Multi-Slot Over-the-Air Computation in Fading Channels

Suhua Tang, *Senior Member, IEEE*, Petar Popovski, *Fellow, IEEE*, Chao Zhang, and Sadao Obana

***Abstract*— IoT systems typically involve separate data collection and processing, and face the scalability issue when the number of nodes increases. For some tasks, only the result of data fusion is needed. Then, the whole process can be realized in an efficient way, integrating the data collection and fusion in one step by over-the-air computation (AirComp). Its shortcoming, however, is signal distortion when channel gains of nodes are different, which cannot be well solved by transmission power control alone in times of deep fading. To address this issue, in this paper, we propose a multi-slot over-the-air computation (MS-AirComp) framework for the sum estimation in fading channels. Compared with conventional data collection (one slot for each node) and AirComp (one slot for all nodes), MS-AirComp is an alternative policy that lies between them, exploiting multiple slots to improve channel gains so as to facilitate power control. Avoiding to obtain instantaneous channel gains of all nodes at the sink is a key point. Specifically, the transmissions are distributed over multiple slots and a threshold of channel gain is set for distributed transmission scheduling. Each node transmits its signal only once, in the slot when its channel gain first gets above the threshold, or in the last slot when its channel gain remains below the threshold. Theoretical analysis gives the closed-form of the computation error in fading channels, based on which the optimal parameters are found. Noticing that computation error tends to be reduced at the cost of more transmission power, a method is suggested to control the increase of transmission power. Simulations confirm that the proposed method can effectively reduce computation error, compared with state-of-the-art methods.**

***Index Terms*— Over-the-air computation, multi-slot, transmission power control, distributed transmission scheduling.**

I. Introduction

INTERNET of Things (IoT) will have billions of devices, many of which will be connected to the Internet by the low power wide area (LPWA) technologies such as NB-IoT and LoRa [1]. Data collection and processing in IoT systems usually are separated, both in the digital domain, and the former involves the one-by-one data transmission. Then, a network with K nodes will involve at least K transmissions per data collection, which takes much time when there are millions of nodes in a LPWA cell. This is a challenging problem, because of limited spectrum bandwidth, real-time requirement, and the transmission collisions in a multiple access channel.

Some sensing tasks actually do not require the individual value from each sensor, if only their fusion, e.g., sum, average, max, etc., is computed correctly. For these tasks, a more efficient method is possible. Recently, a new policy, called over-the-air computation (AirComp) [2], [3], was investigated, which usually integrates the data collection and fusion in the analog domain using uncoded transmissions [4], in the way similar to physical layer network coding [5]. All nodes transmit their signals simultaneously in a coordinated way and the data fusion (sum) is computed over the air. In addition, it has been proven that the computation error of the analog transmission can be made much smaller than that of digital methods when using the same amount of resources [6].

Although AirComp generally only supports the sum operation, by proper pre-processing and post-processing, it can be extended to support any kind of nomographic functions [7], [8], [9]. Recently, deep neural network has also been exploited in pre-processing and post-processing, which enables more advanced processing of sensor data by approximating any function via deep models learned from data [10].

AirComp only takes one slot for data collection, which is very efficient. But to ensure the unbiased data fusion, each node has to pre-amplify its data so that signals arriving at the sink are aligned in their signal magnitude [11], [12]. The pre-amplification usually uses a principle of channel inversion (use a power inversely proportional to the channel coefficient) to not only mitigate the difference in channel gains but also to get an a priori agreed magnitude of the received signals. But this requires a high transmission power in deep fading.

Accurate channel coefficient per node usually is required at the sink, but its collection from many nodes is also a challenging problem. AirComp is used for estimating channel coefficients before actual computation in [13]. But it is still time consuming and the accuracy is limited by noise. A more aggressive effort is to exploit blind AirComp without channel coefficient [14], at the cost of degraded performance.

To overcome channel fading without incurring much overhead in collecting instantaneous channel coefficient, in this

Manuscript received 30 August 2021; revised 28 January 2022, 14 June 2022, 20 November 2022, and 24 January 2023; accepted 8 February 2023. Date of publication 22 February 2023; date of current version 11 October 2023. This work was supported in part by the SCAT Foundation, Japan. The work of Petar Popovski was supported by the Velux Foundation, Denmark, through the Villum Investigator Grant WATER. The work of Chao Zhang was supported by the Key Research and Development Program of Shaanxi under Grant 2023-YBGY-251. The associate editor coordinating the review of this article and approving it for publication was G. C. Ferrante. *(Corresponding author: Suhua Tang.)*

Suhua Tang is with the Graduate School of Informatics and Engineering, The University of Electro-Communications, Tokyo 182-8585, Japan (e-mail: shtang@uec.ac.jp).

Petar Popovski is with the Department of Electronic Systems, Aalborg University, 9220 Aalborg, Denmark.

Chao Zhang is with the School of Information and Communications Engineering, Xi'an Jiaotong University, Xi'an 710049, China.

Sadao Obana is with The University of Electro-Communications, Tokyo 182-8585, Japan.

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TWC.2023.3245304.

Digital Object Identifier 10.1109/TWC.2023.3245304

paper, we propose a multi-slot over-the-air computation (MS-AirComp) framework for the sum estimation in fading channels. Conventional data collection and AirComp are two extremes. The former uses one slot for each node and the latter uses one slot for all nodes. MS-AirComp is an alternative policy that lies between them, exploiting multiple slots to take advantage of time-varying channel gains so as to facilitate power control. It is necessary to avoid transmission when a node is in deep fading. This is also feasible when considering the time diversity due to the random variation of channel gain [15]. Then, the over-the-air computation is distributed over N slots, where N is much less than the number of nodes. Specifically, the sink node sets a threshold of channel gain, based on the channel statistics. Each node transmits its signal only once in the N slots, either in the slot when its channel gain first gets above the predefined threshold, or in the last slot if its channel gain remains below the threshold. The transmission power control policy is the same as in previous methods [11], [12], either the channel inversion policy if the power is less than the constraint or using the maximal value otherwise.

The main contributions of this paper are:

- Shaping channel gain by exploiting time diversity. This paper improves channel gains by using multiple slots and setting a threshold of channel gain. This helps to avoid deep fading and facilitate power control to achieve signal magnitude alignment.
- Distributed transmission scheduling. Transmission scheduling of each node is performed in a distributed way, using a threshold of channel gain, which the sink computes based on the channel statistics instead of instantaneous values. This helps reduce the overhead of feeding back instantaneous channel gain from each node to the sink, which is a key feature.
- Closed-form of the mean squared error (MSE) of the computation result; hereafter referred to as computation MSE. Theoretical analysis of the computation MSE enables to find optimal parameters given the statistics of channel gains, and the simulation results, based on these optimal parameters, are consistent with the analysis.
- Revealing the tradeoff between the computation MSE and transmission power and presenting a simple solution for this issue.

Numerical analysis and Monte Carlo simulations illustrate the promising performance of the proposed method. Compared with state-of-the-art methods, the proposed method greatly reduces the computation MSE, by utilizing channel gain diversity at the cost of only a few slots. By a refined tradeoff between the computation MSE and transmission power, the proposed method effectively reduces the computation MSE while consuming almost the same transmission power as the previous method.

Main notations used in the analysis are listed in Table I. $\mathbb{C}$ and $\mathbb{R}^+$ denote the set of complex, positive real numbers, respectively, and $\mathbb{E}$ is the expectation operation.

The rest of this paper is organized as follows: Sec. II reviews the basic AirComp method and related work. Sec. III first presents the proposed framework, and analyzes the computation MSE. On this basis, optimal parameters are found, transmission power is analyzed, and some numerical results are illustrated. Sec. IV shows the results achieved by Monte Carlo simulation, and points out the necessity of a tradeoff between the computation MSE and transmission power. Then, in Sec. V, a simple method is suggested to control the increase of transmission power. Finally, Sec. VI concludes this paper and points out future work.

TABLE I
MAIN NOTATIONS FOR ANALYSIS

Notation	Meaning	More details
K	Number of sensor nodes	Default value is 100
g_k	Channel gain of node k	
$\overline{g}_k$	Average channel gain of node k	Default value is 10
h_k	Channel coefficient of node k	Complex value. Assumed to be positive real in the analysis for simplicity
b_k	Tx-scaling factor for node k	
a	Rx-scaling factor at the sink	
$P_{\max}$	Normalized maximal Tx power	Default value is 10
g_{th}	Threshold of channel gain	
P_{th}	Threshold of signal receive rate	

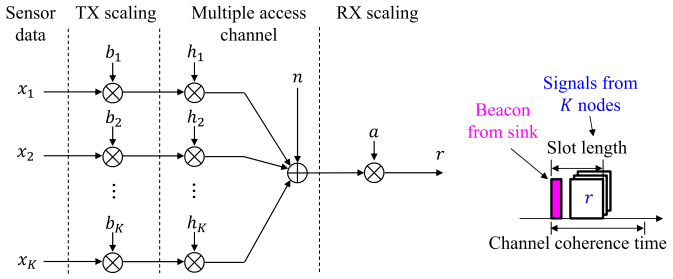


Fig. 1. Basic AirComp model with K nodes and 1 sink. All nodes transmit their signals simultaneously in one slot.

II. RELATED WORK

Here, we review the basic AirComp method, and previous efforts on improving its performance.

A. Basic AirComp Model

In a typical AirComp model [11] for a sensor network, K nodes each transmit the analog signals to a common sink simultaneously, triggered by a control signal from the sink, as shown in Fig. 1. All the transmissions are synchronized so that all signals arrive at the sink at the same time. This is called single-slot transmission, where a slot equals the transmission time of a signal. Both the nodes and the sink have a single antenna each. To deal with the difference in channel gains, the pre-processed signal at the k-th node, $x_k \in \mathbb{C}$ and $|x_k| \leq v$, with zero mean and unit variance ($\mathbb{E}\left[|x_k|^2\right] = 1$), is amplified by its Tx-scaling factor $b_k \in \mathbb{C}$ and sent to the sink.

The sink applies a Rx-scaling factor $a \in \mathbb{C}$ to the received signal to get the computation result as

$$r = a\left(\sum_{k=1}^{K} h_k b_k x_k + n\right), \tag{1}$$

where $h_k \in \mathbb{C}$ is the channel coefficient between node k and the sink, and $n \in \mathbb{C}$ is the additive white Gaussian noise

(AWGN) at the sink with zero mean and variance being σ^2. Channel coefficient h_k is assumed to be known by node k and by the sink, and the latter is a key drawback of AirComp because collecting channel coefficients of all nodes at the sink is time consuming and will greatly degrade the efficiency of AirComp. Signal x_k and channel coefficient h_k change with time. Block fading is assumed, i.e., channel coefficients change per slot and are constant in a slot. Because parameters a and b_k are optimized per slot, the time index is neglected to simplify the analysis. With the maximal power constraint,[1] $|b_k x_k|^2$ should be no more than P', the maximal power. Let $P_{\max}$ denote P'/v^2. Then, $|b_k|^2 \leq P'/v^2 = P_{\max}$.

The error between r and the target sum $\sum_{k=1}^{K} x_k$ is measured by the computation MSE defined in (2), where $\mathbb{E}[\cdot]$ is the expectation operation with respect to signal x_k and the noise, and it is assumed that signals and noise are non-correlated.

$$\mathrm{MSE} = \mathbb{E}\left[\left| r - \sum_{k=1}^{K} x_k \right|^2\right] = \sum_{k=1}^{K} |a h_k b_k - 1|^2 + \sigma^2 |a|^2. \quad (2)$$

When $h_k b_k$, $\forall k$, equals to a constant c, all signals have the same magnitude (i.e., it is equal to c before Rx-scaling and ac after Rx-scaling), which is called *signal magnitude alignment*. In such cases, if a is set to $1/c$, $a h_k b_k = 1$ and (1) gives the result $r = \sum_k x_k + an$, using the same weight for all signals without a bias on any one, which is an unbiased estimation. Because b_k is computed as $1/(a h_k)$, this policy of power control is called channel inversion. Although this removes the signal distortion (The first part in MSE equals to 0), it requires to use a large a when some node is in deep fading (with very small h_k), which leads to an increase of the noise power, the second part of MSE.

Generally, MSE is minimized by taking a tradeoff between signal distortion and noise power. This policy aligns most signals to the same signal magnitude ($a h_k b_k = 1$) while allows a small number of weak signals to be misaligned ($a h_k b_k < 1$) [11], [12], which is called *partial signal magnitude alignment* in this paper.

This AirComp is a centralized method because the sink decides the timing of the computation, and obtains instantaneous channel coefficient (h_k) from each node in advance for computing the optimal parameter (a), which causes a significant overhead.

B. Dealing With Errors in AirComp

Several factors may affect the performance of AirComp. Timing synchronization usually is a necessity of AirComp for avoiding signal distortion. By modulating the sensor data in a series of random signal pulses, AirComp can be realized by a coarse block-synchronization [16]. In [17], AirShare is proposed, in which the sink synchronizes the clock of all nodes before the actual transmission, by using multiple frequencies.

Errors may also be caused by noise and channel fading. Distortion outage for AirComp was first investigated in [18]. In deep fading, the magnitudes of some signals cannot be aligned with that of other signals, under the constraint of maximal transmission power. To solve this problem, the work in [11] and [12] minimizes the computation MSE by jointly optimizing the transmission power and a Rx-scaling factor at the sink node. An active relay model based on amplify-and-forward is studied in [19] and [20]. In [21], reconfigurable intelligent surface is exploited to control the phase of reflected signals so that multipath components of all signals constructively add together at the sink. Antenna array is investigated in [13] and [22] to support vector-valued AirComp. It is further combined with wireless power transfer in [23].

Recently, federated learning, a distributed learning method in which all nodes share a common model to locally process data, has attracted much attention. AirComp, as a key component, is used in aggregating model update created by each node. Researchers have studied how to deal with the impact of fading and noise. The authors in [24] investigated direct model update based on the noisy distorted gradient. Precoding and scaling operations are suggested to mitigate the effect of the noisy channel to accelerate the convergence of the learning process [25]. In federated learning, it is not necessary to receive model update from all nodes. Therefore, in [26] and [27], at each iteration, only nodes with a channel gain large enough are selected to transmit their model update, and the centralized control of AirComp enables this selection. Fast fading is taken into account in [28], for the purpose of one-shot approximation of function values, considering sub-Gaussian noise and correlated channels.

C. Dealing With Fading in Multiple Access Channel

In the multiple access channel, it is efficient to exploit multi-user diversity to deal with channel fading, letting nodes with high channel gains transmit their signals first. This is studied for ALOHA networks in [29] and [30], and for CSMA networks in [31]. Generally, a threshold of channel gain is set for nodes to facilitate distributed transmission scheduling. But these methods only let one node transmit its signal each time in order to avoid transmission collisions.

D. A Short Comparison

Considering the specific task of model aggregation in federated learning, the policy of node selection in [26] and [27] relies on the centralized control, both to obtain channel gains and to get the correct number of nodes involved in the transmission. But it is impractical for the sink to collect channel gain of all nodes in some cases. Transmission power control [11], [12] is necessary in AirComp, but its effect is limited in times of deep fading. In comparison, the proposed method improves channel gain by leveraging the time diversity, letting nodes schedule their transmissions only when their channel gains are high enough. This distributes over-the-air computation into multiple slots, improving system performance at the cost of only a few slots meanwhile avoiding the overhead of collecting instantaneous channel gains at the sink.

[1] If $|b_k x_k|^2$ is greater than the maximal transmission power, the transmitted signal will be distorted after the power amplification.

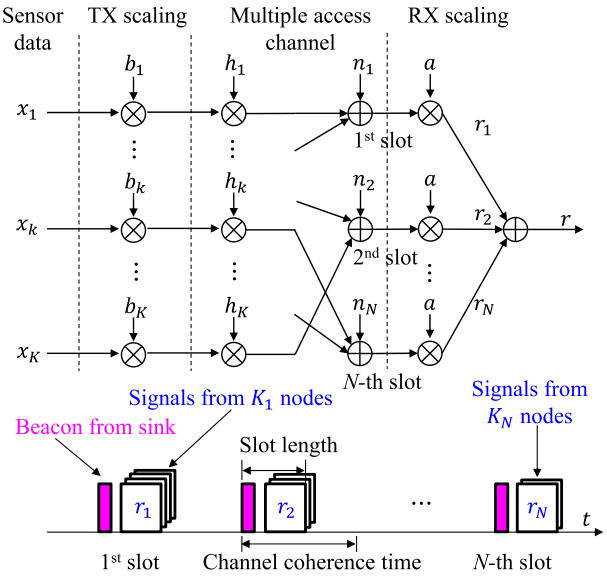


Fig. 2. Proposed system model with K nodes and 1 sink. All nodes should transmit their signals to the common sink within N slots, but each node may select which slot to transmit based on its channel gain.

III. Multi-Slot AirComp

Here, we present the proposed MS-AirComp framework and analyze the computation MSE and the optimal solution for the general model in Sec. III-A, discuss the impact of the signal magnitude alignment in Sec. III-B under a transmission policy defined by probability functions, and present the probability functions for the proposed threshold-based slot selection method in Sec. III-C. For a comparison, we also present the method that always selects the best slot with all channel gains known in advance as an upper bound, do some further analysis such as average channel gain and transmission power in Sec. III-D, discuss the impact of channel estimation error in Sec. III-E, and show some numerical results in Sec. III-F.

A. System Model

This paper focuses on the sum estimation of sensor data obtained by all nodes in a sensor network. It can be easily extended to support other computations by proper pre-processing and post-processing.

Fig. 2 shows the system model. We investigate a multi-slot over-the-air computation (MS-AirComp) framework, extending the AirComp model from single-slot (single transmission chance) to multi-slot (multiple transmission chances) and distributing the transmission to multiple slots. For simplicity, we mainly consider a single frequency channel, but the proposed method may be extended to multi-frequency systems. Specifically, there are N slots for the whole transmission. Each slot involves a beacon from the sink (with pilot signals for each node to estimate the channel coefficient and synchronize its clock with the sink), a fixed space (for each node to transit from receiving to transmitting), and a superposition of signals simultaneously transmitted from nodes on the same frequency as the beacon signal uses. Each node has the chance to choose a slot with a high channel gain to transmit its signal, $x_k \in \mathbb{C}$ and $|x_k| \leq v$, with zero mean and unit variance ($\mathbb{E}\left[|x_k|^2\right] = 1$).

We assume a block-fading channel,[2] where the channel gain is constant in a slot and the slots are separated far enough in time in order to ensure that the channel gain independently varies per slot [15]. Different from the model in Sec. II-A, instantaneous channel coefficient $h_{k,i} \in \mathbb{C}$ in the i-th slot is known by node k (by receiving the beacon signal) but not by the sink. Instead, each node monitors its own link when receiving messages from the sink, estimates probability density function (PDF) of its channel gain and sends it to the sink at a relatively long period. On this basis and exploiting channel reciprocity, the sink decides the optimal parameters. The same Rx-scaling factor $a \in \mathbb{C}$ is used for all slots. Each node monitors N channel coefficients in N slots, but transmits its signal only once. For simplicity, here h_k is used to represent the channel coefficient $h_{k,i}$ of the i-th slot over which node k actually transmits its signal, and $|b_k|^2$ ($b_k \in \mathbb{C}$) corresponds to the transmission power. Each noise sample n_i is produced by the respective slot i.

Then, the signals received over N slots, summed together at the sink, are

$$r = a\left(\sum_{k=1}^{K} h_k b_k x_k + \sum_{i=1}^{N} n_i\right). \tag{3}$$

The signal sum has the same form as in the single-slot case in (1), except that the noise samples in all slots are involved. It is expected that the improvement in channel gain is greater than the increase in noise power. Because node k can always adjust the phase of b_k to ensure that $h_k b_k$ is real and positive, without loss of generality, it is assumed that $h_k \in \mathbb{R}^+$, $b_k \in \mathbb{R}^+$, and $a \in \mathbb{R}^+$ in the analysis, for simplicity. The same assumption was taken in [11], too. Later, in the simulation evaluation, complex parameters are used.

The computation MSE is defined in a similar way as in (2), as follows

$$\begin{aligned} \text{MSE} &= \mathbb{E}\left[\sum_{k=1}^{K} (a\alpha_k - 1)^2\right] + Na^2\sigma^2, \\ \alpha_k &= h_k b_k. \end{aligned} \tag{4}$$

Here α_k changes with h_k and is also random, so the expectation operator (with respect to α_k) is kept.

Assume on each slot, α_k has a PDF $f_{\alpha_k}(x)$ and a cumulative distribution function (CDF) $F_{\alpha_k}(x)$, and their counterparts over N slots, under a selection policy, are $f^N_{\alpha_k}(x)$ and $F^N_{\alpha_k}(x)$, respectively. Define $\psi(x)$ as a function of partial signal magnitude alignment, which changes the magnitude of a signal by adjusting its transmission power. E.g., $\psi(x)$ converts the maximal signal magnitude $\alpha_k = h_k\sqrt{P_{\max}}$ to $h_k b_k$. In this

[2]In wireless sensor networks, although nodes do not move, changes in the environment lead to variations in channel gains. Under typical moving speeds of background objects, the channel has a coherence time that is longer than the packet duration, such that it is constant during a packet transmission and changes afterwards. Therefore, block fading is assumed. Although the channel is coherent within an interval, the slots in the proposed method are separated far enough so that the channel gains between adjacent slots are independent. Another case is to realize the multi-slot transmission by frequency hopping communication. When the sink sends a beacon signal on a selected frequency (outside the coherence bandwidth of the frequency used in the previous transmission), channel gains are independent. Nodes will transmit their signals on the same frequency as the sink does.

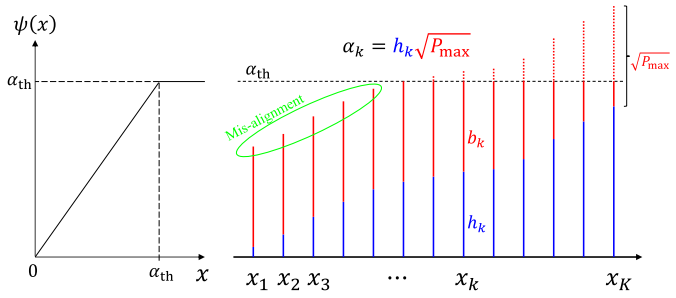


Fig. 3. Partial signal magnitude alignment in AirComp. Signals are sorted in the ascending order of their channel coefficients. Most signals are aligned to a common magnitude $\alpha_{\rm th}$ by adjusting their transmission power, and some weak signals cannot be aligned to $\alpha_{\rm th}$ even using the maximal transmission power. $\psi(x)$ represents the impact of partial signal magnitude alignment.

way, $\psi(x)$ shapes the magnitude of each signal to facilitate the subsequent computation,[3] $a\sum_{k=1}^{K}\psi(\alpha_k)x_k$. Then, the MSE in (4) is computed as the expectation of $(a\psi(\alpha_k)-1)^2$, using the PDF $f^N_{\alpha_k}(x)$, as follows:

$$\text{MSE} = \sum_{k=1}^{K}\mathbb{E}_{\alpha_k\sim f^N_{\alpha_k}}\left[(a\psi(\alpha_k)-1)^2\right] + Na^2\sigma^2. \tag{5}$$

And the optimal parameters a and b_k can be found by

$$\begin{aligned}&\min_{a,b_k}\ \text{MSE},\\ &\text{s.t. } b_k^2 \le P_{\max},\quad k=1,\cdots,K.\end{aligned} \tag{6}$$

Notice that MSE is a quadratic function. Letting the partial differentiation $\frac{\partial \text{MSE}}{\partial a}$ equal to 0, the optimal a can be computed by

$$\begin{aligned}a_{\rm opt} &= \frac{\sum_{k=1}^{K}\int_0^{\infty}\psi(x)f^N_{\alpha_k}(x)dx}{\sum_{k=1}^{K}\int_0^{\infty}\psi^2(x)f^N_{\alpha_k}(x)dx + N\sigma^2}\\ &= \frac{\sum_{k=1}^{K}\mathbb{E}\left[\psi(\alpha_k)\right]}{\sum_{k=1}^{K}\mathbb{E}\left[\psi^2(\alpha_k)\right] + N\sigma^2}.\end{aligned} \tag{7}$$

B. Impact of Signal Magnitude Alignment

This paper takes the policy of partial signal magnitude alignment, as shown in Fig. 3. In the transmission, magnitudes (α_k) of signals from most nodes will be aligned to a common value, which is denoted as $\alpha_{\rm th}$. The Tx-scaling b_k used for the transmission at node k is $b_k = \min(\alpha_{\rm th}/h_k, \sqrt{P_{\max}})$, namely it is equal to $\alpha_{\rm th}/h_k$ if it is not more than $\sqrt{P_{\max}}$ (using the channel inversion policy) and $\sqrt{P_{\max}}$ otherwise (using the maximal power). This power control policy is the same as in the previous work [11]. The difference is that h_k in the actual transmission, after slot selection, is improved in the proposed method.

Considering the above policy of power control, in the following we consider $\alpha_k = \sqrt{P_{\max}}h_k$, the maximal available signal magnitude. With channel gain g_k, $h_k = \sqrt{g_k}$, then $\alpha_k = \sqrt{P_{\max}g_k}$.

At current channel gain g_k, node k needs to check whether $\sqrt{P_{\max}g_k}$, at the maximal transmission power, reaches the threshold $\alpha_{\rm th}$. If affirmative, its signal magnitude will be aligned to $\alpha_{\rm th}$, possibly using a smaller transmission power. Then, $\psi(\alpha_k)$, representing the value of α_k after signal magnitude alignment, corresponds to the function

$$\psi(x) = \begin{cases} x & 0 \le x < \alpha_{\rm th},\\ \alpha_{\rm th} & x \ge \alpha_{\rm th}.\end{cases} \tag{8}$$

This applies to the selected slot where a node transmits its own signal. Using this function, $(a\psi(x)-1)^2$ in (5) becomes $(ax-1)^2$ in the range $(0,\alpha_{\rm th})$, and $(a\alpha_{\rm th}-1)^2$ in the range $[\alpha_{\rm th},\infty)$ with a probability $1-F^N_{\alpha_k}(\alpha_{\rm th})$. Then, the MSE in (5) is simplified as

$$\begin{aligned}\text{MSE} &= \sum_{k=1}^{K}\text{MSE}(k) + Na^2\sigma^2,\\ \text{MSE}(k) &= \int_0^{\alpha_{\rm th}}(ax-1)^2 f^N_{\alpha_k}(x)dx\\ &\quad + (a\alpha_{\rm th}-1)^2\left(1-F^N_{\alpha_k}(\alpha_{\rm th})\right).\end{aligned} \tag{9}$$

Accordingly, the optimal a in (7) can be rewritten as

$$a_{\rm opt} = \frac{\sum_{k=1}^{K}\int_0^{\alpha_{\rm th}} x f^N_{\alpha_k}(x)dx + \alpha_{\rm th}\left(1-F^N_{\alpha_k}(\alpha_{\rm th})\right)}{\sum_{k=1}^{K}\int_0^{\alpha_{\rm th}} x^2 f^N_{\alpha_k}(x)dx + \alpha_{\rm th}^2\left(1-F^N_{\alpha_k}(\alpha_{\rm th})\right) + N\sigma^2}. \tag{10}$$

It is possible to further compute the partial differentiation $\frac{\partial \text{MSE}}{\partial \alpha_{\rm th}} = 0$, but it has no closed-form for $\alpha_{\rm th}$. Therefore, $\alpha_{\rm th}$ is found by minimizing the computation MSE via grid search.

C. Selecting a Slot by a Threshold of Channel Gain

The analysis in the previous section is based on the assumption that a slot is somehow selected, but did not mention how to realize it. Here, we analyze how to set a threshold of channel gain $g_{\rm th}$ for the slot selection (distributed transmission scheduling) in the proposed method.

In each slot, based on the beacon signal from the sink, each node detects channel gain and transmits its signal immediately, if its channel gain is above the threshold and the transmission has not been done in previous slots yet. Otherwise, a node in deep fading (when its channel gain is below the threshold) defers the transmission decision to next slot, expecting a high channel gain in the future. If the channel gain of a node is below the threshold over all slots, the node transmits its signal in the last slot,[4] and the signal arriving at the sink is susceptible to magnitude misalignment.

To ensure a successful transmission, it is required that the channel gain should be above the threshold $g_{\rm th}$, over at least one of the N slots, with a relatively high probability $P_{\rm th}$. On the other hand, if channel gain of node k is below $g_{\rm th}$ over all N slots, node k has to transmit its signal in the N-th slot, with a probability $\left(F_{g_k}(g_{\rm th})\right)^N$, where $F_{g_k}(x)$ denotes the CDF of channel gain g_k per slot. Then,

$$\frac{1}{K}\sum_{k=1}^{K}P_k^s(g_{\rm th}) \ge P_{\rm th} \tag{11}$$

[3]An example of $\psi(x)$ for computing the sum of all signals under the constraint of max transmission power is shown in (8).

[4]This because each node has to finish its transmission, in the last slot if it fails in all previous slots.

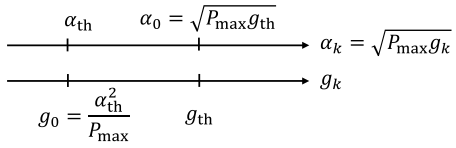


Fig. 4. Relationship between $\alpha_{\rm th}$ (g_0) and α_0 ($g_{\rm th}$).

should be satisfied, where $P_k^s(g_{\rm th}) = 1 - (F_{g_k}(g_{\rm th}))^N$ is the probability that the channel gain of node k is no less than $g_{\rm th}$ over at least one slot.

Generally, at a given gain threshold $g_{\rm th}$, using the maximal power, the signal magnitude $\alpha_0 = \sqrt{P_{\max} g_{\rm th}}$ should be no less than $\alpha_{\rm th}$. Fig. 4 shows this relationship. Then, in the range $[\alpha_0, \infty)$, MSE of the signal part is

$$\text{MSE}_1 = \sum_{k=1}^{K} \left(a\alpha_{\rm th} - 1\right)^2 P_k^s(g_{\rm th}). \tag{12}$$

With a small probability $1 - P_k^s(g_{\rm th})$, the channel gain of node k is below the threshold $g_{\rm th}$ over all slots, and has to transmit its signal in the last slot. Accordingly, in the range $(0, \alpha_0)$, α_k has a conditional distribution

$$\hat{f}_{\alpha_k}(x) = \frac{f_{\alpha_k}(x)}{F_{\alpha_k}(\alpha_0)}, \quad \alpha_k < \alpha_0. \tag{13}$$

This range is further divided into two sub-ranges. In the sub-range $[\alpha_{\rm th}, \alpha_0)$, the error term is $\left(a\alpha_{\rm th} - 1\right)^2$ with a conditional probability $\left(F_{\alpha_k}(\alpha_0) - F_{\alpha_k}(\alpha_{\rm th})\right)/F_{\alpha_k}(\alpha_0)$. In the sub-range $(0, \alpha_{\rm th})$, the error term is $\left(ax - 1\right)^2$. Then, MSE of the signal part in the range $(0, \alpha_0)$ is

$$\begin{aligned} \text{MSE}_2 &= \sum_{k=1}^{K} \text{MSE}(k) \left(1 - P_k^s(g_{\rm th})\right), \\ \text{MSE}(k) &= \left(a\alpha_{\rm th} - 1\right)^2 \frac{F_{\alpha_k}(\alpha_0) - F_{\alpha_k}(\alpha_{\rm th})}{F_{\alpha_k}(\alpha_0)} \\ &\quad + \int_0^{\alpha_{\rm th}} \left(ax - 1\right)^2 \hat{f}_{\alpha_k}(x) dx. \end{aligned} \tag{14}$$

And the overall computation MSE in the whole range is

$$\text{MSE} = \text{MSE}_1 + \text{MSE}_2 + Na^2\sigma^2. \tag{15}$$

The probability functions of α_k in the three ranges are summarized as

$$\begin{aligned} f_{\alpha_k}^N(x) &= \hat{f}_{\alpha_k}(x) \left(1 - P_k^s(g_{\rm th})\right), \alpha_k < \alpha_{\rm th}, \\ 1 - F_{\alpha_k}^N(\alpha_{\rm th}) &= P_k^s(g_{\rm th}) + \frac{F_{\alpha_k}(\alpha_0) - F_{\alpha_k}(\alpha_{\rm th})}{F_{\alpha_k}(\alpha_0)} \left(1 - P_k^s(g_{\rm th})\right), \end{aligned} \tag{16}$$

and the optimal a for the computation MSE can be computed by (10). Then, the optimal $\alpha_{\rm th}$ is found via a grid search.

Under Rayleigh fading, g_k, channel gain of node k, follows an exponential distribution $f_{g_k}(x)$ with an average $\overline{g}_k$, and has a CDF $F_{g_k}(x)$, as follows:

$$f_{g_k}(x) = \frac{1}{\overline{g}_k} \exp\left(-\frac{x}{\overline{g}_k}\right), \tag{17}$$

$$F_{g_k}(x) = 1 - \exp\left(-\frac{x}{\overline{g}_k}\right). \tag{18}$$

Then, $\alpha_k = \sqrt{P_{\max} g_k}$ has a PDF $f_{\alpha_k}(x)$ and a CDF $F_{\alpha_k}(x)$,

$$f_{\alpha_k}(x) = \frac{2x}{P_{\max}\overline{g}_k} \exp\left(-\frac{x^2}{P_{\max}\overline{g}_k}\right), \tag{19}$$

$$F_{\alpha_k}(x) = 1 - \exp\left(-\frac{x^2}{P_{\max}\overline{g}_k}\right). \tag{20}$$

1) Relation Between a and $\alpha_{\rm th}$: Although $\alpha_{\rm th}$ and a are separately computed, they are closely related. Actually, their product, $a\alpha_{\rm th}$, is approximately 1.

According to (9), it is straightforward that $a\alpha_{\rm th}$ approaching 1 will ensure that signals aligned to $\alpha_{\rm th}$ have a computation error close to 0. In fact, it is assumed $a\alpha_{\rm th} = 1$ in [11]. But in the analysis, assuming $a\alpha_{\rm th} = 1$ will make the whole process fail. Once the optimal parameters are determined, it is safe to use the approximation $a\alpha_{\rm th} = 1$ to compute the actual computation MSE.

2) The Whole Process of the Proposed Method:

- For each possible number of slots, N, $g_{\rm th}$ is computed from $P_{\rm th}$ according to (11), which can be simplified in Rayleigh fading.
- For each possible $\alpha_{\rm th}$, the optimal a is computed via (10).
- Parameters $(N, g_{\rm th}, \alpha_{\rm th})$ leading to the minimal computation MSE are found.
- The sink node broadcasts a beacon signal (containing $N, g_{\rm th}, \alpha_{\rm th}$) at the beginning of each slot.
- Each node k monitors the beacon signal, obtains its channel gain g_k, and compares it with the threshold $g_{\rm th}$. If $g_k \geq g_{\rm th}$ and node k has not transmitted its signal yet, it transmits its signal immediately. Otherwise, if the channel gain remains below the threshold over all N slots, node k transmits its signal in the N-th slot. The transmission power of node k is computed as $b_k = \min\{\alpha_{\rm th}/h_k, \sqrt{P_{\max}}\}$, using the channel coefficient h_k.
- At the end of the N-th slot, the signals received over all N slots are summed together at the sink.

In this way, although the sink computes the optimal parameters, each node decides its own transmission slot and transmission power in a distributed way, which both improves channel gain and avoids the overhead of feeding back its instantaneous channel coefficient to the sink.

D. Further Analysis

Before the analysis, we first discuss a generated case, in which channel gains of all slots are known in advance and each node will select the slot with the maximal gain to transmit its signal. This method determines the performance upper bound of the proposed method. According to order statistics [32], $f_{\alpha_k}^N(x)$, the PDF of $\alpha_{\max} = \max_{k=1,\cdots,N} \alpha_k$, is

$$f_{\alpha_{\max}}(x) = N f_{\alpha_k}(x) \left(F_{\alpha_k}(x)\right)^{N-1}. \tag{21}$$

Then, the optimal parameters minimizing the computation MSE in (9) can be found by (10).

Next, we analyze the average channel gain, the probability of misalignment in signal magnitude, and transmission power. In the analysis, the optimal selection method is denoted as OptSel, and the proposed method is denoted as SelFirst.

1) Average Channel Gain: In the proposed method, $f_{\rm th}(x)$, the PDF of g_k in the presence of the gain threshold $g_{\rm th}$, is written as

$$f_{\rm th}(x) = \begin{cases} P_k^s(g_{\rm th})\tilde{f}_{g_k}(x) & g_k \geq g_{\rm th} \\ (1-P_k^s(g_{\rm th}))\,\hat{f}_{g_k}(x) & g_k < g_{\rm th}, \end{cases}$$
$$\tilde{f}_{g_k}(x) = \frac{f_{g_k}(x)}{1-F_{g_k}(g_{\rm th})}, \quad g_k \geq g_{\rm th},$$
$$\hat{f}_{g_k}(x) = \frac{f_{g_k}(x)}{F_{g_k}(g_{\rm th})}, \quad g_k < g_{\rm th}. \tag{22}$$

Then, $\overline{g}_{\rm SelFirst}$, the average channel gain of all nodes, is computed as the average of expected channel gain per node, as follows:

$$\begin{aligned} \overline{g}_{\rm SelFirst} &= \frac{1}{K}\sum_{k=1}^{K}\mathbb{E}_{g_k\sim f_{\rm th}}[g_k] \\ &= \frac{1}{K}\sum_{k=1}^{K} P_k^s(g_{\rm th})\mathbb{E}_{g_k\sim\tilde{f}_{g_k}}[g_k] \\ &\quad + (1-P_k^s(g_{\rm th}))\,\mathbb{E}_{g_k\sim\hat{f}_{g_k}}[g_k], \end{aligned} \tag{23}$$

which is further expressed by the average channel gains in the two ranges $[g_{\rm th},\infty)$ and $(0,g_{\rm th})$.

In the optimal selection method, the probability of channel gain ($g_{\max} = \max_{k=1,\cdots,N} g_k$) is expressed by a single function $f_{g_{\max}}(x)$ in the whole range, which is deduced by using order statistics similar to that in (21). Then, the average channel gain is computed in a similar way as in (23),

$$\overline{g}_{\rm OptSel} = \frac{1}{K}\sum_{k=1}^{K}\mathbb{E}_{g_k\sim f_{g_{\max}}}[g_k],$$
$$f_{g_{\max}}(x) = N f_{g_k}(x)\left(F_{g_k}(x)\right)^{N-1}. \tag{24}$$

2) Probability of Misalignment in Signal Magnitude: In the proposed method, let g_0 denote the channel gain $\frac{\alpha_{\rm th}^2}{P_{\max}}$. The signal of node k arrives at the sink with signal magnitude in alignment when $g_k \geq g_0$. Considering the two ranges $[g_{\rm th},\infty)$ and $[g_0,g_{\rm th})$, the overall probability that the signal from node k is in alignment is

$$P_k^s(g_0) = P_k^s(g_{\rm th}) + (1-P_k^s(g_{\rm th}))\frac{F_{g_k}(g_{\rm th})-F_{g_k}(g_0)}{F_{g_k}(g_{\rm th})}.$$

And the probability that the signal magnitude misalignment occurs in one or more signals is

$$p_{\rm SelFirst}^{\rm mis} = 1-\prod_{k=1}^{K} P_k^s(g_0). \tag{25}$$

When $p = 1-P_k^s(g_0)$ is the same for all nodes, the probability that signals of k nodes are misaligned in signal magnitude is

$$p_k = \binom{K}{k} p^k (1-p)^{K-k}. \tag{26}$$

Then, the average number of signals with magnitude misalignment is $\sum_k k p_k = Kp$.

In the optimal selection method, the probability that the signal of node k arrives at the sink with signal magnitude in alignment is $1-F_{\alpha_k}^N(\alpha_{\rm th})$. Then, the probability that the signal magnitude misalignment occurs in one or more signals is

$$p_{\rm OptSel}^{\rm mis} = 1-\prod_{k=1}^{K}\left(1-F_{\alpha_k}^N(\alpha_{\rm th})\right). \tag{27}$$

When $p = F_{\alpha_k}^N(\alpha_{\rm th})$ is the same for all nodes, the probability that signals of k nodes are misaligned in signal magnitude can be computed in a similar way as in the proposed method.

3) Average Transmission Power: According to Fig. 4, in the range $g_k \in (g_0,\infty)$, the transmission power $p_{g_k}(g_k) = b_k^2 = \alpha_{\rm th}^2/g_k$ is less than $P_{\max}$, and the channel inversion policy is applied for transmission power control. On the other hand, when $g_k \in (0,g_0]$, the maximal power $p_{g_k}(g_k) = P_{\max}$ is used, as follows:

$$p_{g_k}(g_k) = \begin{cases} \alpha_{\rm th}^2/g_k & g_k > g_0 \\ P_{\max} & g_k \leq g_0. \end{cases} \tag{28}$$

Then, in the proposed method, the average transmission power of all nodes can be computed by

$$\begin{aligned} \overline{P}_{\rm SelFirst} &= \frac{1}{K}\sum_{k=1}^{K}\mathbb{E}_{g_k\sim f_{\rm th}}[p_{g_k}(g_k)] \\ &= \frac{1}{K}\sum_{k=1}^{K} P_k^s(g_{\rm th})E_{k1} + (1-P_k^s(g_{\rm th}))\,(E_{k2}+E_{k3}), \\ E_{k1} &= \mathbb{E}_{g_k\sim\tilde{f}_{g_k}}\left[\frac{\alpha_{\rm th}^2}{g_k}\right], \quad g_k \geq g_{\rm th}, \\ E_{k2} &= \int_{g_0}^{g_{\rm th}} \frac{\alpha_{\rm th}^2}{x}\hat{f}_{g_k}(x)dx, \quad g_0 < g_k < g_{\rm th}, \\ E_{k3} &= \int_{0}^{g_0} P_{\max}\hat{f}_{g_k}(x)dx, \quad g_k \leq g_0, \end{aligned} \tag{29}$$

using $f_{\rm th}(x)$ defined in (22). It can be expressed by E_{k1}, E_{k2} and E_{k3}, the average powers in the three ranges of channel gains.

In the optimal selection method, the average transmission power is computed in a similar way as in (29), using $f_{g_{\max}}(x)$ defined in (24), as follows:

$$\begin{aligned} \overline{P}_{\rm OptSel} &= \frac{1}{K}\sum_{k=1}^{K}\mathbb{E}_{g_k\sim f_{g_{\max}}}[p_{g_k}(g_k)] \\ &= \frac{1}{K}\sum_{k=1}^{K}\int_{g_0}^{\infty}\frac{\alpha_{\rm th}^2}{x}f_{g_{\max}}(x)dx \\ &\quad + \int_{0}^{g_0} P_{\max}f_{g_{\max}}(x)dx. \end{aligned} \tag{30}$$

E. Impact of Channel Estimation Error

In a real system, the actual channel has complex channel coefficient. If each node k correctly learns its channel coefficient h_k, its Tx-scaling factor b_k is computed as follows:

$$b_k = \begin{cases} \dfrac{\alpha_{\rm th}}{h_k} & \dfrac{\alpha_{\rm th}}{|h_k|} \leq \sqrt{P_{\max}}, \\ \dfrac{h_k^*}{|h_k|}\sqrt{P_{\max}} & \dfrac{\alpha_{\rm th}}{|h_k|} > \sqrt{P_{\max}}. \end{cases} \tag{31}$$

Due to the estimation error, the instantaneous channel coefficient of node k is estimated as $\hat{h}_k = h_k + \Delta_k$ with a small error Δ_k, and its Tx-scaling factor is computed as $\hat{b}_k$, replacing h_k by $\hat{h}_k$ in (31).

It is reasonable to assume that the distributions of h_k and α_k are accurate enough so that the estimations of a_{opt} and α_{th} are accurate. Then, for signal k which is assumed to be aligned in magnitude, its actual complex magnitude is

$$a_{\text{opt}} h_k \frac{\alpha_{\text{th}}}{\hat{h}_k} = a_{\text{opt}} \alpha_{\text{th}} (1 + \gamma_k),$$
$$\gamma_k \approx -\frac{\Delta_k}{h_k}. \tag{32}$$

For signal k which is assumed to be misaligned in magnitude, its actual complex magnitude is

$$a_{\text{opt}} h_k \frac{\hat{h}_k^*}{|\hat{h}_k|} \sqrt{P_{\max}} = a_{\text{opt}} h_k b_k (1 + \gamma_k),$$
$$1 + \gamma_k = \frac{|h_k|}{|\hat{h}_k|} \frac{\hat{h}_k^*}{h_k^*}, \quad \gamma_k \approx \frac{\Delta_k^*}{2h_k^*} - \frac{\Delta_k}{2h_k}. \tag{33}$$

Here, γ_k denotes how much the magnitude deviates from the ideal value, and decreases with channel estimation error Δ_k. Then, the variation in MSE(k) of signal k due to channel estimation error is

$$|a_{\text{opt}} h_k b_k (1 + \gamma_k) - 1|^2 - |a_{\text{opt}} h_k b_k - 1|^2$$
$$= a_{\text{opt}} h_k b_k (a_{\text{opt}} h_k b_k - 1)(\gamma_k + \gamma_k^*)$$
$$+ (a_{\text{opt}} h_k b_k)^2 \gamma_k \gamma_k^*. \tag{34}$$

For signals expected to be aligned, the variation in MSE is proportional to Δ_k. On the other hand, for signals expected to be misaligned, $\gamma_k + \gamma_k^* = 0$. The variation in MSE is proportional to $|\gamma_k|^2$. This error is small because it is the phase not the magnitude of the Tx-scaling factor that is affected by the channel estimation error. Because most signals are aligned in magnitude, the variation in the overall MSE has the same order as channel estimation error. But its actual value is small when $a_{\text{opt}} h_k b_k - 1$ approaches 0.

F. Numerical Analysis

Here, we do some numerical analysis, using the default setting in Table I unless specified otherwise. All links have the same average channel gain[5] ($\overline{g}_k = 10$) and instantaneous channel gain follows independent block Rayleigh fading.

Fig. 5 shows the average channel gain and the channel gain threshold (SelFirst(g_{th})) under different numbers of slots. At $N = 1$ slot, the average channel gain is 10, in both SelFirst and OptSel. Then, average channel gain increases with the number of slots, although there is a gap between SelFirst and OptSel. The increase of channel gain in SelFirst is because the threshold of channel gain increases with the number of slots. The gap between SelFirst and OptSel is because SelFirst does not know the channel gain of all slots in advance, and setting the threshold cannot ensure to select the optimal slot with the maximal gain, although it does help avoid deep fading.

[5]Here, channel gain involves the fixed gain amplification at the sink, including both the analog amplification which generates the noise and the digital amplification which makes noise power σ^2 equal to 1. This definition of channel gain applies to all simulation evaluations.

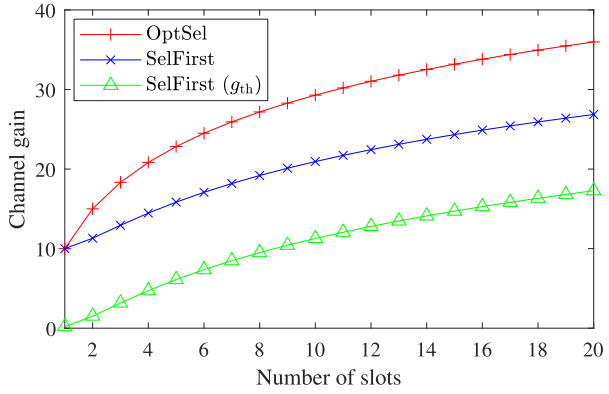


Fig. 5. Channel gain and its threshold under different numbers of slots ($\overline{g}_k = 10$, $P_{\text{th}} = 0.98$, $P_{\max} = 10$).

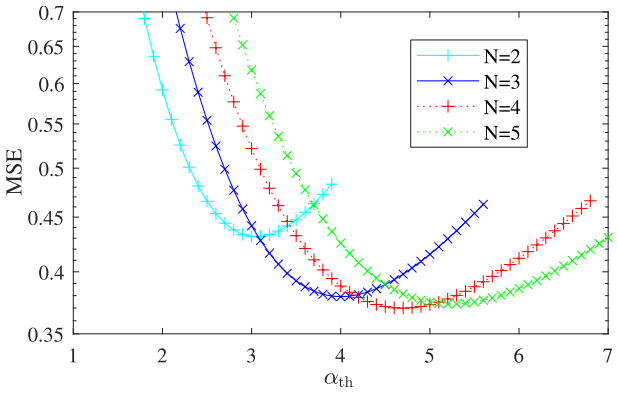


Fig. 6. Variation of computation MSE with respect to α_{th} ($\overline{g}_k = 10$, $P_{\text{th}} = 0.98$, $P_{\max} = 10$).

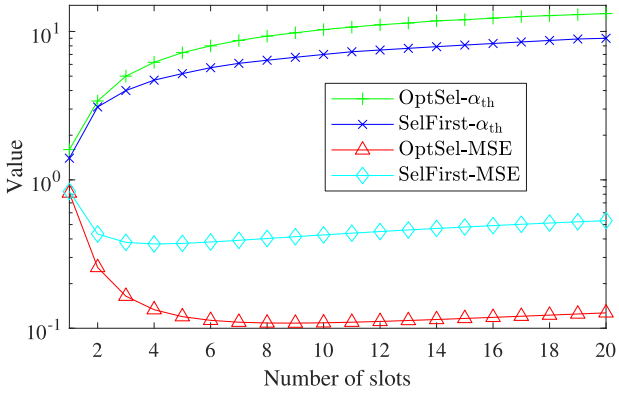


Fig. 7. Change of α_{th} and computation MSE with respect to N, the number of slots ($\overline{g}_k = 10$, $P_{\text{th}} = 0.98$, $P_{\max} = 10$).

Fig. 6 shows how the computation MSE varies with α_{th} in the SelFirst method. Clearly, the computation MSE reaches a minimum at some place, which depends on N, the number of slots.

We further investigate the impact of N, the number of slots. As shown in Fig. 7, with the increase of N, α_{th} increases in both OptSel and SelFirst. The computation MSE first decreases, and after reaching the minimum, increases again, which indicates that some N is optimal. When N is small, the benefit brought by multiple slots in improving channel gains is greater than its cost brought by the increase in noise power, which leads to a decrease in the computation MSE. OptSel has a larger α_{th}, and accordingly a smaller computation MSE.

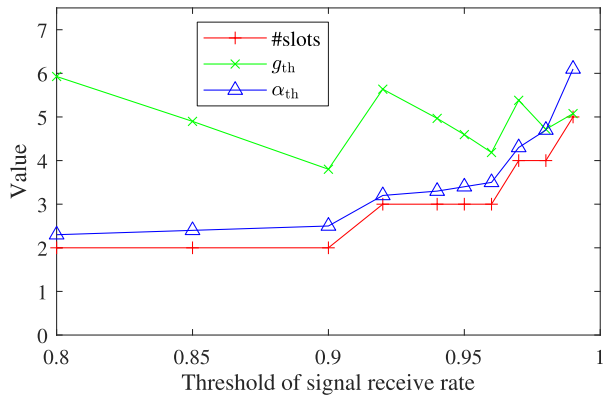


Fig. 8. Optimal parameters: N, $g_{\rm th}$ and $\alpha_{\rm th}$ under different $P_{\rm th}$ ($\overline{g}_k = 10$, $P_{\max} = 10$).

TABLE II
OPTIMAL PARAMETERS UNDER DIFFERENT $P_{\rm th}$

$P_{\rm th}$	0.90	0.92	0.94	0.95	0.96	0.97	0.98	0.99
N	2	3	3	3	3	4	4	5
$\alpha_{\rm th}$	2.5	3.2	3.3	3.4	3.5	4.3	4.7	6.1
$g_{\rm th}$	3.8	5.6	5.0	4.6	4.2	5.4	4.7	5.1

Fig. 8 shows how optimal parameters (N, $g_{\rm th}$, $\alpha_{\rm th}$) change with $P_{\rm th}$. With the increase of $P_{\rm th}$, the number of slots gradually increases. At the same number of slots, $g_{\rm th}$ decreases when $P_{\rm th}$ increases, because a smaller $g_{\rm th}$ ensures that more nodes transmit their signals to meet the requirement of $P_{\rm th}$. $\alpha_{\rm th}$ always increases with $P_{\rm th}$, which helps to reduce a in the MSE equation, and reduce the impact of noise, but at the cost of more transmission power because transmission power is proportional to $\alpha_{\rm th}^2$.

The optimal parameters are shown in Table II.

IV. SIMULATION EVALUATION

Here, we evaluate the proposed SelFirst method by Monte Carlo simulation, and compare it with the OptSel method and the AirComp method [11] described in Sec. II-A. The number of run is 10,000 for each setting.

We mainly consider three metrics, average transmission power, the computation MSE, and the probability of misalignment in signal magnitude. It is assumed that the instantaneous channel gains are known to the sink in AirComp. In comparison, the instantaneous channel gains are not known to the sink in the proposed SelFirst method. To make a fair comparison, OptSel uses the same number of slots as SelFirst does.

First, we use a scenario where all channels follow block Rayleigh fading and have the same average channel gain $\overline{g}_k = 10$. The parameters in Table II are used for SelFirst.

Fig. 9 shows the computation MSE in three methods, under different $P_{\rm th}$. Generally, SelFirst lies between OptSel and AirComp. At $P_{\rm th} = 0.8$, SelFirst and AirComp have almost the same computation MSE. As $P_{\rm th}$ increases, the computation MSE in the SelFirst method gradually decreases and approaches that of OptSel.

Fig. 10 shows average transmission power in three methods. Here, average transmission power in SelFirst and OptSel increases with $P_{\rm th}$. This is consistent with the result of $\alpha_{\rm th}$ in Fig. 8. Surprisingly, both SelFirst and OptSel consume more power than AirComp. It is clear that a tradeoff between the computation MSE and transmission power is necessary.

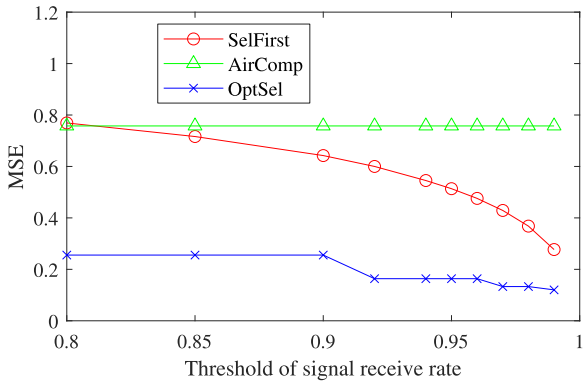


Fig. 9. Computation MSE under different $P_{\rm th}$ ($\overline{g}_k = 10$, $P_{\max} = 10$).

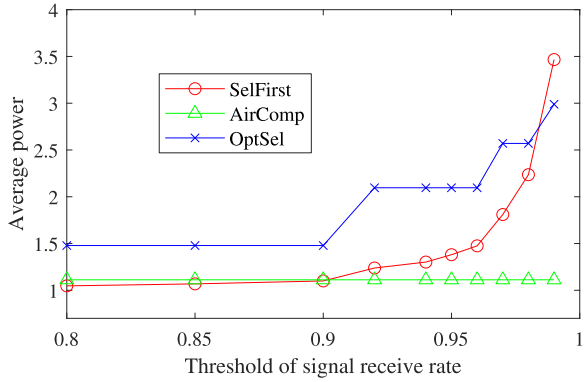


Fig. 10. Average transmission power under different $P_{\rm th}$ ($\overline{g}_k = 10$, $P_{\max} = 10$).

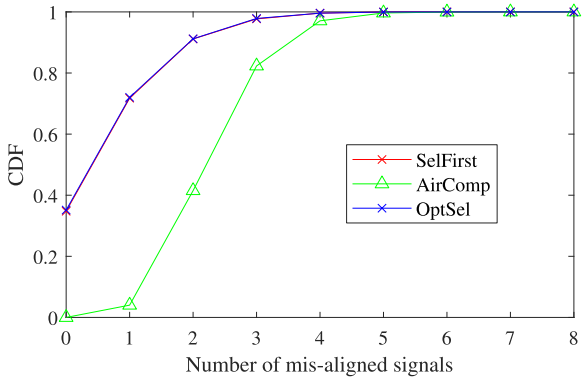


Fig. 11. Cumulative probability of signal misalignment in up to k signals ($\overline{g}_k = 10$, $P_{\rm th} = 0.98$, $P_{\max} = 10$).

Next, we investigate the probability of misalignment in signal magnitude. Fig. 11 shows the cumulative probability of signal misalignment in up to k signals, and the value k corresponds to the horizontal axis. Here, SelFirst and OptSel have almost the same performance. The average number of signals with signal magnitude misalignment is 1.05 in SelFirst, 1.04 in OptSel, while 2.75 in AirComp. These results are consistent with the analysis in Sec. III-D2.

V. TRADEOFF BETWEEN TRANSMISSION POWER AND MSE

The results in Fig. 9 and Fig. 10 have shown that a tradeoff is necessary between the computation MSE and transmission power.

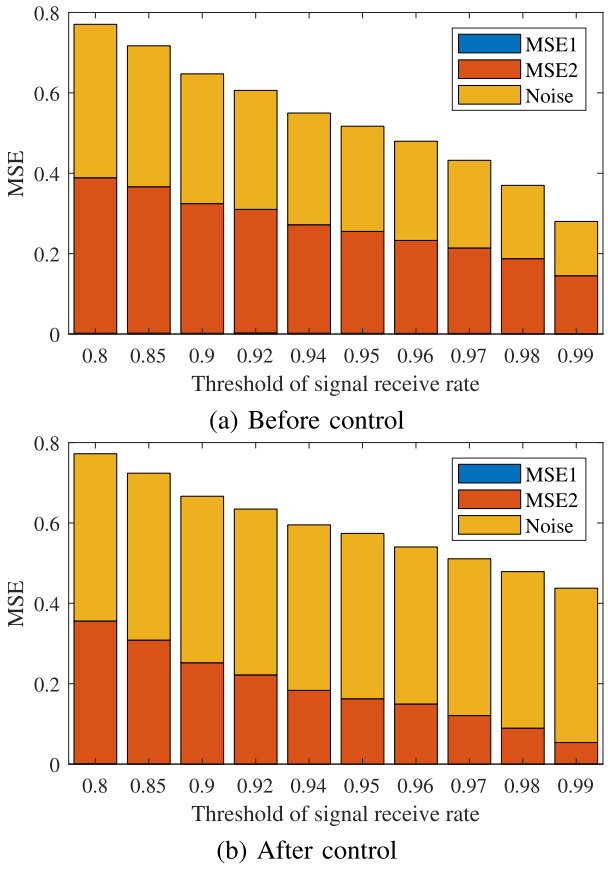


Fig. 12. Different components of computation MSE under different $P_{\rm th}$ ($\bar{g}_k = 10$, $P_{\max} = 10$). MSE1 is very small and invisible in the figure.

In the proposed SelFirst method, we investigate the three parts of MSE, MSE1 (signals with channel gain above the threshold), MSE2 (signals with channel gain below the threshold) and noise in (15). Fig. 12(a) shows the result. Generally, MSE1 and MSE2 are reduced by transmission power control. Here, MSE1 is almost negligibly small. By setting a threshold of channel gain ($g_{\rm th}$ increases with $P_{\rm th}$), MSE2 is further reduced. But surprisingly, noise is also reduced, although it has a factor of N (N increases with $P_{\rm th}$).

To explain this, we look back at (4). MSE of the signal part depends on $a\alpha_{\rm th}$, and MSE of the noise part is $Na^2\sigma^2$. Reducing the MSE of the noise part will lead to a small a. Because $a\alpha_{\rm th}$ approaches 1.0 (to reduce the computation error of the signal part, as discussed in Sec. III-C1), this will lead to a large $\alpha_{\rm th}$ and accordingly a large transmission power (which is proportional to $\alpha_{\rm th}^2$).

To solve this problem, we propose that the optimization is focused on the signal part, and avoid the over-reduction of the noise part. Specifically, $\max\{Na^2, \beta\}\sigma^2$ is used instead of $Na^2\sigma^2$ in (15) so that the optimization for the noise part will stop once the noise power is reduced to $\beta\sigma^2$, where β is an adjustable parameter. It should be noted that this is only used for finding the optimal parameters, not for the computation of MSE in the experiment. By setting β to 0.4 in SelFirst, the corresponding MSE parts are shown in Fig. 12(b). Here, the noise part is almost fixed while MSE of the signal part is reduced more compared with the result in Fig. 12(a).

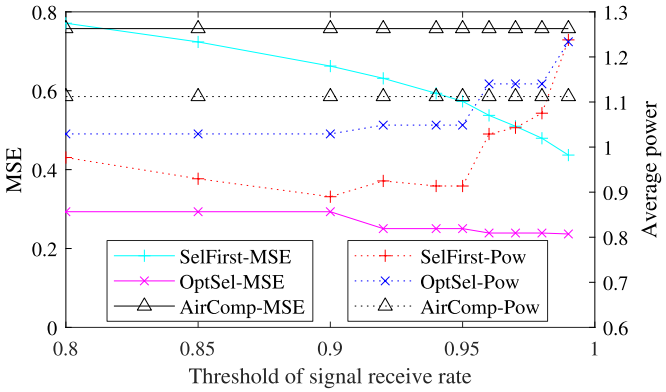


Fig. 13. Computation MSE and average transmission power under different $P_{\rm th}$ ($\bar{g}_k = 10$, $P_{\max} = 10$).

TABLE III
OPTIMAL PARAMETERS UNDER DIFFERENT $P_{\rm th}$ WHEN CONSIDERING THE TRANSMISSION POWER

$P_{\rm th}$	0.90	0.92	0.94	0.95	0.96	0.97	0.98	0.99
N	2	3	3	3	4	4	4	5
$\alpha_{\rm th}$	2.2	2.7	2.7	2.7	3.2	3.2	3.2	3.6
$g_{\rm th}$	3.8	5.6	5.0	4.6	6.0	5.4	4.7	5.1

The overall computation MSE and average transmission power of the three methods are shown in Fig. 13. Here, $\beta = 0.25$ is used in OptSel. As $P_{\rm th}$ increases, the reduction of the computation MSE by SelFirst becomes smaller compared with the result in Fig. 9, but the quick increase of transmission power is avoided, which confirms that the simple method is effective in balancing transmission power and the computation MSE. At $P_{\rm th} = 0.98$, SelFirst consumes almost the same transmission power as AirComp, but reduces the computation MSE by 36.7%.

The optimal parameters for Fig. 13 are shown in Table III. Compared with Table II, $a_{\rm th}$ is reduced much, but other parameters (N and $g_{\rm th}$) remain almost unchanged, which indicate that the optimization of transmission power does not lead to more slots (larger delay).

A. Evaluation in Real Scenarios With Non Equal Gains

Next we evaluate the three methods in a general scenario where all nodes have different average channel gains. Specifically, all nodes are uniformly distributed in a square area (200m $\times$ 200m), and the sink node is located at the center. The mean value of average channel gains in dB is 10, and each node has a channel gain depending on its location, which also varies with time according to block Rayleigh fading.

Fig. 14 shows how average transmission power and computation MSE change with $P_{\rm th}$ in three methods, where $K = 100$ nodes are used. This has a similar trend as in Fig. 13, but the variation of average channel gains has a little negative impact on the performance. At $P_{\rm th} = 0.98$, SelFirst still consumes almost the same transmission power as AirComp, but reduces the computation MSE only by 23.6%. This result is still promising considering that the sink needs to know instantaneous channel gains of all nodes in AirComp but only needs to know channel statistics in SelFirst.

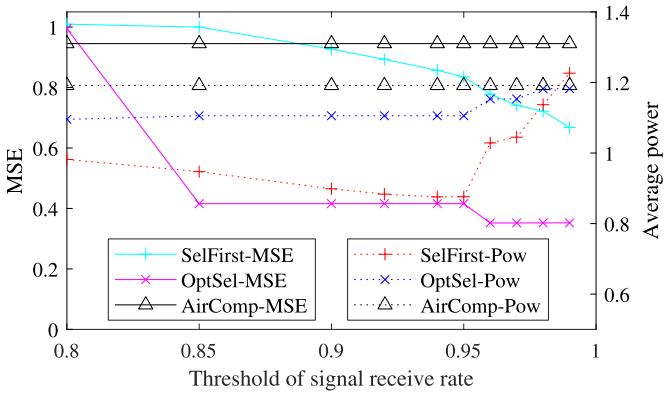


Fig. 14. Computation MSE and average transmission power under different $P_{\rm th}$ in a general scenario where average channel gain of a node varies with its location (The mean value of average channel gains in dB is 10, $K = 100$, $P_{\max} = 10$).

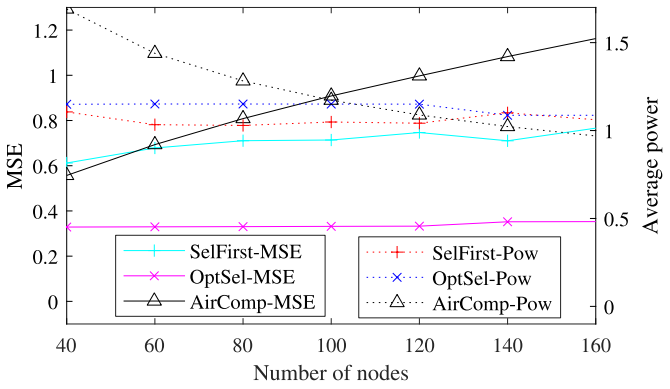


Fig. 15. Computation MSE and average transmission power under different numbers of nodes in a general scenario where average channel gain of a node varies with its location (The mean value of average channel gains in dB is 10, $P_{\rm th} = 0.98$, $P_{\max} = 10$).

Next, we fix $P_{\rm th}$ to 0.98 and change the number of nodes (but ensure that the mean value of average channel gains in dB is 10). Fig. 15 shows how average transmission power and computation MSE change with the number of nodes. Generally, the computation MSE increases with the number of nodes in AirComp, because it becomes more difficult to align signals. The computation MSE of OptSel is almost unchanged, by always selecting the optimal slot. The computation MSE of SelFirst lies between that of AirComp and OptSel, and increases more slowly compared with that of AirComp. Considering that SelFirst consumes less or almost the same average transmission power compared with AirComp, we can say that SelFirst is more scalable with the number of nodes.

VI. Conclusion

To deal with fast channel fading in over-the-air computation, this paper has extended the conventional AirComp method, by distributing data transmission/fusion to multiple slots. A threshold of channel gain is set to make a node transmit its signal when its channel gain gets above the threshold in any slot, or in the last slot if its channel gain remains below the threshold. This helps to avoid deep fading and improve channel gains, which further facilitate the transmission power control for the alignment of signal magnitudes. Theoretical analysis gives the closed-form of the computation MSE, which helps to find optimal parameters. The transmission scheduling at each node is conducted in a distributed way, and works well in the fast fading environment, which avoids the overhead of feeding back instantaneous channel coefficients from each node to the sink. By avoiding over-reducing MSE of the noise part, the proposed method reduces the computation MSE while consuming less or almost the same power as AirComp.

Signals of nodes far from the sink may be misaligned in signal magnitude. In the future, we will try to apply a relay model to solve this problem, and further improve system performance by using multiple antennas at the sink.

References

[1] R. S. Sinha, Y. Wei, and S.-H. Hwang, "A survey on LPWA technology: LoRa and NB-IoT," *ICT Exp.*, vol. 3, no. 1, pp. 14–21, Mar. 2017. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S2405959517300061

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

[3] G. Zhu, J. Xu, K. Huang, and S. Cui, "Over-the-air computing for wireless data aggregation in massive IoT," *IEEE Wireless Commun.*, vol. 28, no. 4, pp. 57–65, Aug. 2021.

[4] J. J. Xiao, S. Cui, Z. Q. Luo, and A. J. Goldsmith, "Linear coherent decentralized estimation," *IEEE Trans. Signal Process.*, vol. 56, no. 2, pp. 757–770, Feb. 2008.

[5] P. Popovski and H. Yomo, "Physical network coding in two-way wireless relay channels," in *Proc. IEEE Int. Conf. Commun.*, Jun. 2007, pp. 707–712.

[6] C.-Z. Lee, L. P. Barnes, and A. Ozgur, "Over-the-air statistical estimation," in *Proc. IEEE Global Commun. Conf.*, Dec. 2020, pp. 1–6.

[7] R. C. Buck, "Approximate complexity and functional representation," *J. Math. Anal. Appl.*, vol. 70, pp. 280–298, Jul. 1979.

[8] M. Goldenbaum, H. Boche, and S. Stańczak, "Nomographic functions: Efficient computation in clustered Gaussian sensor networks," *IEEE Trans. Wireless Commun.*, vol. 14, no. 4, pp. 2093–2105, Apr. 2015.

[9] O. Abari, H. Rahul, and D. Katabi, "Over-the-air function computation in sensor networks," 2016, *arXiv:1612.02307*.

[10] H. Ye, G. Y. Li, and B.-H.-F. Juang, "Deep over-the-air computation," in *Proc. IEEE Global Commun. Conf.*, Dec. 2020, pp. 1–6.

[11] W. Liu, X. Zang, Y. Li, and B. Vucetic, "Over-the-air computation systems: Optimization, analysis and scaling laws," *IEEE Trans. Wireless Commun.*, vol. 19, no. 8, pp. 5488–5502, Aug. 2020.

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

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

[14] J. Dong, Y. Shi, and Z. Ding, "Blind over-the-air computation and data fusion via provable Wirtinger flow," *IEEE Trans. Signal Process.*, vol. 68, pp. 1136–1151, 2020.

[15] A. Goldsmith, *Wireless Communication*. New York, NY, USA: Cambridge Univ. Press, 2005.

[16] M. Goldenbaum and S. Stanczak, "Robust analog function computation via wireless multiple-access channels," *IEEE Trans. Commun.*, vol. 61, no. 9, pp. 3863–3877, Sep. 2013.

[17] O. Abari, H. Rahul, D. Katabi, and M. Pant, "AirShare: Distributed coherent transmission made seamless," in *Proc. IEEE Conf. Comput. Commun. (INFOCOM)*, Apr. 2015, pp. 1742–1750.

[18] C.-H. Wang, A. S. Leong, and S. Dey, "Distortion outage minimization and diversity order analysis for coherent multiaccess," *IEEE Trans. Signal Process.*, vol. 59, no. 12, pp. 6144–6159, Dec. 2011.

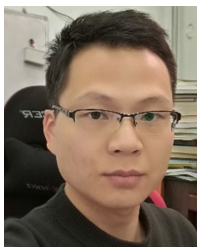
[19] S. Tang, H. Yin, C. Zhang, and S. Obana, "Reliable over-the-air computation by amplify-and-forward based relay," *IEEE Access*, vol. 9, pp. 53333–53342, 2021.

[20] S. Tang, H. Yomo, C. Zhang, and S. Obana, "Node scheduling for AF-based over-the-air computation," *IEEE Wireless Commun. Lett.*, vol. 11, no. 9, pp. 1945–1949, Sep. 2022.

[21] W. Fang, Y. Jiang, Y. Shi, Y. Zhou, W. Chen, and K. B. Letaief, "Over-the-air computation via reconfigurable intelligent surface," *IEEE Trans. Commun.*, vol. 69, no. 12, pp. 8612–8626, Dec. 2021.
[22] D. Wen, G. Zhu, and K. Huang, "Reduced-dimension design of MIMO over-the-air computing for data aggregation in clustered IoT networks," *IEEE Trans. Wireless Commun.*, vol. 18, no. 11, pp. 5255–5268, Nov. 2019.
[23] X. Li, G. Zhu, Y. Gong, and K. Huang, "Wirelessly powered data aggregation for IoT via over-the-air function computation: Beamforming and power control," *IEEE Trans. Wireless Commun.*, vol. 18, no. 7, pp. 3437–3452, Jul. 2019.
[24] T. Sery and K. Cohen, "On analog gradient descent learning over multiple access fading channels," *IEEE Trans. Signal Process.*, vol. 68, pp. 2897–2911, 2020.
[25] T. Sery, N. Shlezinger, K. Cohen, and Y. C. Eldar, "COTAF: Convergent over-the-air federated learning," in *Proc. IEEE Global Commun. Conf.*, Dec. 2020, pp. 1–6.
[26] M. M. Amiri and D. Gündüz, "Federated learning over wireless fading channels," *IEEE Trans. Wireless Commun.*, vol. 19, no. 5, pp. 3546–3557, May 2020.
[27] K. Yang, T. Jiang, Y. Shi, and Z. Ding, "Federated learning via over-the-air computation," *IEEE Trans. Wireless Commun.*, vol. 19, no. 3, pp. 2022–2035, Mar. 2020.
[28] M. Frey, I. Bjelaković and S. Stańczak, "Over-the-air computation in correlated channels," *IEEE Trans. Signal Process.*, vol. 69, pp. 5739–5755, Aug. 2021.
[29] X. Qin and R. A. Berry, "Distributed approaches for exploiting multiuser diversity in wireless networks," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 392–413, Feb. 2006.
[30] S. Adireddy and L. Tong, "Exploiting decentralized channel state information for random access," *IEEE Trans. Inf. Theory*, vol. 51, no. 2, pp. 537–561, Feb. 2005.
[31] S. Tang, "Distributed multiuser scheduling for improving throughput of wireless LAN," *IEEE Trans. Wireless Commun.*, vol. 13, no. 5, pp. 2770–2781, May 2014.
[32] H. A. David and H. N. Nagaraja, *Order Statistics*, 3rd ed. Hoboken, NJ, USA: Wiley, 2003.

Suhua Tang (Senior Member, IEEE) received the B.S. degree in electronic engineering and the Ph.D. degree in information and communication engineering from the University of Science and Technology of China, in 1998 and 2003, respectively. From October 2003 to March 2014, he was with the Adaptive Communications Research Laboratories, ATR, Japan. Since April 2014, he has been with the Graduate School of Informatics and Engineering, The University of Electro-Communications, Japan. His research interests include green communications, ad hoc and sensor networks, inter-vehicle communications, and high precision positioning.

Petar Popovski (Fellow, IEEE) received the Dipl.-Ing. and M.Sc. degrees in communication engineering from the University of Ss. Cyril and Methodius in Skopje and the Ph.D. degree from Aalborg University in 2005. He is a Professor at Aalborg University, where he heads the Section on Connectivity and a Visiting Excellence Chair with the University of Bremen. He has authored the book "*Wireless Connectivity: An Intuitive and Fundamental Guide*." His research interests are in the area of wireless communication and communication theory. He was a Member-at-Large with the Board of Governors in IEEE Communication Society from 2019 to 2021. He received an ERC Consolidator Grant in 2015, the Danish Elite Researcher Award in 2016, the IEEE Fred W. Ellersick Prize in 2016, the IEEE Stephen O. Rice Prize in 2018, the Technical Achievement Award from the IEEE Technical Committee on Smart Grid Communications in 2019, the Danish Telecommunication Prize in 2020, and the Villum Investigator Grant in 2021. He serves as the Vice-Chair for the IEEE Communication Theory Technical Committee and the Steering Committee of IEEE TRANSACTIONS ON GREEN COMMUNICATIONS AND NETWORKING. He was the General Chair of IEEE SmartGridComm 2018 and IEEE Communication Theory Workshop 2019. He is currently the Editor-in-Chief of IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS.



Chao Zhang received the B.Sc. and Ph.D. degrees in electronics and information engineering from the University of Science and Technology of China (USTC), Hefei, Anhui, China, in 2005 and 2010, respectively. From December 2018 to December 2019, he was a Visiting Professor with Iowa State University, Ames, IA, USA. Since July 2010, he has been with Xi'an Jiaotong University, Xi'an, China, where he is currently an Associate Professor with the Faculty of Electronic and Information Engineering, School of Information and Communications Engineering. His research interests include networking and wireless communications, green communications, satellite communications, and 6G related topics.

Sadao Obana received the B.E., M.E., and Ph.D. degrees from Keio University, Tokyo, Japan, in 1976, 1978, and 1993 respectively. After joining KDDI (former KDD) in 1978, he was engaged in research and development in the field of packet exchange systems, network architecture, open systems interconnection (OSI) protocols, database, distributed processing, network management, and intelligent transport systems (ITS). In 2004, he was the Director of the Adaptive Communications Research Laboratories, Advanced Telecommunication Research Institute International (ATR). From 2011 to 2018, he was a Professor with the Graduate School of Informatics and Engineering, The University of Electro-Communications, Japan, where he is currently an Executive Director. He is a member of the Institute of Electronics, Information and Communication Engineers (IEICE) and a fellow of the Information Processing Society of Japan (IPSJ). He received the Award of Minister of Education, Culture, Sports, Science and Technology in 2001.