation (OAC)*, can provide significant reduction in both complexity and bandwidth requirement for function computation over wireless networks [1]. When the receiver is interested in the average value of the samples available at the multiple transmitters, OAC can exploit the superposition property of the wireless medium, and the data samples transmitted as analog values over the shared wireless channel get aggregated "over the air". This provides significant reduction in the required bandwidth, and simplifies the transmitter-receiver design. In [2], [3], the authors show that OAC can be exploited for a large class of nomographic functions, that is, real-valued functions that can be represented as post-processed sums of pre-processed real-valued inputs. The authors in [2]–[4] propose OAC schemes over fast fading MACs, which utilize multiple channel uses to obtain asymptotically optimal reconstructions. Recently, OAC received significant attention in the context of federated edge learning (FEEL) [5]–[7], where the goal is to aggregate the model updates from multiple devices with localized datasets. However, these works assume that the channel bandwidth for transmission matches the source bandwidth; that is, the number of data samples available at the transmitters and the available channel uses are the same. Our goal in this paper is to benefit from the available bandwidth to improve the fidelity of the computed function value.

We emphasize that OAC is a joint source-channel coding scheme with the objective to minimize the reconstruction error in the function value. For point-to-point communication of a Gaussian source over a Gaussian channel, it is well-known that uncoded analog transmission achieves the optimal mean squared error (MSE), when the source and channel bandwidths match [8], [9]. When this is not the case, we have to process the source so that it "fits" the channel bandwidth. Linear bandwidth expansion maps do not provide any gain from the extra bandwidth under an average power constraint, as they simply distribute the power over the larger bandwidth. Accordingly, direct linear approaches to source-channel mapping, referred to as block pulse amplitude modulation (BPAM) [10] are highly sub-optimal, and non-linear modulation schemes like pulse-position modulation (PPM), frequency-position modulation (FPM), frequency modulation (FM), phase modulation (PM), and hybrid digital-analog techniques [11], provide better performance. Non-linear mappings will also be needed for efficient OAC when the channel bandwidth is larger than the source bandwidth. However, developing bandwidth expansion maps for multi-user OAC is highly non-trivial precisely due to the linear nature of the computed nomographic functions with respect to the inputs.

In this paper, we propose an OAC scheme that modulates the source to fit the available channel bandwidth in a wideband channel. Unlike existing bandwidth expansion schemes that depend on the ergodicity of the fast-fading channels [2], [3], our scheme works for block fading channels, and exploits the fact that practical applications often communicate vectors of data values over multiple parallel channels, e.g., OFDM or multiple antenna channels. Communication over multiple channels has also been considered in previous works [12], [13]. In [12], each device transmits its measurement over multiple frequency subchannels to benefit from multi-channel

This work was supported by the European Research Council (ERC) Starting Grant BEACON (grant agreement no. 677854), and UK EPSRC (EP/T023600/1) under CHIST-ERA (CHISTERA-18-SDCDN-001).

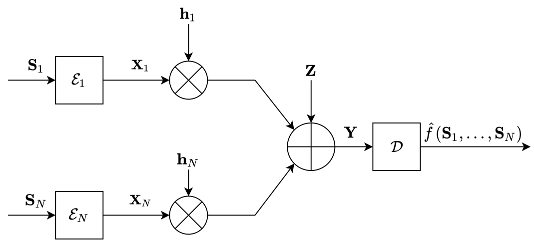

Fig. 1: Communication system model

diversity. In [13], transmission over multiple time slots with fast fading is considered, where each device can choose to transmit in one time slot with high channel gain. We consider two cases: CSI available only at the transmitters (CSI-T), and CSI available only at the receiver (CSI-R). We also evaluate our proposed schemes for FEEL. In FEEL, a central parameter server (PS) sends a global model to all the available devices, which perform one iteration (or one epoch) of training of the received model with their local datasets, and compute the updated local model parameters. Then, they transmit these updated local model parameters to the PS to be aggregated to update the global model. OAC can be used for the aggregation of model updates at the PS, which can significantly reduce the bandwidth requirement and latency of FEEL, which are essential to enable many edge applications.

Our results show that the proposed OAC schemes outperform even ideal capacity achieving digital schemes when the CSI is available at the transmitter, while also providing a simple communication system design. Moreover, the distortion does not scale with the number of participating devices. When the CSI is available only at the receiver, the proposed scheme, though outperformed by both the OAC and digital transmission schemes for the case when CSI is available only at the transmitter, achieves asymptotically perfect function computation. The proposed scheme for CSI at only the receiver scales quadratically with the number of devices.

II. System Model

The communication system, illustrated in Fig. 1, consists of N encoder functions $\mathcal{E}_1, \ldots, \mathcal{E}_N \in \mathcal{S}^p \rightarrow \mathbb{R}^{q \times T}$ corresponding to N devices, $[N] \triangleq \{1, \ldots, N\}$, where q is the number of parallel subchannels, and T is the number of channel uses over which the transmission happens, that map source vectors $\mathbf{S}_n = (S_n(1), \ldots, S_n(p))^T \in \mathcal{S}^p, n \in [N]$, where $\mathcal{S}$ denotes the source alphabet, into channel codewords $\mathbf{X}_1, \ldots, \mathbf{X}_N$, respectively, a decoder function $\mathcal{D}(\cdot)$ at the PS that maps the received signal $\mathbf{Y}$ to estimates of the desired $\mathcal{F}$-valued nomographic functions of the source signals, $f(\cdot) : \mathcal{S}^N \rightarrow \mathcal{F}$. We denote by $\mathbf{F} \in \mathcal{F}^p$ the vector of functions computed on the original values of the source signals, i.e., $F(\rho) = f(S_1(\rho), \ldots, S_N(\rho))$, for $\rho \in [p]$, and by $\hat{\mathbf{F}} \in \mathcal{F}^p$ their reconstruction at the decoder, i.e., $\hat{\mathbf{F}} = \mathcal{D}(\mathbf{Y})$.

Nomographic functions are of the form $\psi(\sum_{n=1}^N \phi_n(S_n))$, where $\phi_n : \mathcal{S} \rightarrow \mathbb{R}$, $\forall n \in [N]$, denotes a pre-processing function at the devices, and $\psi : \mathbb{R} \rightarrow \mathcal{F}$ is a post-processing function at the central receiver.

Channel model: Devices transmit their signals over a fading MAC. The vector of fading coefficients for the link between device n and the PS is denoted by $\mathbf{h}_n \in \mathbb{C}^q$. We assume a block fading model, where the channel remains constant for T channel uses, while the fading coefficients of the parallel channels are assumed to be independent and identically distributed (i.i.d.). Therefore, we assume that $\mathbf{X}_n = (\mathbf{x}_n(1) \cdots \mathbf{x}_n(T)) \in \mathbb{R}^{q \times T}$, where $\mathbf{x}_n(t)$ is a vector of q symbols transmitted over the q parallel channels in the t^{th} channel use. The additive noise vector, denoted by $\mathbf{z} \in \mathbb{C}^q$, is assumed to be i.i.d. according to the complex normal distribution $\mathcal{CN}(0, \sigma_z^2)$. The received signal in the t^{th} use of the ρ^{th} sub-channel, $y(\rho, t)$, is given by:

$$y(\rho, t) = \sum_{n=1}^{N} h_n(\rho) x_n(\rho, t) + z(\rho, t), \tag{1}$$

where $x_n(\rho, t)$ and $z(\rho, t)$ are the transmitted symbol by device n and the noise term, respectively, in the t^{th} use of the ρ^{th} sub-channel.

We consider the following two scenarios:

- *CSI at the transmitter (CSI-T):* The devices have perfect CSI of their respective links to the PS, while the PS (receiver) only has statistical knowledge of the channel. This models the scenario with channel reciprocity, where the PS broadcasts pilot signals to all the devices to estimate their respective channels. This avoids the use of any feedback mechanism by the PS to keep a record of the instantaneous channel states of all the devices.
- *CSI at the receiver (CSI-R):* The CSI is known only to the PS, while the devices do not have any knowledge of their channels.

The power allocation scheme and the channel capacity depend on the available CSI and the average power constraint. The following average power constraint is imposed on each transmitter:

$$\frac{1}{qT} \mathbb{E}\left[||\mathbf{X}_n||_2^2\right] \leq P \qquad \forall n \in [N], \tag{2}$$

where the expectation is over the noise and fading distribution.

Performance metric: The goal is to accurately compute the desired function of the source vectors accurately despite the noise and fading in the channel. Hence, we measure the performance of the proposed schemes by computing the normalized MSE (NMSE):

$$\text{NMSE}\left(\hat{\mathbf{F}}, \mathbf{F}\right) \triangleq \frac{\mathbb{E}\left[||\hat{\mathbf{F}} - \mathbf{F}||_2^2\right]}{||\mathbf{F}||_2^2}, \tag{3}$$

where the expectation is over the noise and fading distribution.

Next, we will consider two alternative transmission schemes, digital transmission with computation at the receiver and analog transmission with OAC, and compare the two in terms of the achieved MSE under different CSI assumptions.

A. Digital Transmission with Receiver Computation

In the digital aggregation scheme, the source vectors are first quantized, and then mapped to a channel codeword. If the length of the source vector is large, it is split into segments of $p \leq q$ values, and each segment is transmitted in a block of q parallel subchannels. The transmission of each segment in independent blocks of channel uses is identical, and without loss of generality, it is sufficient to analyse the transmission in any one particular block. Therefore, we assume that the signal vector to be transmitted by device n in an arbitrary block is $\mathbf{S}_n \triangleq (S_n(1), \ldots, S_n(p))^T$. Encoder n performs vector quantization of $\mathbf{S}_n$ to obtain the quantized vector $\mathbf{v}_n \in \mathbb{Z}^p$, and then maps it to the channel codeword $\mathbf{X}_n \in \mathbb{R}^{q \times T}$.

The decoder first decodes the channel codewords from the transmitters to recover the quantized source signals. In the asymptotic limit of infinite blocklength (i.e., $T \to \infty$, the transmitted codewords can be decoded with a vanishing error probability if the transmission rates lie within the capacity region of the channel. In that case, the only source of error in the computation of the desired function is the error due to quantization. Note that the capacity region only provides an upper bound on the maximum rate at which a source can be transmitted, and is not achievable in practice.

B. OAC

In the proposed OAC scheme, the encoder exploits the fact that the desired function $f(\cdot)$ is a nomographic function. Encoder n pre-processes the source vector $\mathbf{S}_n$ to a vector $\mathbf{s}_n$ of the values $s_n(\rho) = \phi_n(S_n(\rho)), \rho \in [p]$. Encoder n then maps $\mathbf{s}_n \in \mathcal{S}^p$ to the channel codeword $\mathbf{X}_n$, using the encoding function $\mathcal{E}_n : \mathbb{R}^p \to \mathbb{R}^{q \times T}$. The transmitted signals are added over the channel due to the superposition property of the wireless medium. The decoder computes the estimate $\hat{\mathbf{F}} = \mathcal{D}_\rho(\mathbf{Y})$ of the desired function $\mathbf{F}$ using the decoding function $\mathcal{D} : \mathbb{R}^{q \times T} \to \mathcal{F}^p$, by first decoding an estimate $\hat{\mathbf{s}}$ of the sum $\mathbf{s} = \sum_{n=1}^{N} \mathbf{s}_n$, and then computing the functions $\hat{F}(\rho) = \psi(\hat{s}(\rho))$, $\forall \rho \in [p]$.

III. OAC over Fading MAC

A. With CSI-T

When the CSI is available at the transmitters, they can employ a channel inversion scheme similar to the one proposed in [2]. However, unlike [2], we impose an average power constraint on each device, and therefore, each device transmits a symbol in the ρ^{th} subchannel only if $\gamma(\rho) > \gamma_0$, where $\gamma(\rho) \triangleq |h(\rho)|^2$, and γ_0 is a fixed threshold such that the average power constraint is satisfied. We design a matrix $\mathbf{\Pi} \in \mathbb{Z}^{q \times T}$, whose columns consist of permutations of the integers in the set $\{1, \ldots, q\}$. Matrix $\mathbf{\Pi}$ is constructed such that no row contains repetitions of any integer. Note that this is possible only if $T \leq q$.

Device n first pre-processes the source vector $\mathbf{S}_n$ to obtain the vector $\mathbf{s}_n$. The vector $\mathbf{s}_n$ is then padded with $q-p$ zeros. In

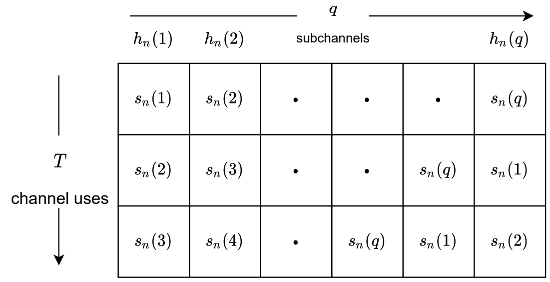

Fig. 2: Transmission of the parameter vector $\mathbf{s}_n$ by device n over q parallel subchannels and T channel uses, using a circular permutation matrix.

the t^{th} channel use and the ρ^{th} sub-channel, device n transmits the symbol:

$$x_n(\rho, t) = \mathbb{1}_{\gamma_n(\rho) > \gamma_0} \frac{\sqrt{\alpha} h_n^*(\rho)}{|h_n(\rho)|^2} s_n(\pi(\rho, t)), \qquad (4)$$

where $\pi(\rho, t)$ is the element in the ρ^{th} row and t^{th} column of matrix $\mathbf{\Pi}$. In other words, each element $s_n(\rho), \rho \in [p]$, is transmitted in a different sub-channel in each of the T channel uses. This is illustrated in Fig. 2. By substituting Eq. (4) in Eq. (1), the received signal at the PS is given by:

$$y(\rho, t) = \sum_{n=1}^{N} \sqrt{\alpha} \mathbb{1}_{\gamma_n(\rho) > \gamma_0} s_n(\pi(\rho, t)) + z(\rho, t). \qquad (5)$$

To obtain the sum, the sub-channel permutations of $\mathbf{\Pi}$ are reversed by the PS to align the elements of the received signal in each time slot. As a result, each time t now effectively has an independent channel coefficient associated with it in each sub-channel ρ, denoted by $h'(\rho, t)$. Thus we have

$$y'(\rho, t) = \sum_{n=1}^{N} \sqrt{\alpha} \mathbb{1}_{\gamma'_n(\rho, t) > \gamma_0} s_n(\rho) + z'(\rho, t), \qquad (6)$$

where $y'(\rho, t), \gamma'_n(\rho, t)$ and $z'(\rho, t)$ are the elements of the corresponding vectors in the ρ^{th} sub-channel and t^{th} time slot after re-alignment. The sum $\mathbf{s}$ is estimated in the following way:

$$\hat{s}(\rho) = \frac{1}{\sqrt{\alpha} T \Pr(\gamma > \gamma_0)} \sum_{t=1}^{T} y'(\rho, t), \qquad (7)$$

which is an unbiased estimate of the sum.

As stated earlier, in the case of a standard MAC, it is difficult to exploit the extra bandwidth for function computation. Here, we exploit the bandwidth to transmit the same signals over different subchannels, which allows the receiver to recover multiple noisy versions of the same sum with different channel gains. The recovered sum will be more accurate as the likelihood of experiencing fading in multiple subchannels is lower. Indeed, in the asymptotic limit of $q, T \to \infty$, the law of large numbers dictates:

$$\hat{s}(\rho) \to \sum_{n=1}^{N} s_n(\rho) \qquad \forall \rho \in [p]. \qquad (8)$$

Finally, the receiver computes the estimates $\hat{F}(\rho) = \psi(\hat{s}(\rho))$, $\forall \rho \in [p]$.

The average power consumption is given by:

$$\frac{1}{qT}\mathbb{E}[||\mathbf{X}_n||_2^2] = \mathbb{E}\left[|x_n(1,1)|^2\right] \tag{9}$$

$$= \mathbb{E}\left[\left(\mathbb{1}_{\gamma_n(1)>\gamma_0}\frac{\sqrt{\alpha}}{|h_n(1)|}s_n(\pi(\rho,t))\right)^2\right] \tag{10}$$

B. With CSI-R

In this section, we describe the scheme for the scenario when the CSI is available perfectly only at the receiver (CSI-R). Since the CSI is not known to the transmitters, devices allocate equal power to all the subchannels by scaling the source signal with a constant α. The value of α is chosen so that the average power constraint is satisfied, and is given by $\alpha = \frac{qP}{||\mathbf{s}_n||_2^2}$. Device n transmits $\mathbf{x}_n(t) = \sqrt{\alpha}\left(s_n(\pi(1,t)) \cdots s_n(\pi(q,t))\right)^T$ in the t^{th} channel use. The signal received by the receiver in the t^{th} channel use and the ρ^{th} sub-channel is given by

$$y(\rho,t) = \sum_{n=1}^{N} \sqrt{\alpha} h_n(\rho) s_n(\pi(\rho,t)) + z(\rho,t). \tag{11}$$

After reversing the permutations of the subchannels, similarly to Section III-A, the PS obtains:

$$y'(\rho,t) = \sum_{n=1}^{N} \sqrt{\alpha} h'_n(\rho,t) s_n(\rho) + z'(\rho,t). \tag{12}$$

The PS then estimates the sum $\mathbf{s}$ as follows:

$$\hat{s}(\rho) = \frac{1}{\sqrt{\alpha}T\mathbb{E}[|h|^2]}\sum_{t=1}^{T}\sum_{n=1}^{N} h'^{*}_n(\rho,t) y'(\rho,t), \tag{13}$$

which, similarly to the CSI-T case, provides T-fold diversity.

IV. Digital transmission with receiver computation

In this section, we describe the baseline digital scheme to evaluate the proposed OAC scheme with CSI-T, where device n uses a uniform lattice quantizer to map the source vector $\mathbf{S}_n$ to the discrete vector $\mathbf{v}_n$.

A. Quantization

When the number of quantization levels is large, the quantization noise is usually modeled as uniformly distributed and independent of the data [14]. Given a $p-$dimensional vector, the quantization noise per dimension of an optimal $p-$dimensional uniform lattice quantizer is given by $\epsilon = G_p V^{2/p}$, where G_p is the normalized second moment, which is a measure of the quantizer efficiency, and V is the volume of a Voronoi cell of the lattice [15]. If we consider a $p-$dimensional cube of volume 1, and if the number of lattice points inside the cube, also called the lattice point density, is 2^Q, then we have $V \approx \frac{1}{2^Q}$. Therefore, we have $\epsilon = G_p 2^{-2Q/p}$. For $p = 1$, we have $G_1 \geq 1/12$, while for $p \to \infty$, the minimum value of $G_p \to 1/2\pi e \approx 0.058550$ [16].

B. Fading MAC with CSI-T

When the CSI is available perfectly only at the transmitters, device n performs channel inversion so that the receiver does not need perfect CSI to correctly decode the transmitted message. Note that to correctly decode the transmitted vectors, the receiver must know whether the fading coefficient is greater or less than γ_0 in any given sub-channel, that is, the receiver requires 1 bit CSI. To satisfy the average power constraint $\frac{1}{qT}\mathbf{E}\left[||\mathbf{X}_n||_2^2\right] \leq P$, device n uses the following power allocation in the ρ^{th} sub-channel and t^{th} channel use:

$$|x_n(\rho,t)|^2 = \mathbb{1}_{\gamma_n(\rho)>\gamma_0}\frac{\alpha}{|h_n(\rho)|^2}, \tag{14}$$

Let the rates at which the devices transmit be denoted by the tuple $(R_1, \ldots, R_N)$. The capacity region of the channel with the above transmission scheme is characterized by the following inequalities:

$$\sum_{a\in\mathcal{A}} R_a \leq \sum_{t=1}^{T}\left(\sum_{\rho=1}^{q}\log\left(1+\sum_{a\in\mathcal{A}}\alpha\mathbb{1}_{\gamma_a(\rho,t)>\gamma_0}\right)\right), \forall\mathcal{A}\subseteq[N], \tag{15}$$

When $q \to \infty$, the inner summation in Eq. (15) is equivalent to the ergodic capacity of a fading MAC with the channel inversion scheme. It is easy to see, from the submodularity of the capacity region, that the operating point with $R_1 = \cdots = R_N = \frac{\sum_{m=1}^{N} R_m}{N}$, lies within the capacity region. Hence, the channel capacity for device n on that operating point is given by $R_n = \frac{\sum_{m=1}^{N} R_m}{N}$. The PS decodes the messages from each device, and then computes the desired function.

V. Experiments

We evaluate the proposed schemes for model aggregation in FEEL, where the desired function is $f(S_1, \ldots, S_N) = \frac{1}{N}\sum_{n=1}^{N} S_n$. We denote the set of data samples available at device n by $\mathcal{D}_n$, and the model update computed by device n on the local data samples $\mathcal{D}_n$, and the parameter vector at the τ^{th} iteration, θ_τ, by $\mathbf{S}_n(\theta_\tau) \in \mathbb{R}^p, n \in [N]$. At each iteration, all the devices deliver their model updates to the PS. The goal of the PS is to update the model parameters as follows:

$$\theta_{\tau+1} = \frac{1}{N}\sum_{n=1}^{N}\mathbf{S}_n(\theta_\tau). \tag{16}$$

A. Reconstruction error analysis

OAC CSI-T: Denoting $\lambda \triangleq \Pr(\gamma > \gamma_0)$ and $\mu \triangleq \sum_{n=1}^{N} s_n(\rho)$, the normalized MSE is given by:

$$\mathrm{NMSE}(\hat{F}(\rho), F(\rho)) = \frac{\mathbb{E}\left[\left(\frac{1}{N}\hat{s}(\rho) - \frac{1}{N}\sum_{n=1}^{N} s_n(\rho)\right)^2\right]}{\left(\frac{1}{N}\sum_{n=1}^{N} s_n(\rho)\right)^2} \tag{17}$$

$$\propto \frac{1-\lambda}{\lambda T} + \frac{\sigma_n^2}{\alpha\mu^2\lambda^2 T}. \tag{18}$$

Note that, if we make the same assumption of infinite parallel channels as in the digital transmission case, the MSE can be

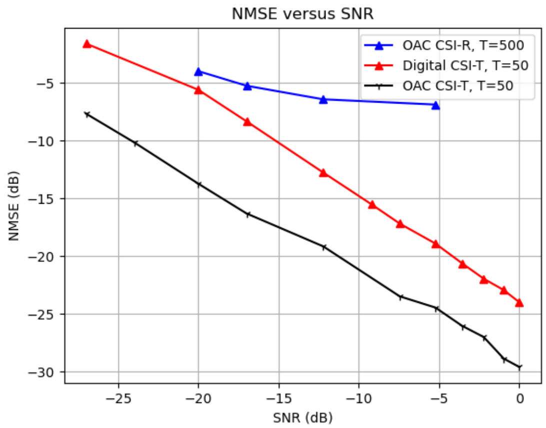

Fig. 3: Comparison of the NMSE achieved by the OAC with CSI-T, OAC with CSI-R, and digital transmission schemes.

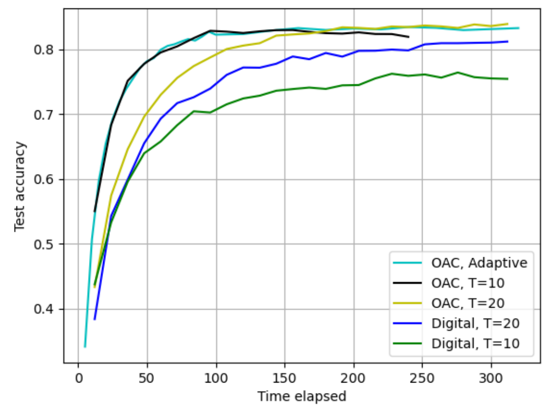

Fig. 4: FEEL: Comparison of testing accuracy of proposed schemes at SNR = -10dB.

made arbitrarily small by increasing the number of channel uses. The scaling parameter α and the threshold γ_0 depend on each other, and are optimized numerically.

OAC CSI-R: Assuming $\sigma_h^2 = \mathbb{E}[|h|^2]$, the normalized MSE is given by:

$$\text{NMSE}\left(\hat{F}(\rho), F(\rho)\right) \propto \frac{1}{T} + \frac{N(N-1)}{T} + \frac{N\sigma_n^2}{\alpha^2 T \sigma_h^2} \quad (19)$$

Notice that the MSE grows quadratically with N, while it is independent of N for the CSI-T case.

B. Synthetic parameter vectors

We first perform experiments on synthetically generated parameter vectors to compute their average. We assume that $N = 100$ devices have parameter vectors of length $p = 5000$, whose elements are sampled from a standard Gaussian distribution. It is also assumed that $q = T$. In Fig. 3, we plot the NMSE versus the SNR for the proposed OAC scheme for CSI-T and CSI-R, and also the digital transmission with receiver computation, assuming CSI-T. It is clear that the OAC CSI-T scheme outperforms both the OAC CSI-R scheme as well as the digital transmission scheme with CSI-T. The superiority of the OAC CSI-T scheme over the OAC CSI-R scheme can be seen by comparing Eq. (18) and (19).

C. FEEL for image classification

To demostrate the effectiveness of the OAC scheme for model aggregation in FEEL, we train a VGG16 model [17] on the CIFAR10 dataset [18] for image classification with 10 classes. We consider a federated setting where $N = 40$ devices with i.i.d. datasets of images from all the classes. At the start, each device is provided with a common initialized model by the PS. Then, each device performs one epoch of training of the local model on its local dataset. After training, the updated local model is delivered to the PS over the wireless channel using the proposed schemes. The PS computes the mean of the received model parameters, and sends them to the devices for the next epoch of training. The CIFAR-10 dataset consists of 60000 32x32 colour images in 10 classes, with 6000 images per class. There are 50000 training images and 10000 test images.

In Fig. 4, we compare the speed of convergence of the model when using the OAC scheme and the digital scheme with CSI-T. We assume that $q = T$. As observed with the experiments on the synthetic parameter vectors, the OAC scheme with CSI-T outperforms the ideal capacity achieving digital scheme. As expected, the test accuracy of the converged model improves when we employ a larger delay T, but it is also interesting to note that the model converges faster in the beginning if a smaller value of delay T is used, indicating that large errors in the aggregated model are tolerable in the initial stages of training, but large errors degrade the performance once model changes become small during the later stages. Taking this into account, we also execute an adaptive delay scheme in which a smaller delay $T = 10$ is used for the first 100 epochs, and then a larger delay $T = 20$ is used for the remaining epochs. We observe that the adaptive delay scheme provides the best convergence time and test accuracy.

VI. Conclusions

It is known that OAC provides significant benefits when trying to recover a function of the signals distributed at multiple transmitters. One challenge in OAC is to exploit all the available channel bandwidth, as it is known that linear bandwidth expansion schemes do not provide any gains. We proposed a scheme that modulates the source to fit the available bandwidth in a wideband channel, and works for block-fading channels. This is achieved by exploiting the channel bandwidth to create diversity for the desired sum. We have shown through numerical experiments, including in a FEEL scenario, that OAC can significantly outperform the digital counterpart even assuming capacity achieving transmission for the latter.

References

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

[2] M. Goldenbaum and S. Stanczak, "Robust analog function computation via wireless multiple-access channels," *IEEE Transactions on Communications*, vol. 61, no. 9, pp. 3863–3877, 2013.

[3] ——, "On the channel estimation effort for analog computation over wireless multiple-access channels," *IEEE Wireless Communications Letters*, vol. 3, no. 3, pp. 261–264, 2014.

[4] J. Dong, Y. Shi, and Z. Ding, "Blind over-the-air computation and data fusion via provable wirtinger flow," *IEEE Transactions on Signal Processing*, vol. 68, pp. 1136–1151, 2020.

[5] M. M. Amiri and D. Gündüz, "Machine learning at the wireless edge: Distributed stochastic gradient descent over-the-air," *IEEE Transactions on Signal Processing*, vol. 68, pp. 2155–2169, Mar. 2020.

[6] M. M. Amiri and D. Gündüz, "Federated learning over wireless fading channels," *IEEE Transactions on Wireless Communications*, vol. 19, no. 5, pp. 3546–3557, May 2020.

[7] H. Sifaou and G. Y. Li, "Robust federated learning via over-the-air computation," 2021.

[8] T. Goblick, "Theoretical limitations on the transmission of data from analog sources," *IEEE Transactions on Information Theory*, vol. 11, no. 4, pp. 558–567, 1965.

[9] M. Gastpar, B. Rimoldi, and M. Vetterli, "To code, or not to code: lossy source-channel communication revisited," *IEEE Transactions on Information Theory*, vol. 49, no. 5, pp. 1147–1158, 2003.

[10] K. H. Lee and D. P. Petersen, "Optimal linear coding for vector channels," *IEEE Transactions on Communications*, vol. 24, pp. 1283–1290, Dec. 1976.

[11] M. Skoglund, N. Phamdo, and F. Alajaji, "Hybrid digital–analog source–channel coding for bandwidth compression/expansion," *IEEE Transactions on Information Theory*, vol. 52, no. 8, pp. 3757–3763, 2006.

[12] T. Qin, W. Liu, B. Vucetic, and Y. Li, "Over-the-air computation via broadband channels," *IEEE Wireless Communications Letters*, vol. 10, no. 10, pp. 2150–2154, 2021.

[13] S. Tang, P. Popovski, C. Zhang, and S. Obana, "Multi-slot over-the-air computation in fading channels," 2021.

[14] C. Boncelet, "Chapter 7 - image noise models," in *The Essential Guide to Image Processing*, A. Bovik, Ed. Boston: Academic Press, 2009, pp. 143–167. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B978012374457900007X

[15] J. H. Conway and N. J. A. Sloane, *Sphere packings, lattices and groups*. Springer Science & Business Media, 2013, vol. 290.

[16] R. Zamir and M. Feder, "On lattice quantization noise," *IEEE Transactions on Information Theory*, vol. 42, no. 4, pp. 1152–1159, 1996.

[17] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *CoRR*, vol. abs/1409.1556, 2014. [Online]. Available: http://arxiv.org/abs/1409.1556

[18] A. Krizhevsky, "Learning multiple layers of features from tiny images," 2009.