

Article

Relay Selection for Over-the-Air Computation Achieving Both Long Lifetime and High Reliability

Jingyang Zhou and Suhua Tang * 

Department of Computer and Network Engineering, The University of Electro-Communications,
Tokyo 182-8585, Japan

* Correspondence: shtang@uec.ac.jp; Tel.: +81-42-443-5308

Abstract: In a general wireless sensor network, a sink node collects data from each node successively and then post-processes the data to obtain useful information. However, conventional methods have a scalability problem: the data collection/processing time increases with the number of nodes, and frequent transmission collisions degrade spectrum efficiency. If only statistical values of the data are needed, using over-the-air computation (AirComp) can efficiently perform data collection and computation. However, AirComp also has its problems: when the channel gain of a node is too low, (i) the transmission power of that node will be high, decreasing the lifetime of that node and the entire network, and (ii) sometimes, the computation error still occurs even though the maximal transmission power is used. To jointly solve these two problems, in this paper we investigate the relay communication for AirComp and study a relay selection protocol. The basic method selects an ordinary node with a good channel condition as a relay node, considering both computation error and power consumption. This method is further enhanced to explicitly consider network lifetime in relay selection. Extensive simulation evaluations confirm that the proposed method helps to prolong the lifetime of the entire network and reduce computation errors as well.

Keywords: Internet of Things; over-the-air computation (AirComp); relay selection



Citation: Zhou, J.; Tang, S. Relay Selection for Over-the-Air Computation Achieving Both Long Lifetime and High Reliability. *Sensors* **2023**, *23*, 3824. <https://doi.org/10.3390/s23083824>

Academic Editors: Abbass Nasser, Christophe Moy, Ali Mansour, Hadi Aggoune, Mohammad Ayaz and Koffi Yao

Received: 28 February 2023

Revised: 30 March 2023

Accepted: 6 April 2023

Published: 8 April 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

With the wide-spread of the Internet of Things, there will be more and more context-aware applications that collect environmental information through sensor nodes (hereinafter referred to as nodes) and make corresponding actions in real time in the future smart society. Nodes will be deployed and connected to the Internet through technologies such as NB-IoT and LoRa [1]. Generally, during the data collection process, the sink node (hereinafter referred to as sink) needs to receive data from each node one by one. When there are tens of thousands of nodes within the coverage of a sink, it will take a lot of time to receive all the data. In addition, the nodes share the same channel in the communication. If the network uses the CSMA (carrier-sense multiple access) protocol for channel access, the increase in the number of nodes will lead to frequent transmission collisions, which further degrades spectrum efficiency.

In some IoT applications, people only want the statistics of the data (e.g., the average temperature or the average noise level in a certain area and so on), not the individual data collected by nodes. For such applications, there is an efficient method called over-the-air computation (AirComp) [2,3]. In AirComp, nodes first preprocess the data, and then all nodes simultaneously transmit the preprocessed data through analog signals [4]. All signals add together at the antenna of the sink, which corresponds to the sum operation. By using analog AirComp, the computation error can be made smaller than that of digital schemes when using the same amount of resources (power and bandwidth) [5]. As well as the sum operation, AirComp also can support any kind of nomographic functions by proper preprocessing and postprocessing [6–8]. Recently, deep AirComp has been studied, using

federated learning in the pre-processing and post-processing stages, which enables more advanced processing of sensor data [9].

Typically, in order to ensure unbiased data fusion, all nodes use transmission power control [10,11] to achieve a consistent signal magnitude when all signals arrive at the sink. Two problems may occur in the original AirComp model:

- For nodes far away from the sink with low channel gain, greater transmission power is required to reach the target magnitude in order to minimize the computation error. This will result in a short lifetime of nodes and the entire network.
- Under the transmission power limit, some nodes cannot reach the target magnitude even though they use the maximum transmission power. This leads to the computation error, which is measured by mean squared error (MSE).

Various methods have been studied to reduce the MSE from different aspects. The basic method is to apply transmission power control [10,11] using high power when the channel gain is low. However, this alone cannot well solve the problem when some nodes are in deep fading or far from the sink.

Conventional methods to channel fading can be applied in AirComp, but simultaneous transmission of multiple signals needs to be taken into account carefully. Time diversity is considered in [12], where the computation is distributed to multiple slots, and each node can independently select one slot to transmit its signal based on its channel gain. The optimal number of slots is formulated, considering the benefit of channel gain improvement and the demerit of increased noise power due to using multiple slots. A counterpart in the frequency domain is studied in [13], where the computation is distributed over multiple channels.

Another effective method to channel fading is path diversity, e.g., relay for network coding [14]. Because AirComp usually works for analog signals, amplify-and-forward based relay is investigated in [15] using a dedicated relay node to help nodes with low channel gains. Two relay policies, Simple Relay Policy (SRP) and Coherent Relay Policy (CRP), are studied. The two relay policies are further refined with a new metric, by which the transmission power is reduced while the MSE is kept low. However, the transmission power of the relay node may be much higher than that of normal nodes, leading to a surge in power consumption. In addition, the impact of the relay node on the network lifetime is not examined in this article. Under the constraint of transmission power, not all nodes can use the relay. In [16], node scheduling, deciding which nodes can use the relay, is studied. Recently, new mechanisms, such as intelligent reflective surfaces, were studied in [17]. The phases of reflective items can be adjusted so that signals add constructively at the sink. However, signals are not amplified at the reflective surface, and usually many reflective items are required to reach a reasonable signal strength. In addition, when multiple signals use the same reflective surface, the control of the surface phase is a non-convex problem.

It is also possible to use multiple antennas at the sink [18]. However, the alignment in channel gain is different from beamforming [19] because in AirComp different signals add together, not necessarily enhancing each other. The control of the antenna array is a non-convex problem, and it is difficult to find the optimal solution. Extending the distance between antennas, simultaneous receiving at multiple sinks is studied in [20], where the interference needs to be taken into account.

The knowledge of channel gains at the sink and the synchronization between nodes and the sink usually are assumed, which, however, is non-trivial. Channel estimation is studied in [18], using an iterative method with multiple antennas. The statistics of channel gains instead of their instantaneous values are exploited in [12]. A more aggressive policy, blind AirComp, is explored in [21].

Typically, it is assumed that signals from all sensors are non-correlated, which simplifies the analysis. However, when sensors are densely deployed, or in some specific applications (e.g., when AirComp is used for model aggregation in federated learning), signals from multiple nodes are correlated. This problem is explored in [22,23], which requires the correlation information and is not well solved yet.

As for power consumption, in [24], Zang et al. try to minimize the sum power of the sensors under the constraint of MSE. The results show that the sensors with poor or good channel conditions should use less power than the ones with moderate channel conditions to reduce the MSE and the sum power of nodes. However, the random variation of channel gains is not taken into account.

To solve the aforementioned problems, in this paper, we investigate relay communication for AirComp, and use relay selection to both extend network lifetime and reduce MSE. To the best of our knowledge, this is the first work on relay selection for AirComp. In this method, each node and the sink are equipped with a single antenna. It is assumed that signals from all nodes are non-correlated and channel gains are known at the sink. We first calculate a set of candidate relay nodes based on channel gains. Then, an evaluation metric, considering transmission power, remaining battery capacity, and MSE, is computed for each candidate, based on which the optimal relay is selected. In addition, the relay node is periodically selected at regular intervals to avoid over-consumption of power at the relay node. This basic method is further enhanced, explicitly considering the network lifetime in the relay selection. Extensive simulation evaluations show that the basic method can extend the network lifetime and reduce MSE, while the enhanced method can further prolong the network lifetime under the same MSE.

In the rest of this paper, Section 2 reviews the related work on the AirComp model, especially the relay transmission methods. Then, Section 3 presents the proposed method, both the selection of candidate relay nodes and the metric for evaluating the suitability of candidate relay nodes. Section 4 describes the simulation setting and analyzes the results and factors affecting the lifetime and MSE. Finally, Section 5 concludes this paper and points out future works.

2. Related Research

2.1. Over-the-Air Computation [10]

A sensor network consisting of K nodes and one sink is considered, as shown in Figure 1. Node k synchronizes its time with the sink using the control signal from the sink, preprocesses its data (s_k) by $\psi(\cdot)$ and transmits its data as an analog signal ($x_k \in \mathbb{C}$, $|x_k| \leq v$, $E\{|x_k|^2\} = 1$) at the same time together with other nodes. These signals are summed up at the sink antenna, which is postprocessed by $\phi(\cdot)$. For simplicity, here, the sum operation is considered: $\psi(t) = t$ is used in the preprocessing and $\phi(t) = at$ is used in the postprocessing, where a is the Rx-scaling parameter. In order to eliminate the difference in channel coefficient (h_k) among nodes, the transmission power is adjusted by a Tx-scaling factor b_k and the magnitudes ($h_k b_k$) of all signals are aligned to the same value if the required power is less than the maximum. The received signal at the sink is

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

where n is additive white Gaussian noise (AWGN) with mean value 0 and variance σ^2 . a , h_k , and b_k are complex numbers. Because it is possible to adjust the phase of b_k to ensure that $h_k b_k$ is positive real, in the analysis, usually it is assumed that a , b_k , and h_k are positive real for simplicity [10]. In the evaluation, complex numbers are used instead.

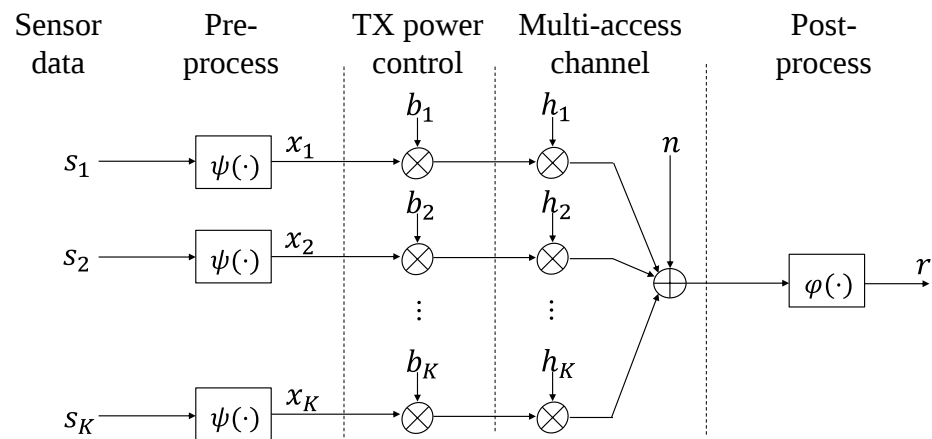


Figure 1. Communication model of AirComp.

Assume all signals are non-correlated and independent of the noise. The mean squared error ($\text{MSE} = \mathbb{E}\{|r - \sum_{k=1}^K x_k|^2\}$), between the received signal r and the target result $\sum_{k=1}^K x_k$, is

$$\text{MSE} = \sum_{k=1}^K |ah_k b_k - 1|^2 + |a|^2 \sigma^2, \quad (2)$$

where $P_k = b_k^2 \leq P_{\max}$. This MSE may be caused by channel fading and noise and is jointly determined by a and b_k .

As shown in Figure 2, when the channel coefficients h_k of some nodes are too low, the target magnitude cannot be reached even with the maximum transmission power. Then, MSE occurs in the received signal.

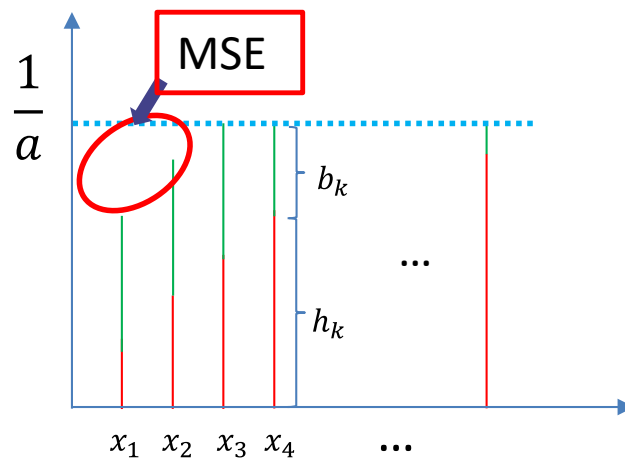


Figure 2. MSE in the signal caused by low channel gain (the length of a line represents the absolute value of a variable).

2.2. Non-Relay Methods for Reducing Power Consumption

In [24], Zang et al. transform the non-convex problem in Equation (2) into a convex problem by setting $\hat{b}_k = ab_k$ as follows:

$$\min_{\{\hat{b}_k\}} \text{MSE} = \sum_{k=1}^K |h_k \hat{b}_k - 1|^2 + \frac{\sigma^2}{P_{\max}} \sum_{k=1}^K |\hat{b}_k|^2. \quad (3)$$

Then, they minimize the sum power of the sensors under the constraint of MSE. The results in [24] show that the sensors with poor or good channel conditions should use less power than the ones with moderate channel conditions to reduce the MSE and the sum power of

nodes. However, they did not consider random channel gains (h_k) but set the channel gains of all the nodes in the simulation to be a series of equal differences, which is not practical.

2.3. Relay Methods

The amplify-and-forward-based relay model in [15] is shown in Figure 3. In addition to nodes and sink d , a dedicated relay node r is added. The nodes are divided into two groups (N_r and $\bar{N}_r$) according to channel gains between nodes and the relay node r /sink d . A general policy is that nodes ($k \in N_r$), far from the sink but close to the relay, use the relay node r , and nodes ($k \in \bar{N}_r$) close to the sink send directly to the sink.

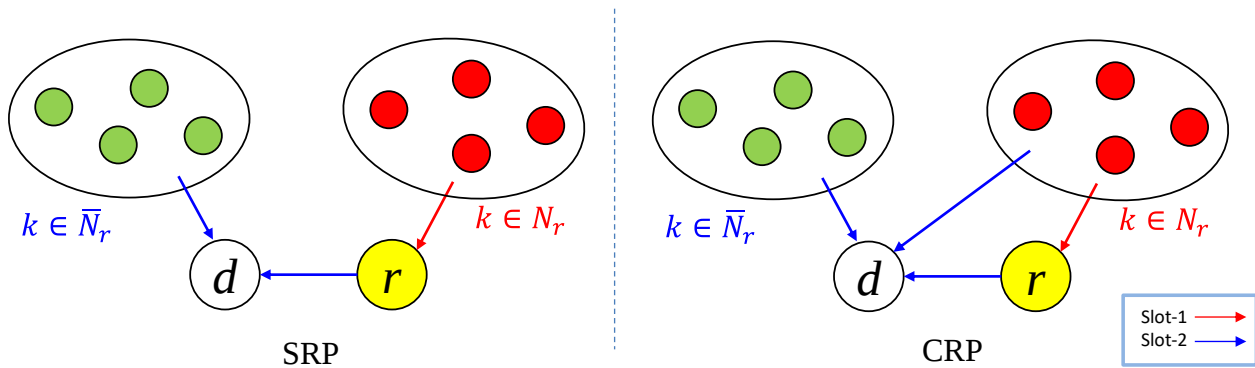


Figure 3. Relay communication policy of SRP and CRP (A circle represents a group of nodes).

The communication is performed in two slots. $b_{k,1}, b_{k,2} \in \mathbb{C}$ reflect the transmission power of node k in two slots. $n_{r,1}$ and $n_{d,2}$ are the AWGN at r in the first slot and at d in the second slot, respectively. $a_{r,1}, a_{d,2} \in \mathbb{C}$ are the receive factors at r and d , respectively.

In the first slot, nodes using relay ($k \in N_r$) transmit their signals to relay node r . The signal received at relay r is

$$s_{r,1} = a_{r,1} \cdot \left(\sum_{k \in N_r} h_{k,r} b_{k,1} x_k + n_{r,1} \right). \quad (4)$$

Because the relay node helps forward multiple signals in the second slot, it tends to consume a lot of power. The transmission power of the relay node is

$$P_{r,TX} = \left| \frac{1}{a_{d,2} h_{r,d}} \right|^2 \cdot \sum_{k \in N_r} |a_{r,1} h_{k,r} b_{k,1}|^2. \quad (5)$$

Based on the difference of communication nodes in the second slot, the relay model can be divided into two types: SRP and CRP.

2.3.1. Relay Communication Policy—SRP

In SRP, in the second slot, nodes not using relay ($k \in \bar{N}_r$) transmit their signals to the sink and the relay node forwards its received signal to the sink. The signal received at the sink is

$$s_{d,2} = a_{d,2} \cdot \left(\sum_{k \in \bar{N}_r \cup \{r\}} h_{k,d} b_{k,2} x_k + n_{d,2} \right). \quad (6)$$

Each node only transmits in one slot. For $k \in N_r$, $b_{k,1} \in \mathbb{C}$, $b_{k,2} = 0$. For $k \in \bar{N}_r$, $b_{k,1} = 0$, $b_{k,2} \in \mathbb{C}$.

2.3.2. Relay Communication Policy—CRP

The CRP communication method is shown on the right side of Figure 3. In the second slot, the CRP allows all nodes to transmit signals to the sink, not just $k \in \bar{N}_r$ and the relay node. Then, the signal received at the sink is

$$s_{d,2} = a_{d,2} \cdot \left(\sum_{k \in N_r \cup \bar{N}_r \cup \{r\}} h_{k,d} b_{k,2} x_k + n_{d,2} \right). \quad (7)$$

For $k \in N_r$, each node transmits its signal in two slots, and there is an overall power constraint per node, $|b_{k,1}|^2 + |b_{k,2}|^2 \leq P_{max}$. For $k \in \bar{N}_r$, $b_{k,1} = 0$, $b_{k,2} \in \mathbb{C}$.

From nodes $k \in N_r$, the sink receives two copies of the same signal in the second slot: one with a magnitude $A_{k,r}$, forwarded by the relay, and the other with a magnitude $A_{k,d}$, sent directly by a node itself. With proper control of the Tx-scale phases, the two copies of the same signal add coherently at the sink, and the overall magnitude is

$$A_k = A_{k,r} + A_{k,d}. \quad (8)$$

With the coherent combination of two copies of the same signal, in CRP, at the first slot, nodes using relay use a part of their power to transmit their signals to the relay node, and the relay node does not need to forward the signal with full magnitude in the second slot. This helps to reduce the transmission power of the relay node.

2.4. The Problems

The above relay methods can effectively reduce the MSE of AirComp through relay transmission. However, there are still two problems:

- The relay node is a dedicated node without power constraint, which increases the system cost.
- Reducing MSE usually is achieved at the cost of increased transmission power at the nodes, which decreases the network lifetime.

3. Proposed Method

In this paper, instead of using a dedicated relay node, we propose a new method to properly select an ordinary node as a relay node in the AirComp network. This method not only considers the power consumption and maximum power constraint of nodes in the network but also reduces the MSE as much as possible. Therefore, we call it relay Selection considering both Lifetime And MSE (SLAM). Here, network lifetime is defined as the difference between the beginning of communications in the network and the time when the first node in the network depletes its energy.

It is already proven that CRP works better than SRP in the amplify-and-forward-based relay [15]. Therefore, in the following we mainly study how to dynamically select a relay for CRP.

3.1. System Model

All nodes and the sink use a single antenna. Nodes synchronize their time with the sink. The sink knows the channel coefficient (h) between itself and all nodes and the channel coefficients among all nodes, and performs centralized relay selection. The estimation of channel gain is left for future work. Block fading is assumed, i.e., the channel coefficient is stable within two slots.

Because AirComp is performed on analog signals, amplify-and-forward (AF)-based relay is used, in the same way as in [15]. However, different from the dedicated relay node used in [15], we chose an ordinary node in the network as a relay node. This choice also caused us to pay attention to the transmission power of the relay node to avoid an

over-increase in its power consumption. Because the relay node has to transmit its own signal as well, its transmission power is

$$P_{r,TX} = \left| \frac{1}{a_{d,2}h_{r,d}} \right|^2 \cdot \left(1 + \sum_{k \in N_r} |a_{r,1}h_{k,r}b_{k,1}|^2 \right). \quad (9)$$

The transmission power of the relay node is decided by the nodes using the relay, under the maximal power constraint. Therefore, it always approaches, but is less than, the maximal power constraint.

The proposed method can be roughly divided into two modules—the calculation of candidate relay nodes and the selection of a relay node from the candidates, as shown in Figure 4.

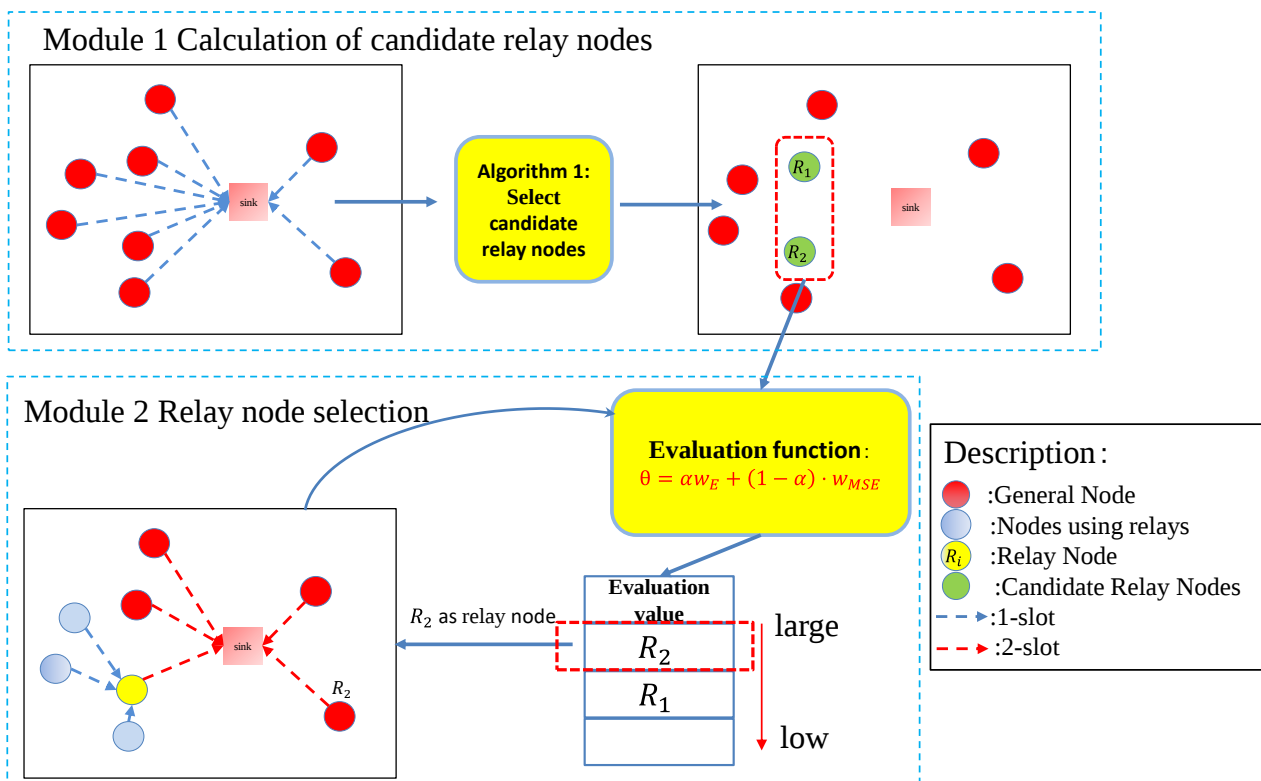


Figure 4. System model of SLAM.

3.2. Calculation of Candidate Relay Nodes

Two conditions are given for deciding whether a node can be used as a relay node for over-the-air computation, as follows:

- The transmission power of a node using the relay is less than that in the basic AirComp model, i.e., the relay node should help to reduce the transmission power of other nodes using relay.
- The relay node's transmission power is less than the maximum value.

We designed Algorithm 1 to select candidate relay nodes. In Algorithm 1, we regarded each ordinary node in turn as a relay node. When regarding node i as a relay node (Line 5), we computed the difference of channel gains to relay and sink as D (Line 7) and sorted nodes in an ascending order of D (Line 8). This is to let nodes far away from the sink but near to the relay use the relay node first. Then, with each possible number of nodes that can use node i as the relay, we recorded the corresponding MSE (Line 9–13). Finally, we found the optimal number of nodes using relay according to the MSE value, and recorded nodes' power and MSE (Line 14). After all nodes were iterated, we sorted MSE _{i}

in an ascending order (Line 16) and selected a number of top nodes as relay candidates R (Line 17).

Algorithm 1: Select candidate relay nodes

```

1. Procedure: Find candidate relay nodes
2. Input:  $h_{k,d}$ : Channel coefficient between sink and node,
            $h_{i,j}$ : Channel coefficient between nodes  $i$  and  $j$ ;
3. Output:  $R$ : A set of candidate relay nodes;
4. For  $i = 1, 2, \dots, K$  do
5.   | Regard node  $i$  as a relay node  $r$ 
6.   | Get  $h_{k,r}$  (channel coefficient between relay  $r$  and node  $k$ );
7.   | Calculate the difference of channel gain to relay and sink,  $D = \{|h_{k,d}| - |h_{k,r}|\}$ ;
8.   | Sort  $D$  in the ascending order;
9.   | For  $j = 1, 2, \dots, K, j \neq i$  do
10.    | | Assume top  $j$  nodes in  $D$  use the relay node;
11.    | | Minimize MSE of CRP, with power constraint for all nodes including relay;
12.    | | Calculate transmission power ( $P_j^i$ ) of all nodes;
13.    | End
14.    | Find optimal number of nodes that can use node  $i$  as relay and record  $\text{MSE}_i$ 
15.  End
16. Sort  $\text{MSE}_i$  in the ascending order;
17. Select a number of top nodes as relay candidates  $R$  based on  $\text{MSE}_i$ ;
18. Return  $R$ 

```

3.3. Relay Selection Considering Both MSE and Remaining Energy

Here, we present the evaluation metric for selecting a node from the candidate relay nodes as the relay.

3.3.1. Evaluation Metric for Selecting Relay

Relay selection should consider both MSE and the remaining energy of candidate relay nodes. Usually, it is preferred to select for the relay a node that both has a large remaining energy (long lifetime) and a small MSE. To this end, we designed a new metric to evaluate the suitability of each candidate as a relay, considering both factors, as follows:

$$\theta_i = \alpha \cdot w_{E,i} + (1 - \alpha) \cdot w_{\text{MSE},i}, \quad (10)$$

$$w_{E,i} = \frac{E_i - \min_{i \in R} E_i}{\max_{i \in R} E_i - \min_{i \in R} E_i}, w_{\text{MSE},i} = - \frac{\text{MSE}_i - \min_{i \in R} \text{MSE}_i}{\max_{i \in R} \text{MSE}_i - \min_{i \in R} \text{MSE}_i},$$

where E_i is the remaining energy of node i , $w_{E,i}$ is the normalized value of E_i , $w_{\text{MSE},i}$ is the normalized value of MSE_i , and α and $1 - \alpha$ denote the weights of remaining energy and MSE, respectively. Here, we multiply the value of MSE by -1 in the evaluation metric to ensure that for both $w_{E,i}$ and $w_{\text{MSE},i}$, a large value means a better suitability.

Using Equation (10), we can compute the metric for each candidate relay node in R . Then, from R , the node with the largest metric θ_i is selected as the relay, as follows:

$$\begin{aligned} & \max_{i \in R, a, b_{k,1}, b_{k,2}} \theta_i, \\ & \text{s.t. } |b_{k,1}|^2 + |b_{k,2}|^2 \leq P_{\max}, k \in N_r, \\ & \quad b_{k,1} = 0, |b_{k,2}|^2 \leq P_{\max}, k \in \bar{N}_r, \\ & \quad P_{r,\text{TX}} \leq P_{\max}. \end{aligned} \quad (11)$$

Relay selection is divided into two steps. The remaining energy per node is not considered when selecting candidate relay nodes, but is considered when selecting one as a relay from the candidate relay nodes. This is due to the following reason: in slow channel fading, the channel gain is stable for a relatively long time, but the remaining energy per

node changes with time. Then, it is possible to compute the candidate relay nodes at a long interval, while periodically selecting a relay at a short interval, so as to distribute the load and power consumption to different nodes.

3.3.2. Problem of the SLAM

We performed some simulations to evaluate the impact of parameter α , using the condition in Table 1. Figure 5 shows how network lifetime and MSE change with α . With a large α , the metric focuses more on remaining energy instead of MSE. Therefore, MSE increases with α . However, the lifetime does not always increase with α . It decreases when α becomes greater than 0.6, which is different from the expectation.

Table 1. Main notations and their default values.

Parameters	Descriptions	Default Values
K	Number of nodes	20
H	Height of antenna	1.5 m
h_{gain}	Antenna gain	1
h_{loss}	Antenna loss	1.5
$Range$	Network range	200 m $\times$ 200 m
$Sink_{pos}$	Position of sink node	(100, 100)
P_{max}	Maximum limit power	10 dBm
σ^2	Noise power	1
J	Battery capacity	10,000 J
P_{idle}	Idle power of node	0.16 mW
T_{trans}	Time of transmit	5 s/min
T_{idle}	Time of idle	55 s/min
f	Frequency of analog wave	2.4 GHz

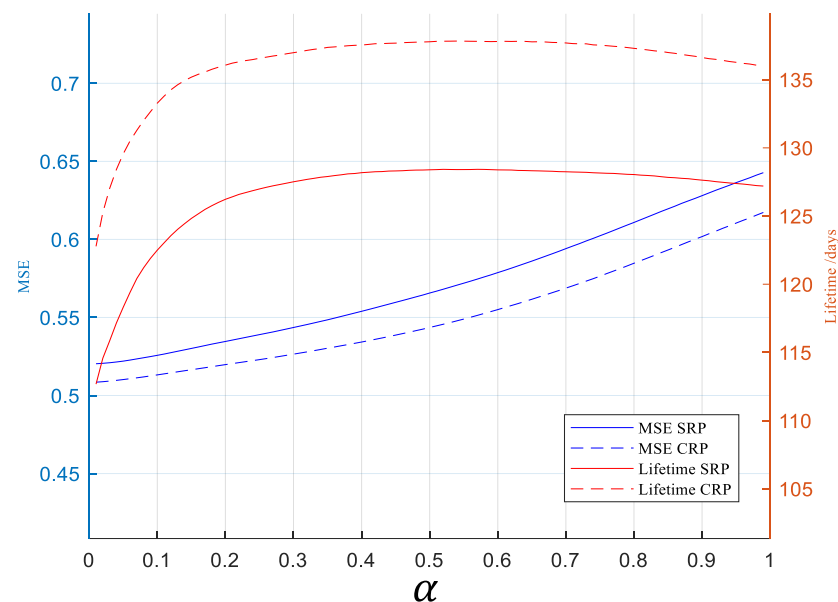


Figure 5. Variation of MSE and lifetime with respect to the parameter α in Equation (10).

During the simulation process, we found that when a node in the network was far away from both the relay node and the sink, the channel gain of this node was very low, causing it to use high transmission power in the communication. As a result, among the nodes in the network, the first node that runs out of energy is not the node that acts as the relay node. The remaining energy of the relay node is not used to improve the network lifetime in this case.

3.4. Extending the Basic Method (SLAM-E)

To solve the problem of the basic method, we designed Algorithm 2 to explicitly predict the remaining lifetime of the network when a relay node is selected.

In Algorithm 2, regarding each candidate in R as a relay in turn, we predicted the network lifetime based on the remaining energy and current power consumption of each node (lines 5 to 9). Here, network lifetime is the minimum of all nodes' lifetime (line 10). By iterating over all the candidate relay nodes in turn, the network lifetimes corresponding to all the candidate relay nodes are obtained.

Then, we updated the metric in Equation (10), replacing $w_{E,i}$ by $w_{L,i}$, as follows.

$$w_{L,i} = \frac{L_i - \min_{i \in R} L_i}{\max_{i \in R} L_i - \min_{i \in R} L_i}. \quad (12)$$

The other process is the same as the basic method in Algorithm 1.

Algorithm 2: Predict the remaining lifetime of the network

1. **Procedure:** Find network lifetime
 2. **Input:** K (Number of nodes), R , E (the set of all nodes' remaining battery);
 3. **Output:** L recording the lifetime of each node in R ;
 4. **For** $i \in R$ **do**
 5. | Regard node i as relay r , and get remaining energy of all nodes ($E_j, j = 1, \dots, K$);
 6. | **For** $j = 1, 2, \dots, K, j \neq i$ **do**
 7. | | Calculate the transmission power of node j (P_j^i);
 8. | | $Lifetime_j^i = E_j / P_j^i$;
 9. | **End**
 10. | Network lifetime computation: $L_i = \min_j Lifetime_j^i$;
 11. **End**
 12. **Return** L ;
-

3.5. Analysis of Computational Complexity

In the network, the number of nodes is K , and the number of candidate relay nodes is $N = |R|$. Here, $|R|$ denotes the number of the candidate relay nodes. Both the basic and extended methods use Algorithm 1 to compute the set of candidate relay nodes. In Algorithm 1, a loop with a loop count K is performed from line 4 to line 15, and another inner loop with a loop count $K - 1$ is performed from line 9 to line 13. In line 11, the CRP algorithm is used, whose time complexity is $T(\cdot) = NK^2 \log K$. Therefore, the time complexity of Algorithm 1 is $T(\cdot) = NK^2 \log K \cdot K(K - 1) = O(NK^4 \log K)$.

In Algorithm 2, an outer loop with a loop count $N = |R|$ is performed from line 4 to line 11, and an inner loop with a loop count $K - 1$ is performed from line 6 to line 9. Therefore, the time complexity of Algorithm 2 is $T(\cdot) = N \cdot K \leq O(K^2)$.

The basic method (SLAM) directly computes the evaluation metric by Equations (10)–(11), and the extra computation is $T(\cdot) = N = O(N)$. The extended method (SLAM-E) needs to predict the network lifetime, and the extra computation is $T(\cdot) = N \cdot K + N = O(NK)$.

Then, the overall computation complexity of both methods is $O(NK^4 \log K)$, decided by Algorithm 1.

4. Simulation Results

The proposed method was verified by simulation, using "MATLAB" (provided by MathWorks Inc.). Network lifetime and MSE are used as the main evaluation metrics. The basic AirComp [10] is also included in the evaluation for a comparison.

4.1. Simulation Parameters

Main simulation parameters are shown in Table 1.

The channel gain between two nodes is mainly determined by their distance. Here, we use the free-space model and the two-ray ground-reflection model to simulate the path loss. The threshold for the signal transmission distance d is

$$d = \frac{4\pi h_t h_r}{\lambda}, \quad (13)$$

where h_t and h_r are the heights of the transmitting and receiving antennas, respectively. λ is the wavelength of the transmitted wave. When the transmission distance is less than d , the free-space model is used, otherwise the two-ray ground-reflection model is used. In addition, block Rayleigh fading is assumed, and the channel is stable within two time slots.

In the evaluation, we considered an area of 200 m by 200 m, and the sink was located at the center of the area (100, 100). In order to evaluate the scalability of the method, the number of nodes was set to 20, 30, and 40, respectively, and nodes were randomly distributed in the experiment area, as shown in Figure 6. On average, the channel gain may be strong enough. However, for nodes at the cell edge or in deep fading, their channel gains will be poor, which requires the use of a large transmission power to reduce the MSE. In such cases, using relay helps to reduce node transmission power and extend network lifetime.

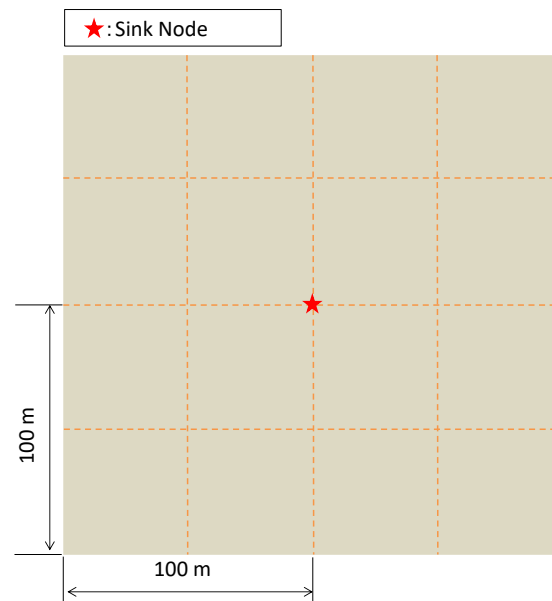


Figure 6. Square area for deploying sensor nodes. The sink is located at the center.

Based on the above conditions, we carried out the simulation experiments 1000 times, and the average results are reported.

4.2. Impact of Parameter α

First, we evaluated the impact of α on network lifetime and MSE, and the results are shown in Figure 7. In both SLAM and SLAM-E, MSE increases with α . As for the lifetime, in the SLAM method, the lifetime, increasing at first, starts to decrease after α becomes greater than 0.6, which is already explained in Figure 5. In comparison, the lifetime in SLAM-E continuously increases with α , indicating that SLAM-E solves the problem of SLAM and effectively prolongs the network lifetime. Hereafter, $\alpha = 0.5$ is chosen for SLAM and SLAM-E unless specified otherwise.

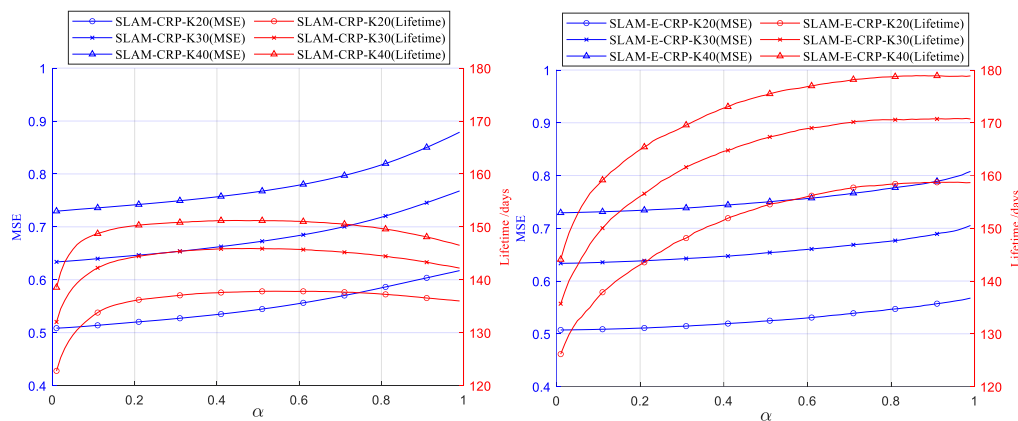


Figure 7. Variation of network lifetime and MSE with respect to α in Equation (10) (SLAM and SLAM-E).

4.3. Result of Lifetime

Next, we evaluated the network lifetime using six settings. Three of these used the SLAM method and the others used the SLAM-E method. In the same method, the number of nodes was set to 20, 30, and 40, respectively. All of them adopt the CRP method when relaying signals.

Figure 8 shows the CDF (cumulative distribution function) of the network lifetime for six settings. The CDF results confirm that SLAM-E prolongs the network lifetime compared with SLAM, under the same number of nodes and the same value of α . Additionally, the lifetime increases with the number of nodes in both SLAM and SLAM-E.

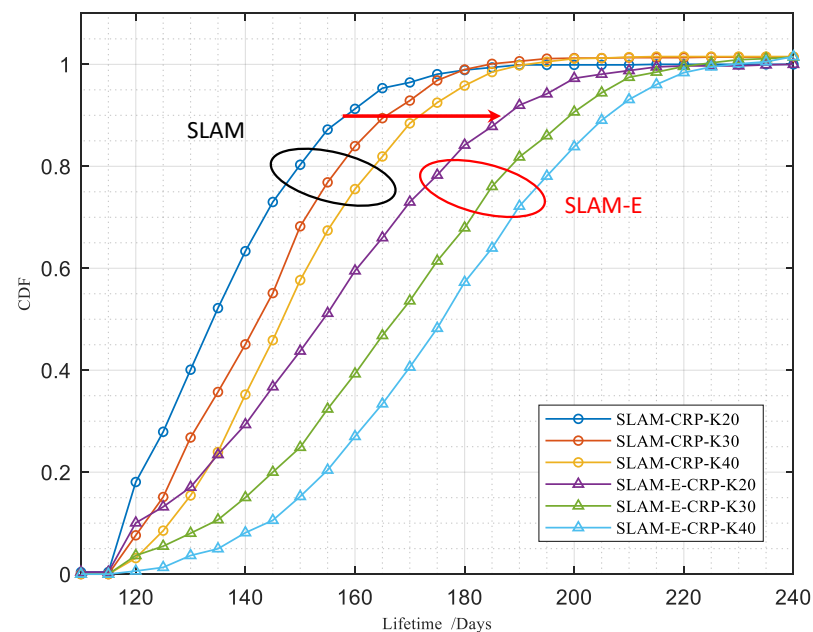


Figure 8. Cumulative distribution of network lifetime ($\alpha = 0.5$, a large circle is used to group results of the same method together).

Figure 9 shows the average network lifetime of four methods. SLAM-CRP improves network lifetime by 28.1% compared with the basic AirComp method, and SLAM-E-CRP further improves the network lifetime by 15.9% compared with SLAM-CRP. Surprisingly, using CRP together with AirComp leads to a worse network lifetime. This is because it focuses more on reducing MSE and nodes consume more power than in AirComp.

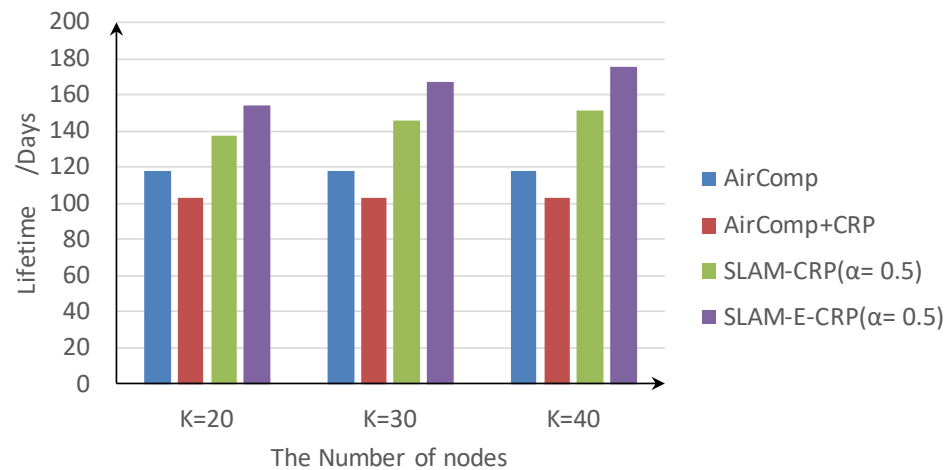


Figure 9. Variation of network lifetime with the number of nodes ($\alpha = 0.5$).

4.4. Result of MSE

Figure 10 shows the CDF of MSE for nine settings. It is clear that MSE increases with the number of nodes. This is straightforward because it becomes more difficult to align signals when there are more nodes. SLAM-E-CRP reduces MSE a little compared with SLAM-CRP, under the same number of nodes.

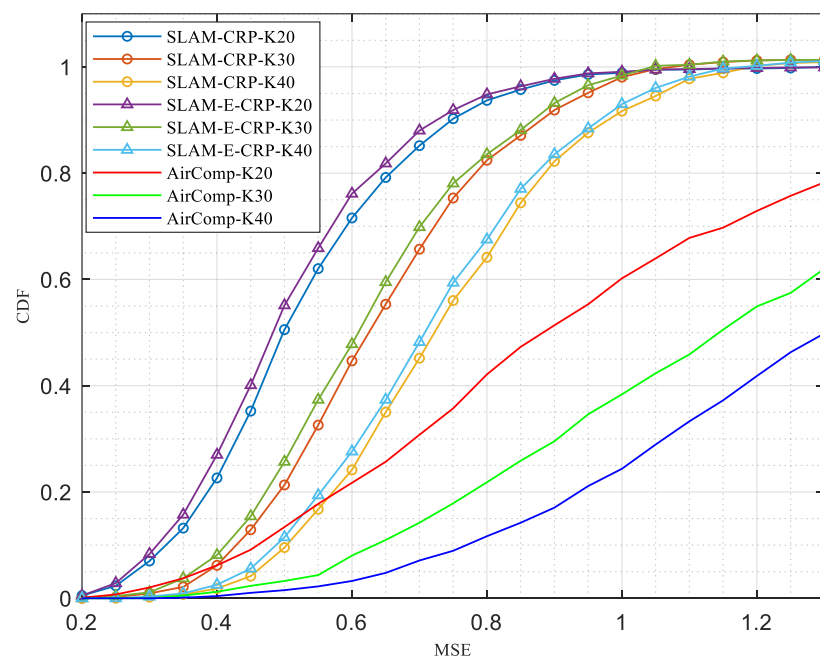


Figure 10. Cumulative distribution of MSE ($\alpha = 0.5$).

Figure 11 shows the average MSE of the four methods. The basic AirComp has the worst MSE. In comparison, the other 3 methods using relay greatly reduce MSE. When the number of nodes is 40, SLAM-CRP reduces the MSE by 45.2% compared with AirComp, and SLAM-E-CRP further reduces the MSE by 2% compared with SLAM-CRP.

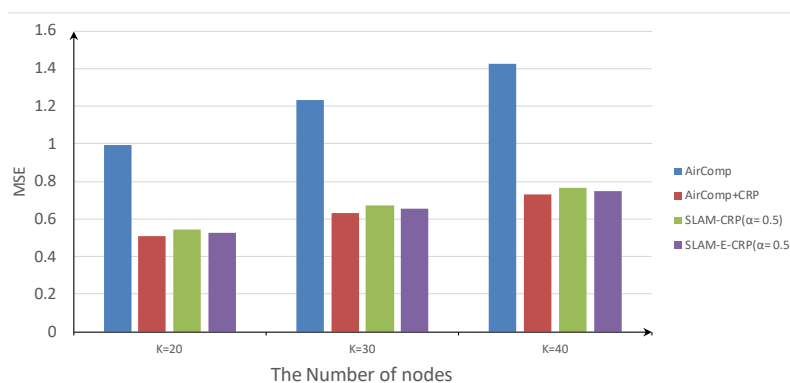


Figure 11. Average MSE of four methods.

The average lifetime and MSE are summarized in Table 2. The MSE of AirComp-CRP is a little less than that of SLAM-CRP and SLAM-E-CRP. This is because AirComp-CRP focuses more on reducing MSE, at the cost of more transmission power (shorter network lifetime). In comparison, SLAM-CRP and its enhancement, SLAM-E-CRP, achieve a better tradeoff between lifetime and MSE, which helps to better exploit the relay node to improve the overall performance.

Table 2. Average lifetime and MSE of four methods ($K = 40$, $\alpha = 0.5$).

Method	Average Lifetime (Day)	Average MSE
AirComp [10]	118	1.424
AirComp-CRP [15]	103.4	0.729
SLAM-CRP	151.2	0.766
SLAM-E-CRP	175.2	0.750

Initially, we thought that the relay node was the bottleneck because its power consumption increases with the number of signals using relay. On this basis, we designed SLAM. However, in the evaluation, we found that some nodes may still use the maximal transmission power even when using the relay node and deplete their energy faster than the relay node because the relay node is dynamically switched. SLAM-E helps to solve this problem, by directly predicting network lifetime. The experiment results confirm that (i) SLAM far outperforms the basic over-the-air computation model, and (ii) SLAM-E further extends the network lifetime while maintaining a similar MSE to that of the SLAM method.

5. Conclusions

In this paper, in order to improve the lifetime of AirComp and reduce MSE, we have proposed a new method (SLAM) to determine the candidate relay nodes and select the optimal relay node. The set of candidate relay nodes is mainly composed of nodes with high channel gains and a lot of remaining energy that can serve as the relay node to help other nodes with low channel gains. A new metric was designed to consider both MSE and lifetime in the selection of the relay node. We further enhanced the basic method to explicitly predict network lifetime in the relay selection, which helps to prolong the network lifetime. Simulation results confirm that the proposed method achieves a better tradeoff between network lifetime and MSE compared with previous methods. In the future, we will investigate the impact of mobile sink, and try to use multiple relay nodes simultaneously to further improve the performance.

Author Contributions: Conceptualization, S.T.; Methodology, J.Z.; Writing—original draft, J.Z.; Writing—review & editing, J.Z. and S.T.; Supervision, S.T. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the SCAT Foundation, Japan.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Sinha, R.S.; Wei, Y.; Hwang, S.-H. A survey on LPWA technology: LoRa and NB-IoT. *ICT Express* **2017**, *3*, 14–21. [\[CrossRef\]](#)
2. Nazer, B.; Gastpar, M. Computation over multiple-access channels. *IEEE Trans. Inf. Theory* **2007**, *53*, 3498–3516. [\[CrossRef\]](#)
3. Zhu, G.; Xu, J.; Huang, K.; Cui, S. Over-the-air computing for wireless data aggregation in massive IoT. *IEEE Wirel. Commun.* **2021**, *28*, 57–65. [\[CrossRef\]](#)
4. Gastpar, M. Uncoded transmission is exactly optimal for a simple Gaussian sensor network. *IEEE Trans. Inf. Theory* **2008**, *54*, 5247–5251. [\[CrossRef\]](#)
5. Lee, C.-Z.; Barnes, L.P.; Ozgur, A. Over-the-air statistical estimation. *IEEE J. Sel. Areas Commun.* **2020**, *40*, 548–561. [\[CrossRef\]](#)
6. Buck, R.C. Approximate complexity and functional representation. *J. Math. Anal. Appl.* **1979**, *70*, 280–298. [\[CrossRef\]](#)
7. Goldenbaum, M.; Boche, H.; Staczak, S. Nomographic functions: Efficient computation in clustered Gaussian sensor networks. *IEEE Trans. Wirel. Commun.* **2015**, *14*, 2093–2105. [\[CrossRef\]](#)
8. Abari, Rahul, H.; Katabi, D. Over-the-Air Function Computation in Sensor Networks. *arXiv* **2016**, arXiv:1612.02307.
9. Ye, H.; Li, G.Y.; Juang, B.-H.F. Deep Over-the-Air Computation. In Proceedings of the GLOBECOM 2020 IEEE Global Communications Conference, Taipei, Taiwan, 7–11 December 2020.
10. Liu, W.; Zang, X.; Li, Y.; Vucetic, B. Over-the-Air Computation Systems: Optimization, Analysis and Scaling Laws. *IEEE Trans. Wirel. Commun.* **2020**, *19*, 5488–5502. [\[CrossRef\]](#)
11. Cao, X.; Zhu, G.; Xu, J.; Huang, K. Optimized power control for over-the-air computation in fading channels. *IEEE Trans. Wirel. Commun.* **2020**, *19*, 7498–7513. [\[CrossRef\]](#)
12. Tang, S.; Popovski, P.; Zhang, C.; Obana, S. Multi-Slot Over-the-Air Computation in Fading Channels. *Accept. IEEE Trans. Wirel. Commun.* **2023**. early access. [\[CrossRef\]](#)
13. Qin, T.; Liu, W.; Vucetic, B.; Li, Y. Over-the-air computation via broadband channels. *IEEE Wirel. Commun. Lett.* **2021**, *10*, 2150–2154. [\[CrossRef\]](#)
14. Pappas, N.; Siris, V.A.; Traganitis, A. Path diversity gain with network coding and multipath transmission in wireless mesh networks. In Proceedings of the 2010 IEEE International Symposium on “A World of Wireless, Mobile and Multimedia Networks” (WoWMoM), Montreal, QC, Canada, 14–17 June 2010; pp. 1–6.
15. Tang, S.; Yin, H.; Zhang, C.; Obana, S. Reliable Over-the-Air Computation by Amplify-and-Forward Based Relay. *IEEE Access* **2021**, *9*, 53333–53342. [\[CrossRef\]](#)
16. Tang, S.; Yomo, H.; Zhang, C.; Obana, S. Node scheduling for AF-based over-the-air computation. *IEEE Wirel. Commun. Lett.* **2022**, *11*, 1945–1949. [\[CrossRef\]](#)
17. Fang, W.; Jiang, Y.; Shi, Y.; Zhou, Y.; Chen, W.; Letaief, K.B. Over-the-air computation via reconfigurable intelligent surface. *IEEE Trans. Commun.* **2021**, *69*, 8612–8626. [\[CrossRef\]](#)
18. Zhu, G.; Huang, K. MIMO over-the-air computation for high mobility multimodal sensing. *IEEE Internet Things J.* **2019**, *6*, 6089–6103. [\[CrossRef\]](#)
19. Navarro-Camba, E.A.; Felici-Castell, S.; Segura-García, J.; García-Pineda, M.; Pérez-Solano, J.J. Feasibility of a Stochastic Collaborative Beamforming for Long Range Communications in Wireless Sensor Networks. *Electronics* **2018**, *7*, 417. [\[CrossRef\]](#)
20. Lan, Q.; Kang, H.S.; Huang, K. Simultaneous signal-and-interference alignment for two-cell over-the-air computation. *IEEE Wirel. Commun. Lett.* **2020**, *9*, 1342–1345. [\[CrossRef\]](#)
21. Dong, J.; Shi, Y.; Ding, Z. Blind over-the-air computation and data fusion via provable wirtinger flow. *IEEE Trans. Signal Process.* **2020**, *68*, 1136–1151. [\[CrossRef\]](#)
22. Liu, W.; Zang, X.; Vucetic, B.; Li, Y. Over-the-air computation with spatial-and-temporal correlated signals. *IEEE Wirel. Commun. Lett.* **2021**, *10*, 1591–1595. [\[CrossRef\]](#)
23. Frey, M.; Bjelakovic, I.; Stanczak, S. Over-the-air computation in correlated channels. *IEEE Trans. Signal Process.* **2021**, *69*, 5739–5755. [\[CrossRef\]](#)
24. Zang, X.; Liu, W.; Li, Y.; Vucetic, B. Over-the-Air Computation Systems: Optimal Design with Sum-Power Constraint. *IEEE Wirel. Commun. Lett.* **2020**, *9*, 1524–1528. [\[CrossRef\]](#)

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.