Bayesian Decision Making via Over-the-Air Soft Information Aggregation

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

***Abstract*—This work formulates a collaborative decision-making framework that exploits over-the-air computation to efficiently aggregate soft information from distributed sensors. This new AirCompFDM protocol approximates the sufficient statistic (SS) of optimum binary hypothesis testing at a server node in this distributed sensing environment under different operation constraints. Leveraging pre/post-processing functions on over-the-air aggregation of sensor log-likelihood ratios, AirCompFDM significantly improves bandwidth efficiency with little detection loss, even from modest numbers of participating sensors and imperfect phase pre-compensation. Without phase pre-compensation, the benefit of over-the-air sensor aggregation diminishes but still can mitigate the effect of channel noise. Importantly, AirCompFDM outperforms the traditional bandwidth-hungry polling scheme, even under low SNR. Furthermore, we analyze the Chernoff information and obtain the approximate effect of sensor aggregation on the probability of detection error that can help develop advanced detection strategies.**

***Index Terms*—Internet of Things, decision-making, collaborative learning, soft information, hypothesis testing.**

I. Introduction

Internet of Things (IoT) broadly covers 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 observations measured by distributed nodes. Of course, the IoT network requires consolidating the most local sensor observations possible to optimize detection performance, i.e., minimize error and maximize correct detection. In this paper, we study the technical challenge of achieving efficient collaborative learning for joint decision-making with data from distributed low-complexity wireless sensors.

Previous works have considered the aggregation of sensor data to improve decision-making at the server. In [1], the authors study a traditional system where the server polls each sensor individually over a multiple-access channel (MAC) with channel capacity R, and determine that R sensors sending one-bit decisions to a server equipped with maximum *a posteriori* (MAP) detector is an optimal detection strategy. However, data fusion relies on polling each sensor individually, with several hidden costs and ideal assumptions. Other works such as [2], [3] attempt to tackle some of these shortcomings, providing detection strategies while lifting some of these assumptions. Regardless, there still exist sizeable computational and bandwidth cost of recovering each sensor signal due to channel estimation and signal separation, the need of pilot signals, network coordination, and other practical considerations.

As an alternative to MAC polling, grant-free access attempts to separate multiple joint transmissions. The server may employ multiple receive antennas to exploit spatial diversity [4]–[6], although signal separation is computationally intensive, requires a significant number of samples of received signal, and the number of receive antennas constrains the number of simultaneous sensor transmissions that can be recovered. Thus, grant-free access schemes have limited applicability in performance- and energy-constrained networks.

However, in many distributed sensing scenarios the server does not need to recover the data of each sensor. Approaches like Federated Learning (FL) [7], [8], which jointly train a shared learning model without directly transmitting local data, have shown that both server and nodes can learn from each other by iteratively sharing model updates of a deep learning neural network, sharing training load and avoiding revealing potentially private data. Nevertheless, FL requires complex network coordination and large bandwidth usage, and thus it is not appropriate for typical IoT deployments.

To overcome these difficulties in collaborative learning, data fusion techniques such as over-the-air computation (AirComp) [9] propose the aggregation of analog signals in the uplink channel, thus providing bandwidth and energy savings by at least a factor of S, the number of active sensors. AirComp requires only a simple access protocol and small computational load, which is ideal for IoT nodes. In our previous work [10], we use AirComp for collaborative decision-making where sensors send direct measurements, and we further demonstrate that detection performance can improve substantially with increasing number of participant sensors in different scenarios, while saving considerable bandwidth with simultaneous transmission of all sensors, and moreover, incurring very low computational and energy costs.

In this paper, we extend our previous results. We now consider a collaborative detection problem where the sensors share *soft information scores* that consolidate their local observations, further improving the decision at the server via over-the-air computation. We call this formulation AirCompFDM. We develop theoretical performance bounds under a Bayesian framework for AirCompFDM, and show that detection per-

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

formance improves for increasing participating sensors with minimal communication load, even in challenging scenarios with one observation per sensor. In particular, for the case of local Gaussian measurements, we also show that our proposed strategy can outperform a MAC polling scheme with perfect sensor signal recovery [1], even in low channel SNR scenarios.

Notations: In the following, vectors will be denoted with small boldface letters, such as $\boldsymbol{z}$. Sets are denoted with calligraphic capital letters. The transpose, element-wise complex conjugation and conjugate transpose are denoted by $\boldsymbol{z}^{\mathsf{T}}$, $\overline{\boldsymbol{z}}$ and $\boldsymbol{z}^{\mathsf{H}}$, respectively. $\mathbf{1}$ represents a vector of ones of appropriate size. Expectation and variance are denoted as $\mathbb{E}\{\cdot\}$ and $\mathrm{Var}\{\cdot\}$, respectively. Finally, $\mathbb{1}[\mathrm{cond}]$ denotes the indicator function of whether cond is true.

II. System Model

Consider a wireless system of single-antenna nodes, where a server node hosts S sensors. In each transmission 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.

A. Over-the-Air Formulation

AirComp aims to compute an estimation or decision from a nomographic function of distributed data collected locally by participating sensors. Each sensor observes a sequence of N observations $\boldsymbol{v}_i = [v_{i,1} \dots v_{i,N}]^{\mathsf{T}}$, for $i \in \mathcal{S} = \{1, \dots, S\}$, of a phenomena H belonging to a discrete set $\{H_0, H_1\}$, which are the two underlying hypotheses with prior probabilities $\pi_0, \pi_1 > 0$, respectively, such that $\pi_0 + \pi_1 = 1$.

At discrete time k, sensors compute a summary message $u_i(k) = u_i(\boldsymbol{v}_i; k)$ and transmit analog signals $x_i(k)$, $i \in \mathcal{S}$, corresponding to a local pre-processing function $\phi_{i,k}$ of the summary message $u_i(k)$, i.e.

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

To save bandwidth with AirComp, the server node receives all wireless signals simultaneously over a shared multiple access channel with individual gains $g_i(k) \in \mathbb{C}/\{0\}$ as:

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

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

The server collects $K \geq 1$ samples of the received signal (2). If the K samples are obtained within a transmission slot, they all experience the same channel gains. Conversely, if the samples are taking in different transmission slots, we assume that channel realizations are independent over slots, resulting in that samples $y(k)$ are independent of each other. In the particular problem of over-the-air decision making, the server decides on an hypothesis using these K samples, collected in the vector $\boldsymbol{y} = [y(1), \dots, y(K)]^{\mathsf{T}}$, by computing a post-processing function $\lambda = \Psi(\boldsymbol{y})$.

B. Sufficient Statistic and Over-the-Air Approximation

If the server had access to all sensor observations from S sensors, the optimal Bayesian test consolidates all observations using the likelihood ratio (LR) or equivalently, the log-likelihood ratio (LLR)

$$\ell(\boldsymbol{v}_1, \cdots, \boldsymbol{v}_S) = \ln\left(\frac{p_{V|H}(\boldsymbol{v}_1, \cdots, \boldsymbol{v}_S|H_1)}{p_{V|H}(\boldsymbol{v}_1, \cdots, \boldsymbol{v}_S|H_0)}\right). \tag{3}$$

As sensor observations are i.i.d., we can rewrite the LLR as

$$\ell(\boldsymbol{v}_1, \cdots, \boldsymbol{v}_S) = \sum_{i \in \mathcal{S}} \ln\left(\frac{p_{V|H}(\boldsymbol{v}_i|H_1)}{p_{V|H}(\boldsymbol{v}_i|H_0)}\right) = \sum_{i \in \mathcal{S}} \ell(\boldsymbol{v}_i), \tag{4}$$

where $\ell(\boldsymbol{v}_i)$ is the LLR computed with local observations only at sensor i. Hence, if the sensors transmit their local LLRs, the server only needs to sum them to form a minimal sufficient statistic (SS) λ for optimal decision-making, and it is only natural to define $u_i(k) = \ell\big(\boldsymbol{v}_i\big)$. Hence, the optimal test is

$$\lambda = \sum_{i \in \mathcal{S}} u_i(k) = \sum_{i \in \mathcal{S}} \ell(\boldsymbol{v}_i) \underset{H_0}{\overset{H_1}{\gtrless}} \ln\left(\frac{\pi_0}{\pi_1}\right) = \gamma. \tag{5}$$

Now, if the server implements over-the-air computation in ideal conditions, i.e. with zero channel noise and equal channels with no fading, the received signal at the server is exactly λ. In other words, AirComp yields the SS of summed LLRs with no additional computation and only needing minimal signal bandwidth, which is indeed the best possible framework for collaborative decision-making.

However, this ideal scenario does not occur in practice, and over-the-air aggregation exhibits amplitude and phase distortions for each sensor signal, plus the effect of channel noise, as presented in (2). In other words, the received signal y is a sample of a noisy and channel-weighted statistic (NCWS), and the technical challenge is to design pre- and post-processing functions ϕ_i and Ψ and a NCWS $\tilde{\lambda}(\boldsymbol{y})$ that approximates the SS, i.e.

$$\tilde{\lambda}(\boldsymbol{y}) = \Psi\Bigg(\sum_{i \in \mathcal{S}} g_i(k) \phi_{i,k}\big(u_i(k)\big) + n(k)\Bigg) \approx \lambda. \tag{6}$$

Hence, our focus will be to study the effects of channel gains and noise in detection performance. In particular, we will devise how to deal with channel phases and/or magnitudes.

C. Compensation in Over-the-Air SS Approximation

Ideally, the system would compensate channel fading by using individual channel inverses, i.e. sensors would use the pre-processing functions $x_i(k) = g_i^{-1} u_i(k)$. Regrettably, sensors usually have limited available energy, and is impractical to assume that they can totally compensate for arbitrary channel magnitudes, which can exhibit significant attenuation in real-world applications. Hence, we do not study full channel compensation and focus on channel phase only.

However, if there exists channel reciprocity such as in a TDD link between server and sensors, the server broadcast used to acquire timing can also be used to estimate channel phase. If we assume that the estimation is exact, the sensors can use the pre-processing functions

$$x_i(k) = \frac{\overline{g_i(k)}}{|g_i(k)|} u_i(k), \tag{7}$$

and denoting $a_i(k) = |g_i(k)| > 0$, the received signal is

$$y(k) = \sum_{i \in \mathcal{S}} a_i(k) u_i(k) + n(k), \tag{8}$$

and as the compensated channels are real, the server makes a decision using NCWS $\tilde{\lambda}_{\mathrm{P}}(\boldsymbol{y}) = K^{-1}\mathrm{Re}\{\mathbf{1}^{\mathsf{T}}\boldsymbol{y}\}$. We call this protocol AirCompFDM-P.

In the opposite scenario where there is no channel reciprocity, sensors cannot estimate channels beforehand and are unable to precompensate their local LLR signals. Thus, sensors do not perform any pre-processing and we simply let $x_i(k) = u_i(k)$, resulting in

$$y(k) = \sum_{i \in \mathcal{S}} g_i(k) u_i(k) + n(k). \tag{9}$$

Regardless, we assume that the server is able to estimate the aggregated channel at each sample, $G_k = \sum_{i \in \mathcal{S}} g_i(k) \in \mathbb{C}$, and compensates for its resulting phase. Under this protocol, denoted AirCompFDM-U, we define the NCWS

$$\tilde{\lambda}_{\mathrm{U}}(\boldsymbol{y}) = \frac{1}{K}\mathrm{Re}\Big\{\sum_{k=1}^{K} \frac{\overline{G_k}}{|G_k|} y(k)\Big\}. \tag{10}$$

Finally, we can consider that phase compensations are not exact and subject to errors, and we instead use quantized phase corrections. For example, if we divide the complex plane into 4 regions with angles $\pm\pi/4$ and $\pm 3\pi/4$ radians, we can compensate phase in the corresponding multiple of $\pi/2$ radians. We call AirCompFDM-M when sensors and server perform quantized phase compensation with M uniform angle partitions of $2\pi/M$ radians, with offset of π/M. We will not analyze this more practical protocol, but will show its performance in numerical experiments.

III. Performance Analysis of AirCompFDM

A. Preliminaries

Let $p_j(y) = p_{Y|H}(y|H_j)$, $j \in \{0,1\}$ be the probability distributions of the received signal under each hypothesis. Using a maximum *a posteriori* (MAP) rule, the optimum probability of error in decision-making is

$$E^* = \int \min\{\mathbb{P}(H_0|y), \mathbb{P}(H_1|y)\} p(y)\,\mathrm{d}y, \tag{11}$$

and it is known that the Chernoff information

$$C^* = -\min_{\alpha \in [0,1]} \log \int p_0^{\alpha}(y) p_1^{1-\alpha}(y)\,\mathrm{d}y \tag{12}$$

yields the best achievable exponent for a Bayesian probability of error:

$$E^* \le \pi_0^{\alpha^*} \pi_1^{1-\alpha^*} \mathrm{e}^{-C^*}. \tag{13}$$

Additionally, we can also obtain the Chernoff information of a particular test or policy $\delta(\boldsymbol{y}) = \mathbb{1}\big[z(\boldsymbol{y}) \ge \eta\big]$, with $z(\boldsymbol{y})$ the test statistic, and η a given threshold. To study test performance, we use the test Chernoff information $C(\delta)$, but as it is usually hard to derive, we also use a convenient lower bound given by the test Battarcharyya coefficient $B(\delta)$. Letting $P_j(\delta) = P(\delta = j|H_j)$, we have that

$$C(\delta) = -\min_{a \in [0,1]} \log \sum_{\delta=0}^{1} P_0^a(\delta) P_1^{1-a}(\delta) \le C^*, \tag{14}$$

$$B(\delta) = -\log \sum_{\delta=0}^{1} P_0^{1/2}(\delta) P_1^{1/2}(\delta) \le C(\delta). \tag{15}$$

Finally, we consider that local observations $\boldsymbol{v}_i$ conditioned on hypothesis H_j, $j \in \{0,1\}$, are normally distributed with means m_j and sensor variance σ^2, with $m_1 > m_0$, and thus the sensor LLRs correspond to

$$u_i = \frac{1}{2\sigma^2}\Big(\|\boldsymbol{v}_i - m_0\mathbf{1}\|^2 - \|\boldsymbol{v}_i - m_1\mathbf{1}\|^2\Big) \tag{16}$$

and by letting $d = (m_1 - m_0)\sqrt{N}/\sigma > 0$ the sensor hypothesis distance, we have that

$$u_i | H_j \sim \mathcal{N}\Big((-1)^{1-j}\frac{d^2}{2}, d^2\Big). \tag{17}$$

B. Exact channel phase precompensation

In this analysis, we assume static channels through K samples of the received signal for simplicity, although the generalization to different channels over transmission bursts is straightforward. Thus, in the following, channel magnitudes are unknown constants for detection purposes, even when they were realizations following a distribution $f_{\mathrm{a}}(a)$. Hence, the NCWS $\tilde{\lambda}_{\mathrm{P}}$ for S participating sensors has the following conditional distributions:

$$\tilde{\lambda}_{\mathrm{P}} | H_j \sim \mathcal{N}\bigg((-1)^{1-j}\frac{Ad^2}{2}, \frac{Pd^2 + 0.5\omega^2}{K}\bigg), \tag{18}$$

where $A = \sum_{i=1}^{S} a_i > 0$ and $P = \sum_{i=1}^{S} a_i^2 > 0$. Note that the server does not need to know or estimate each individual channel gain (which is impossible with over-the-air aggregation and no signal separation), but only the sum of magnitudes A and sum of powers P. The NCWS hypothesis distance for AirCompFDM-P is

$$D_{\mathrm{P}} = \frac{\sqrt{K}|\mu_1 - \mu_0|}{\nu} = \frac{Ad^2\sqrt{K}}{\sqrt{Pd^2 + 0.5\omega^2}}, \tag{19}$$

and the NCWS probability of error is

$$P_{\mathrm{E}}(D_{\mathrm{P}}) = \pi_0 Q\bigg(\frac{D_{\mathrm{P}}}{2} + \frac{\gamma}{D_{\mathrm{P}}}\bigg) + \pi_1 Q\bigg(\frac{D_{\mathrm{P}}}{2} - \frac{\gamma}{D_{\mathrm{P}}}\bigg), \tag{20}$$

with Q the Gaussian Q Function [11] and $\gamma = \ln(\pi_0/\pi_1)$. The Chernoff information for this scenario is given by

$$C^*_{\mathrm{P}} = \frac{D^2_{\mathrm{P}}}{8} = \frac{A^2 d^4 K}{8(Pd^2 + 0.5\omega^2)}. \tag{21}$$

Regrettably, (21) does not really provide insights into detection performance in terms of the number of sensors S, which is implicitly included in A and P. Moreover, A and P are both dependent on channel realizations. To overcome this issue, we perform statistical analysis for D_{P} and D^2_{P} for large enough S. We can rewrite

$$D_{\mathrm{P}} = \frac{A/\sqrt{S}}{\sqrt{P/S + 0.5\omega^2/(Sd^2)}} d\sqrt{K}. \tag{22}$$

By the Strong Law of Large Numbers, $P/S \xrightarrow{a.s.} \mathbb{E}\{a^2\}$ as $S \to \infty$. Since ω^2 is constant, and invoking the Continuous Mapping Theorem (CMT) with the continuous function $h(x) = \sqrt{x}$, we have that

$$\sqrt{\frac{P}{S} + 0.5\frac{\omega^2}{Sd^2}} \xrightarrow{a.s.} \sqrt{\mathbb{E}\{a^2\}} \text{ when } S \to \infty. \tag{23}$$

Using the Central Limit Theorem on the numerator of D_{P}, we obtain

$$\frac{A}{\sqrt{S}} d\sqrt{K} \xrightarrow{d} \Big(W + \sqrt{S}\mathbb{E}\{a\}\Big) d\sqrt{K} \text{ when } S \to \infty, \tag{24}$$

where $W \sim \mathcal{N}\big(0, \mathrm{Var}\{a\}\big)$. Thus, invoking Slutzky's theorem and the CMT with the continuous function $h(x) = x^2$,

$$D^2_{\mathrm{P}} \xrightarrow{d} \frac{\big(W + \sqrt{S}\mathbb{E}\{a\}\big)^2}{\mathbb{E}\{a^2\}} d^2 K \text{ when } S \to \infty. \tag{25}$$

Taking expectation, and observing that $\mathbb{E}\{W^2\} = \mathrm{Var}\{a\}$,

$$\mathbb{E}\{D^2_{\mathrm{P}}\} = \frac{\mathrm{Var}\{a\} + S\mathbb{E}^2\{a\}}{\mathbb{E}\{a^2\}} d^2 K. \tag{26}$$

For large S, channel noise is nullified, equivalent to a high SNR regime. However, as we are interested in the effect of channel SNR in our analysis, we can assume that the convergence theorems invoked in the previous derivation hold with modest values of S, and we do not dismiss the channel noise term in (23). Therefore, we obtain the approximation

$$\mathbb{E}\big\{D^2_{\mathrm{P}}\big\} \approx \frac{\mathrm{Var}\{a\} + S\mathbb{E}^2\{a\}}{\mathbb{E}\{a^2\} + 0.5\omega^2/(d^2 S)} d^2 K. \tag{27}$$

Importantly, (27) shows that the mean squared hypothesis distance (and hence, the mean Chernoff information) has approximately an affine relationship with S, proportional to channel statistics, and the effect of channel noise diminishes with S at rate $1/S$.

The test $\delta_{\mathrm{P}}(\boldsymbol{y}) = \mathbb{1}[\tilde{\lambda}_{\mathrm{P}}(\boldsymbol{y}) \geq A\gamma]$ is the optimal Bayesian test, and it reduces to $\delta_{\mathrm{P}}(y) = \mathbb{1}[\mathrm{Re}\{y\} \geq 0]$ for equally likely hypothesis and $K = 1$. As the conditional distributions (18) have antipodal means, the Battarcharyya coefficient of the test δ_{P} when S sensors collaborate is [1]

$$B(\delta_{\mathrm{P}}) = -\frac{1}{2}\ln\Big[4Q\big(-\sqrt{2}D_{\mathrm{P}}\big)Q\big(\sqrt{2}D_{\mathrm{P}}\big)\Big] \tag{28}$$

and given that we test means of Gaussian distributions, we have that $C(\delta_{\mathrm{P}}) \geq B(\delta_{\mathrm{P}}) \geq C^*_{\mathrm{P}}/2$ [1, Theorem 2]. Hence, test performance should improve with the number of sensors S approximately according to (27), and in consequence, the probability of error of δ_{P} exhibits exponential decay with similar behavior for increasing S.

These theoretical results allow us to formulate optimal detection strategies with AirCompFDM. The network should *always use the largest number of sensors available* to collaborate in detection. Furthermore, the server can *collect less samples* of the incoming signals and still enjoy satisfactory detection performance thanks to source aggregation. Finally, we can design detection strategies based on channel statistics only. Thus, this analysis also applies to scenarios with channels varying throughout transmission bursts, as long as the sampling process ensures that channel realizations $g_i(k)$ are uncorrelated in different bursts.

We can also derive some additional insights. Consider equally likely hypotheses, with test $\delta_{\mathrm{P}}(\boldsymbol{y}) = \mathbb{1}[\tilde{\lambda}_{\mathrm{P}}(\boldsymbol{y}) \geq 0]$. In other words, as all channel magnitudes a_i are positive, we only need to know the sign of the NCWS: different magnitude values will not change the sign of each individual LLR transmitted by the sensors. This suggests that: (1) phase compensation is more important than magnitude compensation; and (2) imperfect phase compensation of AirCompFDM-M should still enjoy significant benefits of aggregation.

C. No phase precompensation

Now we analyze AirCompFDM-U, where we again assume static channel gains over all samples $\boldsymbol{y}$ for simplicity (recall that the extension is straightforward), and drawn from a distribution $f_{\mathrm{g}}(g)$. The conditional distributions of the NCWS with S participating sensors are

$$\tilde{\lambda}_{\mathrm{U}}|H_j \sim \mathcal{N}\bigg((-1)^{1-j}\frac{|G|d^2}{2}, \frac{Pd^2 + 0.5\omega^2}{K}\bigg), \tag{29}$$

where the server only knows or estimates the aggregated sum of gains G and sum of powers P. The NCWS hypothesis distance for AirCompFDM-U is then

$$D_{\mathrm{U}} = \frac{|\mu_1 - \mu_0|}{\nu} = \frac{|G|d^2\sqrt{K}}{\sqrt{Pd^2 + 0.5\omega^2}}, \tag{30}$$

with probability of error $P_E(D_{\mathrm{U}})$ as in (20), and resulting Chernof information

$$C^*_{\mathrm{U}} = \frac{D^2_{\mathrm{U}}}{8} = \frac{|G|^2 d^4 K}{8(Pd^2 + 0.5\omega^2)}. \tag{31}$$

Again, we resort to approximations to obtain an explicit relationship of the Chernoff distance with the number of sensors S. Noting that $|G|^2 = \mathrm{Re}^2\{G\} + \mathrm{Im}^2\{G\}$, we repeat the derivation for large S f the previous section to obtain

$$\mathbb{E}\{D^2_{\mathrm{U}}\} = \frac{\mathrm{Var}\{g\} + S\big[\mathbb{E}^2\{\mathrm{Re}\{g\}\} + \mathbb{E}^2\{\mathrm{Im}\{g\}\}\big]}{\mathbb{E}\{|g|^2\}} d^2 K, \tag{32}$$

and for modest values of S, we approximate as

$$\mathbb{E}\{D_{\mathrm{U}}^2\} \approx \frac{\mathrm{Var}\{g\} + S\big[\mathbb{E}^2\{\mathrm{Re}\{g\}\} + \mathbb{E}^2\{\mathrm{Im}\{g\}\}\big]}{\mathbb{E}\{|g|^2\} + 0.5\omega^2/(Sd^2)} d^2 K. \tag{33}$$

Hence, we again see an approximate affine relationship with S and channel statistics for the Chernoff information. However, note that without precompensation, the aggregation of channel gains can constructive or destructive. This is reflected in the numerator of (33) because if channels have zero mean, e.g. Rayleigh channels, $\mathbb{E}^2\{\mathrm{Re}(g)\} = \mathbb{E}^2\{\mathrm{Im}(g)\} = 0$ and we lose significant collaboration gains, because

$$\mathbb{E}\{D_{\mathrm{U}}^2\} \approx \frac{\mathbb{E}\{|g|^2\}}{\mathbb{E}\{|g|^2\} + 0.5\omega^2/(d^2 S)} d^2 K, \tag{34}$$

but we would still expect the impact of channel noise to diminish at rate $1/S$, albeit that improvement is rather minor.

For AirCompFDM-U, $\delta_{\mathrm{U}}(\boldsymbol{y}) = \mathbb{1}\big[\tilde{\lambda}_{\mathrm{U}}(\boldsymbol{y}) \geq |G|\gamma\big]$ is the optimal Bayesian test, and if hypotheses are equally likely, we have $\delta_{\mathrm{U}}(y) = \mathbb{1}\big[\mathrm{Re}\{\overline{G}y\} \geq 0\big]$ for $K = 1$. In similar fashion as the previous section, we obtain

$$B(\delta_{\mathrm{U}}) = -\frac{1}{2}\ln\Big[4Q\big(-\sqrt{2}D_{\mathrm{U}}\big)Q\big(\sqrt{2}D_{\mathrm{U}}\big)\Big], \tag{35}$$

and because $C(\delta_{\mathrm{U}}) \geq B(\delta_{\mathrm{U}}) \geq C_{\mathrm{U}}^*/2$, we expect test performance to behave in similar fashion as D_{U}^2 with respect to S. Moreover, we can also use this approximation based on channel statistics to design detection strategies when obtaining samples from multiple transmission slots.

IV. Numerical Experiments

To showcase the performance gains of AirCompFDM, we set up several different network settings. Unless otherwise stated, we simulate S sensors and one server, all with a single antenna. We assume each sensor obtains $N = 1$ observation, independent among sensors, and contaminated with i.i.d. Gaussian measurement noise, parameterized by the sensor hypothesis distance $d = |m_1 - m_0|/\sigma$. Without loss of generality, we assume antipodal means, i.e. $m_0 = -m_1 < 0$, and equally likely hypotheses. We simulate i.i.d. Rayleigh channels, i.e. $g_i \sim \mathcal{N}(0, 1/2) + i\mathcal{N}(0, 1/2)$. Channel noise is AWGN with intensity ω^2 corresponding to a given average SNR for a single sensor. The server makes a decision using $K = 1$ samples of the received analog signal. For empirical simulations, we perform 10000 Monte Carlo (MC) simulations with independent realizations of all random variables involved.

Furthermore, we compare to a traditional MAC polling system [1], [2], where sensors send binary local decisions $u_i^{\mathrm{poll}} \in \{0, 1\}$ and the server *perfectly* recovers all sent binary data with rate smaller than the channel capacity, regardless of channel conditions or transmission mechanisms. The server implements a MAP detector over all received sensor decisions $\boldsymbol{u}^{\mathrm{poll}} = [u_1^{\mathrm{poll}}, \cdots, u_S^{\mathrm{poll}}]^\intercal \in \{0, 1\}^S$, with probability of error given by [2]

$$P_{\mathrm{E}}^{\mathrm{MAC}} = 1 - \sum_{\boldsymbol{u}^{\mathrm{poll}}} \max_{j \in \{0,1\}} \big\{P(\boldsymbol{u}^{\mathrm{poll}}|H_j)\pi_j\big\}. \tag{36}$$

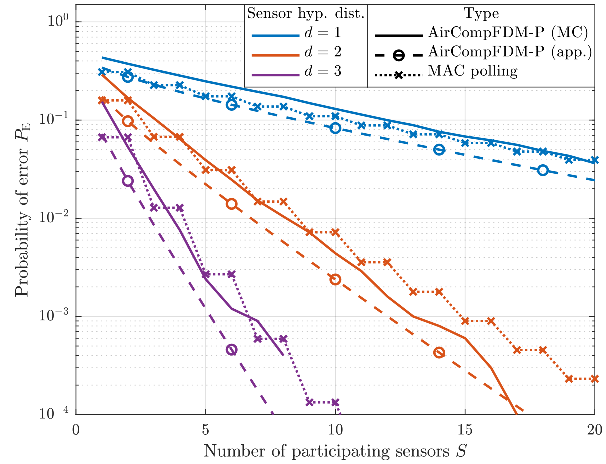

Fig. 1: Probability of error of AirCompFDM-P with respect to participating sensors S, for varying sensor hypothesis distance d, SNR 0dB and $\pi_0 = \pi_1$.

Note that for this MAC polling scheme, channel conditions do not affect the optimality of decision strategies. Therefore, the comparison to AirCompFDM is not totally straightforward, because channel SNR would essentially change the total channel capacity that MAC polling enjoys to aggregate many sensor signals in one time interval. On the other hand, in AirCompFDM we use only a few samples of the received signal, greatly reducing computational load and networking complexity while still enjoying good detection performance.

Fig. 1 shows the probability of error of AirCompFDM-P, with channel SNR of 0dB and $K = 1$ sample of the received signal, for different values of d. First, our approximation $P_{\mathrm{E}}\big(\mathbb{E}\{D_{\mathrm{P}}\}\big)$ predicts lower error than empirical MC simulations, especially for low d, but the gap reduces with increasing S and d, showing that it can help in designing detection strategies. When $S = 1$, MAC polling achieves better detection performance in all cases, due to its perfect recovery of sensor signals. Note that decision-making in MAC polling is equivalent to a majority rule where all binary local decisions are i.i.d. such that $P(u_i^{\mathrm{poll}} = 0|H_0) = P(u_i^{\mathrm{poll}} = 1|H_1)$ -as is the case of Gaussian detection at each sensor-, and thus $P_{\mathrm{E}}^{\mathrm{MAC}}$ only decreases when S is odd. For S even, a tie with $S/2$ counts of each decision is sorted by randomly removing a vote to obtain a majority, with resulting $P_{\mathrm{E}}^{\mathrm{MAC}}$ equal to the one with $S-1$ sensors. Nevertheless, with increasing S, AirCompFDM-P is comparable with MAC polling even using $K = 1$ samples of the received signal. For larger but modest values of d (which can be easily achieved by increasing local observations N), AirCompFDM-P achieves better performance than MAC polling with $S \geq 6$ sensors, with the additional benefit of minimal computational load and bandwidth usage.

Fig. 2 shows the probability of error of a Bayesian detection of AirCompFDM-U, with channel SNR of 0dB and $K = 1$, for different values of sensor hypothesis distance d. As expected, not precompensating phase of zero-mean channels yields worse performance than having channel compensation.

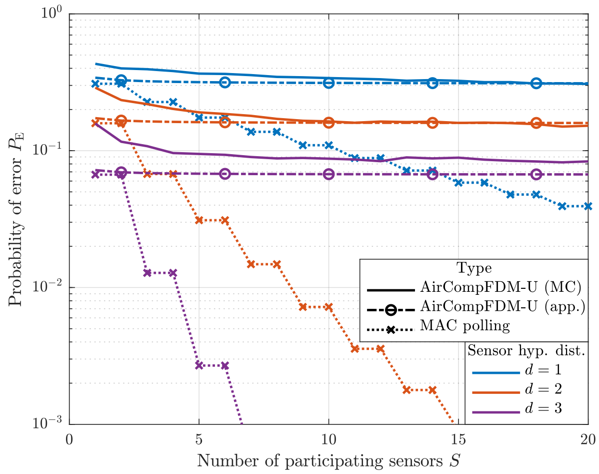

Fig. 2: Probability of error of AirCompFDM-U with respect to participating sensors S, for varying sensor hypothesis distance d, SNR 0dB and $\pi_0 = \pi_1$.

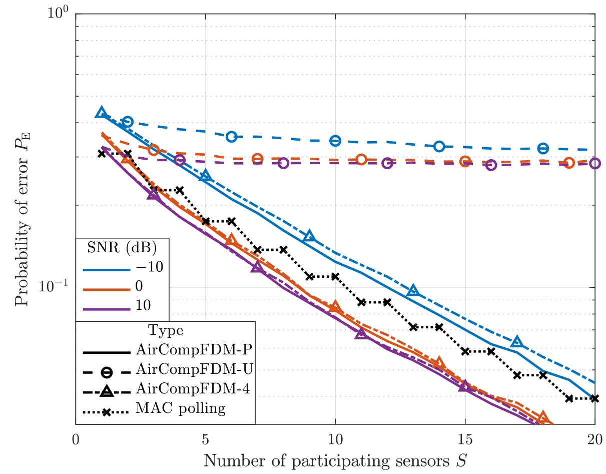

Fig. 3: Probability of error of AirCompFDM with respect to participating sensors S, for different SNR values, with $d = 1$ and $\pi_0 = \pi_1$.

Still, increasing participating sensors does reduce probability of error, albeit with a small effect at this SNR level.

Finally, Fig. 3 shows the probability of error of Monte Carlo AirCompFDM simulations with $d = 1$ and varying SNR values, for different channel compensation mechanisms, and compare to the MAC polling scheme. Evidently, for $S = 1$, AirCompFDM always performs worse than MAC polling, as it is affected by channel gain and noise for any SNR value, although performance improves with increasing SNR. For 0dB of SNR and $S \leq 5$, AirCompFDM-P has similar performance to MAC polling, and for larger number of sensors, AirCompFDM-P shows lower probability of error with minimal network coordination and bandwidth. For high SNR, AirCompFDM-P outperforms MAC polling by the simple collaboration of 2 sensors. Additionally, AirCompFDM-P can compensate for low SNR scenarios by increasing S. For a low SNR of -10 dB and $S = 5$, we outperform MAC polling 2 sensors. On the other hand, we see that AirCompFDM-U indeed accomplishes better channel noise reduction in lower SNR scenarios, being the only improvement obtained by sensor aggregation.

Additionally, in Fig. 3 we also test quantized phase compensation of AirCompFDM-M with $M = 4$ phases as described in Section II-C. AirCompFDM-M enjoys only slightly worse performance as AirCompFDM-P with equally likely hypotheses, and the gap decreases with increasing SNR values. Surprisingly, phase post-compensation at the server was not necessary, and sensor precompensation was enough to provide significant performance gains with S. This exciting result further stresses the computational gains of our proposed AirCompFDM framework in performance-constrained networks.

V. Conclusions

In this work, we develop an over-the-air signal aggregation for collaborative detection in wireless sensor networks. We exploit the natural over-the-air superposition of signals over a shared channel to aggregate distributed sensor data at a server for collaborative binary hypothesis testing, to reduce bandwidth usage while mitigating performance loss. We demonstrate the advantages of this approach under different network access protocols and deployment scenarios, including channel fading, and using a low number of samples and sensor observations, compared to traditional polling of sensors. Our proposed AirCompFDM is a promising framework for collaborate detection in applications such as IoT by balancing the cost of bandwidth usage and detection accuracy.

References

[1] J.-F. Chamberland and V. Veeravalli, "Decentralized detection in sensor networks," *IEEE Trans. Signal Process.*, vol. 51, no. 2, pp. 407–416, 2003.

[2] A. Tarighati and J. Jaldén, "Rate allocation for decentralized detection in wireless sensor networks," in *IEEE 16th Int. Workshop Signal Process. Advances Wireless Commun. (SPAWC)*, 2015, pp. 341–345.

[3] B. Chen and P. Willett, "On the optimality of the likelihood-ratio test for local sensor decision rules in the presence of nonideal channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 2, pp. 693–699, 2005.

[4] 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.

[5] 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.

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

[7] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. Aguera y Arcas, "Communication-Efficient Learning of Deep Networks from Decentralized Data," in *Proc. 20th Int. Conf. Artif. Intell. Statist. (AISTATS)*, vol. 54, Apr. 2017, pp. 1273–1282.

[8] 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.

[9] 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.

[10] C. Feres, B. C. Levy, and Z. Ding, "Over-the-air collaborative learning in joint decision making," in *2022 IEEE Global Commun. Conf. (GLOBECOM)*, 2022, pp. 3581–3586.

[11] B. C. Levy, *Principles of Signal Detection and Parameter Estimation*, 1st ed. Springer, Boston, MA, 2007.