2022 18th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob) | 978-1-6654-6975-3/22/$31.00 ©2022 IEEE | DOI: 10.1109/WIMOB55322.2022.9941709

Over-the-Air Computing under Adaptive Channel State Estimation

Nikos G. Evgenidis*, Vasilis K. Papanikolaou*, Panagiotis D. Diamantoulakis*, and George K. Karagiannidis*
*Department of Electrical and Computer Engineering, Aristotle University of Thessaloniki, GR-54124 Thessaloniki, Greece
e-mails: nevgenid@ece.auth.gr, vpapanikk@auth.gr, padiaman@auth.gr, geokarag@auth.gr

***Abstract*—Over-the-air Computation (AirComp) has attracted significant attention as an efficient way of data fusion by integrating uncoded communication transmissions with computation thanks to the superposition offered by the multiple access channels. However, proper pre-processing and post-processing is required to neutralize the wireless channel effect, in order for AirComp to function successfully. Since, internet-of-things (IoT) type of devices with limited capabilities are the target demographic of AirComp, having perfect channel state information (CSI) available is not always a practical assumption. In this work, we examine the effect of imperfect CSI on the AirComp system and we design a general optimization framework that takes into account both magnitude and phase errors in CSI. On top of that, a pilot retransmission policy is designed that offers a trade-off between cost of retransmissions and gain in the accuracy of the computations. Simulation results show the deterioration caused by the imperfect CSI and also the value of the proposed policy under various system conditions.**

***Index Terms*—over-the-air computing, AirComp, imperfect CSI, optimization framework, MSE minimization**

I. Introduction

In 5G and beyond networks, a paradigm shift has been noticed from human type communications to machine type communications. To deal with the massive amount of distributed data, particularly from internet-of-things type sensors, data aggregation through over-the-air computing (AirComp) has recently emerged as a very attractive technology [1], [2]. AirComp can accommodate various objectives in the network by taking advantage of the superposition property of transmitted waveforms over the multiple access channel (MAC), and proper pre- and post-processing to facilitate a family of functions, called nomographic such as mean, geometric mean, etc [3]. This exploitation can lead to significant increases in latency and computation load at the central processing node, especially when the number of sensors becomes excessively large. As proven in [4], the uncoded AirComp is optimal in the presence of a Gaussian multiple access channel (MAC) with independent data sources. This optimality is in terms of the mean squared error (MSE), which is a basic measure of performance for the discussed system.

AirComp systems face a particular set of challenges that need to be resolved to guarantee their operation. The time synchronization of the devices is one such challenge, where each device has to account for its message's propagation delay to ensure service. Moreover, carrier frequency offset (CFO) issues that are usually satisfied through the use of high quality oscillators need to be addressed with different means, since AirComp devices are usually low-end [5]. On top of that, the transmission through the wireless channel is prone to channel fading and noise. To account for that, a power allocation strategy is required, which is presented in [1], [6], i.e., the optimal power allocation for the transmitting devices in order to achieve a minimum MSE.

The integration of AirComp with popular wireless technologies such as MIMO has gathered a lot of attention from the research community [7]. Also, recently, AirComp has been discussed as an appealing technology to be implemented with federated learning in the network edge [5], [8]–[14]. On top of that, AirComp has shown promise in different scenarios such as in combination with aerial networks with unmanned aerial vehicles (UAVs). More specifically, in [15], the authors optimize the trajectory of the UAV to minimize the time-average MSE, while in [16] they take into account CSI imperfections when data are aggregated into a UAV fusion center. Finally, promising new technologies such as intelligent reconfigurable surfaces (IRS) has been examined as a way to facilitate AirComp with higher performance gains [17], [18].

However, most of these works are focused on the perfect channel state information (CSI) scenario, which cannot easily be guaranteed in practice, when IoT-type devices are concerned. Imperfect CSI can highly deteriorate the AirComp system's performance, as channel information is vital in obtaining the correct message at the receiver. In literature, there are some attempts studying the effect of imperfect CSI and its performance [16], [19]. Nevertheless, none of these look into the general case of imperfect CSI both in magnitude and phase, which is the main motivation of this work. In our paper, we will investigate the problem of the minimization of the MSE and the power allocation that must be used in the transmitting and receiving sides, so that the optimum can be achieved. First, we provide some theoretical analysis based on order statistics in order to get more insight on how the CSI imperfections affect the magnitude and the phase of every channel. The derived results can help us identify the necessary communication conditions that are needed for an AirComp system to work without suffering extensively from a noise-infused CSI. Then, we propose an algorithm that uses alternating optimization for the minimization problem, based on an approximation of the MSE at a worst case scenario. This approach differs from the one in [19] and also takes into consideration the phase difference between the real and the estimated channel, which is obviously the general case to be considered. Furthermore, based on our theoretical analysis we propose a second retransmission round for the weaker estimated channels in an attempt to improve the system's performance.

II. System Model

Let K be the number of transmitting devices in the AirComp system that are such that all of them are independent from one another. For our work we will assume that the receiver and all transmitters have a single antenna. Also let $b_k \in \mathbb{C}$ symbolize the transmitting power at the k-th device and $a \in \mathbb{C}^*$ symbolize the receiver gain factor. For practical reasons we will assume that all devices have a common maximum power magnitude P, so that $|b_k| \leq \sqrt{P}$ for all $k \in \{1, \cdots, K\}$.

We wish to find the computation distortion of the ideal signal $\sum_{k=1}^{K} x_k$, which will be given by

$$\text{MSE} = E\left(|r - \sum_{k=1}^{K} x_k|^2\right), \tag{1}$$

where r is the received signal given by $r = a\left(\sum_{k=1}^{K} b_k h_k x_k + n\right)$, so we have

$$\text{MSE} = E\left(\left|a\left(\sum_{k=1}^{K} b_k h_k x_k + n\right) - \sum_{k=1}^{K} x_k\right|^2\right). \tag{2}$$

All signals and noise are independent with one another and we assume that all signals $x_k \in [-u, u]$ have normalized variance, thus $E(|x_k|^2) = 1, \forall k$. We will also assume that the noise n follows a complex Gaussian distribution with $E(n^2) = \sigma^2$. Taking the expectation with respect to the signals x_k and noise n gives

$$\text{MSE} = E\left(\left|\sum_{k=1}^{K}(ab_k h_k - 1)x_k + an\right|^2\right) \tag{3}$$

and equivalently from the above assumptions

$$\text{MSE} = \sum_{k=1}^{K}\left|ab_k h_k - 1\right|^2 + \sigma^2|a|^2. \tag{4}$$

In order to study possible imperfections in the CSI, we will assume that the error in the estimation of the channel is modeled as additive random variable n_k. Hence, we assume the following:

- $n_k \sim CN(0, \sigma^2) = N(0, \frac{\sigma^2}{2}) + jN(0, \frac{\sigma^2}{2})$, are the noise samples in the kth CSI estimation, uncorrelated and independent between them.
- $h_k \sim CN(0, \sigma_h^2) = N(0, \frac{\sigma_h^2}{2}) + jN(0, \frac{\sigma_h^2}{2})$ because of Rayleigh fading conditions in the channel of transmitter k. Here we assume that $\sigma_h^2 = 2$ for all $|h_k|$ to follow a Rayleigh distribution of mode $\sigma_{Rayleigh} = 1$ as used in the simulations of the paper.

For the estimation of h_k we use symbols of max power $\sqrt{P}$ hence the receiver gets $y_k = \sqrt{P}h_k + n_k$ and assumes that $y_k = \sqrt{P}h'_k$. Consequently, the channel estimation will be given by

$$\sqrt{P}h'_k = \sqrt{P}h_k + n_k \Leftrightarrow h'_k = h_k + \frac{n_k}{\sqrt{P}} = h_k + e_k, \tag{5}$$

where $e_k \sim CN(0, \frac{\sigma^2}{P}) = N(0, \frac{\sigma^2}{2P}) + jN(0, \frac{\sigma^2}{2P})$. Consequently, we can assume that the channel estimation h'_k is distributed as $h'_k \sim CN(0, \frac{\sigma^2 + P\sigma_h^2}{P}) = N(0, \frac{\sigma^2 + P\sigma_h^2}{2P}) + jN(0, \frac{\sigma^2 + P\sigma_h^2}{2P})$.

III. Proposed Policy and Analysis

A. Theoretical Study on AirComp with Imperfect CSI

Assuming the receiver is unaware of the CSI imperfections, the algorithm obtained for an AirComp system with perfect CSI in [1], [6] would lead to a combination of full power and channel inversion methods. Using this algorithm in (4), would result in

$$\text{MSE} = \sum_{k=1}^{i^*}\left|\left(a\sqrt{P}\frac{{h'_k}^H}{|h'_k|}h_k - 1\right)\right|^2 + \sigma^2|a|^2 + \sum_{k=i^*+1}^{K}\left|\left(\frac{h_k}{h'_k} - 1\right)\right|^2, \tag{6}$$

where i^* is the critical number of transmitters that utilize their full power as given in [1]. Moreover, the channel ordering has been used $|h'_1| \leq |h'_2| \leq \cdots \leq |h'_K|$. Since the order has been taken with respect to the estimation of h_k, it is clear that the true order could be different, e.g., $|h_2| \leq |h_1| \leq \cdots \leq |h_K|$, which means that the proposed optimum solution can lead to a different optimum number i^*_{est} of transmitters that use full power, since

$$i^* = \underset{1 \leq i \leq K}{\text{argmax}}\{g_i\} \quad , \tag{7}$$

where

$$g_i = \frac{\sqrt{P}\sum_{k=1}^{i}|h_k|}{\sigma^2 + P\sum_{k=1}^{i}|h_k|^2} \tag{8}$$

and the magnitudes affect the result. Though i^* can be incorrectly estimated the main problem is that the incorrect values of g_i will also affect a, which is given as $a = g_{i^*}$.

Lemma 1: The expected value of the magnitude of the ordered channel gains is given by

$$E(U_r) = \sqrt{\frac{\pi}{4}}\sigma_h\binom{K}{r}r\sum_{k=0}^{r-1}\binom{r-1}{k}\frac{(-1)^k}{(K-r+1+k)^{\frac{3}{2}}}. \tag{9}$$

Proof: The proof is presented in Appendix A. ■

Corollary 1: The expected value of the magnitude of the error given by

$$E(|e_k|) = \int_0^\infty \frac{|e_k|^2}{\sigma_{e_k}^2}\exp\left(-\frac{|e_k|^2}{2\sigma_{e_k}^2}\right)d|e_k| = \sigma_{e_k}\sqrt{\frac{\pi}{2}} = \sqrt{\frac{\sigma^2}{2P}}\sqrt{\frac{\pi}{2}} = \frac{1}{\sqrt{\rho}}\sqrt{\frac{\pi}{4}}, \tag{10}$$

where ρ is the transmit SNR.

Proof: Following the proof of Lemma 1, (10) is easily produced in the same way. ■

Using the above results, we can find statistically how many channels are mostly affected by the CSI imperfection, which is of great value when designing the AirComp system. The main resulting issue is that the order of the estimated channels is very likely to be different from the correct order, which will result in an additional error in the minimization of MSE. Ultimately, the most affected channels will also be more susceptible to greater phase difference during CSI.

B. Optimization Framework

Due to AirComp's operation, imperfections in CSI can greatly hinder performance, even for small levels of noise. Solutions that ignore the possible imperfections that result in optimum controls for the perfect CSI case, by assuming $|h'_k| = |h_k|, \forall k \in \{1, \cdots, K\}$, diverge remarkably from the optimal strategy, which is defined by i^* and a, when imperfect CSI is available.

In order to improve the overall performance, we propose a new optimization framework that will take into account all the extra error terms that occur. In order to do this we assume that the statistical mean of the noise is known to the receiver. First of all, since $a \in \mathbb{C}^*$ we define $\angle a = a_p$ to symbolize its complex phase. By setting $\Delta h_k = \angle(h_k, h'_k)$, using the fact that $h'_k = h_k + e_k$, and (4), the optimization problem can be expressed as

$$\begin{aligned} \min_{a,\mathbf{b}} \quad & \text{MSE} = \textstyle\sum_{k=1}^{K} \left[\left(|a||b_k||h_k| \right)^2 + 1 \right] \\ & -2 \textstyle\sum_{k=1}^{K} \left(|a||b_k||h_k| \cos\left(a_p + \Delta h_k\right) \right) + \sigma^2 |a|^2, \\ \textbf{s.t.} \quad & \text{C}_1 : |b_k| \le \sqrt{P_k}, \forall k. \end{aligned} \tag{11}$$

Theorem 1: The optimal power distribution is given by

$$|b_k| = \min\left(\sqrt{P}, \frac{\cos \Delta h_k}{a|h_k|} \right). \tag{12}$$

Proof: In order to use the phase factor for minimization in (11) we will need to approximate its term and find its extrema. However, from trigonometry, $\cos\left(a_p + \Delta h_k\right) = \cos a_p \cos \Delta h_k - \sin a_p \sin \Delta h_k$ and any approach to approximate this quantity without knowledge about the sign of the phase difference Δh_k cannot be made because the sign of $\sin \Delta h_k$ will be affected. Since the sign of Δh_k is affected by the phase of the noise in the CSI estimation, we cannot make any assumptions about it and thus we cannot further use a_p in minimization. Hence, from now on we will consider a to be a real number and we rewrite the mean squared error as

$$\begin{aligned} \text{MSE} = & \sum_{k=1}^{K} \left[\left(a|b_k||h_k| \right)^2 + 1 \right] \\ & - 2 \sum_{k=1}^{K} \left(a|b_k||h_k| \cos \Delta h_k \right) + \sigma^2 a^2. \end{aligned} \tag{13}$$

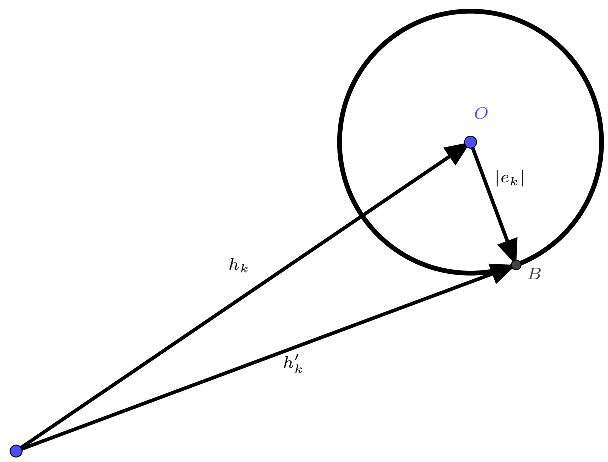

Fig. 1: Worst case approximation of phase difference Δh_k.

In order to find the extrema for every variable $|b_k|$ we take the first order partial derivative to be equal to zero, thus: $\frac{\partial \text{MSE}}{\partial |b_k|} = 0 \Leftrightarrow |b_k| = \frac{\cos \Delta h_k}{a|h_k|}$ and since $|b_k|$ is the power magnitude at device k, it must also be $0 \le |b_k| \le \sqrt{P}$. Consequently, the best power distribution will be given by (12) and, thus, the proof is completed. ■

We can observe that the real MSE is at this point related to the estimations h'_k only through the terms $\cos \Delta h_k$. Assuming the imperfections of CSI due to the noise have a constant magnitude of $|e_k|$, then from Figure 1, the worst case scenario for these terms arise when the phase of the imperfection is such that h'_k becomes tangent to the circle of constant radius $|e_k|$. This approximation will be quite accurate when the conditions are such that the mean estimated error can be assumed less than the mean of the estimated channels, i.e. $E(|h_k|) > E(|e_k|)$, and using it we obtain

$$\cos \Delta h_k^{max} = \frac{\sqrt{|h_k^2| - |e_k^2|}}{|h_k|} = \frac{|h_k'^{worst}|}{|h_k|}. \tag{14}$$

Hence, if we consider the worst case scenario we can make the following approximations $|h'_k|^2 + |e_k|^2 \approx |h_k|^2$ and $\cos \Delta h_k \approx \frac{|h'_k|}{|h_k|}$. We can also assume that $|e_k|^2 \approx E(|e_k|^2) = \frac{E(n^2)}{P} = \frac{\sigma^2}{P} = c$ for all k. While this approach contains its own errors, mostly for large scale fading or great levels of noise, it provides a tight approximation of the sinusoidal term for the greater majority of the involved channels, except for those that are greatly affected by noise. Also, since it covers the worst case in an almost optimal way it will partly compensate for its near optimum behavior in more favorable scenarios.

Using the above mentioned approximation combined with the attainable values of $|b_k|$ given by Theorem 1, we can now express (13) as

$$\mathrm{MSE}_i = \sum_{k=1}^{i} \left| \left(a\sqrt{P}|h'_k| - 1 \right) \right|^2 + a^2(\sigma^2 + iPc) + c \sum_{k=i+1}^{K} \frac{1}{|h'_k|^2 + c}. \quad (15)$$

Lemma 2: The optimal a_i to minimize (15) is given by

$$a_i = \frac{\sqrt{P}\sum_{k=1}^{i}|h'_k|}{(\sigma^2 + iPc) + P\sum_{k=1}^{i}|h'_k|^2}. \quad (16)$$

Proof: Considering (15) as a quadratic polynomial in terms of a and using a well known property for the global extremum of this function, we obtain that the global minimum for every i is given by differentiating (15) in terms of a_i to finally obtain (16), and thus, the proof is completed. ■

At this point, it is observed that (16) will be precise mostly at conditions like relatively high SNR or good channel conditions. This is so, because, due to our approximation the power magnitudes, we will obtain $|b_k|_{\text{opt}} = \frac{\cos\Delta h_k}{a|h_k|} \approx \frac{|h'_k|}{a|h_k|^2} \approx \frac{|h'_k|}{a(|h'_k|^2+c)}$, are points of the function $f(x) = \frac{x}{a(x^2+c)}$ which is an increasing and then decreasing function in terms of x that achieves its maximum at $x = \sqrt{c}$. Hence, for these to be in descending order we need $|h'_1| \geq \sqrt{c}$. Otherwise, we observe that the receiver coefficient a that will be calculated can be such that some of the weaker estimated channels will use the inverse channel method. Finally, expression (16) is quite similar to the optimum value of a for the perfect CSI case, but will always be less than g_i. In other words, the uncertainty caused by the imperfect CSI will force the system to use more transmitting power in an attempt to counter these imperfections.

Corollary 2: At least one device must use its full power during transmission.

Proof: The proof is given in Appendix B. ■

So, it suffices to solve K subproblems for $i \in 1, ..., K$ and, then, compare the minimum values $\mathrm{MSE}_i^{min} = \mathrm{MSE}_i(a_i)$. In order to get a feasible solution for a, it must be $a_i > \frac{|h'_{i+1}|}{(|h'_{i+1}|^2+c)\sqrt{P}}$. If $a_i > \frac{|h'_{i+1}|}{(|h'_{i+1}|^2+c)\sqrt{P}}$ we can always find a better solution to minimize the overall MSE. Thus, we only need to check the values $\mathrm{MSE}_i(a_i)$ when a_i is feasible. This way we can calculate i^* and then estimate a_i and b_k for all k.

It is important to note that this approach differs from the one followed in [19] in its mathematical derivation, but also in the fact that our approximation covers the more general case of phase misalignment as opposed to the phase alignment considered in [19].

C. Pilot Retransmission Policy

From our theoretical analysis it is clear that statistically there will be a number of channels whose estimations will be greatly affected by the noise-induced error during the CSI procedure. In order to limit the effect of this in the MSE of the system we need to have better estimations for the channels. As a result we can consider the possibility of making a second CSI round at least for some channels. Apparently, the most obvious choice is for the weaker channels to retransmit pilot symbols and then use the average of the two estimations as the new correct estimated channel. For this approach we propose the following heuristic algorithm in order to find the number of channels that will need to re-estimate their channels. In order to track the trade-off between the extra resources required for the retransmissions and the resulting MSE, we propose and define the following cost function, called the *Retransmission Policy Cost (RPC)* as

$$\mathrm{RPC} = (\text{Power cost}_k)^b \cdot \left(\frac{E(\mathrm{MSE}_k)}{K} \right)^d, \quad (17)$$

where Power cost_k symbolizes the power resources needed for k retransmissions, $\frac{E(\mathrm{MSE}_k)}{K}$ denotes the MSE for k retransmissions and are our primary concern over the available resources is taken to be their power. Moreover parameters b and d are considered to be weights for power and MSE, respectively. Any combination of parameters can be used to give emphasis to either the cost in terms of resources for the retransmissions or its maximum error tolerance. Withous loss of generality, the time penalty required for the retransmission can be included in the RPC metric, but given that the decrease of $E(\mathrm{MSE})$ is vital for the correct interpretation of the superimposed signals and the fact that the AirComp scheme needs less time than other traditional schemes like NOMA due to the superposition property of MAC, we suppose that the small time latency created by a few retransmission rounds does not affect our system.

IV. Simulation Results

In this section we present the simulation results of our work for an AirComp system. Fading channels have been simulated with Circular-Symmetric Complex Normal distributed variables $CSCN(0, 1)$ to simulate Rayleigh fading channel conditions. Unless otherwise stated, the transmit SNR is set as $\rho = 10$dB. We apply Monte Carlo analysis averaging over 10^4 channel realizations (snapshots). Finally, we define the average per user MSE as $\mathrm{AMSE} = E(\mathrm{MSE})/K$.

In Fig. 2 we look at the performance of the perfect CSI algorithm and the proposed optimization technique for imperfect CSI under imperfect channel estimation. As we can see the perfect CSI algorithm not only achieves worse $E(\mathrm{MSE})/K$ values, but also fails to converge as the number of devices in the system increases. In contrast to this, the proposed technique both achieves better performance and has a diminishing behavior for increasing number of devices.

In Fig. 3 we look at the performance of the retransmission policy discussed in section III. For comparison reasons, the retransmissions have also been simulated for random selections of channels instead of the weaker ones as proposed. As expected both policies achieve better MSE values than the no-retransmission policy due to the better channel estimations for their corresponding re-estimated channels. We can see that the proposed policy achieves a much better improvement rate over the random channel selection policy. It is important to point out that this improvement is greater mainly for the first few weaker channels, which confirms the idea that a re-estimation of these channels can be used to decrease the MSE of the system. We can also observe that though a bigger number of retransmissions can achieve further performance improvement, both policies tend to converge on the same MSE, because a lot of channel re-estimations will now be common.

In order to evaluate in terms of RPC the proposed policy, we design a simple example of an RPC function as follows. To obtain the first, necessary round of channel estimates exactly K power resources are needed. Assuming that each power resource is equal to 1, then, for every retransmission, 1 additional power resource is needed from the corresponding device. Thus, for k retransmissions, Power cost$_k = (K+k)P$ and according to (17) the minimum of RPC is studied. RPC in this case is expressed as

$$\mathrm{RPC} = [(K+k)P]^b \cdot \left(\frac{E(\mathrm{MSE}_k)}{K}\right)^d. \tag{18}$$

Using Monte Carlo analysis we see that such a minimum exists, it is global and it appears that for the statistically weaker channels all retransmission choices achieve a better trade-off between power and MSE than the original proposed scheme without any retransmissions.

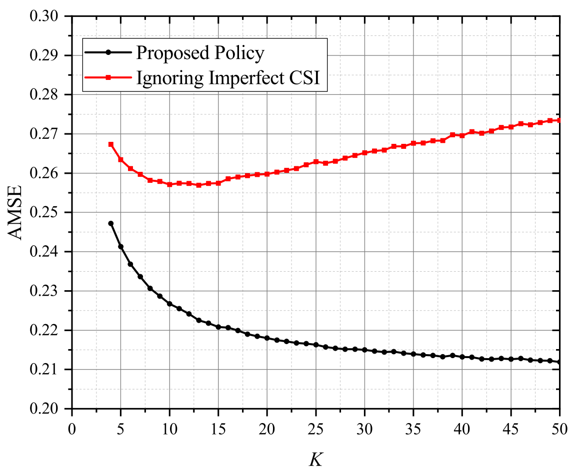

Fig. 2: MSE vs the number of devices in the AirComp system accounting or not for imperfect CSI (Transmit SNR = 10dB).

V. Conclusions

In this paper, an AirComp system under imperfect CSI assumptions was considered. The detrimental effect of imperfect CSI is presented for the MSE, especially when the optimization framework does not account for the imperfections in the channel estimations. We presented a comprehensive analysis on how channel estimation errors affect the AirComp system and a novel optimization framework to minimize MSE under those conditions. In order to counter the effect of the imperfections, an adaptive policy based on pilot retransmission was presented, where the proposed policy shows the potential to greatly improve the performance. Moreover, a new metric was presented alongside the retransmission policy to showcase the efficiency of the approach. Finally, simulation results were presented to validate the effectiveness of the proposed analysis, showcasing that it can offer great insights into the design of AirComp systems.

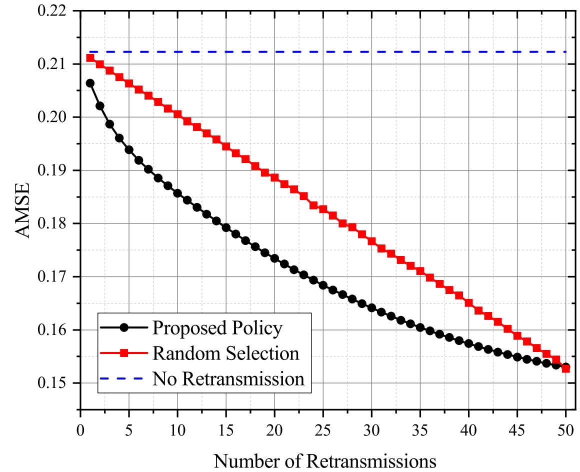

Fig. 3: MSE vs the number of retransmission with various policies (Transmit SNR = 10 dB).

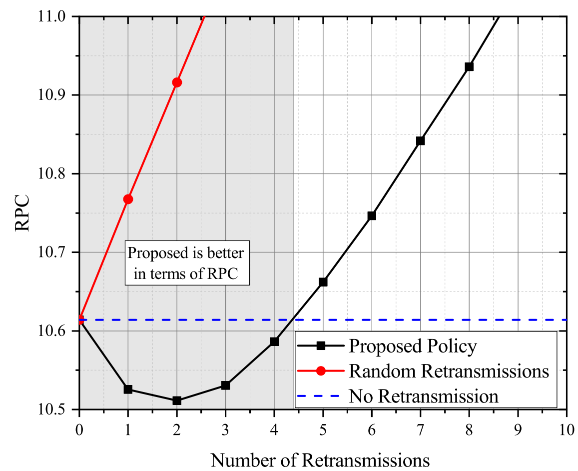

Fig. 4: Retransmission Policy Cost vs the Number of Retransmissions for the proposed policy, the random policy, and the original scheme without retransmission for $b = d = 1$ and $P_k = 1, \forall k$.

Acknowledgment

This paper was supported by the European Union's Horizon 2020 Research and Innovation Program under Agreement 957406. The implementation of the doctoral thesis of Vasilis K. Papanikolaou was co-financed by Greece and the European Union (European Social Fund-ESF) through the Operational Programme "Human Resources Development, Education and Lifelong Learning" in the context of the Act "Enhancing Human Resources Research Potential by undertaking a Doctoral Research" Sub-action 2: IKY Scholarship Programme for PhD candidates in the Greek Universities.

Appendix A
Proof of Lemma 1

We start the proof by presenting an analysis based on order statistics, in order to relate the mean estimation error with the mean of the magnitude of the ordered channels. We denote the correct channel gain ordering in the following way $|h_1^{cor}| \leq |h_2^{cor}| \leq \cdots \leq |h_K^{cor}|$. From the order statistics we can find that every ordered sample $U_r = |h_r^{cor}|$ from a total of K samples has the following probability distribution function (PDF)

$$f_{U_r}(u_r) = \frac{K!}{(K-r)!(r-1)!}\frac{2u_r}{\sigma_h^2} \times \exp\left[-\frac{u_r^2}{\sigma_h^2}(K-r+1)\right]\left[1-\exp\left(-\frac{u_r^2}{\sigma_h^2}\right)\right]^{r-1}. \quad (19)$$

Hence, the expected value $E(U_r)$ can be calculated by

$$E(U_r) = \binom{K}{r} r \int_0^\infty \frac{2u_r^2}{\sigma_h^2}\exp\left[-\frac{u_r^2}{\sigma_h^2}(K-r+1)\right] \times\left[1-\exp\left(-\frac{u_r^2}{\sigma_h^2}\right)\right]^{r-1} du_r. \quad (20)$$

Setting $t = \frac{u_r}{\sigma_h}$ we obtain

$$\begin{aligned} E(U_r) &= 2\sigma_h\binom{K}{r} r \int_0^\infty t^2\exp\left[-t^2(K-r+1)\right] \times\left[1-\exp\left(-t^2\right)\right]^{r-1} dt \\ &= 2\sigma_h\binom{K}{r} r \int_0^\infty t^2 e^{-Kt^2}[e^{t^2}-1]^{r-1}dt. \end{aligned} \quad (21)$$

Using binomial expansion and that $\int_0^\infty \exp(-mt^2) = \sqrt{\frac{\pi}{4m}}$ and after some algebraic manipulations, after the integration, we get

$$\int_0^\infty t^2 e^{-Kt^2}(e^{t^2}-1)^n dt = = \frac{1}{2}\sqrt{\frac{\pi}{4}}\sum_{k=0}^{n}\binom{n}{k}(-1)^k\frac{1}{(K-n+k)^{\frac{3}{2}}}. \quad (22)$$

Then, we can easily get (9) from (21) and (22), and, thus, the proof is concluded.

Appendix B
Proof of Corollary 2

We begin with a proof by contradiction, assuming that a full inverse channel method is utilized. Then, by (16) the overall MSE of the system will be

$$\mathrm{MSE}_0(a_0) = \sigma^2 a_0^2 + c\sum_{k=1}^{K}\frac{1}{|h_k'|^2+c}, \quad (23)$$

where $a_0 > \frac{|h_1'|}{\sqrt{P}(|h_1'|^2+c)}$ for feasibility reasons. Then, taking $a_1 = \frac{|h_1'|}{\sqrt{P}(|h_1'|^2+c)}$ observe that

$$\begin{aligned} \mathrm{MSE}_1(a_1) &= \left|\left(a_1\sqrt{P}|h_1'|-1\right)\right|^2 + a_1^2(\sigma^2+Pc) \\ &\quad + c\sum_{k=2}^{K}\frac{1}{|h_k'|^2+c} = a_1^2\sigma^2 + c\sum_{k=1}^{K}\frac{1}{|h_k'|^2+c} \\ &= \mathrm{MSE}_0(a_1) \end{aligned} \quad (24)$$

and since $a_0 > a_1$ we see that $\mathrm{MSE}_1(a_1) < \mathrm{MSE}_0(a_0)$ which means that we get that the performance of the system is better than the full inverse channel method, and, thus the proof is completed.

References

[1] 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.

[2] 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.

[3] M. Goldenbaum, H. Boche, and S. Stanczak, "Nomographic Functions: Efficient Computation in Clustered Gaussian Sensor Networks," *IEEE Trans. Wireless Commun.*, vol. 14, pp. 2093–2105, 2015.

[4] M. Gastpar, "Uncoded Transmission Is Exactly Optimal for a Simple Gaussian Sensor Network," *IEEE Trans. Inform. Theory*, vol. 54, pp. 5247–5251, 2008.

[5] Y. Shao, D. Gunduz, and S. C. Liew, "Federated Edge Learning With Misaligned Over-the-Air Computation," *IEEE Trans. Wireless Commun.*, vol. 21, no. 6, pp. 3951–3964, Jun. 2022.

[6] X. Cao, G. Zhu, J. Xu, and K. Huang, "Optimized Power Control for Over-the-Air Computation in Fading Channels," Jul. 2020, number: arXiv:1906.06858 arXiv:1906.06858 [cs, eess, math].

[7] X. Zhai, X. Chen, J. Xu, and D. W. Kwan Ng, "Hybrid Beamforming for Massive MIMO Over-the-Air Computation," *IEEE Trans. Commun.*, vol. 69, no. 4, pp. 2737–2751, Apr. 2021.

[8] Y. Sun, S. Zhou, Z. Niu, and D. Gunduz, "Dynamic Scheduling for Over-the-Air Federated Edge Learning With Energy Constraints," *IEEE J. Select. Areas Commun.*, vol. 40, no. 1, pp. 227–242, Jan. 2022.

[9] 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.

[10] N. Zhang and M. Tao, "Gradient Statistics Aware Power Control for Over-the-Air Federated Learning," *IEEE Trans. Wireless Commun.*, vol. 20, no. 8, pp. 5115–5128, Aug. 2021.

[11] X. Fan, Y. Wang, Y. Huo, and Z. Tian, "Joint Optimization of Communications and Federated Learning Over the Air," *IEEE Trans. Wireless Commun.*, vol. 21, no. 6, pp. 4434–4449, Jun. 2022.

[12] C. Xu, S. Liu, Z. Yang, Y. Huang, and K.-K. Wong, "Learning Rate Optimization for Federated Learning Exploiting Over-the-Air Computation," *IEEE J. Select. Areas Commun.*, vol. 39, no. 12, pp. 3742–3756, Dec. 2021.

[13] X. Cao, G. Zhu, J. Xu, Z. Wang, and S. Cui, "Optimized Power Control Design for Over-the-Air Federated Edge Learning," *IEEE J. Select. Areas Commun.*, vol. 40, no. 1, pp. 342–358, Jan. 2022.

[14] T. Sery, N. Shlezinger, K. Cohen, and Y. Eldar, "Over-the-Air Federated Learning From Heterogeneous Data," *IEEE Trans. Signal Process.*, vol. 69, pp. 3796–3811, 2021.

[15] M. Fu, Y. Zhou, Y. Shi, W. Chen, and R. Zhang, "UAV Aided Over-the-Air Computation," *IEEE Trans. Wireless Commun.*, pp. 1–1, 2021.

[16] H. Jung and S.-W. Ko, "Performance Analysis of UAV-Enabled Over-the-Air Computation Under Imperfect Channel Estimation," *IEEE Wireless Commun. Lett.*, vol. 11, no. 3, pp. 438–442, Mar. 2022.

[17] T. Jiang and Y. Shi, "Over-the-Air Computation via Intelligent Reflecting Surfaces," in *Proc. 2019 IEEE Global Commun. Conf. (GLOBECOM)*, 2019, pp. 1–6.

[18] 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.

[19] Y. Chen, G. Zhu, and J. Xu, "Over-the-Air Computation with Imperfect Channel State Information," Feb. 2022, number: arXiv:2202.13666 arXiv:2202.13666 [cs, math].