Over-the-Air Collaborative Learning in Joint Decision Making

Carlos Feres, *Member, IEEE*, Bernard C. Levy, *Life Fellow, IEEE*, Zhi Ding, *Fellow, IEEE*
Department of Electrical and Computer Engineering
University of California, Davis, CA, USA

***Abstract*—We propose an over-the-air learning framework for collaborative decision making in wireless sensor networks. The low complexity framework leverages low-latency sensor transmission for a decision server to coordinate measurement sensors for hypothesis testing through over-the-air aggregation of sensor data over a multiple-access channel. We formulate several collaborative over-the-air hypothesis testing problems under different practical protocols for collaborative learning and decision making. We develop hypothesis tests for these network protocols and deployment scenarios including channel fading. We provide performance benchmark for both basic likelihood ratio test and generalized likelihood ratio test under different deployment conditions. Our results clearly demonstrate gain provided by increasing number of collaborative sensors.**

***Index Terms*—Internet of Things, federated learning, hypothesis testing.**

I. Introduction

Internet of Things (IoT) broadly cover a variety of technologies that effectively deploy and integrate a wide range of devices and sensors to advance a myriad of applications, including smart cities and environmental protection. One important class of IoT applications involve the detection of an underlying critical event based on local observation measured by distributed nodes. Such collaborative learning is a special type of federated learning which, by definition, is machine learning technique that optimizes an algorithm across multiple decentralized devices holding local data samples without exchanging them. In this paper, we study collaborative learning for joint decision making through over-the-air data aggregation in multiple access channel by distributed low-complexity wireless sensors.

Traditional grant-based access networks feature a server node coordinates data transmission by sensor nodes by issuing access grant to facilitate orthogonal channel access. Such centralized orthogonal multiple access (OMA) network consumes significant bandwidth resources, and leads to longer latency. On the other hand, grant-free networks require effective signal separation of multiple sensor transmissions, which is difficult for large number of sensors. Recent works leveraged blind signal recovery by utilizing multiple receive antennas [1]–[3]. However, such approaches are computationally costly and require larger number of receive antennas than the number of participating nodes for multi-sensor signal separation.

However, in many collaborative learning applications it is unnecessary to recover the data of each participating sensor node. Recent works have considered collaborative learning from distributed sources. In particular, Federated Learning (FL) has gained much attention as a joint learning framework [4], [5] for distributed nodes to jointly train a shared learning model by using their own local data without data sharing to protect privacy. In the most popular form of FL, a server collects local gradients to update a shared deep learning neural network (DLN). As such, FL avoids revealing private data and potentially reduces local storage. However, in the context of mobile systems, the need to transmit local updates in a typically large DLN learning model would require complex protocols for wireless access and coordination, and consume large amounts of bandwidth. Such complexity and resource demand make such FL setup less appealing for IoT.

Taking a broader view beyond the resource inefficient DLN framework, different collaborative learning can positively impact the performance and other operational characteristics of an IoT system, such as improving bandwidth efficiency, reduced network coordination, or increased energy efficiency. Applications such as environmental monitoring, where the network distributively senses an underlying event from multiple sensor observations, can be formulated as a collaborative decision making problem that asks multiple sensors for their local measurement and aggregate all the sensor observations to make a decision. In particular, techniques such as over-the-air computation (AirComp) [6], [7] can effectively improve the spectrum efficiency instead of relying on resource-hungry OMA protocols. AirComp based collaborative learning framework takes advantages of the low complexity of IoT nodes, requires simple access protocol, and improves the overall reliability of joint decision making. In particular, AirComp exploits the natural superposition of multiple simultaneous signals, and yields an aggregated signal for the server node to more reliably detect the underlying event in a well formulated hypothesis testing problem with low latency, high bandwidth efficiency, and modest computational load.

In this work, we focus on the collaborative decision making IoT framework as a special case of federated learning. We formulate the collaborative learning process as a hypothesis testing based on over-the-air signal aggregation techniques to reduce latency, simplify network access, improve spectrum efficiency, and improve the accuracy of the decision making process by exploiting the aggregation of multiple sensor data.

Section II presents a general system model for AirCompCL. In Section III we study an scenario where sensors precode their transmitted signals to counter channel phase. Section IV shows

This material is based upon work supported by the National Science Foundation under Grants No. 1824553, No. 2009001 and No. 2029027.

the test design for networks where sensors do not precode their signals. Section V presents the case where non-precoded signals are collected over multiple transmission bursts that are affected by different, random channel realizations. Section VI shows numerical results and Section VII summarizes our work.

Notations: We denote vectors and matrices using lowercase and uppercase bold letters, respectively. $\mathrm{Re}(a)$ and $|a|$ denote the real part and magnitude of scalar a, respectively. We denote transpose, conjugate, and conjugate transpose of $\boldsymbol{z}$ as $\boldsymbol{z}^{\mathsf{T}}$, $\boldsymbol{z}^*$, and $\boldsymbol{z}^{\mathsf{H}}$, respectively. $\mathbf{1}$ and $\boldsymbol{I}$ are a vector of ones and the identity matrix of appropriate size. $\|\cdot\|$ denotes ℓ_2 norm and $\mathbb{E}\{\cdot\}$ denotes expectation.

II. System Model

Consider a TDD wireless system of single-antenna nodes, where a server node hosts S sensors. In each slot, sensors simultaneously transmit (analog) signals to the server over a shared wireless channel. We assume that all sensors have acquired network timing and are synchronized at the server, e.g., via round-trip delay information, such that their transmitted signals would aggregate synchronously at the receiving server. Furthermore, we assume that each burst duration is below the coherence time of wireless channel such that channel gains remain constant within each transmission slot.

AirComp aims to compute an estimation or decision from a nomographic function of distributed data collected locally by participating sensors. Each sensor transmits an analog signal $x_i(t)$, for $i \in \mathcal{S} = \{1, \dots, S\}$, corresponding to a local pre-processing function ϕ_i of the collected signal $u_i(t)$, i.e.

$$x_i(t) = \phi_i\big(u_i(t)\big), \quad i \in \mathcal{S}. \tag{1}$$

The server node receives all wireless signals simultaneously over shared channel with individual gains $g_i \in \mathbb{C}$ as:

$$y(t) = \sum_{i \in \mathcal{S}} g_i x_i(t) + n(t), \tag{2}$$

where $n(t)$ is circularly symmetric complex AWGN with power density ω^2, independent of all channels and signals.

In the particular problem of over-the-air hypothesis testing, the server decides on an hypothesis using K samples of the received signal Eq.(2). A major distinction is whether the sensors directly send their measurements to the server, or if they perform local hypothesis testing and send their own decision to the server. In this work we focus on a centralized decision at the server using the direct transmission of the local measurements of sensors.

Without loss of generality, we focus on binary hypothesis testing as an initial Over-the-Air Collaborative Learning (AirCompCL) framework. Let H_0 and H_1 be two underlying hypotheses regarding the measured events. The i-th sensor observes waveform $\boldsymbol{u}_i$ for duration T under H_0 and H_1 as

$$H_j: \quad u_i(t) = s_j(t) + w_i(t), \quad 0 \le t < T, \tag{3}$$

where $w_i(t)$ represents white Gaussian measurement noise with power density σ_j^2, independent of $s_j(t)$ and independent from other sensor noises. Each sensor may transmit $x_i(t)$ in

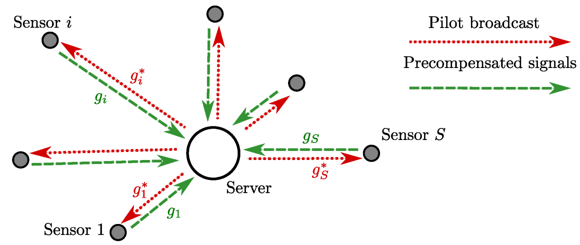


Fig. 1: AirCompCL system using channel precompensation, owing to channel reciprocity. The server node broadcasts pilots that allow sensors to estimate channel and precode their signals.

K ($1 \le K$) opportunities. By using a classic matched filter $s_1(t) - s_0(t)$ at the receiver for the general binary hypothesis [8, Chapter 2], we obtain an equivalent vector representation of K discrete samples for data signals [9, Chapter 2]:

$$\boldsymbol{u}_i | H_j \sim \mathcal{N}(m_j \mathbf{1}, \sigma_j^2 \boldsymbol{I}). \tag{4}$$

The server aims to make a decision based on the collective observation of m_j. Hence, throughout this work we assume that sensor variances are equal in both hypotheses, i.e. $\sigma_0^2 = \sigma_1^2 = \sigma^2$, and define sensor confidence as the squared normalized Gaussian distance, $\xi = |m_1 - m_0|^2/\sigma^2$, and sensor uncertainty as its reciprocal.

Let $\boldsymbol{X} \in \mathbb{C}^{K \times S}$ be a matrix such that its i-th column corresponds to $\phi_i(\boldsymbol{u}_i)$, and let $\boldsymbol{g} = [g_1 \cdots, g_S]^{\mathsf{T}}$. Hence, the samples of the received signal are collected in

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{g} + \boldsymbol{n} \tag{5}$$

with $\boldsymbol{n}$ representing the corresponding receiver noise vector. We also define the channel SNR $\gamma = \mathbb{E}\{|g_i|^2\}|m|^2/\omega^2$, in terms of a single sensor.

For a set of samples $\boldsymbol{y}$, binary hypothesis testing is based in comparing the likelihood ratio (LR) $L(\boldsymbol{y})$ with a threshold τ in a likelihood ratio test (LRT) $L(\boldsymbol{y}) \underset{H_0}{\overset{H_1}{\gtrless}} \tau$.

We can now study this formulation for AirCompCL under different system settings. In particular, we classify different scenarios according to whether there exists TDD channel reciprocity.

III. AirCompCL with Channel Precompensation

We first describe our over-the-air protocol under TDD channel reciprocity, in Fig.1. The server broadcasts timestamps and pilots that allow sensors to estimate reciprocal uplink channel phase and to estimate time-advance. Sensors may precompensate channel phase in their transmission, i.e., $\phi_i(\boldsymbol{u}) = g_i^*/|g_i| \cdot \boldsymbol{u}$. Owing to precompensation and time-advance, the server receives over-the air sum of signals

$$\boldsymbol{y} = \sum_{i \in \mathcal{S}} |g_i| \cdot \boldsymbol{u}_i + \boldsymbol{n}. \tag{6}$$

Without loss of generality, we consider $\boldsymbol{y}$ under hypothesis H_j to follow a real K-dimensional Gaussian distribution $\mathcal{N}(\mu\mathbf{1}, \nu^2\boldsymbol{I}|H_j)$ with

$$\mu_j = \mu|H_j = \sum_{i\in\mathcal{S}} |g_i| \cdot m_j = A \cdot m_j \,, \tag{7a}$$

$$\nu^2 = \sum_{i\in\mathcal{S}} |g_i|^2 \cdot \sigma^2 + 0.5\omega^2 = P \cdot \sigma^2 + 0.5\omega^2. \tag{7b}$$

Note that $\mathbb{E}\{A\} = S \cdot \mathbb{E}\{|g_i|\}$ and $\mathbb{E}\{P\} = S \cdot \mathbb{E}\{|g_i|^2\}$.

The LR over K samples for a binary hypothesis testing is

$$L(\boldsymbol{y}) = \exp\left(\frac{\|\boldsymbol{y}-\mu_0\mathbf{1}\|^2}{2\nu^2} - \frac{\|\boldsymbol{y}-\mu_1\mathbf{1}\|^2}{2\nu^2}\right). \tag{8}$$

During uplink, the server may estimate the reception parameters A, P and/or ω^2, depending on protocol. However, note that $A \geq 0$ and $P \geq 0$. In fact, $A = P = 0$ when no sensors transmit signals. In the following, we study several detection problems depending on the knowledge of the relevant access channel parameters A, P and ω^2.

A. Server with known access channel parameters

We consider the case that the server node has estimated A, P and ω^2. This simple setting serves as baseline to compare with cases where (some of) these parameters are unknown.

First, consider the simple but well known case with $H_0 : m = m_0$ vs $H_1 : m = m_1$, with $m_1 > m_0$. We assume known sensor variance σ^2, which implies known ν^2. Hence, the LR

$$L(\boldsymbol{y}) = \exp\left(\frac{A(m_1-m_0)}{\nu^2}\mathbf{1}^\mathsf{T}\boldsymbol{y} - \frac{A^2(m_0^2-m_1^2)K}{2\nu^2}\right) \tag{9}$$

is monotone increasing function of statistic $t_1(\boldsymbol{y}) = \mathbf{1}^\mathsf{T}\boldsymbol{y}$. Invoking the Karlin-Rubin theorem [10] and setting the probability of false alarm $P_F\big(t_1(\boldsymbol{y}) > \eta_1|H_0\big) = \alpha$, the test

$$\delta_1(\boldsymbol{y},\alpha): \; t_1(\boldsymbol{y}) \underset{H_0}{\overset{H_1}{\gtrless}} \eta_1 = \nu Q^{-1}(\alpha) + Am_0 \tag{10}$$

is UMP of size α for testing H_0 vs. H_1 with power

$$P_D(\alpha) = 1 - Q\big(d - Q^{-1}(\alpha)\big), \tag{11}$$

by defining a normalized distance

$$d = \frac{\sqrt{K}A|m_1-m_0|}{\sqrt{P\sigma^2+0.5\omega^2}}. \tag{12}$$

As expected, d grows with the number of sensors at a rate $\sqrt{S}$ and decreases with noise and aggregated sensor uncertainty.

B. Server with unknown access channel parameters

Here we study the case when the server fails to estimate some or all access channel parameters. Hence, we now design a test for sensor means, i.e. $H_0 : m = m_0$ vs $H_1 : m = m_1 > m_0$ with unknown $A > 0$. Regrettably, no UMP test exists for arbitrary m_0 because the test statistic and size would depend on A. However, it is possible to obtain an UMPI for this detection problem for the special case of $m_0 = 0$. Let us design a test to decide $H_0 : m = 0$ vs. $H_1 : m_1 \neq 0$ with unknown $A > 0$. If either σ^2, P or ω^2 are unknown, ν^2 is unknown, and thus both hypotheses are composite. Hence, we rely on ML estimates of mean and variance:

$$\hat{\mu}_{\mathrm{ML}} = \frac{1}{K}\mathbf{1}^\mathsf{T}\boldsymbol{y}, \qquad \hat{\nu}^2_{\mathrm{ML}} = \frac{1}{K}\|\boldsymbol{y}-\hat{\mu}_{\mathrm{ML}}\mathbf{1}\|^2. \tag{13}$$

The generalized LRT (GLRT) yields

$$\begin{aligned} L(\boldsymbol{y}) &= \frac{f_{\mathrm{y}}(\boldsymbol{y}|\hat{\mu}_{\mathrm{ML}},\hat{\nu}^2_{\mathrm{ML}})}{f_{\mathrm{y}}(\boldsymbol{y}|0,\hat{\nu}^2_{\mathrm{ML}})} \\ &= \exp\left(\frac{\|\boldsymbol{y}\|^2}{\hat{\nu}^2_{\mathrm{ML}}} - \frac{\|\boldsymbol{y}-\hat{\mu}_{\mathrm{ML},1}\mathbf{1}\|^2}{\hat{\nu}^2_{\mathrm{ML}}}\right) \underset{H_0}{\overset{H_1}{\gtrless}} \tau, \end{aligned} \tag{14}$$

and reduces to the test

$$\delta_2(\boldsymbol{y},\eta_2): \quad t_2(\boldsymbol{y}) = \frac{|\mathbf{1}^\mathsf{T}\boldsymbol{y}|}{\sqrt{K}\|\boldsymbol{y}\|} \underset{H_0}{\overset{H_1}{\gtrless}} \eta_2, \tag{15}$$

which is a UMPI test for testing H_1 against H_0, as is invariant under the group composed of scale transformations and the group of symmetries with respect to the hyperplane orthonormal to $\mathbf{1}$, and the LR is a monotone increasing function of $t_2(\boldsymbol{y})$ [9]. The distribution of $t_2(\boldsymbol{y})$ is derived in [9] using hyperspherical coordinates, but equivalently, t_2 under H_0 follows a central folded-t distribution with $K-1$ degrees of freedom (DoF), and under H_1 follows a folded-t distribution with $K-1$ DoF and noncentral parameter $\sqrt{K}Am_1/\hat{\nu}_{\mathrm{ML}}$ [11].

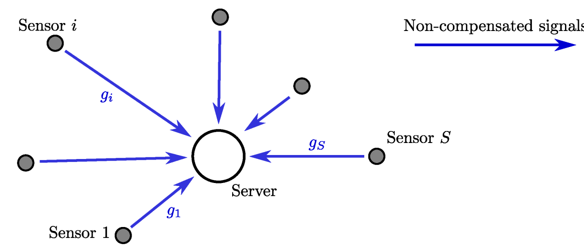


Fig. 2: AirCompCL system with no precompensation. Sensors send their data directly, and server receives signals with channels that could have constructive or destructive mixing.

IV. AirCompCL without Channel Compensation

In IoT deployment where sensors do not utilize channel reciprocity for precompensation, it would be natural for sensors to simply transmit $\phi_i(u) = u$ for $i \in \mathcal{S}$, as depicted in Fig. 2. In this scenario, $\boldsymbol{y}$ is a vector of independent complex random variables with distribution under H_j as $\mathcal{CN}\big(\mu\mathbf{1}, \nu^2\boldsymbol{I}|H_j\big)$, where

$$\mu_j = \mu|H_j = \sum_{i\in\mathcal{S}} g_i \cdot m_j = Gm_j, \quad G \in \mathbb{C}, \tag{16a}$$

$$\nu^2 = \sum_{i\in\mathcal{S}} |g_i|^2 \cdot \sigma^2 + \omega^2 = P\sigma^2 + \omega^2. \tag{16b}$$

In this case, sum of channel gains G at the server can be constructive or destructive. This setting is akin to classical signal detection with unknown gain under Gaussian noise, with the difference being the presence of local sensor noise further scaled by access channel gains. Because of access channel phase, the LR corresponds to

$$L(\boldsymbol{y}) = \exp\left(\frac{\|\boldsymbol{y}-\mu_0\mathbf{1}\|^2}{\nu^2} - \frac{\|\boldsymbol{y}-\mu_1\mathbf{1}\|^2}{\nu^2}\right), \tag{17}$$

depending on the knowledge of parameters G, P and ω^2.

A. Known access channel parameters

Consider first the case of known G, P and ω^2 at the server. Under hypotheses $H_0 : m = m_0$ vs $H_1 : m = m_1 > m_0$ with known sensor noise variance, we may refer back to Section III-A. We use the statistic $t_3(\boldsymbol{y}) = \mathrm{Re}(G^* \mathbf{1}^\mathsf{T} \boldsymbol{y})/|G|$, and establish the Neyman-Pearson (NP) test of size α

$$\delta_3(\boldsymbol{y}, \alpha) : \; t_3(\boldsymbol{y}) \underset{H_0}{\overset{H_1}{\gtrless}} \eta_3 = \nu Q^{-1}(\alpha) + G^* m_0. \tag{18}$$

For a given size α, the power of the test is

$$P_D(\alpha) = 1 - Q\left(\frac{\sqrt{K}|G|(m_1 - m_0)}{\sqrt{P\sigma^2 + \omega^2}} - Q^{-1}(\alpha) \right). \tag{19}$$

B. Unknown access channel parameters

Similarly to the case of Section III-B, as the server does not know the phase of G, it is not possible to obtain an UMP(I) for testing different means unless one of them is zero. Hence, we again test sensor means, i.e. $H_0 : m = m_0 = 0$ vs. $H_0 : m_1 \neq m_0$, with G unknown. If either σ^2, P or ω^2 are unknown, the hypotheses are composite. This detection problem is invariant under the group of complex scalar transformations (i.e. rotations and scalar multiplication), as for any $c \in \mathbb{C} \neq 0$, we have that $\mu' = c\mu$ and $\nu' = |c|\nu$ yields the same problem. For this test, the GLRT approach using ML estimates of all unknown parameters reduces to

$$\delta_4(\boldsymbol{y}, \alpha) : \quad t_4(\boldsymbol{y}) = \frac{|\mathbf{1}^\mathsf{T}\boldsymbol{y}|}{\sqrt{K}\|\boldsymbol{y}\|} \underset{H_0}{\overset{H_1}{\gtrless}} \eta_4. \tag{20}$$

As in Section III-B, this test is UMPI for testing H_1 against H_0. The difference with $t_2(\boldsymbol{y})$ from (15) is that $\boldsymbol{y}$ is now complex such that $|\mathbf{1}^\mathsf{T}\boldsymbol{y}|$ denotes magnitude. Hence, t_4 under H_0 is the magnitude of a complex t-distribution random variable with $K-1$ DoF. Under H_1, t_4 corresponds to the magnitude of a noncentral complex-t random variable with $K-1$ DoF and mean $\sqrt{K} G m_1 / \hat{\nu}_{\mathrm{ML}}$ [12].

V. Fading Access Channels

In the previous sections, we considered that the server obtained K snapshots of local measurements of one short burst transmission from the sensors, during which the fading channels do not change. Nevertheless, the server could also obtain observations occurring in different bursts, subject to different channel realizations regardless of channel precompensation. In this section, we study the effect of accumulating a single observation per burst from $K > 1$ bursts, and in particular, focus on the case with no precompensation from Section IV, in the hopes of being able to counter the effect of channel phase with different channel realizations.

We now consider K transmission bursts with non-precompensated transmissions, where during the k-th burst the sensors transmit a signal $x_i[k] = \phi(u_i[k]) = u_i[k]$, and the server collects one snapshot per burst. Moreover, each burst experiences random i.i.d. channel realizations following a Rayleigh distribution, i.e. $g_i[k] \sim \mathcal{CN}(0, 1)$. To counter channel realizations, the server makes a decision over the K samples obtained in K distinct bursts. Let $\boldsymbol{y} = [y_1 \; \cdots \; y_K]^\mathsf{T}$ be the vector that collects the single snapshot of each burst, and the received signal follows the model

$$y_k = \sum_{i \in \mathcal{S}} g_i[k] \cdot u_i[k] + n_k \,, \tag{21}$$

where the k-th snapshot of $\boldsymbol{u}_i$ follows a complex Gaussian distribution with mean m and variance σ^2, and $\boldsymbol{n}$ is multivariate circularly symmetric AWGN with variance $\omega^2 \boldsymbol{I}$.

Let $z_{i,k} = g_i[k] u_i[k]$ denote the product of the i-th channel realization with their corresponding signal during the k-th burst. For Rayleigh channels, the distribution of each $z_{i,k}$ follows a product-of-Gaussians distribution, and given that the channels have zero mean, the resulting PDF [13]–[15] is

$$z_{i,k}|H_j \sim \frac{2|z_{i,k}|}{\pi\sigma^2} \mathrm{e}^{-\kappa_j^2} \sum_{n=0}^{\infty} \frac{\kappa_j^{2n}}{(n!)^2} \frac{|z_{i,k}|^n}{\sigma^n} \mathrm{K}_n\left(\frac{2|z_{i,k}|}{\sigma^2} \right), \tag{22}$$

where K_v is the Bessel function of the second kind of order v and $\kappa_j = |m_j|/\sigma$. The distribution of z_k corresponds to the convolution of S different copies of distribution (22), convolved with the AWGN distribution. In [16] the authors derive the characteristic function of the inner product of complex Gaussian vectors, which corresponds to the characteristic function of the complex random variable $z_k = \sum_{i \in \mathcal{S}} z_{i,k}$:

$$\Phi_{\mathrm{z_k}}(j\boldsymbol{\zeta})|H_j = \left(1 + \frac{\|\boldsymbol{\zeta}\|^2\sigma^2}{4} \right)^{-S} \exp\left(- \frac{\|\boldsymbol{\zeta}\|^2 S m_j^2}{4 + \|\boldsymbol{\zeta}\|^2\sigma^2} \right),$$

whose PDF can be expressed in the form of a doubly infinite series of Whittaker functions [17], but is not helpful to derive and study the LR in closed form. Moreover, the AWGN still needs to be considered, with a resulting characteristic function

$$\begin{aligned} &\Phi_{\mathrm{y}}(j\boldsymbol{\zeta})|H_j \\ &= \left(1 + \frac{\|\boldsymbol{\zeta}\|^2\sigma^2}{4} \right)^{-S} \exp\left(- \frac{\|\boldsymbol{\zeta}\|^2 S m_j^2}{4 + \|\boldsymbol{\zeta}\|^2\sigma^2} - \frac{\|\boldsymbol{\zeta}\|^2\omega^2}{4} \right). \end{aligned} \tag{23}$$

To obtain the LR, one can perform numerical inversion of (23) for different hypotheses and use spline interpolation, to then construct the LR. Regrettably, such process does not provide analytical insights that helps to design decision tests. However, numerical simulations show that the distribution of z_k resembles a zero-mean Gaussian distribution. As every observation z_k is obtained by the sum of S non-compensated sensor averages and AWGN, the Central Limit Theorem (CLT) allows us to approximate the distribution of y_k as

$$f_{\mathrm{y}}(y|H_j) \approx \mathcal{CN}(0, \nu_j^2) \,, \tag{24}$$

where $\nu_j^2 = \omega^2 + S(\sigma^2 + |m_j|^2)$, thanks to the independence of the i-th channel and signal in each burst [17]. Using these CLT approximations, we can test for different sensor signal means by performing tests for different variances of zero-mean Gaussian received signals over K bursts.

Formally, we test different sensor mean magnitudes $H_0 : |m| = |m_0|$ vs. $H : |m| = |m_1| > |m_0|$ with equal sensor variances, by redefining the detection problem as detecting $H_0 : \nu^2 = \nu_0^2$ vs. $H : \nu^2 = \nu_1^2 > \nu_0^2$ assuming zero-mean Gaussian signals, which we call the approximate problem. Assuming known ω^2 and σ^2, we use the statistic $t_5(\boldsymbol{y}) = \|\boldsymbol{y}\|^2$, which under each hypothesis follows a χ^2 distribution with $2K$ DoF after normalization by the corresponding variance ν_j^2. Hence, the test

$$\delta_5(\boldsymbol{y}, \eta_5) : \quad t_5(\boldsymbol{y}) \underset{H_0}{\overset{H_1}{\gtrless}} \eta_5 = \nu_0^2 \Omega^{-1}(\alpha), \tag{25}$$

where Ω is the CDF of a χ^2_{2K} distribution, is UMP of size α for the approximate problem, with power

$$P_D(\alpha) = \exp\Big(-\frac{\nu_0^2}{\nu_1^2}\Omega^{-1}(\alpha)\Big) \sum_{m=0}^{2K-1} \frac{1}{m!}\Big(-\frac{\nu_0^2}{\nu_1^2}\Omega^{-1}(\alpha)\Big)^m.$$

VI. Numerical Results

This section presents analytical and numerical simulations for the different tests analyzed earlier in the paper. In particular, we study the effect of sensor aggregation in our AirCompCL problem for different numbers of collaborative sensors. In our simulations, we test Rayleigh channels, i.e. $g_i \sim \mathcal{CN}(0, 1)$, and for analytical results, we set

$$\mathbb{E}\{A\} = S\frac{\sqrt{\pi}}{2},\ \mathbb{E}\{P\} = S,\ \mathbb{E}\{G\} = 0,\ \mathbb{E}\{|G|\} = \frac{\sqrt{S\pi}}{2}.$$

Figs. 3 and 4 show the analytical performance for testing different means with and without channel compensation, respectively. Figs. 3a and 4a show the receiver operating characteristic (ROC) of the tests δ_1 and δ_3, under channel SNR of -15dB and -5dB respectively, with different values of sensor confidence and number of users. As expected, the probability of detection increases significantly with the number of users that collaborate in the decision making process, even when that also increases the total sensor uncertainty observed by the server. Nevertheless, for low channel SNR and only $K = 1$ samples, the effect of sensor confidence is small for compensated signals (Fig. 3a). Without channel compensation, the effect of both channel SNR and sensor confidence is stronger, and for low sensor confidence the performance gain of sensor aggregation diminishes. Figs. 3b and 4b show the probability of detection for varying values of channel SNR, at a fixed probability of false alarm of 10%. At low channel SNR, sensor uncertainty does not have strong impact. However, for high channel SNR, sensor uncertainty becomes more dominant in the decision process, especially for non-compensated channels. Importantly, larger number of sensors lowers the required channel SNR to achieve high probability of detection even under significant sensor uncertainty. The use of channel precompensation greatly increases the performance of the tests across parameter values, as expected.

Fig. 5 shows the performance of tests when access channel parameters are unknown. We show the analytical and empirical ROC of precompensated transmissions (test δ_2 of Sec. III-B), and empirical ROC of non-compensated transmissions (test δ_4 of Sec. IV-B. In both cases, performance increases with the number of sources, even for low channel SNR $\gamma = -5\,\text{dB}$ and low number of samples $K = 2$. As expected, the test enjoys significantly worse results with non-compensated transmissions. This is evident given the additional challenge of unknown channel gain (that could be destructive or constructive) and unknown channel phase.

Finally, Fig. 6 shows the ROC of the test δ_5 for different means under the protocol of Section V. We simulate the protocol with a single sample from $K = 2$ different bursts, with low channel SNR and low sensor confidence. We show the empirical ROC obtained using the statistic $\|\boldsymbol{z}\|^2$, the empirical ROC using the LRT computed by numerical inversion of (23), and the analytical ROC using the proposed Gaussian approximation via CLT. For all practical purposes, our analytical Gaussian approximation coincides with both the Monte Carlo simulations of the LRT obtained by numerical inversion of (23), and our empirical tests using the statistic $\|\boldsymbol{z}\|^2$. Moreover, the test improves with increasing number of sources, even in conditions with low channel SNR and sensor confidence, and using only $K = 2$ samples to average over varying channels.

VII. Conclusion

This work investigates over-the-air computation for collaborative detection in wireless IoT networks. We consider low cost sensors in distributed measurement for simultaneous transmission over a shared access channel. The proposed over-the-air collaborative learning framework achieves high spectrum efficiency and enhances decision making through very simple access protocols. We devise three protocols and corresponding hypothesis tests for AirCompCL: (1) sensors that can precompensate their signals and counter channel phase thanks to TDD reciprocity; (2) sensors that do not compensate signals; and (3) sensors that send non-compensated signals across different channel realizations. Our results of the designed tests show significant performance improvement with the number of participating sensors, even when accounting for local sensor noise affecting the measurement accuracy and uncompensated access channel phases. Future work include the study of AirCompCL with local sensor decisions, extensions to M-ary hypothesis testing, consider different sensor noise models, and perform practical experiments in testbed networks.

References

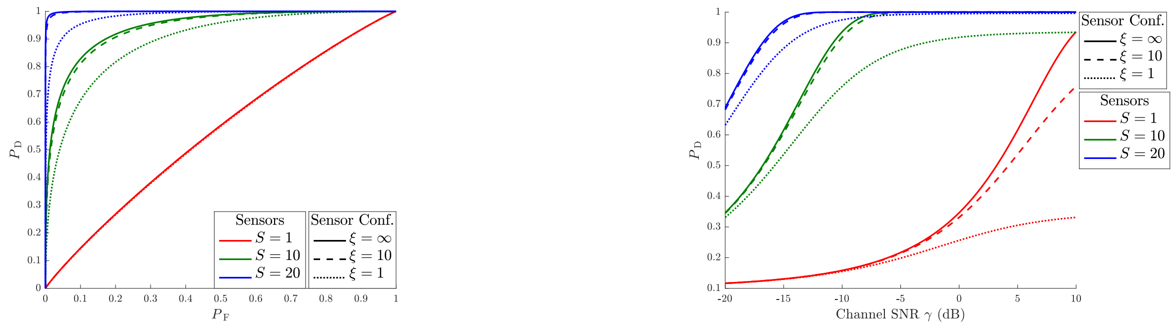
[1] C. Feres and Z. Ding, "Wirtinger Flow Meets Constant Modulus Algorithm: Revisiting Signal Recovery for Grant-Free Access," *IEEE Trans. Signal Process.*, vol. 69, pp. 6515–6529, 2021.

[2] C. Feres and Z. Ding, "A Riemannian Geometric Approach to Blind Signal Recovery for Grant-Free Radio Network Access," *IEEE Trans. Signal Process.*, vol. 70, pp. 1734–1748, 2022.

[3] J. Dong and Y. Shi, "Nonconvex demixing from bilinear measurements," *IEEE Trans. Signal Process.*, vol. 66, no. 19, pp. 5152–5166, Oct 2018.

[4] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y. Arcas, "Communication-Efficient Learning of Deep Networks from Decentralized Data," in *Proc. 20th Int. Conf. Artif. Intell. Statist.*, 2017, pp. 1273–1282.

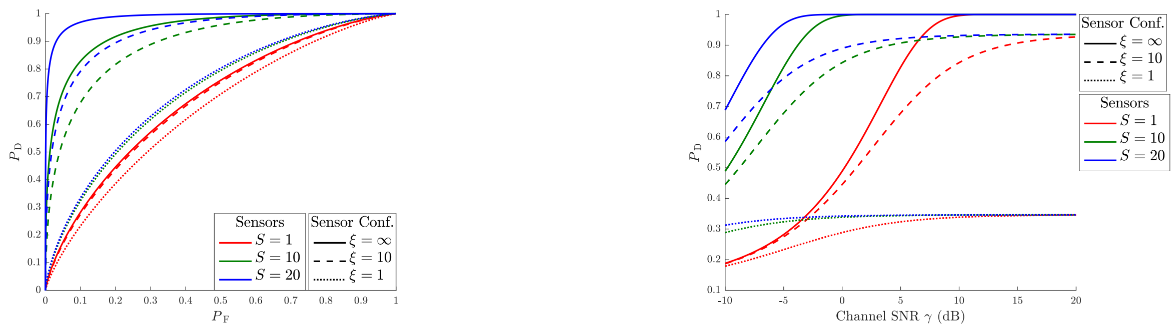
[5] S. Niknam, H. S. Dhillon, and J. H. Reed, "Federated Learning for Wireless Communications: Motivation, Opportunities, and Challenges," *IEEE Commun. Mag.*, vol. 58, no. 6, pp. 46–51, 2020.



(a) ROC of test for different means, with $\gamma = -15\,\mathrm{dB}$.

(b) Probability of detection vs. channel SNR, with $P_F = 0.1$.

Fig. 3: Performance with channel precompensation and known access channel parameters, with $K = 1$ and $m_1 = -m_0 = 0.5$.



(a) ROC of test for different means, with $\gamma = -5\,\mathrm{dB}$.

(b) Probability of detection vs. channel SNR, with $P_F = 0.1$.

Fig. 4: Performance without channel precompensation and known access channel parameters, with $K = 1$ and $m_1 = -m_0 = 0.5$.

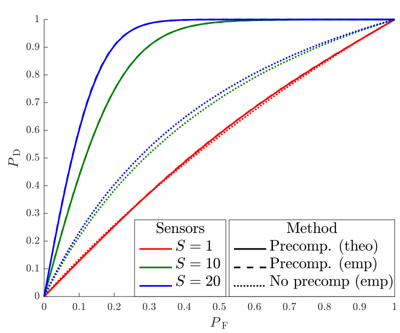


Fig. 5: ROC under unknown access channel parameters, with $\gamma = -5\,\mathrm{dB}$, $\xi = 1$, $K = 2$, $m_0 = 0$ and $m_1 = 1$.

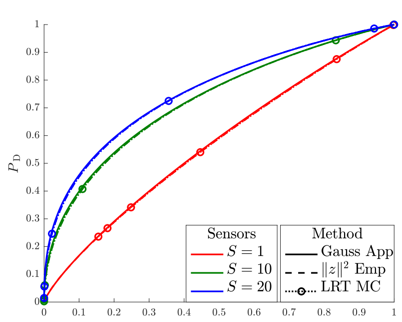


Fig. 6: ROC based on multiple Tx bursts under fading channels: $\gamma = -5\,\mathrm{dB}$, $\xi = 1$, $m_1 = 2m_0 = 2$, and $K = 2$.

[6] 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, 2013.

[7] Y.-S. Lee, K.-H. Lee, and B. C. Jung, "Beamforming Techniques for Over-the-Air Computation in MIMO IoT Networks," *Sensors*, vol. 20, no. 22, 2020.

[8] H. L. V. Trees, *Detection Estimation and Modulation Theory, Part I: Detection, Estimation, and Filtering Theory*. New York, NY, USA: Wiley, 1968.

[9] B. C. Levy, *Principles of Signal Detection and Parameter Estimation*, 1st ed. New York, NY, USA: Springer, 2007.

[10] S. Karlin and H. Rubin, "The Theory of Decision Procedures for Distributions with Monotone Likelihood Ratio," *Ann. Math. Statist.*, vol. 27, no. 2, pp. 272–299, 1956.

[11] S. Psarakis and J. Panaretoes, "The folded t distribution," *Commun. in Statist. - Theory and Methods*, vol. 19, no. 7, pp. 2717–2734, 1990.

[12] W.-J. Yan, L. Yang, X. Yang, and W.-X. Ren, "Statistical modeling for fast Fourier transform coefficients of operational vibration measurements with non-Gaussianity using complex-valued t distribution," *Mech. Syst. Signal Process.*, vol. 132, pp. 293–314, 2019.

[13] N. O'Donoughue and J. M. F. Moura, "On the Product of Independent Complex Gaussians," *IEEE Trans. Signal Process.*, vol. 60, no. 3, pp. 1050–1063, 2012.

[14] S. Guruacharya, B. K. Chalise, and B. Himed, "On the Product of Complex Gaussians With Applications to Radar," *IEEE Signal Process. Lett.*, vol. 26, no. 10, pp. 1536–1540, 2019.

[15] T. Betlehem and A. Coulson, "Distribution of the sum of a complex Gaussian and the product of two complex Gaussians," in *2012 Australian Commun. Theory Workshop (AusCTW)*, 2012, pp. 126–129.

[16] R. K. Mallik and N. C. Sagias, "Distribution of Inner Product of Complex Gaussian Random Vectors and its Applications," *IEEE Trans. Commun.*, vol. 59, no. 12, pp. 3353–3362, 2011.

[17] M. Simon, *Probability Distributions Involving Gaussian Random Variables: A Handbook for Engineers and Scientists*. Boston, MA, USA: Springer, 2002.