

UC Irvine

UC Irvine Previously Published Works

Title

Estimation in Phase-Shift and Forward Wireless Sensor Networks

Permalink

<https://escholarship.org/uc/item/4cf8387b>

Journal

IEEE Transactions on Signal Processing, 61(15)

ISSN

1053-587X

Authors

Jiang, Feng
Chen, Jie
Swindlehurst, A Lee

Publication Date

2013-08-01

DOI

10.1109/tsp.2013.2260333

Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at <https://creativecommons.org/licenses/by/4.0/>

Peer reviewed

Estimation in Phase-Shift and Forward Wireless Sensor Networks

Feng Jiang, Jie Chen, and A. Lee Swindlehurst

Department of Electrical Engineering and Computer Science

University of California at Irvine,

Irvine, CA 92697, USA

Email: {feng.jiang, jie.chen, swindle}@uci.edu

arXiv:1208.4423v2 [cs.IT] 19 Apr 2013

Abstract

We consider a network of single-antenna sensors that observe an unknown deterministic parameter. Each sensor applies a phase shift to the observation and the sensors simultaneously transmit the result to a multi-antenna fusion center (FC). Based on its knowledge of the wireless channel to the sensors, the FC calculates values for the phase factors that minimize the variance of the parameter estimate, and feeds this information back to the sensors. The use of a phase-shift-only transmission scheme provides a simplified analog implementation at the sensor, and also leads to a simpler algorithm design and performance analysis. We propose two algorithms for this problem, a numerical solution based on a relaxed semidefinite programming problem, and a closed-form solution based on the analytic constant modulus algorithm. Both approaches are shown to provide performance close to the theoretical bound. We derive asymptotic performance analyses for cases involving large numbers of sensors or large numbers of FC antennas, and we also study the impact of phase errors at the sensor transmitters. Finally, we consider the sensor selection problem, in which only a subset of the sensors is chosen to send their observations to the FC.

Index Terms

Wireless sensor networks, analog sensor networks, distributed beamforming, phase-only beamforming, sensor management

This work is supported by the Air Force Office of Scientific Research grant FA9550-10-1-0310, and by the National Science Foundation under grant CCF-0916073.

I. INTRODUCTION

A. Background

Wireless sensor networks (WSNs) have been widely studied for detection and estimation problems. Recently, considerable research has focused on the fusion of *analog* rather than encoded digital data in a distributed sensor network to improve estimation performance. The advantages of analog WSNs have been established in [1]–[3], where it was shown that when using distortion between the source and recovered signal as the performance metric, digital transmission (separate source and channel coding) achieves an exponentially worse performance than analog signaling. A number of studies have focused on algorithm development and analysis for analog WSNs with a single-antenna fusion center (FC). In [4], the sensors amplify and forward their observations of a scalar source to the FC via fading channels, and algorithms are developed to either minimize estimation error subject to transmit power constraints or minimize power subject to estimation error constraints. The scalar source model for this problem was generalized to correlated vector sources in [5]. An opportunistic power allocation approach was proposed in [6], and the scaling law with respect to the number of sensors was shown to be the same as the optimal power allocation proposed in [4]. In [7], the asymptotic variance of the best linear unbiased estimator of an analog WSN is derived, together with an analysis of the effect of different assumptions regarding channel knowledge at the sensors. Scaling laws with respect to the number of sensors have been studied in [8] for a diversity-based method (where only the sensor with the best channel transmits), as well as for the coherent multiple access channel (MAC) and orthogonal channel cases, assuming a Gaussian source. In [9], a power optimization problem was formulated to minimize the outage probability of the MSE for the coherent MAC channel. More complicated settings involving analog WSNs with nonlinear measurement models [10] or relays [11], [12] have also been studied.

The results described above all assume that the FC is equipped with only one antenna. Just as multi-antenna receivers can provide significant capacity or diversity gains in communication systems, the estimation performance of a WSN should also benefit from the use of a multi-antenna FC, though prior work on this scenario is limited. A general scenario is investigated in [13], involving vector observations of a vector-valued random process at the sensors, and linearly precoded vector transmissions from the sensors to a multi-antenna FC. Optimal solutions for the precoders that minimize the mean-squared error (MSE) at the FC are derived for a coherent MAC

under power and bandwidth constraints. In [14], single-antenna sensors amplify and forward their observations to a multi-antenna FC, but it is shown that for Rayleigh fading channels, the improvement in estimate variance is upper bounded by only a factor of two compared to the case of a single-antenna FC. The performance of two heuristic algorithms for choosing the gain and phase of the sensor transmissions is also studied. Subsequent results by the same authors in [15], [16], have demonstrated that when the channel undergoes (zero-mean) Rayleigh fading, there is a limit to the improvement in detection performance for a multi-antenna FC as well, but when the channel is Rician, performance improves monotonically with respect to number of antennas.

The term “amplify and forward” is often used to describe analog sensor networks like those discussed above, since each sensor applies a complex gain to the observation before sending it to the FC. For a coherent MAC, one can think of this as a type of distributed transmit beamforming, although it is distinguished from distributed beamforming applications such as those in communications since in a WSN the observed noise is transmitted together with the signal of interest. Some prior research in radar and communications has focused on scenarios where the beamformer weights implement only a phase shift rather than both a gain and a phase. The advantage of using phase shifting only is that it simplifies the implementation and is easily performed with analog hardware. Phase-shift-only beamformers have most often been applied to receivers that null spatial interference [17], [18], but it has also been considered on the transmit side for MISO wireless communications systems [19], which is similar to the problem considered here. For the distributed WSN estimation problem, phase-only sensor transmissions have been proposed in [20], where the phase is a scaled version of the observation itself. Phase-only transmissions were also considered in the context of distributed detection in [15], leading to a problem similar to one of those we consider here.

In addition to the work outlined above, other WSN research has focused on sensor selection problems, particularly in situations where the sensors have limited battery power. In these problems, only a subset of the sensors are chosen to transmit their observations, while the others remain idle to conserve power. The sensor selection problem has been tackled from various perspectives, with the goal of optimizing the estimation accuracy [11], [21], [22] or some heuristic system utility [23], [24]. In [21], the authors investigated maximum likelihood (ML) estimation of a vector parameter by selecting a fixed-size subset of the sensors. An approximate solution

was found by relaxing the original Boolean optimization to a convex optimization problem. A dynamic model is used to describe the parameter of interest in [22], and sensors use the Kalman filter to estimate the parameter. At each time step, a single sensor is selected and the measurement at the selected sensor is shared with all other sensors. A numerical sensor selection algorithm was proposed to minimize an upper bound on the expected estimation error covariance. Instead of the estimation accuracy, a utility function that takes into account the measurement quality or energy cost can also be used as the metric for sensor selection. In [24], each sensor independently optimizes its own operation status based on a utility function which depends on the sensor's own measurement and the predicted operation status of other sensors. A threshold is then found to enable the sensor to switch its status for either energy efficiency or energy consumption, and a power allocation algorithm was proposed to minimize the MSE at FC.

B. Approach and Contributions

In this paper we consider a distributed WSN with single-antenna sensors that observe an unknown deterministic parameter corrupted by noise. The low-complexity sensors apply a phase shift (rather than both a gain and phase) to their observation and then simultaneously transmit the result to a multi-antenna FC over a coherent MAC. One advantage of a phase-shift-only transmission is that it leads to a simpler analog implementation at the sensor. The FC determines the optimal value of the phase for each sensor in order to minimize the ML estimation error, and then feeds this information back to the sensors so that they can apply the appropriate phase shift. The estimation performance of the phase-optimized sensor network is shown to be considerably improved compared with the non-optimized case, and close to that achieved by sensors that can adjust both the transmit gain and phase. We analyze the asymptotic behavior of the algorithm for a large number of sensors and a large number of antennas at the FC. In addition, we analyze the impact of phase errors at the sensors due, for example, to errors in the feedback channel, a time-varying main channel or phase-shifter drift. We also consider a sensor selection problem similar to that in [21], and analyze its asymptotic behavior as well. Some additional details regarding the contributions of the paper are listed below.

- 1) We present two algorithms for determining the phase factors used at each sensor. In the first, we use the semi-definite relaxation presented in [15], [25] to convert the original problem to a semidefinite programming (SDP) problem that can be efficiently solved by interior-

point methods. For the second algorithm, we apply the analytic constant modulus algorithm (ACMA) [26], which provides a considerably simpler closed-form solution. Despite the reduction in complexity, the performance of ACMA is shown via simulation to be only slightly worse than the SDP solution, and close to the theoretical lower bound on the estimate variance. This is especially encouraging for networks with a large number of sensors N , since the SDP complexity is on the order of $N^{3.5}$, while that for ACMA is only on the order of N^2 .

- 2) We separately derive performance scaling laws with respect to the number of antennas and the number of sensors assuming non-fading channels that take path loss into account. For both cases, we derive conditions that determine whether or not the presence of multiple antennas at the FC provides a significant benefit to the estimation performance. Prior work in [14]–[16] has focused on either AWGN channels with identical channel gains, or on fading channels where the channel gains are identically distributed, corresponding to the case where the distances from the sensors to the FC are roughly the same. References [14]–[16] also assume a special case where the noise at each of the sensors has the same variance, although [16] examines how certain upper bounds on performance change when the sensor noise is arbitrarily correlated.
- 3) Using our model for the non-fading case, we are able to elucidate detailed conditions under which the asymptotic estimation performance will improve with the addition of more antennas M at the FC. While [14], [15] showed that performance always improves with increasing M for AWGN channels with identical gains and identically distributed sensor noise, we derive more detailed conditions that take into account the possibility of non-uniform distances between the sensors and FC and non-uniform noise at the sensors.
- 4) We conduct an analysis of the impact of phase errors at the sensors assuming relatively small phase errors with variance $\sigma_p^2 \ll 1$ (square-radians). In particular, we show that the degradation to the estimate variance is bounded above by a factor of $1 + \sigma_p^2$. We note that the effect of errors in the transmit phase at the sensors has previously been considered for the case of $M = 1$ in [7], although using a different phase error model.
- 5) We consider the sensor selection problem separately for low and high sensor measurement noise. For the low measurement noise scenario, we relax the sensor selection problem to a standard linear programming (LP) problem, and we also propose a reduced complexity

version of the algorithm. For the high measurement noise scenario, we show that the estimation error is lower bounded by the inverse of the measurement noise power, which motivates the use of a simple selection method based on choosing the sensors with the lowest measurement noise.

A subset of the above results was presented in an earlier conference paper [27].

C. Organization

The paper is organized as follows. Section II describes the assumed system model. Section III formulates the phase optimization problem and proposes a numerical solution based on SDP as well as a closed-form solution based on the algebraic constant modulus algorithm. In Section IV, the asymptotic performance of the algorithm is analyzed for a large number of sensors and antennas. The effect of phase errors is analyzed in Section V and the sensor selection problem is investigated in Section VI. Simulation results are then presented in Section VII and our conclusions can be found in Section VIII.

II. SYSTEM MODEL

We assume that N single-antenna sensors in a distributed sensor network independently observe an unknown but deterministic complex-valued parameter θ according to the following model for sensor i :

$$y_i = \theta + v_i ,$$

where v_i is complex-valued Gaussian observation noise with variance $\sigma_{v,i}^2$. The noise is assumed to be independent from sensor to sensor. Each sensor phase shifts its observation and transmits the signal $a_i y_i$ to the FC, where $|a_i| = 1$. Assuming a coherent MAC and an FC with M antennas, the vector signal received at the FC can be expressed as

$$\mathbf{y} = \mathbf{H}\mathbf{a}\theta + \mathbf{H}\mathbf{D}\mathbf{v} + \mathbf{n} , \quad (1)$$

where $\mathbf{H} = [\mathbf{h}_1, \dots, \mathbf{h}_N]$ and $\mathbf{h}_i \in \mathbb{C}^{M \times 1}$ is the channel vector between the i th sensor and the FC, $\mathbf{a} = [a_1, \dots, a_N]^T$ contains the adjustable phase parameters, $\mathbf{D} = \text{diag}\{a_1, \dots, a_N\}$, \mathbf{v} is the sensor measurement noise vector with covariance $\mathbf{V} = \mathbb{E}\{\mathbf{v}\mathbf{v}^H\} = \text{diag}\{\sigma_{v,1}^2, \dots, \sigma_{v,N}^2\}$, and \mathbf{n} is complex Gaussian noise at the FC with covariance $\mathbb{E}\{\mathbf{n}\mathbf{n}^H\} = \sigma_n^2 \mathbf{I}_M$, where \mathbf{I}_M is an $M \times M$ identity matrix. Note that since the sensors can only phase shift their observation prior

to transmission, we ignore the issue of power control and assume that the sensors have sufficient power to forward their observation to the FC.

The combined noise term $\mathbf{H}\mathbf{D}\mathbf{v} + \mathbf{n}$ in (1) is Gaussian with covariance $\mathbf{H}\mathbf{V}\mathbf{H}^H + \sigma_n^2\mathbf{I}$, since $\mathbf{D}\mathbf{V}\mathbf{D}^H = \mathbf{V}$ due to the phase-only assumption. Assuming the FC is aware of the channel matrix \mathbf{H} , the noise covariance \mathbf{V} and σ_n^2 , it can calculate the ML estimate of θ using [28]

$$\hat{\theta}_{ML} = \frac{\mathbf{a}^H \mathbf{H}^H (\mathbf{H}\mathbf{V}\mathbf{H}^H + \sigma_n^2 \mathbf{I}_M)^{-1} \mathbf{y}}{\mathbf{a}^H \mathbf{H}^H (\mathbf{H}\mathbf{V}\mathbf{H}^H + \sigma_n^2 \mathbf{I}_M)^{-1} \mathbf{H}\mathbf{a}} .$$

The estimator $\hat{\theta}_{ML}$ is unbiased with variance

$$\text{Var}(\hat{\theta}_{ML}) = (\mathbf{a}^H \mathbf{H}^H (\mathbf{H}\mathbf{V}\mathbf{H}^H + \sigma_n^2 \mathbf{I}_M)^{-1} \mathbf{H}\mathbf{a})^{-1} . \quad (2)$$

Furthermore, since $\|\mathbf{a}\| = N$ when only phase shifts are used at the sensors, it is easy to see that the variance is lower bounded by

$$\text{Var}(\hat{\theta}_{ML}) \geq \frac{1}{N \lambda_{\max}(\mathbf{H}^H (\mathbf{H}\mathbf{V}\mathbf{H}^H + \sigma_n^2 \mathbf{I}_M)^{-1} \mathbf{H})} , \quad (3)$$

where $\lambda_{\max}(\cdot)$ denotes the largest eigenvalue of its matrix argument. Note that the bound in (3) is in general unachievable, since with probability one the given matrix will not have an eigenvector with unit modulus elements.

III. OPTIMIZING THE SENSOR PHASE

In this section we consider the problem of choosing \mathbf{a} to minimize $\text{Var}(\hat{\theta}_{ML})$ in (2). The unit modulus constraint prevents a trivial solution, but as we note below, a direct solution is not possible even without this constraint since the noise covariance would then depend on \mathbf{a} . The general optimization problem is formulated as

$$\begin{aligned} \min_{\mathbf{a}} \quad & \text{Var}(\hat{\theta}_{ML}) \\ \text{s.t.} \quad & |a_i| = 1, \quad i = 1, \dots, N . \end{aligned} \quad (4)$$

Defining $\mathbf{B} = \mathbf{H}^H (\mathbf{H}\mathbf{V}\mathbf{H}^H + \sigma_n^2 \mathbf{I}_M)^{-1} \mathbf{H}$, the problem can be rewritten as

$$\begin{aligned} \max_{\mathbf{a}} \quad & \mathbf{a}^H \mathbf{B}\mathbf{a} \\ \text{s.t.} \quad & |a_i| = 1, \quad i = 1, \dots, N . \end{aligned} \quad (5)$$

Note that this optimization can only determine \mathbf{a} to within an arbitrary phase shift $e^{j\phi}$, but this scaling has no impact on the estimate of θ . In other words, the vector \mathbf{a} and the vector $\mathbf{a}e^{j\phi}$

for arbitrary ϕ will both yield the same estimate $\hat{\theta}_{ML}$. Since the FC is aware of the vector \mathbf{a} determined by the optimization in (5), any arbitrary phase factor present in the $\mathbf{H}\mathbf{a}\theta$ term of the model in (1) will be canceled when the ML estimate of θ is computed. This is also clear from the variance expression in (2), which is insensitive to any phase shift to \mathbf{a} .

If there are only two sensors in the network, a simple closed-form solution to (5) can be obtained. Defining $\mathbf{B} = \begin{bmatrix} a & be^{j\beta} \\ be^{-j\beta} & c \end{bmatrix}$ with $a, b, c > 0$ and $\mathbf{a} = [e^{j\beta_1}, e^{j\beta_2}]$, then $\mathbf{a}^H \mathbf{B} \mathbf{a}$ is calculated as

$$\begin{aligned} \mathbf{a}^H \mathbf{B} \mathbf{a} &= a + c + 2b \cos(\beta_1 - \beta_2 - \beta) \\ &\leq a + c + 2b, \end{aligned} \quad (6)$$

and the equality in (6) can be achieved for any β_1, β_2 that satisfy $\beta_1 - \beta_2 = \beta$. For the general situation where $N > 2$, a solution to (5) appears to be intractable. Instead, in the discussion that follows we present two suboptimal approaches in order to obtain an approximate solution. The first approach is based on an SDP problem obtained by relaxing a rank constraint in a reformulated version of (5), similar to the approach proposed in [15], [25]. The second converts the problem to one that can be solved via the ACMA of [26]. It is worth emphasizing here that if the transmission gain of the sensors was also adjustable, then the corresponding problem would be

$$\begin{aligned} \max_{\mathbf{a}} \quad & \mathbf{a}^H \mathbf{H}^H (\mathbf{H} \mathbf{D} \mathbf{V} \mathbf{D}^H \mathbf{H}^H + \sigma_n^2 \mathbf{I}_M)^{-1} \mathbf{H} \mathbf{a} \\ \text{s.t.} \quad & \mathbf{a}^H \mathbf{a} \leq N, \end{aligned} \quad (7)$$

which also has no closed-form solution due to the dependence on \mathbf{a} (through the matrix \mathbf{D}) inside the matrix inverse. While in general both our SDP solution and (7) require numerical optimizations, we will see in Sections IV-VI that the theoretical analysis of performance and the solution to the sensor selection problem is considerably simpler with the phase-only constraint. The simulations of Section VII will also demonstrate that there is often little performance loss incurred by using phase-shift-only transmissions.

A. SDP Formulation

To begin, we rewrite (5) as follows:

$$\begin{aligned} \max_{\mathbf{a}} \quad & \text{tr}(\mathbf{B}\mathbf{a}\mathbf{a}^H) \\ \text{s.t.} \quad & |a_i| = 1, \quad i = 1, \dots, N. \end{aligned} \quad (8)$$

Making the association $\mathbf{A} = \mathbf{a}\mathbf{a}^H$, problem (8) is equivalent to:

$$\begin{aligned} \max_{\mathbf{A}} \quad & \text{tr}(\mathbf{B}\mathbf{A}) \\ \text{s.t.} \quad & \mathbf{A}_{i,i} = 1, \quad i = 1, \dots, N \\ & \text{rank}(\mathbf{A}) = 1 \\ & \mathbf{A} \succeq 0, \end{aligned} \quad (9)$$

where $\mathbf{A}_{i,i}$ denotes the i th diagonal element of \mathbf{A} . Following the approach of [15], [25], we then relax the rank-one constraint, so that the problem becomes a standard SDP:

$$\begin{aligned} \max_{\mathbf{A}} \quad & \text{tr}(\mathbf{B}\mathbf{A}) \\ \text{s.t.} \quad & \mathbf{A}_{i,i} = 1, \quad i = 1, \dots, N \\ & \mathbf{A} \succeq 0. \end{aligned} \quad (10)$$

Defining $\mathbf{B}_r = \text{real}\{\mathbf{B}\}$, $\mathbf{B}_i = \text{imag}\{\mathbf{B}\}$, and similarly for \mathbf{A}_r and \mathbf{A}_i , we can convert (10) to the equivalent real form

$$\begin{aligned} \max_{\{\mathbf{A}_r, \mathbf{A}_i\}} \quad & \text{tr}(\mathbf{B}_r\mathbf{A}_r - \mathbf{B}_i\mathbf{A}_i) \\ \text{s.t.} \quad & \mathbf{A}_r \mathbf{A}_i = 1, \quad i = 1, \dots, N \\ & \begin{bmatrix} \mathbf{A}_r & -\mathbf{A}_i \\ \mathbf{A}_i & \mathbf{A}_r \end{bmatrix} \succeq 0. \end{aligned} \quad (11)$$

Problem (11) can be efficiently solved by a standard interior-point method [29].

In general, the solution to (11) will not be rank one, so an additional step is necessary to estimate \mathbf{a} . Let \mathbf{A}_r^* , \mathbf{A}_i^* denote the solution to problem (11), then the solution to problem (10) is given by $\mathbf{A}^* = \mathbf{A}_r^* + j\mathbf{A}_i^*$. If $\text{rank}(\mathbf{A}^*) > 1$, we can use a method similar to Algorithm 2 in [30] to extract a rank-one solution, as follows:

- 1) Decompose¹ $\mathbf{A}^* = \mathbf{C}^H \mathbf{C}$, define $\tilde{\mathbf{B}} = \mathbf{C} \mathbf{B} \mathbf{C}^H$, and find a unitary matrix \mathbf{U} that can diagonalize $\tilde{\mathbf{B}}$.
- 2) Let $\mathbf{r} \in \mathbb{C}^{N \times 1}$ be a random vector whose i th element is set to $e^{j\omega_i}$, where ω_i is uniformly distributed over $[0, 2\pi)$.
- 3) Set $\tilde{\mathbf{a}} = \mathbf{C}^H \mathbf{U} \mathbf{r}$, and the solution is given by $\mathbf{a}^* = [a_1^* \cdots a_N^*]^T$, where $a_i^* = e^{j\angle \tilde{a}_i}$ and $\angle z$ represents the phase of a complex number z .

A detailed discussion of the reasoning behind the above rank-one modification can be found in [30].

B. ACMA Formulation

For this discussion, we will assume that $N > M$, which represents the most common scenario. Thus, the $N \times N$ matrix \mathbf{B} in the quadratic form $\mathbf{a}^H \mathbf{B} \mathbf{a}$ that we are trying to maximize is low rank; in particular, $\text{rank}(\mathbf{B}) \leq M < N$. Clearly, any component of \mathbf{a} orthogonal to the columns or rows of \mathbf{B} will not contribute to our goal of minimizing the estimate variance. In particular, if we define the singular value decomposition (SVD) $\mathbf{B} = \mathbf{U} \mathbf{\Sigma} \mathbf{U}^H$, we ideally seek a vector \mathbf{a} such that

$$\begin{aligned} \mathbf{a} &= \sum_{k=1}^m w_k \mathbf{u}_k = \mathbf{U}_m \mathbf{w} \\ |a_i| &= 1, \end{aligned} \quad (12)$$

where $\mathbf{U}_m = [\mathbf{u}_1 \cdots \mathbf{u}_m]$ contains the first $m \leq \text{rank}(\mathbf{B}) \leq M$ singular vectors of \mathbf{B} and $\mathbf{w} = [w_1 \cdots w_m]^T$. The problem of finding the coefficient vector \mathbf{w} of a linear combination of the columns of a given matrix \mathbf{U}_m that yields a vector with unit modulus elements is precisely the problem solved by the ACMA [26].

Our problem is slightly different from the one considered in [26], since there will in general be no solution to (12) even in the absence of noise. However, in our simulation results we will see that the ACMA solution provides performance close to that obtained by the SDP formulation above. Note also that there is a trade-off in the choice of m , the number of vectors in $\text{span}(\mathbf{B})$ to include in the linear combination of (12). A small value of m allows us to focus on forming \mathbf{a} from vectors that will tend to increase the value of $\mathbf{a}^H \mathbf{B} \mathbf{a}$, while a larger value for m provides

¹Since \mathbf{A}^* is the solution to problem (10), \mathbf{A}^* is positive semidefinite.

more degrees of freedom in finding a vector whose elements satisfy $|a_i| = 1$. Another drawback to choosing a larger value for m is that the ACMA solution can only be found if $N > m^2$. As long as M is not too large, one could in principle try all values of $m = 1, \dots, M$ that satisfy $N > m^2$ and choose the one that yields the smallest estimate variance. We will see later in the simulations that a small value for m already provides good performance, so the choice of m is not a significant issue.

The general ACMA approach can be formulated to find multiple solutions to (12), but in our case we only need a single solution, and thus a simplified version of ACMA can be used, as outlined here for a given m . The ACMA is obtained by defining the rows of \mathbf{U}_m as $\mathbf{U}_m^H = [\tilde{\mathbf{u}}_1 \cdots \tilde{\mathbf{u}}_N]$, and then rewriting the constraint $|a_i| = |\tilde{\mathbf{u}}_i^H \mathbf{w}| = 1$ as

$$(\bar{\tilde{\mathbf{u}}}_i \otimes \tilde{\mathbf{u}}_i)^H (\bar{\mathbf{w}} \otimes \mathbf{w}) = 1 ,$$

where $(\bar{\cdot})$ denotes the complex conjugate and \otimes the Kronecker product. Stacking all N such constraints into a single equation results in

$$\mathbf{P} (\bar{\mathbf{w}} \otimes \mathbf{w}) = 0 , \quad (13)$$

where

$$\mathbf{P} = \begin{bmatrix} (\bar{\tilde{\mathbf{u}}}_1 \otimes \tilde{\mathbf{u}}_1)^H & -1 \\ \vdots & \vdots \\ (\bar{\tilde{\mathbf{u}}}_N \otimes \tilde{\mathbf{u}}_N)^H & -1 \end{bmatrix} . \quad (14)$$

If an exact solution to (13) existed, then a vector in the null space of \mathbf{P} would have the form $\left[(\bar{\mathbf{w}} \otimes \mathbf{w})^T \ 1 \right]^T$, and \mathbf{w} could be found by stripping away the 1 and then unstacking the resulting vector into a rank-one matrix (see [26] for more details). In our problem, an exact solution to (13) does not exist, so we use the following approach to obtain an approximation:

- 1) Let \mathbf{q} represent the right singular vector of \mathbf{P} associated with the smallest singular value, and define the vector $\tilde{\mathbf{q}}$ to contain the first m^2 elements of \mathbf{q} .
- 2) Set \mathbf{w} equal to the singular vector of $\tilde{\mathbf{Q}} + \tilde{\mathbf{Q}}^H$ with largest singular value, where the $m \times m$ matrix

$$\tilde{\mathbf{Q}} = \text{vec}^{-1}(\tilde{\mathbf{q}}) \quad (15)$$

is formed by dividing $\tilde{\mathbf{q}}$ into sub-vectors of length m and stacking them together in a matrix.

- 3) Set $\hat{\mathbf{a}} = \mathbf{U}_m \mathbf{w}$. The vector \mathbf{a} is then found by setting the magnitude of all the elements of $\hat{\mathbf{a}}$ equal to unity. In particular, the i -th element of \mathbf{a} is given by

$$a_i^* = e^{j\angle \hat{a}_i}.$$

C. Comparison of Computational Complexity

As discussed in [25], the computational load of the SDP problem in (10) is of the order $O(N^{3.5})$. The additional steps required to take the SDP result and find a rank-one solution require an $O(N^3)$ eigenvalue decomposition, so the overall complexity is dominated by the SDP. For ACMA, the dominant computational step occurs in finding the m principal eigenvectors of the Hermitian matrix \mathbf{B} , which requires only an order $O(mN^2)$ computation [31]. Finding the least dominant singular vector of \mathbf{P} is an $O(N^2) + O(m^4)$ operation, and the remaining steps have relatively trivial complexity. Since $m \ll N$ in typical scenarios, we see that ACMA enjoys a significantly lower computational load compared to the SDP approach. Despite this, we will see that ACMA has performance that is only slightly inferior to using the SDP solution.

IV. ASYMPTOTIC PERFORMANCE ANALYSIS

In this section, we analyze the asymptotic performance achievable using only phase-shifts for the sensor transmissions. We will separately study cases where the number of sensors is large ($N \rightarrow \infty$) or the number of FC antennas is large ($M \rightarrow \infty$). Our analysis will be based on a non-fading channel model that takes path loss into account, similar to models used in [32], [33]. In particular, for the channel between the FC and sensor i , we assume

$$\mathbf{h}_i = \frac{1}{d_i^\alpha} \tilde{\mathbf{h}}_i,$$

where d_i denotes the distance between the i th sensor and the FC, α is the path loss exponent and $\tilde{\mathbf{h}}_i$ is given by

$$\tilde{\mathbf{h}}_i = [e^{j\gamma_{i,1}} \ e^{j\gamma_{i,2}} \ \dots \ e^{j\gamma_{i,M}}]^T,$$

where $\gamma_{i,j}$ is uniformly distributed over $[0, 2\pi)$.

A. Estimation Performance for Large N

From (3) we know that the lower bound on $\text{Var}(\hat{\theta}_{ML})$ depends on the largest eigenvalue of $\mathbf{H}^H(\mathbf{H}\mathbf{V}\mathbf{H}^H + \sigma_n^2\mathbf{I}_M)^{-1}\mathbf{H}$. We begin by deriving a lower bound for this eigenvalue. The (m, n) th element of $\mathbf{H}\mathbf{V}\mathbf{H}^H$ can be expressed as

$$(\mathbf{H}\mathbf{V}\mathbf{H}^H)_{m,n} = \sum_{i=1}^N \frac{e^{j(\gamma_{i,m} - \gamma_{i,n})} \sigma_{v,i}^2}{d_i^{2\alpha}}.$$

According to the strong law of large numbers, as $N \rightarrow \infty$ we have

$$\begin{aligned} \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \frac{e^{j(\gamma_{i,m} - \gamma_{i,n})} \sigma_{v,i}^2}{d_i^{2\alpha}} &\stackrel{(a)}{=} \mathbb{E} \left\{ \frac{\sigma_{v,i}^2}{d_i^{2\alpha}} \right\} \mathbb{E} \{ e^{j(\gamma_{i,m} - \gamma_{i,n})} \} \\ &\stackrel{(b)}{=} \begin{cases} \mathbb{E} \left\{ \frac{\sigma_{v,i}^2}{d_i^{2\alpha}} \right\} & m = n \\ 0 & m \neq n, \end{cases} \end{aligned} \quad (16)$$

where (a) follows from the assumption that $\gamma_{i,m}$, d_i and $\sigma_{v,i}^2$ are independent and (b) is due to the fact that $\gamma_{i,m}$ and $\gamma_{i,n}$ are independent and uniformly distributed over $[0, 2\pi)$. Thus, for sufficiently large N we have

$$\lim_{N \rightarrow \infty} \mathbf{H}\mathbf{V}\mathbf{H}^H = N \mathbb{E} \left\{ \frac{\sigma_{v,i}^2}{d_i^{2\alpha}} \right\} \mathbf{I}_M. \quad (17)$$

Based on (17), we have

$$\begin{aligned} \lim_{N \rightarrow \infty} \lambda_{\max}(\mathbf{H}^H(\mathbf{H}\mathbf{V}\mathbf{H}^H + \sigma_n^2\mathbf{I}_M)^{-1}\mathbf{H}) &= \frac{1}{N \mathbb{E} \left\{ \frac{\sigma_{v,i}^2}{d_i^{2\alpha}} \right\} + \sigma_n^2} \left[\lim_{N \rightarrow \infty} \lambda_{\max}(\mathbf{H}^H\mathbf{H}) \right] \\ &\stackrel{(c)}{=} \frac{N \mathbb{E} \left\{ \frac{1}{d_i^{2\alpha}} \right\}}{N \mathbb{E} \left\{ \frac{\sigma_{v,i}^2}{d_i^{2\alpha}} \right\} + \sigma_n^2}, \end{aligned} \quad (18)$$

where (c) is due to the fact that $\lambda_{\max}(\mathbf{H}^H\mathbf{H}) = \lambda_{\max}(\mathbf{H}\mathbf{H}^H)$. Substituting (18) into (3), we have the following asymptotic lower bound on the estimate variance:

$$\text{Var}(\hat{\theta}_{ML}) \geq \frac{N \mathbb{E} \left\{ \frac{\sigma_{v,i}^2}{d_i^{2\alpha}} \right\} + \sigma_n^2}{N^2 \mathbb{E} \left\{ \frac{1}{d_i^{2\alpha}} \right\}}. \quad (19)$$

For large enough N , the lower bound can be approximated using sample averages:

$$\text{Var}(\hat{\theta}_{ML}) \geq \frac{\sum_{i=1}^N \frac{\sigma_{v,i}^2}{d_i^{2\alpha}} + \sigma_n^2}{N \sum_{i=1}^N \frac{1}{d_i^{2\alpha}}}. \quad (20)$$

Next, we derive an upper bound on the estimate variance and compare it with the lower bound obtained above. The upper bound is obtained by calculating the variance obtained when only

a single antenna is present at the FC. For the given channel model, the optimal choice for the vector of sensor phases is just the conjugate of the channel phases: $\mathbf{a} = [e^{-j\gamma_{1,1}} \dots e^{-j\gamma_{N,1}}]^T$, which when applied to (2) leads to

$$\text{Var}(\hat{\theta}_{ML}) \leq \frac{\sum_{i=1}^N \frac{\sigma_{v,i}^2}{d_i^{2\alpha}} + \sigma_n^2}{\left(\sum_{i=1}^N \frac{1}{d_i^\alpha}\right)^2}. \quad (21)$$

When $N \rightarrow \infty$, both the upper and lower bounds converge to 0, but the ratio of the lower bound in (20) to the upper bound in (21) converges to

$$\lim_{N \rightarrow \infty} \frac{\left(\sum_{i=1}^N \frac{1}{d_i^\alpha}\right)^2}{N \sum_{i=1}^N \frac{1}{d_i^{2\alpha}}} = \frac{\left(\mathbb{E}\left\{\frac{1}{d_i^\alpha}\right\}\right)^2}{\mathbb{E}\left\{\frac{1}{d_i^{2\alpha}}\right\}} = 1 - \frac{\text{Var}\left\{\frac{1}{d_i^\alpha}\right\}}{\mathbb{E}\left\{\frac{1}{d_i^{2\alpha}}\right\}}. \quad (22)$$

Interestingly, we see that if $\text{Var}\left\{\frac{1}{d_i^\alpha}\right\} \ll \mathbb{E}\left\{\frac{1}{d_i^\alpha}\right\}$, the gap between the upper and lower bound is very small, and the availability of multiple antennas at the FC does not provide much benefit compared with the single antenna system when $N \rightarrow \infty$. On the other hand, if $\text{Var}\left\{\frac{1}{d_i^\alpha}\right\} \rightarrow \mathbb{E}\left\{\frac{1}{d_i^{2\alpha}}\right\}$, the potential exists for multiple antennas to significantly lower the estimate variance.

B. Estimation Performance for Large M

Using the matrix inversion lemma, we have

$$\begin{aligned} \mathbf{H}^H (\mathbf{H}\mathbf{V}\mathbf{H}^H + \sigma_n^2 \mathbf{I}_M)^{-1} \mathbf{H} &= \mathbf{H}^H \left(\frac{1}{\sigma_n^2} \mathbf{I}_M - \frac{1}{\sigma_n^4} \mathbf{H} \left(\mathbf{V}^{-1} + \frac{1}{\sigma_n^2} \mathbf{H}^H \mathbf{H} \right)^{-1} \mathbf{H}^H \right) \mathbf{H} \\ &= \frac{1}{\sigma_n^2} \mathbf{H}^H \mathbf{H} - \frac{1}{\sigma_n^4} \mathbf{H}^H \mathbf{H} \left(\mathbf{V}^{-1} + \frac{1}{\sigma_n^2} \mathbf{H}^H \mathbf{H} \right)^{-1} \mathbf{H}^H \mathbf{H}. \end{aligned} \quad (23)$$

Furthermore, the (m, n) th element of $\mathbf{H}^H \mathbf{H}$ is given by

$$(\mathbf{H}^H \mathbf{H})_{m,n} = \frac{1}{d_m^\alpha d_n^\alpha} \sum_{i=1}^M e^{j(\gamma_{n,i} - \gamma_{m,i})}. \quad (24)$$

Similar to (16), as $M \rightarrow \infty$ we have

$$\lim_{M \rightarrow \infty} \frac{1}{M} \sum_{i=1}^M e^{j(\gamma_{n,i} - \gamma_{m,i})} = \begin{cases} 1 & m = n \\ 0 & m \neq n, \end{cases} \quad (25)$$

and thus

$$\lim_{M \rightarrow \infty} \mathbf{H}^H \mathbf{H} = M \text{diag} \left\{ \frac{1}{d_1^{2\alpha}} \dots \frac{1}{d_N^{2\alpha}} \right\}. \quad (26)$$

Substituting (26) into (23), we have

$$\lim_{M \rightarrow \infty} \mathbf{H}^H (\mathbf{H} \mathbf{V} \mathbf{H}^H + \sigma_n^2 \mathbf{I}_M)^{-1} \mathbf{H} = \text{diag} \left\{ \frac{M}{d_1^{2\alpha} \sigma_n^2 + M \sigma_{v,i}^2}, \dots, \frac{M}{d_N^{2\alpha} \sigma_n^2 + M \sigma_{v,N}^2} \right\},$$

and thus

$$\lim_{M \rightarrow \infty} \text{Var}(\hat{\theta}_{ML}) = \frac{1}{M \sum_{i=1}^N \frac{1}{d_i^{2\alpha} \sigma_n^2 + M \sigma_{v,i}^2}}. \quad (27)$$

Note that this asymptotic expression is independent of the choice of \mathbf{a} . Here, for large M , the benefit of having multiple antennas at the FC hinges on the relative magnitude of $M \sigma_{v,i}^2$ versus $d_i^{2\alpha} \sigma_n^2$. If $M \sigma_{v,i}^2 \ll d_i^{2\alpha} \sigma_n^2$, a reduction in variance by a factor of M is possible. In this case, where the SNR at the FC is low but the signals sent from the sensors are high quality, the coherent gain from the combination of the relatively noise-free sensor signals helps increase the SNR at the FC. On the other hand, when $M \sigma_{v,i}^2 \gg d_i^{2\alpha} \sigma_n^2$, performance is asymptotically independent of M . Here, the coherent gain not only applies to θ but also to the sensor noise, which is stronger in this case.

V. IMPACT OF IMPERFECT PHASE

The previous sections have assumed that the FC can calculate the vector \mathbf{a} and feed the phase information back to the sensors error free. Whether the feedback channel is digital or analog, there are about to be errors either in the received feedback at the sensors or in how the phase shift is actually implemented. Furthermore, the wireless channel may change during the time required for calculation and feedback of \mathbf{a} , so even if the phase shifts are implemented perfectly at the sensors, they may no longer be valid for the current channel. In this section, we evaluate the impact of errors in the sensor phase shifts on the estimation accuracy.

Define the phase shift for the i th sensor as $a_i = e^{j\alpha_i}$, and assume that

$$\alpha_i = \alpha_i^* + \Delta_i,$$

where α_i^* is the optimal phase and Δ_i is a Gaussian perturbation (in radians) with zero mean and variance σ_p^2 . Define $\mathbf{E} = \mathbf{H}^H (\mathbf{H} \mathbf{V} \mathbf{H}^H + \sigma_n^2 \mathbf{I})^{-\frac{1}{2}}$, so that $\text{Var}(\hat{\theta}_{ML})$ can be expressed as

$$\text{Var}(\hat{\theta}_{ML}) = \frac{1}{\|\mathbf{a}^H \mathbf{E}\|^2} = \frac{1}{\sum_{i=1}^M |\mathbf{a}^H \mathbf{e}_i|^2}, \quad (28)$$

where \mathbf{e}_i is the i th column of \mathbf{E} . Let $e_{i,j}e^{j\beta_j}$ be a polar coordinate representation of the j th element of \mathbf{e}_i , so that

$$\begin{aligned} |\mathbf{a}^H \mathbf{e}_i|^2 &= \left| \sum_{j=1}^M e_{i,j} e^{\alpha_j^* + \Delta_j + \beta_j} \right|^2 \\ &= \sum_{j=1}^M e_{i,j}^2 + \sum_{l=1}^M \sum_{\substack{m=1 \\ m \neq l}}^M e_{i,l} e_{i,m} \cos(\alpha_l^* + \Delta_l + \beta_l - \alpha_m^* - \Delta_m - \beta_m). \end{aligned} \quad (29)$$

Define $\delta_{l,m}^i = \Delta_l - \Delta_m$ and $\tau_{l,m}^i = \alpha_l^* + \beta_l - \alpha_m^* - \beta_m$. If we assume $\sigma_p^2 \ll 1$, (29) may be approximated via a 2nd order Taylor series as follows:

$$\begin{aligned} |\mathbf{a}^H \mathbf{e}_i|^2 &\approx \sum_{j=1}^N e_{i,j}^2 + \sum_{l=1}^N \sum_{\substack{m=1 \\ m \neq l}}^N e_{i,l} e_{i,m} \left(\cos(\tau_{l,m}^i) - \sin(\tau_{l,m}^i) \delta_{l,m}^i - \frac{\cos(\tau_{l,m}^i)}{2} (\delta_{l,m}^i)^2 \right) \\ &= \sum_{j=1}^N e_{i,j}^2 + \sum_{l=1}^N \sum_{\substack{m=1 \\ m \neq l}}^N e_{i,l} e_{i,m} \cos(\tau_{l,m}^i) - \sum_{l=1}^N \sum_{\substack{m=1 \\ m \neq l}}^N e_{i,l} e_{i,m} \left(\sin(\tau_{l,m}^i) \delta_{l,m}^i + \frac{\cos(\tau_{l,m}^i)}{2} (\delta_{l,m}^i)^2 \right). \end{aligned} \quad (30)$$

Substituting (30) into (28), we have

$$\text{Var}(\hat{\theta}_{ML}) \approx \frac{1}{\sum_{i=1}^M \left(\sum_{j=1}^N e_{i,j}^2 + \sum_{l=1}^N \sum_{\substack{m=1 \\ m \neq l}}^N e_{i,l} e_{i,m} \cos(\tau_{l,m}^i) - \sum_{l=1}^N \sum_{\substack{m=1 \\ m \neq l}}^N e_{i,l} e_{i,m} \left(\sin(\tau_{l,m}^i) \delta_{l,m}^i + \frac{\cos(\tau_{l,m}^i)}{2} (\delta_{l,m}^i)^2 \right) \right)}.$$

In the previous equation, the effect of the phase error is confined to the second double sum inside the outermost parentheses. If we define $\hat{\theta}_{ML}^P$ to be the estimate obtained with no phase errors, then

$$\text{Var}(\hat{\theta}_{ML}^P) = \frac{1}{\sum_{i=1}^M \left(\sum_{j=1}^N e_{i,j}^2 + \sum_{l=1}^N \sum_{\substack{m=1 \\ m \neq l}}^N e_{i,l} e_{i,m} \cos(\tau_{l,m}^i) \right)}, \quad (31)$$

which is deterministic and does not depend on the random phase errors. We can then obtain the following approximation

$$\text{Var}(\hat{\theta}_{ML}) \stackrel{(f)}{\approx} \text{Var}(\hat{\theta}_{ML}^P) \left(1 + \frac{\sum_{i=1}^M \left(\sum_{l=1}^N \sum_{\substack{m=1 \\ m \neq l}}^N e_{i,l} e_{i,m} \left(\sin(\tau_{l,m}^i) \delta_{l,m}^i + \frac{\cos(\tau_{l,m}^i)}{2} (\delta_{l,m}^i)^2 \right) \right)}{\sum_{i=1}^M \left(\sum_{j=1}^N e_{i,j}^2 + \sum_{l=1}^N \sum_{\substack{m=1 \\ m \neq l}}^N e_{i,l} e_{i,m} \cos(\tau_{l,m}^i) \right)} \right),$$

where (f) is due to the first order Taylor approximation $(1 - \frac{x}{y})^{-1} \approx 1 + \frac{x}{y}$ for $x \ll y$. We use the ratio of $\text{Var}(\hat{\theta}_{ML})$ to $\text{Var}(\hat{\theta}_{ML}^P)$ to measure the effect of the phase error, which yields

$$\frac{\text{Var}(\hat{\theta}_{ML})}{\text{Var}(\hat{\theta}_{ML}^P)} \approx \left(1 + \frac{\sum_{i=1}^M \left(\sum_{l=1}^N \sum_{\substack{m=1 \\ m \neq l}}^N e_{i,l} e_{i,m} \left(\sin(\tau_{l,m}^i) \delta_{l,m}^i + \frac{\cos(\tau_{l,m}^i)}{2} (\delta_{l,m}^i)^2 \right) \right)}{\sum_{i=1}^M \left(\sum_{j=1}^N e_{i,j}^2 + \sum_{l=1}^N \sum_{\substack{m=1 \\ m \neq l}}^N e_{i,l} e_{i,m} \cos(\tau_{l,m}^i) \right)} \right).$$

Note that the only term in the above expression that is random is the numerator on the right-hand side.

Taking the expectation of the ratio with respect to the phase perturbations Δ_i , we have

$$\begin{aligned} \mathbb{E} \left\{ \frac{\text{Var}(\hat{\theta}_{ML})}{\text{Var}(\hat{\theta}_{ML}^P)} \right\} &= \left(1 + \frac{\sum_{i=1}^M \left(\sum_{l=1}^N \sum_{\substack{m=1 \\ m \neq l}}^N e_{i,l} e_{i,m} \left(\sin(\tau_{l,m}^i) \mathbb{E} \{ \delta_{l,m}^i \} + \frac{\cos(\tau_{l,m}^i)}{2} \mathbb{E} \{ (\delta_{l,m}^i)^2 \} \right) \right)}{\sum_{i=1}^M \left(\sum_{j=1}^N e_{i,j}^2 + \sum_{l=1}^N \sum_{\substack{m=1 \\ m \neq l}}^N e_{i,l} e_{i,m} \cos(\tau_{l,m}^i) \right)} \right) \\ &\stackrel{(h)}{=} \left(1 + \frac{\sum_{i=1}^M \sum_{l=1}^N \sum_{\substack{m=1 \\ m \neq l}}^N e_{i,l} e_{i,m} \cos(\tau_{l,m}^i) \sigma_p^2}{\sum_{i=1}^M \left(\sum_{j=1}^N e_{i,j}^2 + \sum_{l=1}^N \sum_{\substack{m=1 \\ m \neq l}}^N e_{i,l} e_{i,m} \cos(\tau_{l,m}^i) \right)} \right), \end{aligned} \quad (32)$$

where in (h) we exploit the fact that $\mathbb{E} \{ \delta_{l,m}^i \} = 0$ and $\mathbb{E} \{ (\delta_{l,m}^i)^2 \} = 2\sigma_p^2$. Since

$$\sum_{l=1}^N \sum_{\substack{m=1 \\ m \neq l}}^N e_{i,l} e_{i,m} \cos(\tau_{l,m}^i) \leq (N-1) \sum_{l=1}^N e_{i,l}^2,$$

the ratio in (32) is approximately upper bounded by

$$\mathbb{E} \left\{ \frac{\text{Var}(\hat{\theta}_{ML})}{\text{Var}(\hat{\theta}_{ML}^P)} \right\} \leq 1 + \left(1 - \frac{1}{N} \right) \sigma_p^2. \quad (33)$$

We see from (33) that the impact of the phase errors increases with N , but in all cases the degradation in the estimate variance is approximately bounded above by a factor of $1 + \sigma_p^2$.

VI. SENSOR SELECTION

As mentioned earlier, in situations where it is desired to use only a subset of the sensors to estimate the parameter (*e.g.*, in order to conserve power at the sensors), the FC needs a method to perform the sensor selection. Assuming only $K < N$ of the sensors are to be selected for transmission to the FC, an optimal solution to the problem would require solving the following maximization:

$$\begin{aligned} \max_{\mathbf{a}, \mathbf{x}} \quad & \mathbf{x}^T \mathbf{D}^H \mathbf{H}^H (\mathbf{H} \mathbf{V} \mathbf{X} \mathbf{H}^H + \sigma_n^2 \mathbf{I}_M)^{-1} \mathbf{H} \mathbf{D} \mathbf{x} \\ \text{s.t.} \quad & \sum_{i=1}^N x_i = K \\ & x_i \in \{0, 1\} \\ & |a_i| = 1, \end{aligned} \quad (34)$$

where $\mathbf{D} = \text{diag} \{a_1, \dots, a_N\}$, $\mathbf{x} = [x_1, \dots, x_N]^T$ is the selection vector and $\mathbf{X} = \text{diag} \{x_1, \dots, x_N\}$. Even if one chooses one of the suboptimal approaches described in Section III for estimating \mathbf{a} ,

solving for \mathbf{x} in (34) requires an exhaustive search over all possible K -sensor combinations and is in general NP-hard. Instead, in this section we derive conditions under which much simpler selection strategies can be applied. We consider the following two cases: (1) low sensor noise relative to the noise at the FC, $\sigma_{v,i}^2 \ll \sigma_n^2$, and (2) relatively high sensor noise $\sigma_{v,i}^2 \gg \sigma_n^2$. For (1), we derive a LP solution as well as a simpler greedy algorithm, and for (2) we show that the problem reduces to choosing the sensors with the lowest measurement noise.

A. Algorithms for High FC Noise

Let \mathbf{a} be the phase vector obtained using one of the algorithms in Section III assuming all N sensors are active. When $\sigma_{v,i}^2 \ll \sigma_n^2$, we ignore the term $\mathbf{H}\mathbf{V}\mathbf{X}\mathbf{H}^H$ in (34), and the problem simplifies to

$$\begin{aligned} \max_{\mathbf{x}} \quad & \mathbf{x}^T \mathbf{D}^H \mathbf{H}^H \mathbf{H} \mathbf{D} \mathbf{x} \\ \text{s.t.} \quad & \sum_{i=1}^N x_i = K \\ & x_i = \{0, 1\} . \end{aligned} \tag{35}$$

Define $\mathbf{F} = \mathbf{D}^H \mathbf{H}^H \mathbf{H} \mathbf{D}$ so that (35) can be rewritten as

$$\begin{aligned} \max_{\mathbf{x}} \quad & \mathbf{x}^T \mathbf{Re}\{\mathbf{F}\} \mathbf{x} \\ \text{s.t.} \quad & \sum_{i=1}^N x_i = K \\ & x_i = \{0, 1\} . \end{aligned} \tag{36}$$

Since $x_i^2 = x_i$, (36) is equivalent to

$$\begin{aligned} \max_{x_i} \quad & \sum_{i=1}^N \mathbf{F}_{i,i} x_i + 2 \sum_{i=1}^{N-1} \sum_{j=i+1}^N \mathbf{Re}\{\mathbf{F}_{i,j}\} x_i x_j \\ \text{s.t.} \quad & \sum_{i=1}^N x_i = K \\ & x_i = \{0, 1\} , \end{aligned} \tag{37}$$

where $\mathbf{F}_{i,j}$ denotes the (i,j) th element of matrix \mathbf{F} . By linearizing the term $x_i x_j$ [34], (37) is equivalent to

$$\max_{x_i, y_{ij}} \sum_{i=1}^N \mathbf{F}_{i,i} x_i + 2 \sum_{i=1}^{N-1} \sum_{j=i+1}^N \operatorname{Re}\{\mathbf{F}_{i,j}\} y_{ij} \quad (38a)$$

$$s.t. \sum_{i=1}^N x_i = K \quad (38b)$$

$$1 - x_i - x_j + y_{ij} \geq 0 \quad (38c)$$

$$x_i - y_{ij} \geq 0 \quad (38d)$$

$$x_j - y_{ij} \geq 0 \quad (38e)$$

$$y_{ij} \geq 0 \quad (38f)$$

$$x_i = \{0, 1\}, \quad (38g)$$

where the constraints (38c)-(38g) lead to $y_{ij} = x_i x_j$.

Note that all of the constraints in (38) are linear, except for (38g). If we relax the constraint in (38g), the condition $0 \leq x_i \leq 1$ is implicitly included in (39c)-(39f), and we are left with a LP problem in standard form [34]:

$$\max_{x_i, y_{ij}} \sum_{i=1}^N \mathbf{F}_{i,i} x_i + 2 \sum_{i=1}^{N-1} \sum_{j=i+1}^N \operatorname{Re}\{\mathbf{F}_{i,j}\} y_{ij} \quad (39a)$$

$$s.t. \sum_{i=1}^N x_i = K \quad (39b)$$

$$1 - x_i - x_j + y_{ij} \geq 0 \quad (39c)$$

$$x_i - y_{ij} \geq 0 \quad (39d)$$

$$x_j - y_{ij} \geq 0 \quad (39e)$$

$$y_{ij} \geq 0. \quad (39f)$$

To find the $x_i = \{0, 1\}$ solution needed for sensor selection, one can take the result of (39) and simply set the K largest elements to one and the rest to zero. If desired, once the K sensors have been selected, the phase vector \mathbf{a} for these K sensors can be recomputed based on a reduced dimension version of the algorithms in Section III.

The above LP problem has $\frac{N(N-1)}{2} + N$ variables and $2N(N-1) + 1$ constraints, and thus will require on the order of $\left(\frac{N(N-1)}{2} + N\right)^2 (2N(N-1) + 1)$ arithmetic operations [29]. A simpler greedy algorithm is presented below that only requires $O(KN)$ operations, and that achieves performance close to the LP approach. The greedy algorithm is based on the following observation:

$$\begin{aligned} \mathbf{x}^T \mathbf{D}^H \mathbf{H}^H \mathbf{H} \mathbf{D} \mathbf{x} &= \sum_{i=1}^K \sum_{j=1}^K \bar{a}_i a_j \mathbf{h}_i^H \mathbf{h}_j \\ &= \sum_{i=1}^{K-1} \sum_{j=1}^{K-1} \bar{a}_i a_j \mathbf{h}_i^H \mathbf{h}_j + \|\mathbf{h}_K\|^2 + 2\text{Re} \left\{ \sum_{j=1}^{K-1} \bar{a}_K a_j \mathbf{h}_K^H \mathbf{h}_j \right\}. \end{aligned}$$

The idea behind the greedy algorithm is to add sensors one at a time based on those for which the last two terms in the above sum are the largest. The steps of the algorithm are detailed below.

Greedy Sensor Selection Algorithm

- 1) Select the first sensor as the one with the strongest channel: $i = \arg \max_k \|\mathbf{h}_k\|^2$, and initialize the active sensor set as $\mathcal{S} = \{i\}$.
- 2) While $|\mathcal{S}| \leq K$, perform the following:
 - a) Solve

$$i = \arg \max_{k \notin \mathcal{S}} \|\mathbf{h}_k\|^2 + 2\text{Re} \left\{ \sum_{j \in \mathcal{S}} \bar{a}_k a_j \mathbf{h}_k^H \mathbf{h}_j \right\}.$$

- b) Update $\mathcal{S} = \mathcal{S} \cup i$.

As with the LP algorithm, once the K sensors are selected, an updated solution for the associated K elements of \mathbf{a} can be obtained.

B. Algorithm for High Sensor Noise

When $\sigma_{v,i}^2 \gg \sigma_n^2$ and assuming that $N > M$ (the case of interest when sensor selection is necessary), the original criterion can be simplified to

$$\begin{aligned} \mathbf{a}^H \mathbf{H}^H (\mathbf{H} \mathbf{V} \mathbf{H}^H)^{-1} \mathbf{H} \mathbf{a} &= \mathbf{a}^H \mathbf{V}^{-\frac{1}{2}} \mathbf{V}^{\frac{1}{2}} \mathbf{H}^H (\mathbf{H} \mathbf{V} \mathbf{H}^H)^{-1} \mathbf{H} \mathbf{V}^{\frac{1}{2}} \mathbf{V}^{-\frac{1}{2}} \mathbf{a} \\ &= \mathbf{a}^H \mathbf{V}^{-\frac{1}{2}} \mathbf{P}_{VH} \mathbf{V}^{-\frac{1}{2}} \mathbf{a}, \end{aligned}$$

where $\mathbf{P}_{VH} = \mathbf{V}^{\frac{1}{2}} \mathbf{H}^H (\mathbf{H} \mathbf{V} \mathbf{H}^H)^{-1} \mathbf{H} \mathbf{V}^{\frac{1}{2}}$ is a rank M projection matrix. Ideally, to maximize the criterion function, one should attempt to find a vector of the form $\mathbf{V}^{-\frac{1}{2}} \mathbf{a}$ that lies in the

subspace defined by \mathbf{P}_{VH} . Assuming the vector \mathbf{a} can approximately achieve this goal, the lower bound on variance is approximately achieved and we have

$$\frac{1}{\mathbf{a}^H \mathbf{V}^{-\frac{1}{2}} \mathbf{P}_{VH} \mathbf{V}^{-\frac{1}{2}} \mathbf{a}} \approx \frac{1}{\mathbf{a}^H \mathbf{V}^{-1} \mathbf{a}} = \frac{1}{\sum_{i=1}^N \frac{1}{\sigma_{v,i}^2}}. \quad (40)$$

With respect to the sensor selection problem, this suggests that when $\sigma_{v,i}^2 \gg \sigma_n^2$, the K sensors with the smallest values of $\sigma_{v,i}^2$ should be chosen.

VII. SIMULATION RESULTS

Here we present the results of several simulation examples to illustrate the performance of the proposed algorithms. In all cases, the path loss exponent α was set to 1, and each result is obtained by averaging over 300 channel realizations. The sensors are assumed to lie in a plane at random angles with respect to the FC, uniformly distributed over $[0, 2\pi)$. The distances of the sensors to the FC will be specified separately below. To evaluate the performance without feedback, \mathbf{a} is set to a vector of all ones. In some of the simulations, we will compare the performance of the proposed algorithms with that obtained by (7), where both the sensor gain and phase can be adjusted. In these simulations, we use the active-set method to optimize (7), and we use several different initializations in order to have a better chance of obtaining the global optimum. When the ACMA algorithm is implemented, the subspace dimension was set at $m = 2$.

In the first two examples, we study the estimation performance for $M = 4$ FC antennas with increasing N for a case where the sensor measurement noise $\sigma_{v,i}^2$ is uniformly distributed over $[0.01, 0.1]$ and the FC noise σ_n^2 is set to 0.1. Fig. 1 shows the results assuming that the sensor distances d_i are uniformly distributed in the interval $[3, 20]$, while in Fig. 2 $d_i = 11.5$ for all sensors. In both cases, even though the lower bound of (3) is not achievable, we see that the performance of the proposed SDP and ACMA methods is nonetheless reasonably close to the bound, and not significantly worse than the performance obtained by optimizing both the phase and gain. As N gets larger in Fig. 1, the estimation error for all of the methods (except the no-feedback case) falls within the asymptotic lower and upper bounds of (20) and (21). When $N = 50$, the ratio $\text{Var} \left\{ \frac{1}{d_i^\alpha} \right\} / \mathbb{E} \left\{ \frac{1}{d_i^{2\alpha}} \right\}$ is 0.304 for Fig. 1, and the ratio between the lower and upper bound is 0.702, which is in excellent agreement with the value of $1 - 0.304$ predicted by Eq. (22). Since the upper bound in (21) corresponds to the case of $M = 1$, one may suppose that

the gap in Fig. 1 between the bounds of (20) and (21) indicates that the presence of multiple antennas at the FC could provide a benefit for large N . However, the performance of SDP and ACMA is approaching the upper bound more tightly, indicating that there is no benefit from having multiple antennas in this case. In Fig. 2 where the d_i are all equal, the asymptotic bounds in (20) and (21) are identical, and asymptotically we expect no benefit from multiple antennas at the FC. We see again that for large N the performance of the SDP and ACMA methods is essentially at the predicted bound. When the d_i are equal and $\frac{\sigma_{v,i}^2}{d_i^\alpha} \ll \sigma_n^2$, the matrix $\mathbf{H}^H(\mathbf{H}\mathbf{V}\mathbf{H}^H + \sigma_n^2\mathbf{I}_M)^{-1}\mathbf{H}$ asymptotically approaches a scaled identity matrix, so in this case the performance of the proposed phase-shift only algorithms even approaches the lower bound of Eq. (3).

Fig. 3 illustrates the performance for $N = 4$ with an increasing number of FC antennas M when $\sigma_{v,i}^2$ is uniformly distributed over $[0.001, 0.01]$ and $\sigma_n^2 = 0.1$. In this example, for most of the sensors we have $M\sigma_{v,i}^2 \ll d_i^{2\alpha}\sigma_n^2$, so in this case we see an improvement as the number of FC antennas increases. However, the benefit of optimizing the transmit phase (and gain for that matter) is reduced as M increases.

In Fig. 4, we investigate the effect of phase errors for two cases, $\sigma_p^2 = 0.1$ and $\sigma_p^2 = 0.2$ assuming the same noise parameter settings as in the first two examples. For each channel realization, results for 3000 different phase error realizations were obtained and averaged to obtain the given plot. The ratio of the variance obtained by the SDP algorithm with and without phase errors is plotted for $M = 2, 4, 6$ for both values of σ_p^2 , and the approximate bound of (33) is also shown. The results show that the performance degradation increases with N , and that (33) provides a reasonable indication of performance for large N . Fig. 4 also shows that increasing the number of FC antennas improves the robustness of the algorithm to imprecise sensor phase.

In Fig. 5, we compare the performance of the three different sensor selection algorithms discussed in the paper (LP, greedy and min-sensor-noise) as a function of σ_n^2 assuming $M = 4$ antennas, $N = 35$ sensors and the sensor noise is uniformly distributed over $[0.001, 0.01]$. The sensor distances d_i are uniformly distributed in the interval $[3, 20]$. Three sets of curves are plotted, one for $K = 5$ selected sensors, one for $K = 10$, and one corresponding to when all the sensor nodes are used (the solid curve, obtained using the SDP algorithm). After the sensor selection, the proposed SDP is used to re-optimize the selected sensor nodes' phase parameters. For small σ_n^2 such that $\sigma_{v,i}^2 \gg \sigma_n^2$, we see as predicted that the best performance is obtained by

simply selecting the K sensors with the smaller measurement noise. On the other hand, again in agreement with our analysis, the LP and greedy algorithms achieve the lowest estimation error for larger values of σ_n^2 . Interestingly, the greedy algorithm provides performance essentially identical to the LP approach at a significantly reduced computational cost.

VIII. CONCLUSIONS

In this paper, we investigated a distributed network of single antenna sensors employing a phase-shift and forward strategy for sending their noisy parameter observations to a multi-antenna FC. We presented two algorithms for finding the sensor phase shifts that minimize the variance of the estimated parameter, one based on a relaxed SDP and a closed-form heuristic algorithm based on the ACMA approach. We analyzed the asymptotic performance of the phase-shift and forward scheme for both large numbers of sensors and FC antennas, and we derived conditions under which increasing the number of FC antennas will significantly benefit the estimation performance. We also analyzed the performance degradation that results when sensor phase errors of variance σ_p^2 are present, and we showed that for large N the variance will approximately increase by a factor of $1 + \sigma_p^2$ provided that $\sigma_p^2 \ll 1$ square radian. The sensor selection problem was studied assuming either low or high sensor noise with respect to the noise at the FC. For low sensor noise, two algorithms were proposed, one based on linear programming with a relaxed integer constraint, and a computationally simpler greedy approach. For high sensor noise, we showed that choosing the sensors with the smallest noise variances was approximately optimal. Simulation studies of the proposed algorithms illustrate their advantages and the validity of the asymptotic analyses.

REFERENCES

- [1] M. Gastpar, "Uncoded transmission is exactly optimal for a simple gaussian "sensor" network," *IEEE Trans. Info. Theory*, vol. 54, no. 11, pp. 5247–5251, Nov. 2008.
- [2] M. Gastpar and M. Vetterli, "Power, spatio-temporal bandwidth, and distortion in large sensor networks," *IEEE J. Sel. Areas Commun.*, vol. 23, no. 4, pp. 745–754, Apr. 2005.
- [3] —, "Source channel communication in sensor networks," in *Proc. ACM/IEEE IPSN 2003*, Apr. 2003, pp. 162–177.
- [4] S. Cui, J.-J. Xiao, A. J. Goldsmith, Z.-Q. Luo, and H. V. Poor, "Estimation diversity and energy efficiency in distributed sensing," *IEEE Trans. Signal Process.*, vol. 55, no. 9, pp. 4683–4695, Sep. 2007.
- [5] I. Bahceci and A. K. Khandani, "Linear estimation of correlated data in wireless sensor network with optimum power allocation and analog modulation," *IEEE Trans. Commun.*, vol. 56, no. 7, pp. 1146–1156, Jul. 2008.

- [6] J. Matamoros and C. Anton-Haro, "Scaling law of an opportunistic power allocation scheme for amplify-and-forward wireless sensor networks," *IEEE Commun. Letter*, vol. 15, no. 2, pp. 169–171, Feb. 2011.
- [7] M. K. Banavar, C. Tepedelenlioglu, and A. Spanias, "Estimation over fading channels with limited feedback using distributed sensing," *IEEE Trans. Signal Process.*, vol. 58, no. 1, pp. 414–425, Jan. 2010.
- [8] A. S. Leong and S. Dey, "On scaling laws of diversity schemes in decentralized estimation," *IEEE Trans. Info. Theory*, vol. 57, no. 7, pp. 4740–4759, Jul. 2011.
- [9] C.-H. Wang, A. S. Leong, and S. Dey, "Distortion outage minimization and diversity order analysis for coherent multiaccess," *IEEE Trans. Signal Process.*, vol. 59, no. 12, pp. 6144–6159, Dec. 2011.
- [10] J. Fang and H. Li, "Power constrained distributed estimation with correlated sensor data," *IEEE Trans. Signal Process.*, vol. 57, no. 8, pp. 3292–3297, Aug. 2009.
- [11] G. Thatte and U. Mitra, "Sensor selection and power allocation for distributed estimation in sensor networks: Beyond the star topology," *IEEE Trans. Signal Process.*, vol. 56, no. 7, pp. 2649–2661, Jul. 2008.
- [12] K. Zarifi, S. Zaidi, S. Affes, and A. Ghayeb, "A distributed amplify-and-forward beamforming technique in wireless sensor networks," *IEEE Trans. Signal Process.*, vol. 59, no. 8, pp. 3657–3674, Aug. 2011.
- [13] J.-J. Xiao, S. Cui, Z.-Q. Luo, and A. J. Goldsmith, "Linear coherent decentralized estimation," *IEEE Trans. Signal Process.*, vol. 56, no. 2, pp. 757–770, Feb. 2008.
- [14] A. D. Smith, M. K. Banavar, C. Tepedelenlioglu, and A. Spanias, "Distributed estimation over fading macs with multiple antennas at the fusion center," in *Proc. IEEE Asilomar 2009*, Nov. 2009, pp. 424–428.
- [15] M. K. Banavar, A. D. Smith, C. Tepedelenlioglu, and A. Spanias, "On the effectiveness of multiple antennas in distributed detection over fading macs," *IEEE Trans. Wireless Commun.*, vol. 11, no. 5, pp. 1744–1752, May 2012.
- [16] —, "Distributed detection over fading macs with multiple antennas at the fusion center," *arXiv:1001.3173*, 2010.
- [17] S. T. Smith, "Optimum phase-only adaptive nulling," *IEEE Trans. Signal Process.*, vol. 47, no. 7, pp. 1835–1843, Jul. 1999.
- [18] P. J. Kajenski, "Phase only antenna pattern notching via a semidefinite programming relaxation," *IEEE Trans. Antennas Propag.*, vol. 60, no. 5, pp. 2562–2565, May 2012.
- [19] M. Xia, W. Wen, and S.-C. Kim, "Opportunistic cophasing transmission in miso systems," *IEEE Trans. Commun.*, vol. 57, no. 12, pp. 3764–3770, Dec. 2009.
- [20] C. Tepedelenlioglu and A. B. Narasimhamurthy, "Universal distributed estimation over multiple access channels with constant modulus signaling," *IEEE Trans. Signal Process.*, vol. 58, no. 9, pp. 4783–4794, Sep. 2010.
- [21] S. Joshi and S. Boyd, "Sensor selection via convex optimization," *IEEE Trans. Signal Process.*, vol. 57, no. 2, pp. 451–462, Feb. 2009.
- [22] V. Gupta, T. H. Chung, B. Hassibi, and R. M. Murray, "On a stochastic sensor selection algorithm with applications in sensor scheduling and sensor coverage," *Automatica*, vol. 42, pp. 251–260, Feb. 2006.
- [23] F. Bian, D. Kempe, and R. Govindan, "Utility-based sensor selection," in *Proc. ACM/IEEE IPSN 2006*, Apr. 2006, pp. 11–18.
- [24] V. Krishnamurthy, "Decentralized activation in dense sensor networks via global games," *IEEE Trans. Signal Process.*, vol. 56, no. 10, pp. 4936–4950, Oct. 2008.
- [25] Z.-Q. Luo and W. Yu, "An introduction to convex optimization for communications and signal processing," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 8, pp. 1426–1438, Aug. 2006.
- [26] A.-J. van der Veen and A. Paulraj, "An analytical constant modulus algorithm," *IEEE Trans. Signal Process.*, vol. 44, no. 5, pp. 1136–1155, May 1996.

- [27] F. Jiang, J. Chen, and A. L. Swindlehurst, "Phase-only analog encoding for a multi-antenna fusion center," in *IEEE ICASSP 2012*, Mar. 2012, pp. 2645–2648.
- [28] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. NJ: Prentice Hall, 1993.
- [29] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge University Press, 2004.
- [30] Y. Zhang and A. M.-C. So, "Optimal spectrum sharing in MIMO cognitive radio networks via semidefinite programming," *IEEE J. Sel. Areas Commun.*, vol. 29, no. 2, pp. 362–373, Feb. 2011.
- [31] G. Golub and C. V. Loan, *Matrix Computations*. Baltimore, MD: The Johns Hopkins University Press, 1989.
- [32] G. Kramer and M. Gastpar, "Capacity theorems for wireless relay channels," in *Proc. Allerton Conf. Communications, Control and Computing, 2003*, Oct. 2003, pp. 1074–1083.
- [33] S. A. Jafar, "The ergodic capacity of phase-fading interference networks," *IEEE Trans. Info. Theory*, vol. 57, no. 12, pp. 7685–7694, Feb. 2011.
- [34] A. Billionnet and A. Faye, "A lower bound for a constainted quadratic 0-1 minimization problem," *Discrete Applied Mathematics*, vol. 74, no. 2, pp. 135–146, Apr. 1997.

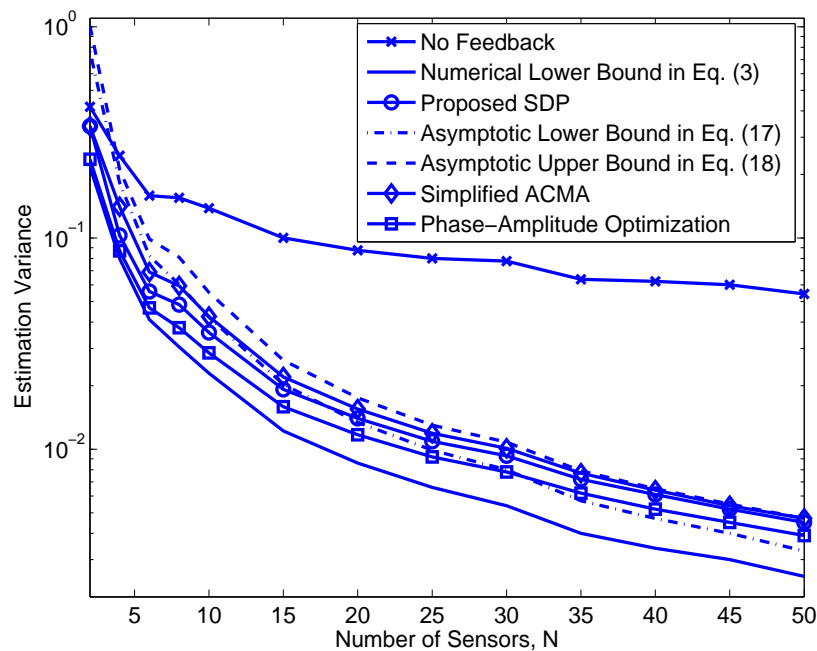


Fig. 1. Performance of the proposed algorithms with an increasing number of sensors for a low measurement noise scenario ($\sigma_n^2 = 0.1$, $\sigma_{v,i}^2$ uniformly distributed over $[0.01, 0.1]$, d_i uniformly distributed over $[3, 20]$ and $M = 4$).

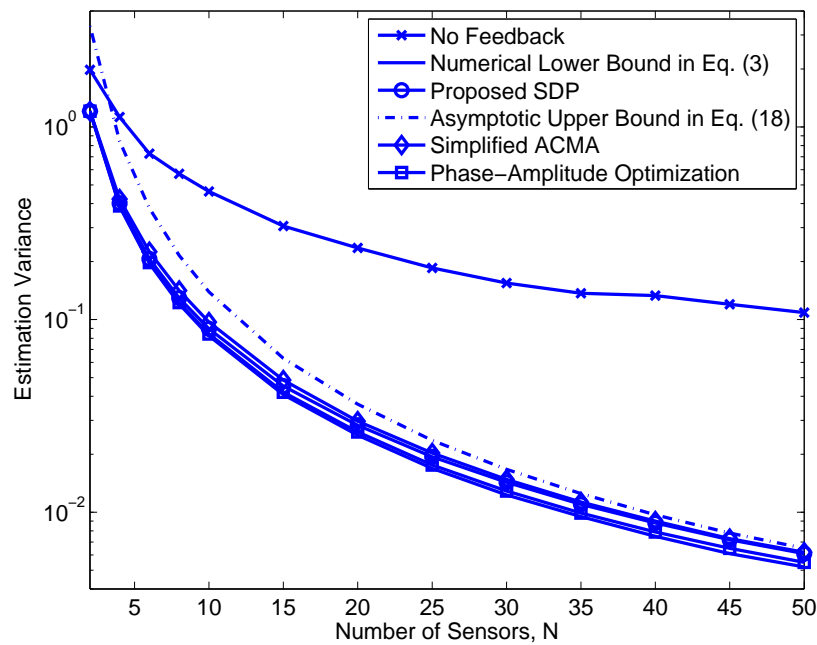


Fig. 2. Performance of the proposed algorithms with an increasing number of sensors for a low measurement noise scenario ($\sigma_n^2 = 0.1$, $\sigma_{v,i}^2$ uniformly distributed over $[0.01, 0.1]$, $d_i = 11.5$ and $M = 4$).

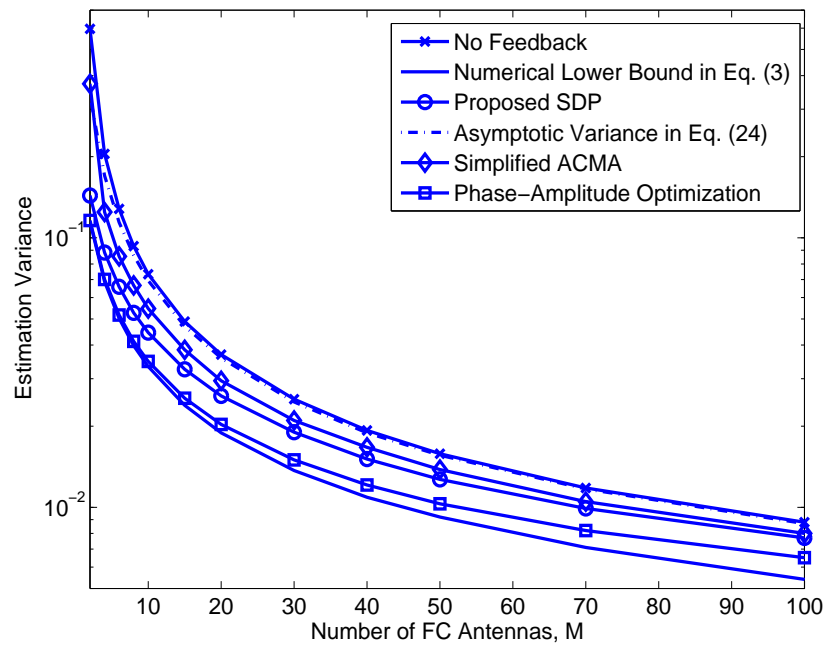


Fig. 3. Performance of the proposed algorithms with an increasing number of antennas ($\sigma_n^2 = 0.1$, $\sigma_{v,i}^2$ uniformly distributed over $[0.001, 0.01]$, d_i uniformly distributed over $[3, 20]$ and $N = 4$).

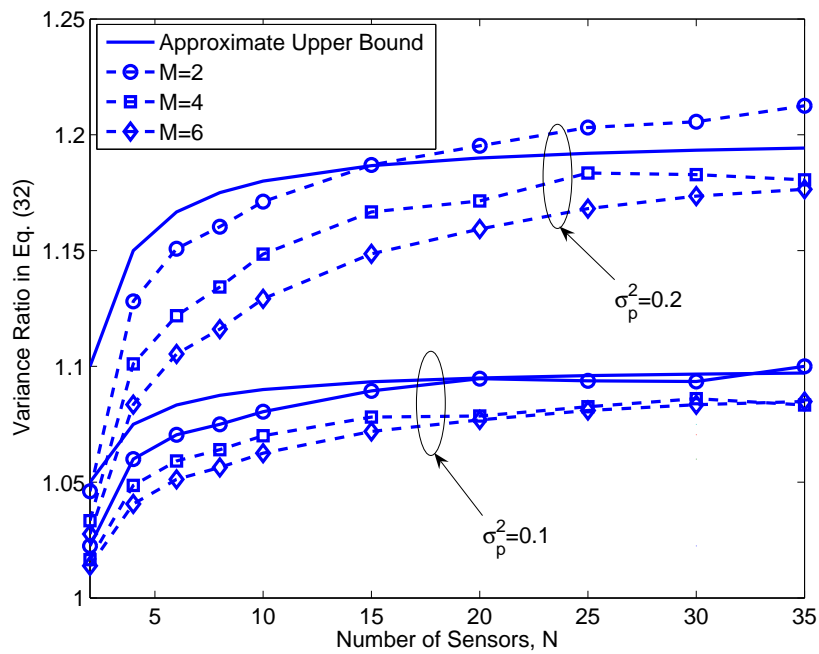


Fig. 4. Effect of phase errors on algorithm performance ($\sigma_n^2 = 0.1$, $\sigma_{v,i}^2$ uniformly distributed over $[0.01, 0.1]$ and d_i uniformly distributed over $[3, 20]$).

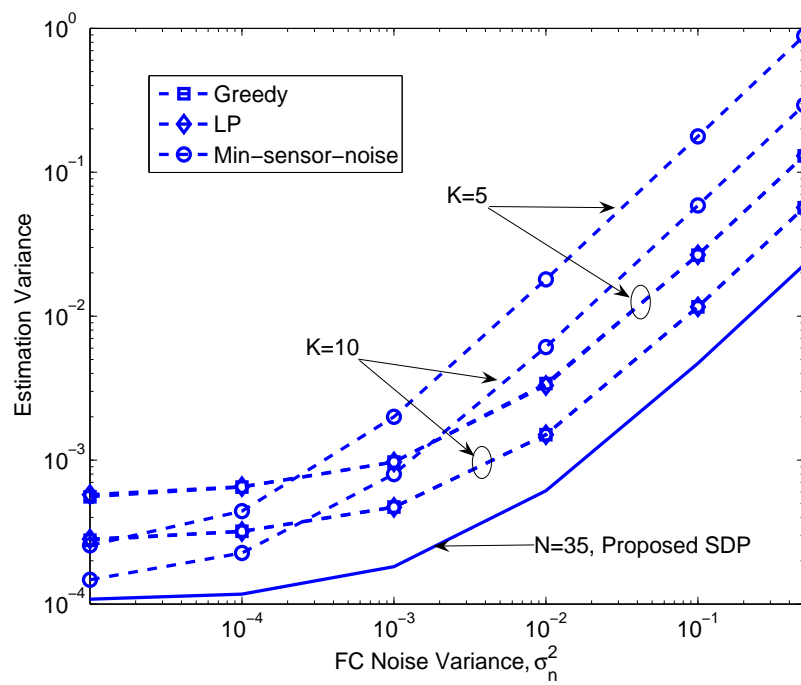


Fig. 5. Performance comparison between different sensor selection algorithms ($N = 35$, $M = 4$, $\sigma_{v,i}^2$ uniformly distributed over $[0.001, 0.01]$ and d_i uniformly distributed over $[3, 20]$).