Over-the-Air Computing with Imperfect CSI: Design and Performance Optimization

Nikos G. Evgenidis, Vasilis K. Papanikolaou, Member, IEEE,

Panagiotis D. Diamantoulakis, Senior Member, IEEE, and George K. Karagiannidis, Fellow, IEEE

Abstract-Over-the-air computing (AirComp) has recently attracted considerable attention as an efficient method of data fusion by integrating uncoded communication transmissions with computation thanks to the signal superposition offered by the multiple access channels. However, appropriate processing is required to neutralize the wireless channel effect. As, internet-ofthings (IoT) applications through low-cost devices is the main target of AirComp, perfect availability of channel state information (CSI) is not always practical, there is the need to investigate the effect of imperfect CSI on AirComp. Specifically, we present novel closed-form expressions for tight approximations that can be used to design and evaluate AirComp systems. Furthermore, we design a general optimization framework that takes into account both magnitude and phase errors in the CSI. Finally, a pilot retransmission policy is designed, that offers trade-off between resources cost and the gain in the accuracy of the computations. In order to validate its application, a utility function of the cost of retransmission is introduced, namely, Retransmission Policy Cost (RPC), which can incorporate the power or throughput cost opposing to the expected gain of the selected policy. Simulations show the deterioration caused by the imperfect CSI and highlight the added 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 wireless networks, there is a paradigm shift from human to machine type communications. The former is characterized from a large amount of data requested in bursts by individual users, while the latter from a small amount of data, albeit continuously and from a huge number of devices. As a result, today's systems produce an enormous amount of data at extremely high rates [2]. To deal with the massive amount of distributed data, wireless

This paper was presented in part at the 18th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob) [1].

data aggregation (WDA) through over-the-air computing (Air-Comp) has recently emerged as an attractive technology [3]–[6]. AirComp can achieve several goals in the network by exploiting the superposition property of transmitted waveforms over the multiple access channel (MAC), and appropriate preand post-processing to produce a family of functions, called *nomographic* such as arithmetic mean, geometric mean, etc. The pioneering works [7], [8] and [9] provide a detailed presentation on the computable nomographic functions over MACs. This exploitation can lead to a significant reduction in latency and computation load at the central processing node, especially when the number of sensors becomes too large.

1

A. Literature Review

As shown in [10], the uncoded AirComp is optimal in the presence of a Gaussian MAC with independent data sources. This optimality is expressed in terms of the mean squared error (MSE), which is a fundamental measure of the performance for the discussed system. In [11] and [12], the authors present a thorough analysis on the performance of analog computation over wireless MACs. The integration of AirComp with popular wireless technologies such as MIMO has received a lot of attention from the research community [13]–[17]. In particular, the authors in [13] investigated the joint hybrid beamforming of a MIMO AirComp system in order to minimize the MSE in a cost-effective way. In [14], the aim, was to minimize the power consumption of the transmitting devises under the constraint of minimum required MSE. In [15] an AirComp system combined with MIMO was studied for the simultaneous calculation of different output target functions. In [16], the authors design a zero-forcing beamforming transmit scheme, they implement selection combining at the receiver side and track an MSE outage metric. Finally, in [17], the authors propose the study of an integrated sensing, communication, and computation over-the-air system with MIMO and they examine algorithms that allow the multiple functionalities of this system to coexist. In addition, reconfigurable intelligent surfaces (RIS) have also been examined as a way to facilitate AirComp with higher performance gains [18], [19].

Recently, AirComp has been discussed as an attractive technology to be implemented with federated learning in the network edge [20]–[30]. The idea behind this is that the gradients, which are produced in distributed machines using local models, can be aggregated via AirComp on the base station of the network, thus updating the global model in an efficient way. A similar idea has been considered for

N. G. Evgenidis and P. D. Diamantoulakis are with the Wireless Communications and Information Processing Group (WCIP), Department of Electrical and Computer Engineering, Aristotle University of Thessaloniki (e-mails: nevgenid@ece.auth.gr, padiaman@auth.gr).

V. K. Papanikolaou is with the Institute for Digital Communications, Friedrich-Alexander-University Erlangen-Nuremberg (FAU), 91054 Erlangen, Germany (e-mail: vasilis.papanikolaou@fau.de).

G. K. Karagiannidis is with the Wireless Communication and Information Processing Group (WCIP), Department of Electrical and Computer Engineering, Aristotle University of Thessaloniki, 54 124, Thessaloniki, Greece and with the Artificial Intelligence & Cyber Systems Research Center, Lebanese American University (LAU), Lebanon (e-mail: geokarag@auth.gr).

The work of G. K. Karagiannidis was supported by the European Union's Horizon 2020 Research and Innovation Program under Agreement 957406.

AirComp as a means of training deep neural networks [31]. In addition, AirComp has shown promise in a number of different scenarios, such as in combination with aerial networks with unmanned aerial vehicles (UAVs). More specifically, in [32], the authors optimize the trajectory of the UAV to minimize the time-average MSE, while in [33] they take into account CSI imperfections when aggregating data into a UAV fusion center. Another interesting technology that has been considered is the use of wireless powered transfer (WPT). Since WDA is a basic target for IoT sensor networks, WPT can assist on the energy efficiency and functionality of these type of networks, as studied in [34]. An interesting application was studied in [35], where AirComp was proposed to compute the control signal of a controller.

AirComp systems face a special set of challenges that need to be resolved to ensure their uninterrupted and successful operation. Precise time synchronization of the transmitting devices is one of them, since each device must take into account its message's propagation delay. Moreover, carrier frequency offset (CFO) issues, which are usually solved by the use of high quality oscillators, need to be addressed with different means, since AirComp devices are usually required to be low-cost [20]. On top of that, the transmission over the wireless channel is susceptible to path loss, fading, and noise. To account for this, a power allocation strategy is required, such as those presented in [3], [36], i.e., the optimal power allocation for the transmitting devices in order to achieve a minimum MSE under the assumption of common maximum power transmission. Finally, in [37] the same challenge was addressed when a total maximum power transmission constraint was imposed. The authors provided a closed-form solution and important insights into the operation of such an AirComp system.

B. Motivation and Contributions

Most of the aforementioned works focus on the perfect channel state information (CSI) scenario, which is not easily guaranteed in practice, when IoT devices are utilized. Imperfect CSI can highly deteriorate the AirComp system's performance, as channel information is crucial in obtaining the correct message at the receiver. In the literature, there have been some attempts to study the effect of imperfect CSI and its performance [33], [38]. However, none of them look into the general case of imperfect CSI both in magnitude and phase. Also, none of the previous works has provided theoretical analysis of the performance of the policies and utility functions discussed to be used in more practical related applications of over-the-air computing systems.

More specifically, the main contributions of this paper can be summarized as follows:

• First, we provide a theoretical analysis based on order statistics to gain more insights into how CSI imperfections affect the channel magnitude and phase. The presented analysis is based on the proposed policy of the perfect CSI case but the derived results can help identify the conditions necessary for an AirComp system to operate without suffering excessively from noisy CSI. We also extract useful approximations that can be used to partially estimate the performance of this policy under imperfect CSI.

- Next, we study the problem of minimizing the MSE through power allocation at the transmitter and gain factor selection at the receiver. We then 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 that in [38] mainly by the fact that it considers the phase difference between the real and the estimated channel, which is obviously the general case to be considered. Also in contrast to [38] no assumption of accurate phase estimation is made, which also leads to the study of inaccuracies caused by phase differences. For better comparison, performance analysis in terms of power consumption for the two policies, is also provided.
- Finally, based on the theoretical analysis we propose a retransmission round for the weaker estimated channels, as an attempt to improve the system's performance. We then address the problem of finding the optimal number of retransmissions with the aim to ensure the efficient use of the system's resources. For this purpose a utility function that considers the trade-off between MSE minimization and resource cost has been introduced and studied. Also a closed-form has been derived that allows to evaluate the mean value of the overhead, produed by the utilization of the extra retransmission round, which proves its usefulness since it takes into account practical details of an over-the-air system.

To our knowledge, this is the first time that a utility function has been used to further improve the performance of an over-the-air computing system besides solving an optimization problem.

C. Structure

The rest of this paper is organized in the following way: Section II describes the system model, the objective function to be minimized, and the CSI estimation procedure. In Section III we analyze the performance of the extra terms that arise from the use of the proposed policy under perfect CSI knowledge, due to the considered CSI imperfections. In Section IV we investigate the optimization of MSE and in Section IV-B we look at the use of the retransmission policy and its performance enhancement on the system. In Section V we present the simulation results along with extra comments and discussions. Finally, in Section VI we draw some conclusions from our work and note some future research opportunities that arise in this field.

II. SYSTEM MODEL

We consider an AirComp system consisting of one receiver that acts as a fusion center and multiple transmitting devices. Let K be the number of transmitting devices in the AirComp system, where each of them is independent. Assume that we wish to calculate a function $f : \mathbb{R}^K \to \mathbb{R}$ of all the transmitted data, denoted as $f(x_1, x_2, \dots, x_K)$. If f is a nomographic



Fig. 1: Nomographic representation of target functions and the distortion created by the channel effect and noise.

function of the data, then there are suitable pre-processing functions $\varphi_k : \mathbb{R} \to \mathbb{R}, \forall k \in \{1, \dots, K\}$ and a postprocessing function $\psi : \mathbb{R} \to \mathbb{R}$ such that the target function f can be expressed as [7]–[9]

$$f(x_1, x_2, \cdots, x_K) = \psi\left(\sum_{k=1}^K \varphi_k(x_k)\right).$$
(1)

As shown in Fig. 1, function f can be precisely computed over an ideal MAC, which is the basic idea behind AirComp.

In the context of this paper and without loss of generality, we assume that the receiver and all transmitters are equipped with a single antenna. Then, let $b_k \in \mathbb{C}$ be the transmitting factor at the k-th device such that $|b_k|$ symbolizes the transmitted power and $\angle b_k$ symbolizes the phase of the transmitting signal. Similarly $a \in \mathbb{C}^*$ symbolizes the receiver gain factor. All transmitting devices are assumed to have a common maximum power magnitude P, so that $|b_k| \leq \sqrt{P}$ for all $k \in \{1, \dots, K\}$.



Fig. 2: Model of an AirComp system made up of K transmitting devices and 1 fusion centre.

Next, we assume that the target function f is the arithmetic mean of the input data. Hence, the pre- and post-processing functions are linear on the data. We aim to calculate the com-

putation distortion between the ideal signal $r_{\text{ideal}} = \sum_{k=1}^{K} x_k$ and the received signal given by

$$r = a\left(\sum_{k=1}^{K} b_k h_k x_k + n\right).$$
⁽²⁾

The distortion between the two signals y and \overline{y} is given by the mean squared error of the two signals as

$$MSE = \mathbb{E}\left[\left|\overline{y} - y\right|^2\right].$$
 (3)

Hence, (3) results into

$$MSE = \mathbb{E}\left[\left|r - r_{ideal}\right|^{2}\right],$$
(4)

and substituting r and r_{ideal} in (4) we obtain

1

$$MSE = \mathbb{E}\left[\left|a\left(\sum_{k=1}^{K}b_{k}h_{k}x_{k}+n\right)-\sum_{k=1}^{K}x_{k}\right|^{2}\right].$$
 (5)

Furthermore, we assume that all signals and noise are independent with to each other and all signals $x_k \in [-u, u]$ have normalized variance, $\mathbb{E}[|x_k|^2] = 1, \forall k$. We also assume that the noise *n* follows a complex Gaussian distribution with $\mathbb{E}[n^2] = \sigma^2$. Taking the expectation with respect to the signals x_k and noise *n* gives

$$MSE = \mathbb{E}\left[\left|\sum_{k=1}^{K} (ab_k h_k - 1)x_k + an\right|^2\right]$$
(6)

and equivalently from the above assumptions

$$MSE = \sum_{k=1}^{K} |ab_k h_k - 1|^2 + \sigma^2 |a|^2.$$
(7)

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

$$n_k \sim CN(0, \sigma^2) = N\left(0, \frac{\sigma^2}{2}\right) + jN\left(0, \frac{\sigma^2}{2}\right), \quad (8)$$

are the noise samples in the k-th CSI estimation which are uncorrelated and independent and

$$h_k \sim CN(0, \sigma_h^2) = N\left(0, \frac{\sigma_h^2}{2}\right) + jN\left(0, \frac{\sigma_h^2}{2}\right), \quad (9)$$

because of Rayleigh fading condition.

For the estimation of h_k , pilot symbols of maximum transmitting power \sqrt{P} , which are known a priori to both transmitter and receiver, are utilized. Since AirComp systems are mainly designed with IoT devices in mind, where low complexity is desired, the least square (LS) channel estimation is one of the most practical CSI methods to be used. Hence, the baseband equivalent at the receiver is $y_k = \sqrt{P}h_k + n_k$ and so the receiver assumes that $y_k = \sqrt{P}h'_k$. Consequently, the channel estimation is

$$\sqrt{P}h'_k = \sqrt{P}h_k + n_k \Leftrightarrow h'_k = h_k + \frac{n_k}{\sqrt{P}} = h_k + e_k, \quad (10)$$

where $e_k \sim CN(0, \frac{\sigma^2}{P}) = N(0, \frac{\sigma^2}{2P}) + jN(0, \frac{\sigma^2}{2P})$ is the estimation error due to the presence of noise [39], [40]. As

seen from the CSI estimation model, we assume that estimation is made on both magnitude and phase of the channel. Different CSI models that estimate only the magnitude or statistical properties of the channel have not been considered because they would only worsen the system's performance due to less available knowledge on the receiver. Another reason for the chosen CSI estimation method is that because of the vast number of transmitting devices on over-the-air computing systems, insufficient CSI estimation will compromise all of them and as a result, there will be greater MSE distortion. Consequently, we can assume that the channel estimation h'_k is distributed as

$$h'_{k} \sim CN\left(0, \frac{\sigma^{2} + P\sigma_{h}^{2}}{P}\right)$$
$$= N\left(0, \frac{\sigma^{2} + P\sigma_{h}^{2}}{2P}\right) + jN\left(0, \frac{\sigma^{2} + P\sigma_{h}^{2}}{2P}\right).$$
(11)

III. AIRCOMP WITH IMPERFECT CSI

Assuming the receiver is unaware of the CSI imperfections, the algorithm obtained for an AirComp system with perfect CSI in [3], [36] would lead to a combination of full power and channel inversion methods. Applying this algorithm in (7), would result in

$$MSE = \underbrace{\sum_{k=1}^{i^{*}} \left| \left(a\sqrt{P} \frac{h_{k}^{\prime H}}{|h_{k}^{\prime}|} h_{k} - 1 \right) \right|^{2}}_{\text{Full Power Terms}} + \underbrace{\sum_{k=i^{*}+1}^{K} \left| \left(\frac{h_{k}}{h_{k}^{\prime}} - 1 \right) \right|^{2}}_{\text{Inverse Channel Terms}},$$
(12)

where i^* is the critical number of transmitting devices that utilize their full power as given in [3], whereas the rest of the transmitting devices utilize the inverse channel method. Moreover, the ascending channel ordering has been assumed, without loss of generality, as $|h'_1| \leq |h'_2| \leq \cdots \leq |h'_K|$. Since the order has been taken with respect to the estimation of h_k , it is clear that the order of the actual channel gains could be different, e.g., $|h_2| \leq |h_1| \leq \cdots \leq |h_K|$. The number of transmitting devices i^*_{est} that use their full power can be calculated based on the imperfect CSI as

$$i_{\text{est}}^* = \underset{1 \le i \le K}{\operatorname{argmax}} \{ g_i(\mathbf{h}') \}, \tag{13}$$

where \mathbf{h}' is the vector containing the estimated CSI and

$$g_i(\mathbf{h}') = \frac{\sqrt{P} \sum_{k=1}^i |h'_k|}{\sigma^2 + P \sum_{k=1}^i |h'_k|^2}.$$
 (14)

Due to the estimated magnitudes of the channel gain terms in (14), $i_{\text{est}}^* \neq i^*$ in general, resulting in suboptimal performance. On top of that, the main issue stemming from (14) is that the incorrect values of g_i will also affect a, which, in general, is given as $a = g_{i^*}$.

Notice that, as thoroughly studied in [3], when the minimum MSE is achieved by the corresponding $g_i(\mathbf{h}')$, it holds that

$$\frac{1}{\sqrt{P}|h_{i_{\text{est}}^*+1}|} \le g_{i_{\text{est}}^*}(\mathbf{h}') \le \frac{1}{\sqrt{P}|h_{i_{\text{est}}^*}|}.$$
 (15)

This is important because it is a necessary condition that must hold from a feasibility point of view, since a should be such that at least the right hand side of (15) is satisfied for $i_{\text{est}}^* = i^*$ to be the critical number.

4

In this context, it is of vital importance to study the channel ordering, since the ones with the lowest gain are more prone to estimation imperfections, leading both (13) and (14) to miscalculations. The following lemma, combined with (17), provide an answer to that.

Lemma 1: The expected value of the magnitude of the ordered channel gains for the r-th ordered channel is given by

$$\mathbb{E}[|h_r|] = \sqrt{\frac{\pi}{4}} \sigma_h \binom{K}{r} r \sum_{k=0}^{r-1} \binom{r-1}{k} \frac{(-1)^k}{(K-r+1+k)^{\frac{3}{2}}}.$$
(16)

Proof: The proof is provided in Appendix A. Moreover, the expected value of the magnitude of the error is given by [41]

$$\mathbb{E}[|e_k|] = \int_0^\infty \frac{|e_k|^2}{\sigma_{e_k}^2} \exp\left(-\frac{|e_k|^2}{2\sigma_{e_k}^2}\right) d|e_k| = \sigma_{e_k} \sqrt{\frac{\pi}{2}} = \sqrt{\frac{\sigma^2}{2P}} \sqrt{\frac{\pi}{2}} = \frac{1}{\sqrt{\rho}} \sqrt{\frac{\pi}{4}}, \quad (17)$$

where ρ is the transmit SNR.



Fig. 3: Mean ordered channels compared to expected error magnitude for K = 40, 50 transmitting devices (Transmit SNR = 10dB).

In order to better illustrate the above results, Fig. 3 presents the mean value of the order channel gains for different K.

As it is evident from this figure and calculated from (16), (17) the channels with the lowest gain can be comparable to the magnitude of the estimation error. Therefore, the corresponding channels' estimation is highly affected, since it is the result of the complex addition of the true channel value and the error. Using (16) and (17), the number of the channels that are mostly affected by the CSI imperfection can be found, which is of great value when designing the AirComp system.

Authorized licensed use limited to: Aristotle University of Thessaloniki. Downloaded on March 10,2024 at 17:07:21 UTC from IEEE Xplore. Restrictions apply. © 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

However, when imperfect CSI is available there are more issues to take into account. The most important of them is that the order of the estimated channels is very likely to be different from the correct one, which leads to an additional error in the minimization of MSE. Ultimately, the most affected channels will also be more susceptible to greater phase difference during CSI.

In contrast to the perfect CSI case, (12) contains an extra term that is made up from the use of the inverse channel technique. In order to understand the effect of this term, a theoretical analysis to extract approximations of its lower and upper bounds are presented. From the system model described in Section II, it will be $|e_k|^2 \sim \text{Exponential}(\frac{P}{\sigma^2})$ and the ordered statistics $U_k = |h'_k|^2$ are derived from the general probability distribution function (PDF) $|h'|^2 \sim \text{Exponential}(\frac{P}{\sigma^2 + P\sigma_h^2})$. First, the following lemmas are presented to aid in finding the aforementioned bounds.

Lemma 2: The expected value of the inverse of the ordered estimated channel gains is given by

$$\mathbb{E}\left[\frac{1}{U_r}\right] = \lambda \binom{K}{r} r \sum_{k=0}^{r-1} \binom{r-1}{k} (-1)^{k+1} \ln\left(K - r + 1 + k\right),\tag{18}$$

where r > 1 and $\lambda = \frac{P}{\sigma^2 + P\sigma_h^2}$ is the parameter of the exponential distribution that $|h'|^2$ follows.

Proof: The proof is presented in Appendix B.

Lemma 3: The expected value of the inverse of the ordered estimated channel gains is given by

$$\mathbb{E}\left[\frac{1}{U_r^2}\right] = \lambda K \left(\mathbb{E}\left[\frac{1}{U_{r-1,K-1}}\right] - \mathbb{E}\left[\frac{1}{U_{r,K}}\right]\right), \quad (19)$$

where r > 2 and the second index shows the number of total samples in the ordered statistics.

Proof: The proof is presented in Appendix C. Let expression

$$S = \sum_{k=i^*+1}^{K} \left| \left(\frac{h_k}{h'_k} - 1 \right) \right|^2 = \sum_{k=i^*+1}^{K} \left| \frac{e_k}{h'_k} \right|^2$$
(20)

be the term under discussion. We are interested in calculating the expected value of S, because of the effect it will have in the overall performance of the perfect CSI policy under imperfect CSI. Whenever $i^* \geq 2$ from the law of total expectation we have

$$\mathbb{E}[S] = \sum_{m=2}^{K-1} \mathbb{E}[S|i^* = m] P\{i^* = m\},$$
(21)

where $P\{\cdot\}$ denotes the probability.

Then, a direct upper bound for this sum can be computed as shown in the following lemma.

Lemma 4: An upper bound can be computed for (21) as follows

$$\mathbb{E}[S] \le \sum_{m=2}^{K-1} P\{i^* = m\} \sum_{k=i^*+1}^{K} \sqrt{\mathbb{E}[|e_k|^4]} \mathbb{E}\left[\frac{1}{|h'_k|^4}\right], \quad (22)$$

where it is well known that

$$\mathbb{E}[|e_k|^4] = \frac{2}{\lambda_{|e_k|^2}} = \frac{2}{\rho^2},$$
(23)

where $\lambda_{|e_k|^2} = \frac{P}{\sigma^2}$ and ρ is the transmit SNR.

Proof: The bound can easily be computed by taking the Cauchy-Schwarz inequality in (21) multiple times.

From Lemma 2, Lemma 3 and (23) we can evaluate the upper bound described by (22). Since the Cauchy-Schwarz inequality was used multiple times to derive a closed-form expressions, the bound will not be too tight. In order to get a better view of the behavior of S we can examine a different set of bounds. Due to the channel ordering, it is trivial to show that S is upper bounded by S' and lower bounded by S'' that are given as

$$S' = \sum_{k=i^*+1}^{K} \left| \frac{e_k}{h'_{k-1}} \right|^2$$
(24)

and

$$S'' = \sum_{k=i^*+1}^{K-1} \left| \frac{e_k}{h'_{k+1}} \right|^2,$$
(25)

respectively.

Theorem 1: In cases where the noise is sufficiently small, the expected value of S' and S'' can be tightly approximated by the following expressions

$$\mathbb{E}[S'] = \frac{\sigma^2}{P} \sum_{m=2}^{K-1} P\{i^* = m\} \sum_{r=i^*}^{K-1} \mathbb{E}\left[\frac{1}{U_r}\right]$$
(26)

and

$$\mathbb{E}[S''] = \frac{\sigma^2}{P} \sum_{m=1}^{K-2} P\{i^* = m\} \sum_{r=i^*+2}^{K} \mathbb{E}\left[\frac{1}{U_r}\right]$$
(27)

Proof: Due to the correlation between e_k and h'_k , the presented analysis relies on the assumption that $|e_k|^2 \approx \mathbb{E}[|e_k|^2] = \frac{\sigma^2}{P}$, since $|e_k|^2 \sim \text{Exponential}(\frac{P}{\sigma^2})$. It is important to note that this is based on the practical assumption that the statistical mean of the noise is known to the receiver. Therefore, our assumption is a consequence of the p.d.f. that the error e_k follows due to the presence of noise and it is straightforward to get that $\mathbb{E}[|e_k|^2] = \frac{\mathbb{E}[|n_k|^2]}{P} = \frac{\sigma^2}{P}$ because of (8). Also as assumed, for high values of transmit SNR, where $P \gg \sigma^2$, the channel estimation becomes almost independent of the error. Indeed if we carefully examine (16), (17) we can see that for increasing transmit SNR, the mean value of the error will decrease by a factor of $1/\sqrt{\rho}$.

We start by taking into account the inequalities (24) and (25). (24) holds for $1 \le i^* \le K - 1$ and (25) holds for $1 \le i^* \le K - 2$ and $i^* = K$ results in the use of the full power technique. This means that

$$U_{k} = \frac{1}{\lambda} \sum_{j=1}^{k} \frac{Z_{j}}{K - j + 1},$$
(28)

where Z_j are random variables that follow the normalized exponential PDF [42] and $\lambda = \frac{P}{\sigma^2 + P\sigma_h^2}$. Observe that by the definition in (28) and trivial inequalities

$$\frac{1}{\lambda} \sum_{j=1}^{k} \frac{Z_j}{K-k+1} \ge U_k \ge \frac{1}{\lambda} \sum_{j=1}^{k} \frac{Z_j}{K}$$
(29)

Authorized licensed use limited to: Aristotle University of Thessaloniki. Downloaded on March 10,2024 at 17:07:21 UTC from IEEE Xplore. Restrictions apply. © 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information. holds. Also, it is known that the random variable $X_k = \frac{1}{\sum_{j=1}^k Z_j}$ follows the inverse Gamma distribution and has mean value equal to $\frac{1}{k-1}$ [43]. Obviously, this holds for k > 1 and equivalently $i^* \geq 2$. Using the law of total expectations for the left-hand side (LHS) of (26) we get

$$\mathbb{E}[S'] = \sum_{m=2}^{K-1} \mathbb{E}[S'|i^* = m] P\{i^* = m\},$$
(30)

where $P\{\cdot\}$ denotes the probability. From (30) and assuming sufficiently small noise we can get

$$\mathbb{E}[S'] \approx \frac{\sigma^2}{P} \sum_{m=2}^{K-1} P\{i^* = m\} \sum_{r=i^*}^{K-1} \mathbb{E}\left[\frac{1}{U_r}\right].$$
 (31)

Following the same way for (25) we get

$$\mathbb{E}[S''] \approx \frac{\sigma^2}{P} \sum_{m=1}^{K-2} P\{i^* = m\} \sum_{r=i^*+2}^{K} \mathbb{E}\left[\frac{1}{U_r}\right] \qquad (32)$$

and the proof is completed.

Calculating sums with the expression $\mathbb{E}\left[\frac{1}{U_r}\right]$ can be computationally difficult in the general case. Therefore, we aim to further simplify the aforementioned approximations of $\mathbb{E}[S]$, leading to closed-form expressions being obtained that approximate (26) and (27).

Corollary 1: Expressions (26) and (27) can be further relaxed with the help of harmonic numbers as

$$\frac{\mathbb{E}[S']}{K} \le \frac{1}{1+\rho\sigma_h^2} \sum_{m=2}^{K-1} P\{i^* = m\} (H_{K-2} - H_{i^*-2}) \quad (33)$$

and

$$\frac{\mathbb{E}[S'']}{K} \ge \frac{1}{1+\rho\sigma_h^2} \sum_{m=1}^{K-2} P\{i^* = m\} \left(H_{K-1} - H_{i^*} - \frac{K-i^*-1}{K}\right)$$
(34)

where ρ denotes the transmit SNR and the harmonic numbers are defined as $H_k = \ln k + \gamma + \frac{1}{2k} + O\left(\frac{1}{k^2}\right)$, with γ denoting the Euler-Mascheroni constant [44].

Proof: By the right-hand side (RHS) of (29) we can write

$$\mathbb{E}[S'] \le \frac{\sigma^2}{P} \lambda \sum_{m=2}^{K-1} P\{i^* = m\} \sum_{r=i^*}^{K-1} \frac{K}{r-1}$$
(35)

which leads to (33). Also, By the LHS of (29)

$$\mathbb{E}[S''] \ge \frac{\sigma^2}{P} \lambda \sum_{m=1}^{K-2} P\{i^* = m\} \sum_{r=i^*+2}^{K} \frac{K-r+1}{r-1} \qquad (36)$$

which equivalently leads to (34) and the proof is completed.

As it can be observed from the derived approximations, the increase of transmit SNR will reduce the effect of the inverse channel terms. However, we can see that for values of i^* similar to the ones derived in the perfect CSI case where i^* is considerably less than K, the harmonic numbers H_{K-2} , H_{K-1} will dominate in (33) and (34), respectively. Considering the asymptotic approach of the harmonic numbers, we can see that the behavior of both approximations slowly diverge for increasing values of K. From the above, it can be seen that in order to neutralize the uncertainty caused by imperfect CSI, more transmitting devices should transmit with maximum power, i.e., i^* is larger.

Remark 1: There are a few useful observations that we can use to evaluate the performance of the given approximations in terms of our original assumption and how it will affect them.

- With the increase of transmit SNR, ρ , the effect of the error on the estimation of the channels is reduced, resulting in better approximations.
- With the increase of the number of transmitting devices on the system, K, more channels will be affected by errors in estimation, as showcased in Fig. 3, resulting in less accurate approximations.

IV. OPTIMIZATION FRAMEWORK

As it was shown in the previous section, imperfections in CSI can deteriorate the 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 $|h'_k| = |h_k|, \forall k \in \{1, \dots, K\}$, diverge remarkably from the optimal strategy.

While the mathematical analysis in Section III produced some interesting approximations that can be effectively utilized in the design of an AirComp system, in order to improve the overall performance, we propose a novel optimization framework that takes into account all the extra error terms. To this end, we make the practical assumption 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 denote 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 of minimizing the MSE is , expressed as

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

$$\end{split}$$

Theorem 2: The optimal power distribution is given by

$$|b_k| = \min\left(\sqrt{P}, \frac{\cos\Delta h_k}{a|h_k|}\right).$$
(38)

Proof: In order to use the phase factor for minimization in (37) we will need to approximate the term $\cos(a_p + \Delta h_k)$ and find its extrema. However, from trigonometry, $\cos(a_p + \Delta h_k) = \cos a_p \cos \Delta h_k - \sin a_p \sin \Delta h_k$ and, as such, 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. Due to the randomness of Δh_k , choosing a specific a_p could worsen the MSE. Therefore, with no prior knowledge of Δh_k , on average, the best option is to set $a_p = 0$. Hence, from now on we



Fig. 4: Worst case approximation of phase difference Δh_k .

will consider a to be a real number and we rewrite the mean squared error as

$$MSE = \sum_{k=1}^{K} \left[(a|b_k||h_k|)^2 + 1 \right] - 2 \sum_{k=1}^{K} (a|b_k||h_k| \cos \Delta h_k) + \sigma^2 a^2.$$
(39)

In order to find the extrema for every variable $|b_k|$ we take the first order partial derivative to be equal to zero, thus: $\frac{\partial MSE}{\partial |b_k|} = 0 \Leftrightarrow |b_k| = \frac{\cos \Delta h_k}{a|h_k|}$ 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 (38) 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 Fig. 4, 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. $\mathbb{E}[|h_k|] > \mathbb{E}[|e_k|]$, and using it we obtain

$$\cos \Delta h_k^{\max} = \frac{\sqrt{|h_k^2| - |e_k^2|}}{|h_k|} = \frac{|h_k^{\text{worst}}|}{|h_k|}.$$
 (40)

Hence, if we consider the worst case scenario, where the maximum phase difference is achieved, we have the following relations regarding the relation between the real channel conditions, the estimation error and the channel estimation itself [41]

$$|h'_k|^2 + |e_k|^2 = |h_k|^2, (41a)$$

$$\cos \Delta h_k = \cos \Delta h_k^{\max} = \frac{|h'_k|}{|h_k|}.$$
(41b)

In order to approximate the estimation error, due to the knowledge of its statistics alone, we will use the mean value as the best unbiased estimator, thus

$$|e_k|^2 \approx \mathbb{E}[|e_k|^2] = \frac{\mathbb{E}[|n|^2]}{P} = \frac{\sigma^2}{P} = c,$$
 (42)

for all k. This approach provides a tight approximation of the sinusoidal term for the greater majority of the involved channels, except for those that are greatly affected by the noise. On top of that, this method accounts for the worst case so the actual performance can possibly be better in a more favorable setting.

Using the approximations (41a), (41b) and (42), combined with the attainable values of $|b_k|$ given by Theorem 2, (39) can now be expressed as

$$MSE_{i}(a) = \sum_{k=1}^{i} \left| \left(a\sqrt{P} |h'_{k}| - 1 \right) \right|^{2} + a^{2}(\sigma^{2} + iPc) + c \sum_{k=i+1}^{K} \frac{1}{|h'_{k}|^{2} + c},$$
(43)

where $i \in \{1, \dots, K\}$ symbolizes the possible values of the critical number and the receiver gain, a, is the variable of interest. Following this, the optimal scaling factor a_i is given in Lemma 5.

Lemma 5: The optimal a that minimizes (43) is given by

$$a = a_i = \frac{\sqrt{P} \sum_{k=1}^{i} |h'_k|}{(\sigma^2 + iPc) + P \sum_{k=1}^{i} |h'_k|^2}.$$
 (44)

Proof: Considering (43) as a quadratic polynomial in terms of a and using a well known property for the global extremum of this function, we get the global minimum for every i, that is given by differentiating (43) in terms of a_i to finally get (44), and thus, the proof is completed.

At this point, it is observed that (44) will be a tight approximation for high SNR or good channel conditions. This is so, because, due to our approximation the power magnitudes, which will be $|b_k|_{\text{opt}} = \frac{\cos \Delta h_k}{a|h_k|} \approx \frac{|h'_k|}{a|h_k|^2} \approx \frac{|h'_k|}{a(|h'_k|^2+c)}$ by (38), 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 $|h'_1| \ge \sqrt{c}$. Otherwise, we observe that the receiver coefficient a that will be calculated can lead some of the weaker estimated channels to use the inverse channel method. Finally, (44) 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 D.

So, it suffices to solve K subproblems for $i \in 1, ..., K$ and, then, compare the minimum values $MSE_i^{min} = MSE_i(a_i)$. Theorem 2 dictates that in order to get a feasible solution for a_i , it must hold that

$$\frac{|h'_i|}{(|h'_i|^2 + c)\sqrt{P}} \ge a_i > \frac{|h'_{i+1}|}{(|h'_{i+1}|^2 + c)\sqrt{P}}.$$
(45)

If (45) is not satisfied, the following lemma can be used to search for feasible solutions.

 $\begin{array}{c|c} Lemma & 6: \quad \text{Let} & \mathcal{I}_i & \text{be the interval} \\ \left(\frac{|h'_{i+1}|}{(|h'_{i+1}|^2+c)\sqrt{P}}, \frac{|h'_i|}{(|h'_i|^2+c)\sqrt{P}}\right]. & \text{Then, if } a_i < \frac{|h'_{i+1}|}{(|h'_{i+1}|^2+c)\sqrt{P}} \\ \text{there exists } a \in \mathcal{I}_{i+1} \text{ such that a better MSE can be achieved} \\ \text{and if } a_i \geq \frac{|h'_i|}{(|h'_i|^2+c)\sqrt{P}} \text{ there exists } a \in \mathcal{I}_{i-1} \text{ such that a better MSE can be achieved.} \end{array}$

Proof: The proof is given in Appendix E.

The feasibility condition along with Lemma 6 effectively show that whenever a_i is not feasible with regards to power allocation, then there is another value of a that would achieve better MSE. Since at least one user must use its full power as proved in Corollary 2 and there are finite intervals of the form \mathcal{I}_i , a global minimum for MSE exists and its optimal receiver factor a will be such that it equals a feasible value of a_i . Thus, we only need to check the values $MSE_i(a_i)$ when a_i is feasible. This way we can calculate i^* and then estimate a_i and b_k for all k.

Theorem 3: The critical number for the imperfect CSI policy is given by

$$i^* = \underset{1 \le i \le K}{\operatorname{argmin}} \operatorname{MSE}(a_i), \tag{46}$$

Correspondingly, from Lemma 5, the optimum a will be given as $a^* = a_{i^*}$.

It is important to note that this approach differs from the one followed in [38] 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 [38].

A. Power Consumption

In this subsection, we compare the power efficiency of the imperfect CSI policy obtained by the optimization framework and the perfect CSI policy that ignores channel imperfections. Let the critical number of the perfect CSI policy be i_1^* and the critical number of the imperfect CSI policy be i_2^* . We define the total power consumption \mathcal{P} to be given by

$$\mathcal{P}_{tot,1} = \sum_{k=1}^{K} |b_k|^2 = Pi_1^* + \frac{1}{g_{i_1^*}^2} \sum_{k=i_1^*+1}^{K} \frac{1}{|h_k'|^2} \qquad (47)$$

for the perfect CSI policy and by

$$\mathcal{P}_{tot,2} = \sum_{k=1}^{K} |b_k|^2 = Pi_2^* + \frac{1}{a^{*2}} \sum_{k=i_2^*+1}^{K} \frac{|h_k'|^2}{(|h_k'|^2 + c)^2} \quad (48)$$

for the imperfect CSI policy. To examine the power efficiency of each policy, we introduce the following lemma.

Lemma 7: If
$$g_i < \frac{1}{|h'_{i+1}|\sqrt{P}}$$
 then $a_i < \frac{|h'_{i+1}|}{(|h'_{i+1}|^2 + c)\sqrt{P}}$
Proof: The proof is given in Appendix F.

With the help of Lemma 7, the following theorem is presented.

Theorem 4: The imperfect CSI policy always consumes more power than the perfect CSI policy that ignores imperfections, i.e. $\mathcal{P}_{tot,2} > \mathcal{P}_{tot,1}$.

Proof: Let the number of devices transmitting with full power when the perfect CSI policy is used be i_1 and the number of devices transmitting with full power when the imperfect CSI policy is used, i_2 . Similarly to Lemma 6 and also accordingly with [3], whenever there exists $i_1 > i$ such that $MSE_{i_1} < MSE_i$ for the perfect CSI policy, there will also exist $i_2 > i$ such that $MSE_{i_2} < MSE_i$ for the imperfect CSI policy. The minimum i for which $g_i < \frac{1}{|h'_{i+1}|\sqrt{P}}$ is not satisfied will be the critical number $i_1 = i_1^*$ for which the minimum MSE is achieved for the perfect CSI policy, which means that necessarily $i_2^* \ge i_1^*$ will hold. By the definition in (13), i_1^* is such that $g_{i_1^*} > g_{i_2^*}$ and whenever $i_2^* > i_1^*$, as proven in [3], it will also hold that $g_{i_2^*} > \frac{1}{|h'_{i_2^*}|\sqrt{P}}$. Also notice that for feasibility reasons it will be $a^* < \frac{|h'_{i_2^*}|}{(|h'_{i_2^*}|^2 + c)\sqrt{P}}$. From these inequalities, we derive that

these inequalities, we derive that

$$\frac{g_{i_1^*}}{a^*} > \frac{|h_{i_2^*}'|^2 + c}{|h_{i_2^*}'|^2} > \frac{|h_k'|^2 + c}{|h_k'|^2}, \ \forall k \in \{i_2^* + 1, \cdots, K\}.$$
(49)

If the equality $i_2^* = i_1^*$ holds, then by the definitions in (14), (44) it will be

$$\frac{g_{i_2^*}}{a^*} = 1 + \frac{i_2^* P c}{\sigma^2 + P \sum_{k=1}^{i_2^*} |h_k'|^2} > \frac{|h_{i_2^*+1}'|^2 + c}{|h_{i_2^*+1}'|^2}$$
(50)

which after simplifications holds because $g_{i_2^*} \geq \frac{1}{|h'_{i_2^*+1}|\sqrt{P}}$. Thus, from (50) we are back in (49). Due to full power transmission we get

$$Pi_{2}^{*} > Pi_{1}^{*} + \frac{1}{g_{i_{1}^{*}}^{2}} \sum_{k=i_{1}^{*}+1}^{i_{2}^{*}} \frac{1}{|h_{k}'|^{2}}$$
(51)

and by applying (49) for $i_2^* + 1 \le k \le K$ and summing up we get

$$\frac{1}{a^{*2}} \sum_{k=i_{2}^{*}+1}^{K} \frac{|h_{k}'|^{2}}{(|h_{k}'|^{2}+c)^{2}} > \frac{1}{g_{i_{1}}^{2}} \sum_{k=i_{2}^{*}+1}^{K} \frac{1}{|h_{k}'|^{2}}.$$
 (52)

Combining (51) and (52) completes the proof of Theorem 4.

B. Pilot Retransmission Policy

From the aforementioned 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 better estimations for the channels are desired. As a result, we can consider the possibility of making a second CSI estimation round, at least for some channels. One option is for the transmitting devices with 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 reestimate 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 utility function, called the Retransmission Policy Cost (RPC) as

$$\zeta(k) = \mathcal{C}(k)^d \left(\frac{\mathbb{E}[\text{MSE}_k]}{K}\right)^J, \qquad (53)$$

where C(k) symbolizes the cost in resources needed for k retransmissions, $\frac{\mathbb{E}[MSE_k]}{K}$ denotes the MSE for k retransmissions and our primary concern over the available resources is taken to be the selected cost. Moreover, parameters $\{d, f\} \in \mathbb{R}^2$ are considered to be weights for the cost and the 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. It is important to highlight that the proposed RPC can be changed and its weight parameters can be adjusted whenever CSI estimation takes place to better capture the current state and needs of the system. Without loss of generality, the time penalty required for the retransmission can be included in the RPC metric, but given that the decrease of $\mathbb{E}[\text{MSE}]$ is vital for the correct interpretation of the superimposed signals, the added overhead due to the added retransmissions is negligent for practical systems.

In order to evaluate in terms of RPC the proposed policy, we consider two special cases of interest.

Case 1) Power cost related RPC: In this case the primary cost is considered to be the power resources available to the transmitting devices. 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, $C_{power}(k) = (K + k)P$ and according to (53) the minimum of RPC is studied. RPC is then expressed as

$$\zeta(k) = \left[(K+k)P \right]^d \left(\frac{\mathbb{E}[\text{MSE}_k]}{K} \right)^f.$$
 (54)

Case 2) Throughput cost related RPC: Another interesting use of the RPC function would be to consider the lost information per transmission, i.e., the added overhead to the transmission, as the cost of the retransmissions. Then, we can use the ergodic capacity to find the information lost for every retransmission. In this context, since the channel estimation is performed for each device individually, for comparison we will assume each device to be transmitting its data in a typical one-to-one communication link. Considering this, for k retransmissions, we define the throughput cost as $C_{thr}(k) = \mathbb{E}[C_{tr,total}] + \sum_{l=1}^{k} \mathbb{E}[C_{tr,l}]$, where $\mathbb{E}[C_{tr,total}] =$ $\sum_{l=1}^{K} \mathbb{E}[C_{tr,l}]$ symbolizes the total information lost for the initial CSI estimation round, before any retransmissions take place. In this scenario, RPC is expressed as

$$\zeta(k) = \left(\mathbb{E}[C_{\text{tr,total}}] + \sum_{l=1}^{k} \mathbb{E}[C_{\text{tr},l}]\right)^{d} \left(\frac{\mathbb{E}[\text{MSE}_{k}]}{K}\right)^{f}.$$
 (55)

For the calculation of the values of ζ , we will use the following lemma in order to evaluate the throughput cost for every ordered channel.

Lemma 8: The ergodic capacity of the ordered channel gains is given by

$$\mathbb{E}[C_{\text{tr},r}] = \frac{r}{2\ln 2} \binom{K}{r} \sum_{m=0}^{r-1} \binom{r-1}{m} \frac{e^b}{b\rho\sigma_h^2} (-1)^{m+1} \text{Ei}(-b),$$

(56)

where $b = \frac{K-r+1+m}{\rho\sigma_h^2}$ and $\operatorname{Ei}(x) = -\int_{-x}^{\infty} \frac{e^{-t}}{t} dt$ denotes the exponential integral function.

Proof: The proof is presented in Appendix G.

It is worth pointing out that since (56) calculates the mean information loss, it essentially gives a measure of the extra overhead that will be caused by a retransmission for every channel. Therefore, the *Throughput cost related RPC* can be used to take into account the additional overhead of the system.

By definition, RPC expresses a trade-off between a considered cost in resources and the improved MSE achieved by the use of these resources. Then, according to (54) and (55), the preferable number of retransmissions will be given by

$$k = \underset{0 \le k \le K}{\operatorname{argmin}} \{\zeta(k)\},\tag{57}$$

where k = 0 would result in no retransmission.

V. SIMULATION RESULTS

In this section, simulation results are presented to validate the aforementioned analysis. The fading channels have been modeled as Circular-Symmetric Complex Normal distributed variables CSCN(0, 1) to simulate Rayleigh fading 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 MSE per user as AMSE = $\mathbb{E}[MSE]/K$ and the average per user \mathcal{P} as $\overline{\mathcal{P}} = \mathbb{E}[\mathcal{P}]/K$.

In Fig. 5, the performance of the perfect CSI algorithm and the proposed optimization technique for imperfect CSI under imperfect channel estimation is presented. As it can be observed, ignoring CSI imperfections not only achieves worse AMSE values, but also fails to converge to a solution as the number of transmitting 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 transmitting devices. Apart from that, it is significant to notice that this behavior is mainly a result of the inverse channel term achieved by the use of the algorithm that ignores the imperfect CSI.

It is worth noting that the robustness that the proposed policy exhibits for the increasing number of transmitting devices can be quite important in practical applications of over-the-air computing systems, depending on the use of the resulting target function itself. For example, in federated learning studies where the target function is itself a parameter that updates a global training model, the diverging behavior of the perfect CSI algorithm under imperfect CSI can be of pivotal significance to the functionality of the model since it could lead to more iterations taking place or divergence of the parameters of the model.

In Fig. 6, the comparison between the mean value of the inverse channel terms, the proposed bound (22), as well as the derived approximation expressions, as given by (26), (27), (34), and (18) combined with Lemma 2 is illustrated. Apart from confirming the validity of the theoretical analysis of this approach, it is important to observe that the growth rate is indeed logarithmic. On top of that, an important remark stems from the fact that the tight lower approximation seems to provide an asymptotic approximation of the term under investigation. Hence, studying the statistical probability of i^*



Fig. 5: MSE performance and distinct terms vs the number of transmitting devices in the AirComp system accounting or not for imperfect CSI (Transmit SNR = 10dB).

and combining it with (32) and (18), especially when the noise is sufficiently low, can provide useful insights for the design of a system that is going to use the policy that does not account for the CSI imperfections.



Fig. 6: Inverse channel terms vs the number of transmitting devices in the AirComp system for the various approximations (Transmit SNR = 10dB).

In Fig. 7 we look at the AMSE performance of different used policies for varying transmit SNR. Firstly, it is important to notice that the full power policy not only achieves large AMSE, but it fails to improve its performance for increasing SNR in contrast to every other discussed policy. This behavior is expected because in the full power policy no channel cancellation is expected which is necessary to correctly estimate the target function of interest. Also it is important to note that the inverse channel policy cannot reach the performance level of the perfect and imperfect CSI policies under imperfect CSI. This is a result of two factors, the non-



Fig. 7: AMSE vs transmit SNR for different policies and CSI conditions for K = 50 transmitting devices.

vanishing inverse channel term described by (12), which would be omitted in the case of perfect CSI, and the presence of noise since the receiver factor in this policy increases its effect. Concerning the proposed policy we can see that it outperforms the perfect CSI policy under imperfect CSI, especially in the low SNR region. This behavior is to be expected since lower transmit SNR results in greater estimations errors and thus, it is imperative to take this effect into consideration. We would like to point out that the performance improvement even for more practical scenarios like 10dB, is more than 20%, which pinpoints the usefulness of the proposed policy. As it is obvious, in the perfect CSI case the AMSE is smaller than in the imperfect CSI case. In order to get a more favorable AMSE under the latter, one could use the proposed RPC functions to achieve further performance improvements.

As it can be seen in Fig. 8, the perfect CSI policy does indeed require less power than the imperfect CSI policy regardless of the transmit SNR, validating Theorem 4. However, it is worth noting that for increasing K the latter shows the same converging behavior as the former, which means that the power per device is ultimately reduced in both cases. These results, however, do not take into account the MSE performance.

In Fig. 9 the retransmission policy, which was discussed in Section IV-B, is presented in terms of average MSE. For any retransmitting device we consider the new estimation of the channel to be given by $h''_k = \frac{h'_{k,1} + h'_{k,2}}{2}$, where $h'_{k,1}$ is the initial estimation of k-th channel and $h'_{k,2}$ is the re-estimation made by the proposed retransmission. 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

Authorized licensed use limited to: Aristotle University of Thessaloniki. Downloaded on March 10,2024 at 17:07:21 UTC from IEEE Xplore. Restrictions apply. © 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information



Fig. 8: \mathcal{P} performance vs the number of transmitting devices in the AirComp system accounting or not for imperfect CSI.

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.

From Fig. 10, we also notice that for the available resources at the transmitting devices, the preferable technique would be to make use of them in order to estimate more channels rather than better estimating a smaller number of channels. In order to observe this we simulate two scenarios, a first one where only 1 retransmission is made per channel and a second one where 2 retransmissions are made per channel. For the latter, the new estimation of the *k*-th channel is $h''_k = \frac{h'_{k,1} + h'_{k,2} + h'_{k,3}}{3}$. Comparing the two resulting curves for an even amount of resources (so that the 2 retransmissions can be made per channel) we observe that the single round retransmission achieves better performance in terms of AMSE.

In Fig. 11a the power related RPC metric is presented, i.e., $C_{power}(k)$ is examined by means of (54), with regards to the number of retransmissions. By (57) we can see a minimum exists and it is global. For the global minimum observed at k = 2 we can derive from Fig. 9 that an improvement of almost 7% is already achievable for only two channel re-estimations. It is important to point out that 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. Notice that the number of channels that were statistically found to be more error-prone as shown in Fig. 3. This also confirms the idea, on which the retransmission policy was based.

For comparison, in Fig. 11b we consider the proposed RPC with respect to power resources cost against a no retransmissions system which has greater maximum transmission power as a result of its conservation of power resources for no additional pilot symbols. In this context every pilot symbol



Fig. 9: MSE vs the number of retransmission with various policies (Transmit SNR = 10 dB).



Fig. 10: MSE vs the number of retransmission for 1 or 2 retransmission rounds for every channel (Transmit SNR = 10 dB).

is transmitted with maximum power, \sqrt{P} , as explained in Section II. For the no retransmissions scheme for k additional pilot symbols, the would-be consumed power is uniformly distributed among all K transmitting devices, thus the new maximum transmission power P'_{max} becomes

$$P_{\max}' = \sqrt{P} + k \frac{\sqrt{P}}{K} = \sqrt{P} \left(\frac{K+k}{K}\right).$$
 (58)

In Fig. 11b, the black curves show the AMSE of the two policies studied in Fig. 11a for increasing number of retransmissions, while the red curve shows the AMSE of the no retransmission system for increasing maximum transmission power as described by (58). Studying Fig. 11b, we can observe that for a few retransmissions the MSE improvement gain is greater than the gain for a similar system with no



(a) Power cost-related RPC vs the Number of Retransmissions

(b) MSE vs the number of Retransmissions or transmit SNR.

Fig. 11: The proposed policy, the random policy, and the original scheme without retransmission for d = f = 1 and $P_k = 1$, $\forall k$ when power consumption is considered as cost.

retransmissions and greater available power transmission at each user. Therefore, we can conclude that the utilized resources for some additional pilot symbols have greater impact than the maximum transmit SNR of each user. This behavior is expected, because as explained in Section III the main difficulty that arises in an imperfect CSI scenario is due to the fact that the worst channels' estimation will be greatly affected by the imperfections and thus, a more accurate estimation of them should be preferable over some additional transmission power. Although this may seem counter-intuitive, we would like to point out that in over-the-air systems, unlike other communication models, greater transmission power does not guarantee better results since the objective is the calculation of a target function and not the detection of the received symbols. In fact, as we can see from Fig. 7, the maximum power policy is the most energy-greedy, but clearly fails to achieve the MSE levels of our proposed optimal policy.

In Fig. 12 we present the throughput RPC for an increasing number of retransmissions as well as the corresponding mean throughput loss. Similarly to Fig. 11a, with the power RPC, in Fig. 12, with the throughput RPC expressed in (55), we can see that by (57) a minimum exists, it is global and it appears that for the statistically weaker channels all retransmission choices achieve a better trade-off between throughput and MSE than the original proposed scheme without any retransmissions. Thus, it is possible to choose the number of retransmissions in such a way, so that both previously discussed RPC functions will achieve better trade-offs. It is important to notice that in the throughput related RPC due to the ordered channels, every calculated term has already a different weight and, hence, every term has a different contribution in $\mathbb{E}[C_{tr,total}]$. This is evident by examining the gap in mean throughput loss between consecutive points. As expected, the lower the channel gain is, the smaller the corresponding mean throughput loss is, which



Fig. 12: Retransmission Policy Cost vs the Number of Retransmissions for the proposed policy and the original scheme without retransmission for d = f = 1 and $P_k = 1$, $\forall k$ when the throughput loss is considered as cost.

justifies the proposed policy and also explains the behavior of the RPC curve.

Regarding the RPC utility function cases, it is worth emphasizing on the fact that depending on the application of the over-the-air computing system, an AMSE threshold could be necessary to ensure functionality. In this case, further AMSE decrease can be obligatory. Also the fact that the overhead caused by every retransmission, which is expressed as the loss of transmitted information for every retransmission, can be explicitly calculated by Lemma 8 allows various RPC functions to be used while considering fundamental operating parameters of the system. Thus, RPC could offer a systematic way to handle the system's performance while simultaneously

making considerate usage of its resources to comply with operational constraints of the system's application of interest.

VI. CONCLUSIONS AND FUTURE EXTENSIONS

In this paper, an AirComp system under imperfect CSI assumptions was considered. The detrimental effect of imperfect CSI was presented for the MSE, and insightful approximations are derived to aid in the design of such systems. Moreover, a comprehensive analysis on how channel estimation errors affect the AirComp system was presented and a novel optimization framework to minimize MSE under those conditions was proposed. 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 utility function was presented alongside the retransmission policy to showcase the efficiency of the approach. Extensive 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. We note that the findings of this paper delve deeper into the design of more practical AirComp systems and they can form the basis for possible extensions in multiple-input multiple-output (MIMO) cases and federated learning.

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 $|h_1^{\text{cor}}| \leq |h_2^{\text{cor}}| \leq \cdots \leq |h_K^{\text{cor}}|$. From the order statistics we can find that every ordered sample $U_r = |h_r^{\text{cor}}|$ from a total of K samples has the following PDF

$$f_{U_r}(u_r) = \frac{K!}{(K-r)!(r-1)!} \frac{2u_r}{\sigma_h^2} \\ \times \exp\left[-\frac{u_r^2}{\sigma_h^2}(K-r+1)\right] \left[1 - \exp\left(-\frac{u_r^2}{\sigma_h^2}\right)\right]^{r-1}.$$
(59)

Hence, the expected value $\mathbb{E}[U_r]$ can be calculated by

$$\mathbb{E}[U_r] = \binom{K}{r} r \int_0^\infty \frac{2u_r^2}{\sigma_h^2} \exp\left[-\frac{u_r^2}{\sigma_h^2}(K-r+1)\right] \\ \times \left[1 - \exp\left(-\frac{u_r^2}{\sigma_h^2}\right)\right]^{r-1} du_r.$$
(60)

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

$$\mathbb{E}[U_r] = 2\sigma_h \binom{K}{r} r \int_0^\infty t^2 \exp\left[-t^2(K-r+1)\right] \\ \times \left[1 - \exp\left(-t^2\right)\right]^{r-1} dt \\ = 2\sigma_h \binom{K}{r} r \int_0^\infty t^2 e^{-Kt^2} [e^{t^2} - 1]^{r-1} dt.$$
(61)

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

$$\int_{0}^{\infty} t^{2} e^{-Kt^{2}} (e^{t^{2}} - 1)^{n} dt =$$

$$= \frac{1}{2} \sqrt{\frac{\pi}{4}} \sum_{k=0}^{n} \binom{n}{k} (-1)^{k} \frac{1}{(K - n + k)^{\frac{3}{2}}}.$$
(62)

Then, we can easily get (16) from (61) and (62), and, thus, the proof is concluded.

APPENDIX B Proof of Lemma 2

Using ordered statistics for U_r we know that the PDF is given by

$$f_{U_r}(u_r) = \binom{K}{r} r \lambda e^{-\lambda u_r (K-r+1)} (1 - e^{-\lambda u_r})^{r-1} du_r.$$
(63)

For r = 1 it is straightforward that

$$\mathbb{E}\left[\frac{1}{U_1}\right] = \lambda K \int_0^\infty \frac{e^{-tK}}{t} dt = +\infty, \tag{64}$$

since the function $g(t) = \frac{e^{-t\kappa}}{t}$ is continuous in $I = (0, +\infty)$, $g(t) > 0, \forall t \in I$ and $\lim_{t\to\infty} g(t) = 0$, $\lim_{t\to 0^+} = +\infty$. Setting $t = \lambda u_r$ the expected value will be equal to

$$\mathbb{E}\left[\frac{1}{U_r}\right] = \lambda \binom{K}{r} r \int_0^\infty \frac{1}{t} e^{-tK} (e^t - 1)^{r-1} dt.$$
(65)

For convenience let $I_r = \int_0^\infty \frac{1}{t} e^{-tK} (e^t - 1)^{r-1} dt$ and also m = r - 1. Applying integration by parts and binomial expansion

$$I_{r} = \ln(t)e^{-tK}(e^{t} - 1)^{m} \Big|_{0}^{\infty}$$

- $\int_{0}^{\infty} \ln(t) \left[e^{-tK}(e^{t} - 1)^{m}\right]' dt$
= $\sum_{k=0}^{m} {m \choose k} (-1)^{k} \left[(K - m + k) \int_{0}^{\infty} \ln(t)e^{-t(K - m + k)} dt \right]$
= $\sum_{k=0}^{m} {m \choose k} (-1)^{k} \left(\int_{0}^{\infty} \ln\left(\frac{z}{K - m + k}\right)e^{-z} dz \right).$ (66)

Notice that the first term on the integration by parts is equal to 0 because for $t \to +\infty$ we can prove that

$$\lim_{t \to +\infty} \left[\ln (t) e^{-tK} (e^t - 1)^m \right] = \lim_{t \to +\infty} \ln (t) P(e^{-t})$$
$$= \lim_{t \to +\infty} \frac{\ln t}{\frac{1}{P(e^{-t})}} = \lim_{t \to +\infty} \frac{\frac{1}{t}}{-\frac{e^{-t}P'(e^{-t})}{P^2(e^{-t})}}$$
$$= \lim_{u \to 0^+} \frac{P^2(u)}{u \ln (u) P'(u)} = \lim_{u \to 0^+} \frac{P(u)}{\ln (u)} \lim_{u \to 0^+} \frac{P(u)}{u P'(u)}$$
$$= 0 \times \frac{1}{K - m} = 0$$

and similarly for $t \to 0^+$ we can prove that

$$\lim_{t \to 0^+} \left[\ln (t) e^{-tK} (e^t - 1)^m \right] = \lim_{t \to 0^+} \left[\ln (t) (e^t - 1)^m \right]$$
$$= \lim_{t \to 0^+} \frac{\ln t}{\frac{1}{(e^t - 1)^m}} = \lim_{t \to 0^+} \frac{\frac{1}{t}}{-\frac{me^t (e^t - 1)^{m-1}}{(e^t - 1)^{2m}}}$$
$$= -\lim_{u \to 1} \frac{(u - 1)^{m+1}}{mu \ln u} = -\lim_{u \to 1} \frac{1}{mu} \lim_{u \to 1} \frac{(u - 1)^{m+1}}{\ln u}$$
$$= -\frac{1}{m} \lim_{u \to 1} (m + 1)u(u - 1)^m = 0$$

where by P we symbolized the polynomial described as $P(x) = \sum_{k=0}^{m} {m \choose k} (-1)^k x^{K-m+k}$ with $\deg(P(x)) = K$ and $\min \{\deg(P(x)) = K\} = K - m$. Breaking the logarithmic term in (66) would result in two integrals that can both be easily calculated by considering the well-known fact that $\gamma = -\int_0^\infty \ln (z)e^{-z}dz$. Hence, (66) will result in

$$I_r = \sum_{k=0}^{m} {m \choose k} (-1)^k \left(-\gamma - \ln \left(K - m + k\right)\right)$$
$$= \sum_{k=0}^{m} {m \choose k} (-1)^{k+1} \ln \left(K - m + k\right)$$
(67)

with (67) holding by the binomial expansion theorem, since

$$\sum_{k=0}^{m} \binom{m}{k} (-1)^{k} (-\gamma) = (-\gamma) \sum_{k=0}^{m} \binom{m}{k} 1^{m-k} (-1)^{k} = 0.$$
(68)

Substituting (67) back in (64) we can get (18) and the proof is completed.

APPENDIX C Proof of Lemma 3

From (63) and setting $t = \lambda u_r$ we get

$$\mathbb{E}\left[\frac{1}{U_r^2}\right] = \lambda^2 \binom{K}{r} r \int_0^\infty \frac{1}{t^2} e^{-tK} (e^t - 1)^{r-1} dt.$$
 (69)

For convenience let $J_r = \int_0^\infty \frac{1}{t^2} e^{-tK} (e^t - 1)^{r-1} dt$ and also m = r - 1. Applying integration by parts and binomial expansion

$$J_{r} = -\frac{1}{t}e^{-tK}(e^{t}-1)^{m}\Big|_{0}^{\infty} + \int_{0}^{\infty}\frac{1}{t}\left[e^{-tK}(e^{t}-1)^{m}\right]'dt$$

$$= -K\int_{0}^{\infty}\frac{1}{t}\left[e^{-tK}(e^{t}-1)^{m}\right]dt$$

$$+ m\int_{0}^{\infty}\frac{1}{t}\left[e^{-t(K-1)}(e^{t}-1)^{m-1}\right]dt$$

$$= -K\frac{1}{\lambda\binom{K}{r}r}\mathbb{E}\left[\frac{1}{U_{r,K}}\right] + m\frac{1}{\lambda\binom{K-1}{m}m}\mathbb{E}\left[\frac{1}{U_{m,K-1}}\right],$$

(70)

where the second index in the mean shows the number of samples in the ordered statistics. Notice that the first term on the integration by parts is equal to 0 because for $t \to +\infty$ we can prove that

$$\lim_{t \to +\infty} \frac{e^{-tK}(e^t - 1)^m}{t} = \lim_{t \to +\infty} \frac{e^{-t(K-m)}(1 - e^{-t})^m}{t} = 0$$

and similarly for $t \to 0^+$ we can prove that

$$\lim_{t \to 0^+} \frac{e^{-tK}(e^t - 1)^m}{t} = \lim_{t \to 0^+} \frac{(e^t - 1)^m}{t} = 0$$

by L'Hospital's Rule. Substituting (70) in (69) we get

$$\mathbb{E}\left[\frac{1}{U_r^2}\right] = \lambda K \left(\mathbb{E}\left[\frac{1}{U_{r-1,K-1}}\right] - \mathbb{E}\left[\frac{1}{U_{r,K}}\right]\right).$$
(71)

APPENDIX D Proof of Corollary 2

We begin with a proof by contradiction, assuming that a full inverse channel method is utilized. Then, by (14) the overall MSE of the system will be

$$MSE_0(a_0) = \sigma^2 a_0^2 + c \sum_{k=1}^K \frac{1}{|h'_k|^2 + c},$$
(72)

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

$$MSE_{1}(a_{1}) = \left| \left(a_{1}\sqrt{P}|h_{1}'| - 1 \right) \right|^{2} + a_{1}^{2}(\sigma^{2} + Pc) + c \sum_{k=2}^{K} \frac{1}{|h_{k}'|^{2} + c} = a_{1}^{2}\sigma^{2} + c \sum_{k=1}^{K} \frac{1}{|h_{k}'|^{2} + c} = MSE_{0}(a_{1})$$
(73)

and since $a_0 > a_1$ we see that $MSE_1(a_1) < MSE_0(a_0)$ 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.

Appendix E Proof of Lemma 6

For the first case notice that when $a = a' = \frac{|h'_{i+1}|}{(|h'_{i+1}|^2 + c)\sqrt{P}}$

it is trivial to show that

$$MSE_{i+1}(a') = \sum_{k=1}^{i+1} \left(a'\sqrt{P}|h'_k| - 1 \right)^2 + a'^2 [\sigma^2 + (i+1)Pc] + c \sum_{k=i+2}^{K} \frac{1}{|h'_k|^2 + c} = MSE_i(a') + \left(\frac{|h'_{i+1}|^2}{|h'_{i+1}|^2 + c} - 1 \right)^2 + c \frac{|h'_{i+1}|^2}{(|h'_{i+1}|^2 + c)^2} - c \left(\frac{1}{|h'_{i+1}|^2 + c} \right) = MSE_i(a').$$
(74)

Since $a_i < \frac{|h'_{i+1}|}{(|h'_{i+1}|^2+c)\sqrt{P}}$ and $MSE_i(a)$ is an increasing function in the interval $[a_i, \infty)$, we conclude that for $a \in \mathcal{I}_i$ due to (74) it will be $MSE_i(a) > MSE_i(a') = MSE_{i+1}(a')$. For the second case observe that since $a_i \geq \frac{|h'_i|}{(|h'|^2+c)\sqrt{P}}$ and

For the second case observe that since $a_i \ge \frac{|h'_i|}{(|h'_i|^2 + c)\sqrt{P}}$ and $MSE_i(a)$ is a decreasing function in the interval $[0, a_i]$ there exists $a' \in \left(\frac{|h'_i|}{(|h'_i|^2 + c)\sqrt{P}}, a_i\right] \subset \mathcal{I}_{i-1}$ such that $MSE_i(a) >$

 $\text{MSE}_i(a')$ because for feasibility reasons when the critical setting $t = \frac{u_r^2}{\sigma_h^2}$ we obtain number is *i* by (38) it must be $a \leq \frac{|h'_i|}{(|h'_i|^2 + c)\sqrt{P}}$. Then, we get get

$$MSE_{i}(a') = \sum_{k=1}^{i} \left(a'\sqrt{P}|h'_{k}| - 1 \right)^{2} + a'^{2}(\sigma^{2} + iPc) + c \sum_{k=i+1}^{K} \frac{1}{|h'_{k}|^{2} + c} > \sum_{k=1}^{i-1} \left| \left(a'\sqrt{P}|h'_{k}| - 1 \right) \right|^{2} + a'^{2}(\sigma^{2} + (i-1)Pc) + c \sum_{k=i}^{K} \frac{1}{|h'_{k}|^{2} + c} = MSE_{i-1}(a').$$
(75)

Thus, by (75) it will be $MSE_i(a) > MSE_{i-1}(a')$ and the proof is complete.

APPENDIX F PROOF OF LEMMA 7

By the definitions of a_i, g_i in (14), (44) we can work out that

$$g_i = \frac{a_i \sqrt{P} \sum_{k=1}^i |h'_k|}{\sqrt{P} \sum_{k=1}^i |h'_k| - a_i i P c}.$$
 (76)

Then, if $g_i < \frac{1}{|h'_{i+1}|\sqrt{P}}$ by (76) we can prove that

$$a_i < \frac{\sqrt{P}\sum_{k=1}^{i} |h'_k|}{P|h'_{i+1}|\sum_{k=1}^{i} |h'_k| + iPc}.$$
(77)

Finally, observe that

$$\frac{\sqrt{P}\sum_{k=1}^{i}|h'_{k}|}{P|h'_{i+1}|\sum_{k=1}^{i}|h'_{k}|+iPc} < \frac{|h'_{i+1}|}{(|h'_{i+1}|^{2}+c)\sqrt{P}},$$
(78)

because (78) after cross-multiplication and simplifying is equivalent to $\sum_{k=1}^{i} |h'_{k}| < i|\hat{h}'_{i+1}| = \sum_{k=1}^{i} |\hat{h'_{i+1}}|$ which holds due to the estimated channels' ordering.

APPENDIX G **PROOF OF LEMMA 8**

The capacity per symbol transmission of every channel is known to be given by

$$C_{\rm tr,i} = \frac{1}{2} \log_2 \left(1 + \rho |h_i|^2 \right) \tag{79}$$

Using order statistics we can find that every ordered sample $U_r = |h_r^{\rm cor}|$ from a total of K samples has the PDF described by (59). Hence, the expected value $\mathbb{E}[U_r]$ can be calculated by

$$\mathbb{E}[C_{\mathrm{tr},r}] = \frac{1}{2} \binom{K}{r} r \int_0^\infty \frac{2u_r}{\sigma_h^2} \exp\left[-\frac{u_r^2}{\sigma_h^2}(K-r+1)\right] \\ \times \left[1 - \exp\left(-\frac{u_r^2}{\sigma_h^2}\right)\right]^{r-1} \log_2\left(1 + \rho u_r^2\right) du_r.$$
(80)

$$\mathbb{E}[U_r] = \frac{1}{2} \binom{K}{r} r \int_0^\infty \exp\left[-t(K-r+1)\right] \\ \times \left[1 - \exp\left(-t\right)\right]^{r-1} \log_2\left(1 + t\rho\sigma_h^2\right) dt \\ = \frac{1}{2} \binom{K}{r} r \int_0^\infty e^{-Kt} (e^t - 1)^{r-1} \log_2\left(1 + t\rho\sigma_h^2\right) dt.$$
(81)

15

Using binomial expansion (81) can be rewritten as

$$\mathbb{E}[U_r] = \frac{1}{2} \binom{K}{r} r \sum_{m=0}^{r-1} \left[\binom{r-1}{m} (-1)^m I_{r,m} \right], \qquad (82)$$

where by $I_{r,m}$ we denote the following integral

$$T_{r,m} = \int_0^\infty e^{-Kt} e^{t(r-1-m)} \log_2 \left(1 + t\rho\sigma_h^2\right) dt.$$
 (83)

The last one can be rigorously calculated in a few steps. Substituting $y = 1 + t\rho\sigma_h^2$ and setting for convenience $b = \frac{K-r+1+m}{\rho\sigma_h^2}$, (83) is equivalent with

$$I_{r,m} = \frac{1}{\rho \sigma_h^2} \int_1^\infty e^{-b(y-1)} \log_2{(y)} dy.$$
 (84)

Then, changing to the natural logarithm and setting z = byfrom (84) we have

$$I_{r,m} = \frac{1}{K - r + 1 + m} \frac{1}{\ln 2} e^b \times \left[\int_b^\infty e^{-z} \ln(z) dz + \int_b^\infty e^{-z} \ln\left(\frac{1}{b}\right) dz \right].$$
(85)

Finally, by part integration of the first integral in (85) we obtain the following

$$I_{r,m} = \frac{1}{K - r + 1 + m} \cdot \frac{1}{\ln 2} e^b$$

$$\times \left\{ \left[-e^{-z} \ln\left(z\right) \right]_b^\infty - \ln\left(\frac{1}{b}\right) \left[e^{-z} \right]_b^\infty + \int_b^\infty \frac{e^{-z}}{z} dz \right\}$$

$$= \frac{1}{K - r + 1 + m} \cdot \frac{1}{\ln 2} e^b \int_b^\infty \frac{e^{-z}}{z} dz. \tag{86}$$

Combining relations (80), (81), (83) and (86) we can conclude (56).

REFERENCES

- [1] N. G. Evgenidis, V. K. Papanikolaou, P. D. Diamantoulakis, and G. K. Karagiannidis, "Over-the-Air Computing under Adaptive Channel State Estimation," in Proc. 18th Int. Conf. Wireless Mobile Comput., Netw. Commun. (WiMob), 2022, pp. 223-228.
- [2] Ericsson, "Ericsson Mobility Report," Tech. Rep., Nov. 2022. [Online]. Available: https://www.ericsson.com/en/reports-andpapers/mobility-report/reports/november-2022
- [3] 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.
- [4] 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.
- N. A. Mitsiou, P. S. Bouzinis, P. D. Diamantoulakis, [5] R. Schober, and G. K. Karagiannidis, "Accelerating Distributed Optimization via Over-the-Air Computing," 2022. [Online]. Available: https://arxiv.org/abs/2212.14003

- [6] P. S. Bouzinis, N. A. Mitsiou, P. D. Diamantoulakis, D. Tyrovolas, and G. K. Karagiannidis, "Intelligent Over-the-Air Computing Environment," *IEEE Wireless Commun. Lett.*, vol. 12, no. 1, pp. 134–137, 2023.
- [7] M. Goldenbaum, H. Boche, and S. Stańczak, "Harnessing interference for analog function computation in wireless sensor networks," *IEEE Trans. Signal Process.*, vol. 61, no. 20, pp. 4893–4906, Oct 2013.
- [8] —, "Analyzing the space of functions analog-computable via wireless multiple-access channels," in *Proc. 8th Int. Symp. Wireless Commun. Syst.*, Nov 2011, pp. 779–783.
- [9] 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.
- [10] M. Gastpar, "Uncoded Transmission Is Exactly Optimal for a Simple Gaussian Sensor Network," *IEEE Trans. Inf. Theory*, vol. 54, pp. 5247– 5251, 2008.
- [11] 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.
- [12] —, "On the Channel Estimation Effort for Analog Computation over Wireless Multiple-Access Channels," *IEEE Wireless Commun. Lett.*, vol. 3, no. 3, pp. 261–264, June 2014.
- [13] 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.
- [14] X. Zhai, X. Chen, and Y. Cai, "Power Minimization for Massive MIMO Over-the-Air Computation With Two-Timescale Hybrid Beamforming," *IEEE Wireless Commun. Lett.*, vol. 10, no. 4, pp. 873–877, April 2021.
- [15] 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.
- [16] L. Chen, N. Zhao, Y. Chen, F. R. Yu, and G. Wei, "Over-the-Air Computation for IoT Networks: Computing Multiple Functions With Antenna Arrays," *IEEE Internet of Things J.*, vol. 5, no. 6, pp. 5296– 5306, Dec. 2018.
- [17] X. Li, F. Liu, Z. Zhou, G. Zhu, S. Wang, K. Huang, and Y. Gong, "Integrated Sensing, Communication, and Computation Over-the-Air: MIMO Beamforming Design," *IEEE Trans. Wireless Commun.*, pp. 1– 1, 2023.
- [18] T. Jiang and Y. Shi, "Over-the-Air Computation via Intelligent Reflecting Surfaces," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, 2019, pp. 1–6.
- [19] W. Fang, Y. Jiang, Y. Shi, Y. Zhou, W. Chen, and K. B. Letaief, "Overthe-Air Computation via Reconfigurable Intelligent Surface," *IEEE Trans. Commun.*, vol. 69, no. 12, pp. 8612–8626, Dec. 2021.
- [20] 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.
- [21] Y. Sun, S. Zhou, Z. Niu, and D. Gunduz, "Dynamic Scheduling for Over-the-Air Federated Edge Learning With Energy Constraints," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 1, pp. 227–242, Jan. 2022.
- [22] K. Yang, T. Jiang, Y. Shi, and Z. Ding, "Federated Learning via Overthe-Air Computation," *IEEE Trans. Wireless Commun.*, vol. 19, no. 3, pp. 2022–2035, Mar. 2020.
- [23] 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.
- [24] 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.
- [25] 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. Sel. Areas Commun.*, vol. 39, no. 12, pp. 3742–3756, Dec. 2021.
- [26] X. Cao, G. Zhu, J. Xu, Z. Wang, and S. Cui, "Optimized Power Control Design for Over-the-Air Federated Edge Learning," *IEEE J. Sel. Areas Commun*, vol. 40, no. 1, pp. 342–358, Jan. 2022.
- [27] 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.
- [28] X. Cao, G. Zhu, J. Xu, and S. Cui, "Transmission Power Control for Over-the-Air Federated Averaging at Network Edge," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 5, pp. 1571–1586, May 2022.
- [29] K. Xu, H. H. Yang, Z. Zhao, W. Hong, T. Q. Quek, and M. Peng, "Pruning analog over-the-air distributed learning models with accuracy loss guarantee," in *Proc. IEEE Int. Conf. Commun.*, May 2022, pp. 5202– 5207.

- [30] G. Zhu, Y. Du, D. Gündüz, and K. Huang, "One-bit over-the-air aggregation for communication-efficient federated edge learning: Design and convergence analysis," *IEEE Transactions on Wireless Communications*, vol. 20, no. 3, pp. 2120–2135, March 2021.
- [31] H. Ye, G. Y. Li, and B.-H. F. Juang, "Deep over-the-air computation," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec 2020, pp. 1–6.
- [32] M. Fu, Y. Zhou, Y. Shi, W. Chen, and R. Zhang, "UAV Aided Overthe-Air Computation," *IEEE Trans. Wireless Commun.*, pp. 1–1, 2021.
- [33] H. Jung and S.-W. Ko, "Performance Analysis of UAV-Enabled Overthe-Air Computation Under Imperfect Channel Estimation," *IEEE Wireless Commun. Lett.*, vol. 11, no. 3, pp. 438–442, Mar. 2022.
- [34] 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, July 2019.
- [35] P. Park, P. D. Marco, and C. Fischione, "Optimized over-the-air computation for wireless control systems," *IEEE Commun. Lett.*, vol. 26, no. 2, pp. 424–428, Feb 2022.
- [36] 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].
- [37] X. Zang, W. Liu, Y. Li, and B. Vucetic, "Over-the-Air Computation Systems: Optimal Design With Sum-Power Constraint," *IEEE Wireless Commun. Lett.*, vol. 9, no. 9, pp. 1524–1528, Sep. 2020.
- [38] 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].
- [39] A. T. Asyhari and A. Guillén i Fàbregas, "Nearest Neighbor Decoding in MIMO Block-Fading Channels With Imperfect CSIR," *IEEE Transactions on Information Theory*, vol. 58, no. 3, pp. 1483–1517, March 2012.
- [40] M. Biguesh and A. Gershman, "Training-based MIMO channel estimation: a study of estimator tradeoffs and optimal training signals," *IEEE Transactions on Signal Processing*, vol. 54, no. 3, pp. 884–893, March 2006.
- [41] A. Papoulis and S. Pillai, Probability, Random Variables, and Stochastic Processes, ser. McGraw-Hill series in electrical engineering: Communications and signal processing. Tata McGraw-Hill, 2002. [Online]. Available: https://books.google.gr/books?id=g6eUoWOlcQMC
- [42] B. C. Arnold, N. Balakrishnan, and H. N. Nagaraja, A First Course in Order Statistics (Classics in Applied Mathematics). USA: Society for Industrial and Applied Mathematics, 2008.
- [43] Introduction to Probability and Statistics for Engineers and Scientists. Elsevier, 2021. [Online]. Available: https://doi.org/10.1016/c2018-0-02166-0
- [44] R. P. Boas and J. W. Wrench, "Partial sums of the harmonic series," *The American Mathematical Monthly*, vol. 78, no. 8, pp. 864–870, 1971. [Online]. Available: http://www.jstor.org/stable/2316476



Nikos G. Evgenidis received the Diploma (5 years) in Electrical and Computer Engineering from the Aristotle University of Thessaloniki, Greece, in 2022, where he is currently pursuing his PhD. He is also a member of the Wireless and Communications and Information Processing (WCIP) group. His major research interests include semantic communications, over-the-air computing, non-orthogonal multiple access, machine learning and optimization theory.



Vasilis K. Papanikolaou (Member, IEEE) received the Diploma (5 years) and PhD degrees in Electrical and Computer Engineering from the Aristotle University of Thessaloniki, Greece, in 2018 and 2023, respectively, where he was a member of the Wireless and Communications and Information Processing (WCIP) group. He is currently a Humboldt Post-Doctoral Research Fellow with the Institute for Digital Communications, Friedrich-Alexander-University Erlangen-Nurnberg (FAU), Erlangen, Germany. He was a Visiting Research Assistant with Lancaster

University, UK, with Khalifa University, Abu Dhabi, UAE, and with the Northumbria University, Newcastle-upon-Tyne, UK. His major research interests include optical wireless communications, wireless power transfer, multiple access schemes, and optimization theory. He was an Exemplary Reviewer of the IEEE Wireless Communications Letters in 2019, 2020, and 2022 and of the IEEE Open Journal of the Communications Society in 2021 (top 3% of reviewers).



Panagiotis D. Diamantoulakis received the Diploma (five years) and the Ph.D. degree from the Department of Electrical and Computer Engineering, Aristotle University of Thessaloniki, Thessaloniki, Greece, in 2012 and 2017, respectively. Since 2022 he is a Postdoctoral Fellow with the Department of Applied Informatics, University of Macedonia, Thessaloniki, Greece. Since 2017, he has been a Postdoctoral Fellow with Wireless Communications and Information Processing (WCIP) Group, AUTH and since 2021, he has been a Visiting Assistant

Professor with the Key Lab of Information Coding and Transmission, Southwest Jiaotong University, Chengdu, China. His research interests include optimization theory and applications in wireless networks, optical wireless communications, and goal-oriented communications. He is a Working Group Member of the Newfocus COST Action "European Network on Future Generation Optical Wireless Communication Technologies." He serves as an Editor of IEEE Open Journal of the Communications Society, Physical Communications (Elsevier), and Frontiers in Communications and Networks, while during 2018-2023 he has been an Editor of IEEE Wireless Communications Letters in 2020, and an Exemplary Reviewer of the IEEE Communications Letters in 2014 and the IEEE Transactions on Communications in 2017 and 2019 (top 3% of reviewers).



George K. Karagiannidis (Fellow, IEEE) is currently Professor in the Electrical and Computer Engineering Dept. of Aristotle University of Thessaloniki, Greece and Head of Wireless Communications and Information Processing (WCIP) Group. He is also Faculty Fellow in the Cyber Security Systems and Applied AI Research Center, Lebanese American University. His research interests are in the areas of Wireless Communications Systems and Networks, Signal processing, Optical Wireless Communications, Wireless Power Transfer and Applications

and Communications, and Signal Processing for Biomedical Engineering. Dr. Karagiannidis was in the past Editor in several IEEE journals and from 2012 to 2015 he was the Editor-in Chief of IEEE Communications Letters. From September 2018 to June 2022 he served as Associate Editor-in Chief of IEEE Open Journal of Communications Society. Currently, he is in the Steering Committee of IEEE Transactions on Cognitive Communications & Networks. Recently, he received three prestigious awards: The 2021 IEEE ComSoc RCC Technical Recognition Award, the 2018 IEEE ComSoc SPCE Technical Recognition Award and the 2022 Humboldt Research Award from Alexander von Humboldt Foundation. Dr. Karagiannidis is one of the highly-cited authors across all areas of Electrical Engineering, recognized from Clarivate Analytics as Web-of-Science Highly-Cited Researcher in the eight consecutive years 2015-2022.