Massive MIMO for Decentralized Estimation of a Correlated Source

Amirpasha Shirazinia, *Member, IEEE*, Subhrakanti Dey, *Senior Member, IEEE*, Domenico Ciuonzo, *Member, IEEE*, and Pierluigi Salvo Rossi, *Senior Member, IEEE*

Abstract—We consider a decentralized multi-sensor estimation problem where L sensor nodes observe noisy versions of a correlated random source vector. The sensors amplify and forward their observations over a fading coherent multiple access channel (MAC) to a fusion center (FC). The FC is equipped with a large array of N antennas and adopts a minimum mean-square error (MMSE) approach for estimating the source. We optimize the amplification factor (or equivalently transmission power) at each sensor node in two different scenarios: a) with the objective of total power minimization subject to mean square error (MSE) of source estimation constraint, and b) with the objective of minimizing MSE subject to total power constraint. For this purpose, based on the well-known favorable propagation condition (when $L \ll N$) achieved in massive multiple-input multiple-output (MIMO), we apply an asymptotic approximation on the MSE and use convex optimization techniques to solve for the optimal sensor power allocation in a) and b). In a), we show that the total power consumption at the sensors decays as $1/N$, replicating the power savings obtained in massive MIMO mobile communications literature. We also show several extensions of the aforementioned scenarios to the cases where sensor-to-FC fading channels are correlated, and channel coefficients are subject to estimation error. Through numerical studies, we also illustrate the superiority of the proposed optimal power allocation methods over uniform power allocation.

Index Terms—Decentralized estimation, wireless sensor networks, massive MIMO, coherent MAC, convex optimization, power allocation.

I. INTRODUCTION

THIS section summarizes prior work in the field, and provides paper contributions as well as outline and mathematical notations.

A. Background

Wireless sensor networks have recently attracted much research interest due to their practical popularity in accomplishing

A. Shirazinia and S. Dey are with the Signals and Systems Division, Department of Engineering Sciences, Uppsala University, Uppsala 751 21, Sweden (e-mail: amirpasha.shirazinia@signal.uu.se; subhrakanti.dey@signal.uu.se).

D. Ciuonzo is with University of Naples "Federico II", DIETI, I-80125 Naples, Italy (e-mail: domenico.ciuonzo@ieee.org).

P. Salvo Rossi is with the Department of Electronics and Telecommunications, Norwegian University of Science and Technology, Trondheim 7491, Norway (e-mail: salvorossi@ieee.org).

decentralized tasks, such as monitoring, sensing, computation and communication. In this field, many different schemes for decentralized sensing and information transmission, detection and estimation of sources using multiple sensors have been proposed, see, e.g., [2]–[15]. One popular transmission technique in wireless sensor networks is analog amplify and forwarding [16], where sensors transmit over fading channels a scaled version of their analog measurements to a fusion center (FC) for estimation or detection purposes. Analog amplify and forwarding technique has been shown to be optimal in some situations [17]. Therefore, this technique has been investigated under different multiple access schemes such as coherent multiple access [11], [12], [16], and orthogonal multiple access [13]–[15]. Analog transmissions over coherent MAC normally provide a better performance compared to orthogonal MAC that requires more resources such as time and bandwidth; however, this performance improvement is achieved at the price of precise phase synchronization at sensor nodes.

Depending on the application area, the aforementioned works can be also categorized into two different contexts: decentralized detection [5], [7], [9], [10] and decentralized estimation [2]–[4], [6], [11]–[15], [18] of either correlated or uncorrelated sources. Moreover, from a technological point of view, these works can be divided into two groups: when the FC is equipped with a single antenna [5], [6], [13]–[15] or multiple antennas [7]–[10], [12], [18]. It is well-known that using multiple antennas can increase spectral efficiency of a wireless system through spatial multiplexing. With respect to these works, it should be highlighted that the focus of the current paper is on decentralized estimation of a correlated source over coherent MACs where the FC is equipped with large arrays of antennas.

Indeed, there has been a vast interest recently to equip the FC (or the base station in cellular communication framework) with large arrays of antennas, also known as the massive multiple-input multiple output (MIMO) framework [19]–[21]. The use of arrays with massive number of antennas in wireless communication does not only increase spectral efficiency, but it can also improve energy efficiency of MIMO system [22], [23]. This improvement has been obtained at the expense of employing a large arrays of antennas in a fixed volume, which makes its practical hardware implementation rather complex. However, recent research has shown that massive MIMO arrays are quite realistic at typical radio frequencies [24], [25], and low-complexity effective signal processing approaches have been also proposed to deal with the large dimension of antenna arrays [26].

With the assumption of employing large number of antenna arrays, known results in MIMO communication systems can be

considerably simplified [21], [27]. As a result, it provides analytical solutions to problems that would otherwise be mathematically intractable. In wireless sensor networks, the massive MIMO framework has been also recently employed for decentralized detection and estimation [28]–[30]. For example, in [28], maximum likelihood (ML) estimation of a *scalar deterministic* source is considered under the assumption of massive MIMO. Further, while the work [29] mainly considers a decentralized detection problem, the authors also show that similar results can be achieved for decentralized estimation of a *random scalar* source.

B. Contributions and Main Results

In the present paper, within the massive MIMO framework, we focus on the decentralized estimation problem in a more general case where the source is modeled as a *vector* comprised of *random* and *correlated* components. Our main objective is to optimally design the sensors' amplification factors with respect to minimizing sensor power consumptions or maximizing estimation accuracy subject to relevant constraints. In our setting, L sensor nodes observe noisy versions of a possibly correlated random source vector. The sensor nodes amplify the observations according to their power budget, and forward them over coherent fading MACs to a FC equipped with a large number of antennas, denoted by N. The FC estimates the source by adopting a minimum mean square error (MMSE) estimator, which, by definition, gives the lowest possible MSE.

We optimize the sensor amplification factor (or equivalently the transmission power allocation) at each sensor node:

a) with the objective of total power minimization subject to a maximum MSE constraint (incurred by using the MMSE estimator), and

b) with the objective of minimizing MSE of source estimation subject to a total power constraint.

For the purpose of optimization, we apply an asymptotic approximation based on the *favorable propagation conditions in the massive MIMO literature* in order to simplify the MSE expression when $N \gg L$ [19]. Using the resulting approximation, our contributions are as follows:

- We show that the power allocation optimization problems, mentioned earlier, are convex, which, in general, can be solved numerically using well-known convex optimization techniques in polynomial time.
- We analyze the optimization problems, and under some conditions derive closed-form solutions to them. Our analysis reveals that, in the present framework, as the number of antennas at the FC, N, increases, the sensors can decrease their total power consumption with a factor proportional to $1/N$ in order to satisfy a targeted MSE.
- It is also seen that using the massive MIMO approximation, the optimal sensor power allocation only depends on the distance-based attenuation components and not on the randomly-varying fading gains.
- We extend the optimal power allocation to three new cases, where sensor-to-FC channel gains are correlated, sensor-to-FC channel additive noise components are correlated, and channel gains are subject to estimation error. For instance, our analysis reveals that under channel correlation, the performance, in terms of total power or MSE, can suffer from potential degradation.

- We illustrate, via numerical results, significant gains obtained by the proposed optimal power allocation methods compared to uniform power allocation.

The main contributions with respect to existing literature on massive MIMO-based decentralized estimation can be highlighted as follows. We consider decentralized estimation of a *general correlated random vector* in a massive MIMO framework over coherent MAC using the MMSE estimator, whereas [28] studied ML estimation of a *scalar deterministic source* and [29] considered MMSE estimation of a *scalar random* source. It should be also mentioned that a similar decentralized estimation problem with correlated noise and multiple antennas at the FC was also considered in [12]. However, the following differences with respect to our formulation are present: first, the unknown source to be estimated in [12] is a random scalar. Second, a non-asymptotic analysis (i.e., not massive MIMO) has been considered with coherent MAC, and finally, the authors considered MSE of the best linear unbiased estimator (BLUE) as the relevant metric.

C. Paper Outline

The remainder of the paper is organized as follows. In Section II, we begin with system description, and then proceed with problem formulation in Section III. We analyze power-minimization problem and MSE-minimization problem in Section IV, where each consists of the study of uncorrelated source and correlated source cases. In Section V, we will analyze optimal power allocations, where sensor-to-FC channel gain and additive noise components are correlated, and where MIMO channels are assumed to be a generic model. Performance comparison of the proposed optimization schemes with other methods via numerical simulations are illustrated in Section VI, followed by concluding remarks in Section VII.

D. Notation

We denote column vectors and matrices by bold lower-case and upper-case letters, respectively. The matrix trace is denoted by $\mathrm{Tr}(\cdot)$, and matrix/vector conjugate-transpose (resp. transpose) by $(\cdot)^H$ (resp. $(\cdot)^\top$). The notation $\mathrm{diag}(a_1, a_2, \ldots, a_n)$ is used for a diagonal matrix whose diagonal elements are $a_1, \ldots, a_n$. We use $\mathbb{E}[\cdot]$ to denote the expectation operator. $[\mathbf{A}]_{ij}$ means the element of the matrix $\mathbf{A}$ at ith row and jth column. The vector of all zeros of size n, and the identity matrix of size $n \times n$ are denoted by $\mathbf{0}_n$ and $\mathbf{I}_n$, respectively. The notation $\mathbf{X} \succ \mathbf{0}$ (resp. $\mathbf{X} \succeq \mathbf{0}$) means that the matrix $\mathbf{X}$ is positive definite (resp. positive semi-definite). $[\cdot]^+$ denotes $\max\{0, \cdot\}$ and $(\cdot)^\star$ means optimality in some sense. The circularly-symmetric Gaussian distribution is denoted by $\mathcal{CN}$. Finally, we denote equality in an asymptotic sense by $\overset{a}{=}$.

II. System Description

We study the decentralized estimation task over a coherent MAC shown in Fig. 1. Based on the studied system model in Fig. 1, the received signal at the FC can be written as

$$\mathbf{y} = \mathbf{HD}\boldsymbol{\theta} + \underbrace{\mathbf{HDn} + \mathbf{v}}_{\triangleq \mathbf{w}}. \tag{1}$$

where each component of the vector $\mathbf{y} \in \mathbb{C}^N$ is a coherent sum of received signals from all sensors. In (1), the source is characterized by the vector $\boldsymbol{\theta} = [\theta_1, \ldots, \theta_L]^\top \in \mathbb{C}^L$, where $\boldsymbol{\theta} \sim$
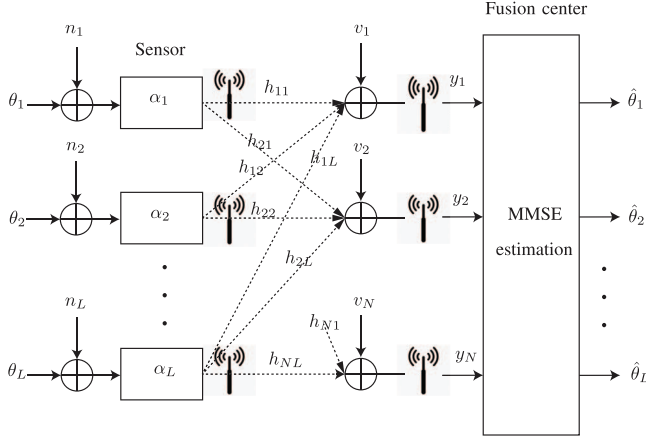
Fig. 1. System model for decentralized MMSE estimation over coherent MAC with analog forwarding.

$\mathcal{CN}(\mathbf{0}_L, \mathbf{C}_\theta)$ and $\mathbf{C}_\theta \in \mathbb{C}^{L \times L}$ is the positive definite source covariance matrix which is not necessarily diagonal. The lth source-to-sensor noise component is denoted by n_l. Stacking up all the noise components, we denote the source-to-sensor noise vector by $\mathbf{n} = [n_1, \ldots, n_L]^\top$, where $\mathbf{n} \sim \mathcal{CN}(\mathbf{0}_L, \mathbf{C}_n)$ and $\mathbf{C}_n \triangleq \mathrm{diag}\left(\sigma_{n_1}^2, \ldots, \sigma_{n_L}^2\right)$ is the source-to-sensor noise covariance matrix. Further, α_l is the complex amplification gain at the lth sensor node (to be designed) and $\mathbf{D} \triangleq \mathrm{diag}(\alpha_1, \ldots, \alpha_L)$. We denote the channel matrix between the L sensors and the FC equipped with N antennas by $\mathbf{H} \in \mathbb{C}^{N \times L}$, where $[\mathbf{H}]_{il}$ corresponds to the channel gain between the lth sensor and the ith antenna at the FC. The channel matrix $\mathbf{H}$ models independent fast fading (small-scale fading) and log-normal shadow fading (large-scale fading). The generic coefficient $[\mathbf{H}]_{il}$ is then expressed as

$$[\mathbf{H}]_{il} = \frac{1}{\sqrt{d_l^{2\beta}}}[\mathbf{G}]_{il}, \quad i = 1, \ldots, N, \quad l = 1, \ldots, L, \quad (2)$$

where d_l is the distance between the lth sensor to the FC, and 2β is the pathloss exponent. Furthermore, $[\mathbf{G}]_{il}$ are independent and identically distributed (i.i.d.) random variables drawn from $\mathcal{CN}(0, 1)$. Based on the above assumptions, we have

$$\mathbf{H} = \mathbf{G}\mathbf{\Gamma}^{1/2}, \quad (3)$$

where $\mathbf{\Gamma} \triangleq \mathrm{diag}(\gamma_1, \ldots, \gamma_L)$ such that $\gamma_l \triangleq d_l^{-2\beta}$.

Note that the channel matrix $\mathbf{H}$, comprising of complex zero-mean Gaussian entries, implies that there are non-line of sight wireless transmissions between sensors and the FC, which arise due to obstacles (fixed or mobile) resulting in multi-path fading propagation and a rich scattering environment. Application examples for such a scenario include outdoor (e.g., in distributed tracking) or residential wireless sensor networks studied, e.g., in [31]–[33].

The additive Gaussian noise at the FC is denoted by $\mathbf{v} = [v_1, \ldots, v_N]^\top$ with distribution $\mathcal{CN}(\mathbf{0}_N, \mathbf{C}_v)$, where $\mathbf{C}_v = \sigma_v^2 \mathbf{I}_N$. The FC provides an estimate of the source vector from the received signal vector $\mathbf{y}$ in (1). We assume that the FC has perfect knowledge of the source and noise statistics as well as the channel gain matrix $\mathbf{H}$. This may be, in principle, achieved by pilot transmissions from the sensors at the beginning of each fading block or from the FC followed by sensors

informing the FC about their channel estimates assuming time-division-duplex (TDD) transmission and channel reciprocity. The perfect channel state assumption at the FC may be unrealistic in large MIMO systems due to pilot contamination problems or simply imperfect channel estimation. For this reason, this assumption will be relaxed later in Section V.D, where we will consider system design with noisy channel estimates.

Remark 1: In coherent MAC—unlike orthogonal MAC which requires more resources such as time or bandwidth—all sensor transmissions occur simultaneously but require distributed phase synchronization, also known as distributed beamforming at the sensor transmitters. This implies that the sensors need the knowledge of the phase of the complex channels so that they can cancel it before transmission allowing coherent reception at the FC. This synchronization might be difficult to achieve in a sensor network with large number of sensors/receive antennas due to the resulting large communication overhead. However, challenges in distributed beamforming involved in developing massive MIMO have been identified in [34], and significant technical progress in design—such as synchronization, reducing overhead, aggregate feedback etc.—in order to achieve distributed beamforming has been reported.

Given the model under investigation, the minimum mean-square error (MMSE) estimator can be applied in order to give the lowest possible MSE. In the next section, we show the resulting MSE and discuss our design method for optimally allocating power to sensors.

III. PROBLEM FORMULATION

Exploiting the MMSE estimator at the FC, the estimated vector $\widehat{\boldsymbol{\theta}} \triangleq [\widehat{\theta}_1, \ldots, \widehat{\theta}_L]^\top$ is obtained as the conditional mean [35, Ch. 15]

$$\widehat{\boldsymbol{\theta}} = \mathbb{E}[\boldsymbol{\theta}|\mathbf{H}, \mathbf{y}] = \left(\mathbf{C}_\theta^{-1} + \mathbf{D}^H\mathbf{H}^H\mathbf{C}_w^{-1}\mathbf{H}\mathbf{D}\right)^{-1}\mathbf{D}^H\mathbf{H}^H\mathbf{C}_w^{-1}\mathbf{y},$$
$$(4)$$

which gives the following MSE[1]

$$\begin{aligned} \mathrm{MSE} &\triangleq \mathbb{E}\left[\left.\|\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}\|_2^2\right| \mathbf{H}\right] \\ &= \mathrm{Tr}\left\{\left(\mathbf{C}_\theta^{-1} + \mathbf{D}^H\mathbf{H}^H\mathbf{C}_w^{-1}\mathbf{H}\mathbf{D}\right)^{-1}\right\}, \quad (5) \end{aligned}$$

where $\mathbf{C}_w = \mathbb{E}[\mathbf{w}\mathbf{w}^H] = \mathbf{H}\mathbf{D}\mathbf{C}_n\mathbf{D}^H\mathbf{H}^H + \mathbf{C}_v$. By using the matrix inversion lemma on $\mathbf{C}_w^{-1}$, (5) can be rewritten as

$$\begin{aligned} \mathrm{MSE} = \mathrm{Tr}\Big\{ &\left(\mathbf{C}_\theta^{-1} + \sigma_v^{-2}\mathbf{D}^H\mathbf{H}^H\mathbf{H}\mathbf{D} - \sigma_v^{-4}\mathbf{D}^H\mathbf{H}^H\mathbf{H}\mathbf{D} \right. \\ &\left. \times \left(\mathbf{C}_n^{-1} + \sigma_v^{-2}\mathbf{D}^H\mathbf{H}^H\mathbf{H}\mathbf{D}\right)^{-1}\mathbf{D}^H\mathbf{H}^H\mathbf{H}\mathbf{D}\right)^{-1}\Big\}. \quad (6) \end{aligned}$$

Now, since $[\mathbf{G}]_{il} \overset{\text{i.i.d.}}{\sim} \mathcal{CN}(0, 1), (l = 1, \ldots, L, i = 1, \ldots, N)$, then using the massive MIMO framework, as $N \to \infty$ while L remains fixed ($L \ll N$) the so-called *favorable propagation conditions* hold [20]. It is known that under such a condition,

$$\mathbf{H}^H\mathbf{H} = \mathbf{\Gamma}^{1/2}\mathbf{G}^H\mathbf{G}\mathbf{\Gamma}^{1/2} \overset{a}{=} N\mathbf{\Gamma}. \quad (7)$$

---

[1]We note that due to the Gaussian nature of the system (see e.g., (4)), the linear MMSE (LMMSE) and the MMSE estimators are both linear and equivalent under the setup of this work ([35], Chap. 15).
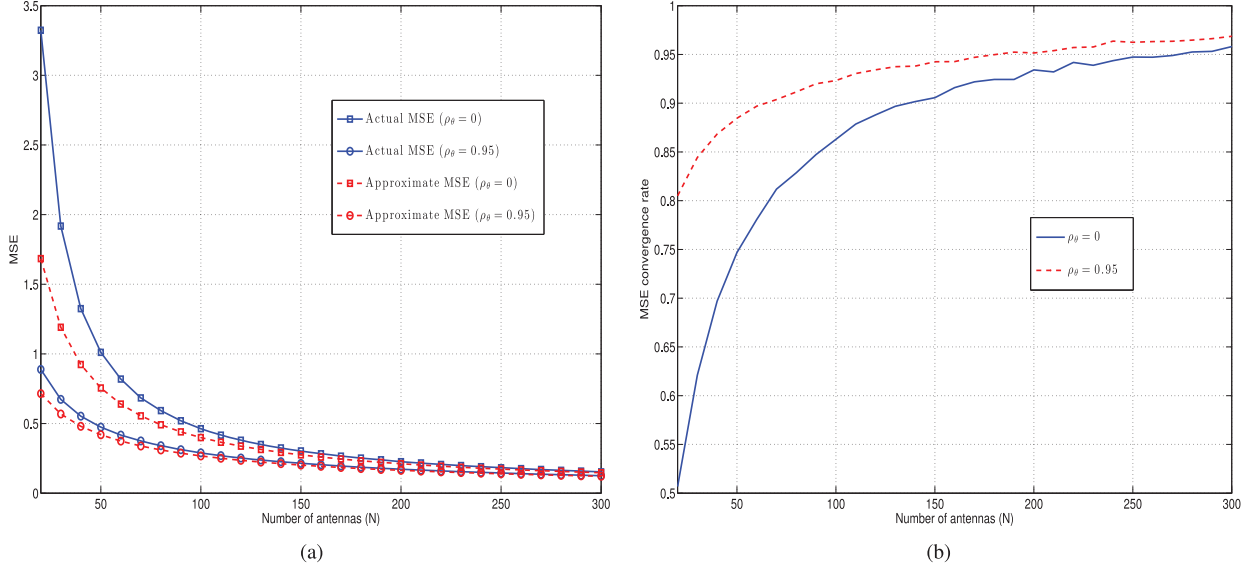
Fig. 2.   Left: Comparison of actual MSE (6) (averaged over 100 realizations of the channel) and approximate MSE using massive MIMO approximation (8) for two scenarios when the source is uncorrelated $\rho_\theta = 0$, and highly correlated $\rho_\theta = 0.95$. Right: Comparison of convergence rate of approximate MSE to the actual MSE for the uncorrelated and correlated scenarios. (a) Actual MSE vs. approximate MSE, (b) Convergence rate of MSE.

Next, we define $\mathbf{\Lambda} \triangleq \mathbf{D}^H \mathbf{\Gamma} \mathbf{D} = \mathrm{diag}(\lambda_1, \ldots, \lambda_L)$ where $\lambda_l \triangleq \gamma_l |\alpha_l|^2, \forall l$. Hence, the MSE in (6) can be asymptotically approximated as

$$\mathrm{MSE} \stackrel{a}{=} \mathrm{Tr} \left\{ \left(\mathbf{C}_\theta^{-1} + N\sigma_v^{-2} \mathbf{\Lambda} \right. \right.$$
$$\left. \left. - N^2 \sigma_v^{-4} \mathbf{\Lambda} \left(\mathbf{C}_n^{-1} + N\sigma_v^{-2}\mathbf{\Lambda} \right)^{-1} \mathbf{\Lambda} \right)^{-1} \right\}. \quad (8)$$

We note that, from now on, whenever we use the term MSE, we refer to the asymptotic MSE expressed by (8).

Remark 2: Using the massive MIMO approximation (7), the MSE formulation is simplified (cf. (8)). Further, under this approximation, the optimal design of the sensor amplification factors α_l reduces to optimizing only the (squared) modulus of α_l, and not the complex-valued factor α_l. This implies that optimization of the phase shift component is irrelevant in the large-array regime, which has been already observed in [29] for a scalar source. Finally, as a byproduct, the obtained power allocation problems (as we will see later in the subsequent sections) become mathematically more tractable and easier to solve given the simplifications arising from the massive MIMO approximation.

Another central design criterion that needs to be considered is the average total power consumed by sensors for transmission, which can be formulated as

$$P_{\mathrm{tot}} = \mathbb{E}\left[||\mathbf{D}(\boldsymbol{\theta} + \mathbf{n})||_2^2 \right] = \sum_{l=1}^{L} |\alpha_l|^2 \left[\mathbf{C}_\theta + \mathbf{C}_n \right]_{ll}$$
$$= \sum_{l=1}^{L} \frac{\lambda_l}{\gamma_l} \left[\mathbf{C}_\theta + \mathbf{C}_n \right]_{ll}. \quad (9)$$

In the following example, we offer insights into the accuracy of the MSE approximation (8) when large number of antennas are used at the FC.

Example 1: In Fig. 2, we compare the actual MSE in (6) and approximate MSE in (8) using (real) uniform power allocation, i.e., $\alpha_l^2 [\mathbf{C}_\theta + \mathbf{C}_n]_{ll} = \frac{P_{\mathrm{tot}}}{L}, \forall l$, with $P_{\mathrm{tot}} = 0.01$ W. For this study, we consider $L = 15$ sensors, and assume a homogenous scenario, where $\forall l \in \{1, \ldots, L\}$, the source-to-sensor noise variances $\sigma_{n_l}^2 = 10^{-3}$ W. Further, $\sigma_v^2 = 10^{-6}$ W. In order to determine the actual MSE, we draw the entries of the matrix $\mathbf{G}$ from $\mathcal{CN}(0, 1)$, and then plot average MSE over 100 realizations of $\mathbf{G}$. We also set the pathloss exponent to $2\beta = 2$, and sensor-to-FC distances d_l are chosen randomly according to a uniform distribution ranging from 20 to 70 meters. We consider two scenarios, where the source is either correlated or uncorrelated. We use the exponential correlation model (see later Section VI for details) with correlation coefficient $0 \leq \rho_\theta < 1$ in order to generate the source covariance matrix. Therefore, $\rho_\theta = 0$ refers to an uncorrelated source, whereas $\rho_\theta = 0.95$ refers to a highly correlated source. We observe from Fig. 2(a) that as N increases the gap between the actual and approximate MSE decreases such that at $N = 150$, the difference between the actual and approximate MSEs is approximately 0.028 for $\rho_\theta = 0$, and approximately 0.012 for $\rho_\theta = 0.95$. Hence, it can be seen that source correlation not only decreases MSE, but also increases convergence rate of the approximate MSE to the actual MSE with respect to N. The latter feature is illustrated in Fig. 2(b), where the MSE convergence rate is defined as the fraction of approximate MSE to the actual MSE, and is upper-bounded by 1. The purpose of this figure is to show how fast the curves approach to 1. For example, we have numerically shown that the source correlation can speed up the convergence rate.[2]

---

[2]It is worth-pointing out that the convergence rate may also depend on several parameters such as source correlation, available power budget, noise level, power allocation, etc. For instance, in [36], the authors studied the convergence of $\mathbf{H}^H \mathbf{H}$ to $N\mathbf{\Gamma}$ in a massive MIMO system under different convergence criteria. However, a rigorous convergence rate analysis of MSE falls beyond the scope of the current work.

In what follows, we pose two optimization problems dealing with sensor power allocation. The first problem minimizes the total power consumption by the sensors subject to reconstruction MSE constraint. Therefore, this optimization problem is desirable when power is limited, due to sensors' battery limitations, and we are imposed to guarantee a worst-case MSE. In the second problem, the MSE is minimized subject to total power constraint. Hence, this problem formulation is posed when the estimation accuracy, in terms of MSE, is more demanding rather than power consumption of the sensors. More specifically, the first optimization problem is stated as follows

$$\begin{cases} \underset{\{\lambda_l \geq 0\}_{l=1}^{L}}{\text{minimize}} & P_{\text{tot}} \\ \text{subject to} & \text{MSE} \leq \bar{d}, \end{cases} \quad \text{(P1)}$$

where P_{tot} and MSE are specified by (9) and (8), respectively. Further, $\bar{d}$ is a maximum user-defined distortion threshold, and has to be chosen such that $\text{Tr}\left\{ (\mathbf{C}_\theta^{-1} + \mathbf{C}_n^{-1})^{-1} \right\} < \bar{d} \leq \text{Tr}\{\mathbf{C}_\theta\}$. Note that the lower-bound on $\bar{d}$ corresponds to the case of transmission over an ideal noiseless channel, i.e., all measurements are directly available at the FC, however with measurement noise at sensors still present.

Inspired by similar results in massive MIMO-based cellular communication, it is straightforward to show[3] as the number of antennas N grows under the MSE constraint at the FC, then the total optimal decays proportionally with $1/N$. Thus, we can conclude the following:

Remark 3: The optimal total power consumed by sensors decays with a factor proportional to $1/N$.

The second optimization problem which is investigated is stated as follows

$$\begin{cases} \underset{\{\lambda_l \geq 0\}_{l=1}^{L}}{\text{minimize}} & \text{MSE} \\ \text{subject to} & P_{\text{tot}} \leq \bar{P}. \end{cases} \quad \text{(P2)}$$

We note that after solving the optimization problems (P1) and (P2) for λ_l, one can recover the complex gain α_l with amplitude $|\alpha_l|$ and an arbitrary phase as discussed in Remark 2. It can be easily shown (see e.g., [37, Problem 3.26]) for the convexity of the MSE and discussions on matrix concavity in [37, p. 110]) that the optimization problems (P1) and (P2) are convex in the variables $\lambda_l, l = 1, \ldots, L$. Hence, they can be solved numerically using standard convex optimization techniques [37].

Remark 4: In (P2), only the total power consumption (for all L sensors) has been constrained. In many wireless sensor network applications, the total power constraint has some physical implications. Typically, this constraint is imposed to guarantee a fair comparison when the number of sensor nodes can be variable, or there are strict constraints on how much interference can the sensor network generate on other wireless networks operating within its transmission range. Other application scenarios, where this constraint is critical, are discussed in [3] as well. Instead, depending on applications, one can constrain individual power consumption per sensor *in addition* to the total power constraint. In this case, similar analysis can be pursued to find optimal power allocations, albeit at the expense of the additional complexity of L extra power constraints rather than one total power constraint. This kind of optimization problem, for a

---

³This can be shown easily by using a change of variable, e.g., $\widetilde{\lambda}_l \triangleq N\lambda_l/\sigma_v^2$, and then by rewriting an equivalent optimization problem with the new variable $\widetilde{\lambda}_l$.

---

slightly different decentralized estimation problem over orthogonal channels with a single antenna at th FC, has already been studied, e.g., in [13].

Although it is more desirable to address the convex optimization problems (P1) and (P2) analytically, or provide closed-form solutions, this is not, however, feasible in some cases. Therefore, in such situations, it is more convenient to rewrite the optimization problems in a way to be solved in a computationally more efficient manner. A typical approach is to use the semidefinite programming (SDP) method that can be solved via the low, polynomial complexity interior point method [38]. The optimizations (P1) and (P2) can be expressed as SDPs by introducing slack random variables and by using Schur's complement [11], [37]. More precisely, optimization problem (P1) can be equivalently solved as

$$\underset{\{\lambda_l \geq 0\}_{l=1}^{L}, \mathbf{X}, \mathbf{Y}}{\text{minimize}} \quad \sum_{l=1}^{L} \lambda_l \frac{1}{\gamma_l} [\mathbf{C}_\theta + \mathbf{C}_n]_{ll}$$

$$\text{subject to} \quad \text{Tr}\{\mathbf{Y}\} \leq \bar{d}$$

$$\begin{bmatrix} \mathbf{C}_\theta^{-1} + N\sigma_v^{-2}\boldsymbol{\Lambda} - N^2\sigma_v^{-4}\mathbf{X} & \mathbf{I}_L \\ \mathbf{I}_L & \mathbf{Y} \end{bmatrix} \succeq \mathbf{0}$$

$$\begin{bmatrix} \mathbf{X} & \boldsymbol{\Lambda} \\ \boldsymbol{\Lambda} & \mathbf{C}_n^{-1} + N\sigma_v^{-2}\boldsymbol{\Lambda} \end{bmatrix} \succeq \mathbf{0}, \quad \text{(10)}$$

where $\mathbf{X} \succeq \mathbf{0}$ is a $L \times L$ diagonal matrix, and $\mathbf{Y} \succeq \mathbf{0}$ is a Hermitian $L \times L$ matrix. Further, $\boldsymbol{\Lambda} = \text{diag}(\lambda_1, \ldots, \lambda_L)$.

Similarly, the optimization problem (P2) can be equivalently solved as the following SDP

$$\underset{\{\lambda_l \geq 0\}_{l=1}^{L}, \widetilde{\mathbf{X}}, \widetilde{\mathbf{Y}}}{\text{minimize}} \quad \text{Tr}\{\widetilde{\mathbf{X}}\}$$

$$\text{subject to} \quad \sum_{l=1}^{L} \lambda_l \frac{1}{\gamma_l} [\mathbf{C}_\theta + \mathbf{C}_n]_{ll} \leq \bar{P}$$

$$\begin{bmatrix} \mathbf{C}_\theta^{-1} + N\sigma_v^{-2}\boldsymbol{\Lambda} - N^2\sigma_v^{-4}\widetilde{\mathbf{Y}} & \mathbf{I}_L \\ \mathbf{I}_L & \widetilde{\mathbf{X}} \end{bmatrix} \succeq \mathbf{0}$$

$$\begin{bmatrix} \widetilde{\mathbf{Y}} & \boldsymbol{\Lambda} \\ \boldsymbol{\Lambda} & \mathbf{C}_n^{-1} + N\sigma_v^{-2}\boldsymbol{\Lambda} \end{bmatrix} \succeq \mathbf{0}, \quad \text{(11)}$$

where $\widetilde{\mathbf{X}} \succeq \mathbf{0}$ is a Hermitian $L \times L$ matrix, and $\widetilde{\mathbf{Y}} \succeq \mathbf{0}$ is a $L \times L$ diagonal matrix.

In the next sections, we analyze the optimization problems (P1) and (P2).

IV. ANALYSIS OF (P1) AND (P2)

We analyze the problems (P1) and (P2) in two cases where the source is either *correlated* or *uncorrelated*. In the correlated case, we offer an approach which yields to a set of non-linear equations. Moreover, we relax the original problem which leads to an approximate closed-form solution. In the uncorrelated case, we also give a closed-form solution to the optimization problem.

A. Analysis of (P1): Correlated Source

1) Exact Analysis: First, we consider the general case where the elements of the source vector are correlated, i.e., $\mathbf{C}_\theta$ is not necessarily diagonal. In order to solve for λ_l in (P1), we write the Karush-Kuhn-Tucker (KKT) conditions [37]. It should be mentioned that a similar approach has been taken in [5], [14], [14] for solving an optimization problem with a single antenna

FC. However, here we provide an analytical approach to the problem of interest for the sake of completeness. Moreover, the structure of the optimal power allocation that we derive in this subsection offers insight into its dependence on the design variables and parameters.

We first introduce the Lagrange multiplier $\mu \geq 0$, and then write the Lagrangian (by neglecting the constraints $\lambda_l \geq 0$ for the moment) as shown in (12) at the bottom of the page.

Taking the partial derivative of (12) with respect to $\lambda_l (l = 1, \ldots, L)$, we obtain (13), shown at the bottom of the page, where we used the fact that for $\mathbf{A} \succ \mathbf{0}$, we have $\frac{d\mathbf{A}^{-1}}{dt} = -\mathbf{A}^{-1} \frac{d\mathbf{A}}{dt} \mathbf{A}^{-1}$ in which t is an element of the matrix $\mathbf{A}$. Letting (13) equal zero, and reconsidering $\lambda_l \geq 0$, it yields the following set of implicit non-linear equations for $l = 1, 2, \ldots, L$

$$\lambda_l = \left[\left\{ \frac{\mu \sigma_v^2 \sigma_{n_l}^{-4}}{N/\gamma_l [\mathbf{C}_\theta + \mathbf{C}_n]_{ll}} \right. \right.$$
$$\left. \left. \times \left[\left(\mathbf{C}_\theta^{-1} + \mathrm{diag}\left(\ldots, \frac{N\lambda_l}{\sigma_v^2 + N\sigma_{n_l}^2 \lambda_l}, \ldots \right) \right)^{-2} \right]_{ll} \right\}^{1/2} \right.$$
$$\left. - \frac{\sigma_v^2}{N\sigma_{n_l}^2} \right]^+ \tag{14}$$

Since $\mathbf{C}_\theta$ is not diagonal, for an arbitrary Lagrange multiplier μ, (14) can be solved using non-linear iterative numerical methods, or computer solvers (e.g., fsolve in MATLAB). These methods are typically faster than standard convex solvers (e.g., CVX [39]) for solving SDP-type programming as developed in (10). The optimal μ is determined such that the MSE constraint in (P1) is satisfied with equality at the optimal solution which can be proved by complementary slackness theorem [37].

2) Approximate Analysis: We propose an approximate approach in order to solve $\{\lambda_l\}_{l=1}^L$ when N is sufficiently large. For this purpose, we first assume that $\lambda_l > 0, \forall l$. Hence, using the matrix inversion lemma, we write the MSE in (8) as

$$\mathrm{MSE} = \mathrm{Tr}\left\{ \left(\mathbf{C}_\theta^{-1} + \left(\mathbf{C}_n + \frac{1}{N}\sigma_v^2 \boldsymbol{\Lambda}^{-1} \right)^{-1} \right)^{-1} \right\}.$$

Next, we expand the above MSE expression via Taylor series as a function of $\frac{1}{N}$ around $\frac{1}{N} = 0$, which yields

$$\mathrm{MSE} = \mathrm{Tr}\left\{ (\mathbf{C}_\theta^{-1} + \mathbf{C}_n^{-1})^{-1} \right\}$$
$$+ \frac{1}{N}\mathrm{Tr}\left\{ (\mathbf{C}_\theta^{-1} + \mathbf{C}_n^{-1})^{-2} \mathrm{diag}\left(\ldots, \frac{\sigma_v^2}{\sigma_{n_l}^4 \lambda_l}, \ldots \right) \right\}$$
$$+ \mathcal{O}\left(\frac{1}{N^2} \right), \tag{15}$$

where the first term represents the MSE expression at $\frac{1}{N} = 0$, and the coefficient of $1/N$ in the second term represents the first-order partial derivative of the MSE, with respect to $1/N$, at $1/N = 0$. Note that we have implicitly assumed that sensors always amplify their observations with positive (non-zero) gains, i.e., $\lambda_l > 0$. The reason for expanding the Taylor series around $1/N = 0$ is due to the massive MIMO framework with a large number of antennas N, and when N is sufficiently large, we can neglect the second and higher order terms in the series.

Note that $\mathbf{C}_n = \mathrm{diag}(\sigma_{n_1}^2, \ldots, \sigma_{n_L}^2)$, and for brevity, we define $\mathbf{Q} \triangleq (\mathbf{C}_\theta^{-1} + \mathbf{C}_n^{-1})^{-1}$. Then, by neglecting second-order moments (in $\frac{1}{N}$) of (15) onwards, and plugging it back into the constraint in (P1), we denote the approximate MSE by

$$\widetilde{\mathrm{MSE}} \triangleq \mathrm{Tr}\{\mathbf{Q}\} + \frac{1}{N}\mathrm{Tr}\left\{ \mathbf{Q}^2 \mathrm{diag}\left(\ldots, \frac{\sigma_v^2}{\sigma_{n_l}^4 \lambda_l}, \ldots \right) \right\}. \tag{16}$$

Now, we solve the following optimization problem for $\{\lambda_l\}_{l=1}^L$

$$\begin{array}{ll} \underset{\{\lambda_l > 0\}_{l=1}^L}{\mathrm{minimize}} & \sum_{l=1}^L \lambda_l \frac{1}{\gamma_l}[\mathbf{C}_\theta + \mathbf{C}_n]_{ll} \\ \mathrm{subject\ to} & \widetilde{\mathrm{MSE}} \leq \bar{d}, \end{array} \tag{P3}$$

which has the closed-form solution stated as follows.

Theorem 1: The optimal solution to Problem (P3) is given by

$$\lambda_l^\star = \frac{1}{N}\left[\frac{\frac{[\mathbf{Q}^2]_{ll}^{1/2} \sigma_v^2 \sqrt{\gamma_l}}{\sigma_{n_l}^2 [\mathbf{C}_\theta + \mathbf{C}_n]_{ll}^{1/2}} \sum_{m=1}^L \frac{d_m^\beta [\mathbf{C}_\theta + \mathbf{C}_n]_{mm}^{1/2} [\mathbf{Q}^2]_{mm}^{1/2}}{\sigma_{n_m}^2}}{\bar{d} - \sum_{m=1}^L [\mathbf{Q}]_{mm}} \right]. \tag{17}$$

Proof: The proof is given in Appendix A. □

$$\mathcal{L}(\boldsymbol{\Lambda}, \mu) = \sum_{l=1}^L \frac{\lambda_l}{\gamma_l}[\mathbf{C}_\theta + \mathbf{C}_n]_{ll} + \mu \left(\mathrm{Tr}\left\{ \left(\mathbf{C}_\theta^{-1} + N\sigma_v^{-2}\boldsymbol{\Lambda} - N^2 \sigma_v^{-4}\boldsymbol{\Lambda} (\mathbf{C}_n^{-1} + N\sigma_v^{-2}\boldsymbol{\Lambda})^{-1} \boldsymbol{\Lambda} \right)^{-1} \right\} - \bar{d} \right) \tag{12}$$

$$\frac{\partial \mathcal{L}(\boldsymbol{\Lambda}, \mu)}{\partial \lambda_l} = -\mu \mathrm{Tr}\left\{ \left(\mathbf{C}_\theta^{-1} + \mathrm{diag}\left(\ldots, \frac{N\lambda_l}{\sigma_v^2 + N\sigma_{n_l}^2 \lambda_l}, \ldots \right) \right)^{-2} \right.$$
$$\left. \times \mathrm{diag}\left(0, \ldots, 0, \frac{N\sigma_v^2}{(\sigma_v^2 + N\sigma_{n_l}^2 \lambda_l)^2}, 0, \ldots, 0 \right) \right\} + \frac{1}{\gamma_l}[\mathbf{C}_\theta + \mathbf{C}_n]_{ll} \tag{13}$$

In Section VI, we will numerically illustrate that such an approximate solution is nearly-optimal. By studying (17) in Theorem 1, it can be observed that the total power consumed by sensors decays with a factor proportional to $1/N$, which conforms with our finding in Remark 3. It should be noted that since an approximation of MSE, i.e., $\widetilde{\text{MSE}}$, is used in the constraint of (P3), $\lambda_l, \forall l$, might not satisfy the exact MSE by equality. In this case, we can easily rescale the resulting $\lambda_l^\star$ in (17) in order to satisfy the original MSE constraint by equality.

B. Analysis of (P1): Uncorrelated Source

Here we assume that the elements of the source vector $\boldsymbol{\theta}$ are uncorrelated, such that the source covariance matrix $\mathbf{C}_\theta \triangleq \text{diag}(\sigma_{\theta_1}^2, \ldots, \sigma_{\theta_L}^2)$. In this case, the optimization problem (P1) can be simplified into

$$
\begin{aligned}
&\underset{\{\lambda_l \geq 0\}_{l=1}^L}{\text{minimize}} && \sum_{l=1}^L \lambda_l \frac{1}{\gamma_l}\left(\sigma_{\theta_l}^2 + \sigma_{n_l}^2\right) \\
&\text{subject to} && \sum_{l=1}^L \frac{1}{\frac{1}{\sigma_{\theta_l}^2} + \frac{\lambda_l}{\sigma_{n_l}^2 \lambda_l + \sigma_v^2/N}} \leq \bar{d} \quad \text{(P4)}
\end{aligned}
$$

The optimal solution for $\lambda_l, \quad l = 1, 2, \ldots, L$ can be obtained in closed-form analytically, and is stated as follows.

Theorem 2: Provided $\frac{\sigma_{\theta_l}^2 \sqrt{\gamma_l}}{(\sigma_{\theta_l}^2 + \sigma_{n_l}^2)^{1/2}}$ is ordered decreasingly in $l \in \{1, \ldots, L\}$, $\exists$ a unique $M^\star$ such that

$$
M^\star = \max \left\{ M \in \{1, \ldots, L\} : \frac{\sigma_{\theta_l}^2 \sqrt{\gamma_l}}{(\sigma_{\theta_l}^2 + \sigma_{n_l}^2)^{1/2}} \right.
$$
$$
\left. > \frac{\bar{d} - \sum_{m=1}^M \frac{1}{\sigma_{\theta_m}^2 + 1/\sigma_{n_m}^2} - \sum_{m=M+1}^L \sigma_{\theta_m}^2}{\sum_{m=1}^M \frac{\sigma_{\theta_m}^2 d_m^\beta}{(\sigma_{\theta_m}^2 + \sigma_{n_m}^2)^{1/2}}} \right\}. \tag{18}
$$

Then, the optimal solution is given by

$$
\lambda_l^\star = \frac{1}{N} \left[\frac{\frac{\sigma_{\theta_l}^2 \sigma_v^2 \sqrt{\gamma_l}}{(\sigma_{\theta_l}^2 + \sigma_{n_l}^2)^{3/2}} \sum_{m=1}^{M^\star} \frac{\sigma_{\theta_m}^2 d_m^\beta}{(\sigma_{\theta_m}^2 + \sigma_{n_m}^2)^{1/2}}}{\bar{d} - \sum_{m=1}^{M^\star} 1/[1/\sigma_{\theta_m}^2 + 1/\sigma_{n_m}^2] - \sum_{m=M^\star+1}^L \sigma_{\theta_m}^2} \right.
$$
$$
\left. - \frac{\sigma_v^2}{\sigma_{\theta_l}^2 + \sigma_{n_l}^2} \right], \quad (19)
$$

for $l = 1, \ldots, M^\star$, and $\lambda_l^\star = 0$ for $l = M^\star + 1, \ldots, L$. Further, the amplitude of the optimal amplification gain for sensor l becomes $|\alpha_l^\star| = \sqrt{\lambda_l^\star/\gamma_l}$.

Proof: The proof follows by using KKT conditions, and from similar steps as in ([14], Example 2). However, for the sake of completeness, we give the details of the proof in Appendix B. $\square$

From Theorem 2, it can be also realized that the total power consumed by all sensors decays with a factor proportional to $1/N$ whose decaying rate is determined by (19).

C. Analysis of (P2): Correlated Source

1) Exact Analysis: Similar to the analysis in Section IV-A1, by introducing the Lagrange multiplier $\mu \geq 0$, we write the Lagrangian as

$$
\mathcal{L}(\boldsymbol{\Lambda}, \mu) = \mu \left(\sum_{l=1}^L \frac{\lambda_l}{\gamma_l}[\mathbf{C}_\theta + \mathbf{C}_n]_{ll} - \bar{P} \right)
$$
$$
+ \text{Tr} \left\{ \left(\mathbf{C}_\theta^{-1} + N\sigma_v^{-2}\boldsymbol{\Lambda} - N^2\sigma_v^{-4}\boldsymbol{\Lambda} \right. \right.
$$
$$
\left. \left. \times \left(\mathbf{C}_n^{-1} + N\sigma_v^{-2}\boldsymbol{\Lambda} \right)^{-1} \boldsymbol{\Lambda} \right)^{-1} \right\}. \tag{20}
$$

Taking the partial derivative of (20) with respect to λ_l ($l = 1, \ldots, L$), we obtain

$$
\frac{\partial \mathcal{L}(\boldsymbol{\Lambda}, \mu)}{\partial \lambda_l} = -\text{Tr} \left\{ \left(\mathbf{C}_\theta^{-1} + \text{diag}\left(\ldots, \frac{N\lambda_l}{\sigma_v^2 + N\sigma_{n_l}^2 \lambda_l}, \ldots \right) \right)^{-2} \right.
$$
$$
\left. \times \text{diag}\left(0, \ldots, 0, \frac{N\sigma_v^2}{(\sigma_v^2 + N\sigma_{n_l}^2 \lambda_l)^2}, 0, \ldots, 0 \right) \right\}
$$
$$
+ \frac{\mu}{\gamma_l}[\mathbf{C}_\theta + \mathbf{C}_n]_{ll} \tag{21}
$$

and letting (21) equal zero, it yields (since $\lambda_l \geq 0$)

$$
\lambda_l = \left[\left\{ \frac{\sigma_v^2 \sigma_{n_l}^{-4}}{\frac{N\mu}{\gamma_l}[\mathbf{C}_\theta + \mathbf{C}_n]_{ll}} \right. \right.
$$
$$
\left. \times \left[\left(\mathbf{C}_\theta^{-1} + \text{diag}\left(\ldots, \frac{N\lambda_l}{\sigma_v^2 + N\sigma_{n_l}^2 \lambda_l}, \ldots \right) \right)^{-2} \right]_{ll} \right\}^{1/2}
$$
$$
\left. - \frac{\sigma_v^2}{N\sigma_{n_l}^2} \right]^+. \tag{22}
$$

Finally, μ is determined in order to satisfy the power constraint in (P2) with equality.

2) Approximate Analysis: Similar to the analysis in Section IV-A2, we can derive approximate closed-form solution for Problem (P2) when N is sufficiently large. To do so, we follow the approximation of MSE in (15). Hence, the problem can be shown to be simplified into

$$
\begin{aligned}
&\underset{\{\lambda_l > 0\}_{l=1}^L}{\text{minimize}} && \text{Tr}\left\{ \mathbf{Q}^2 \text{diag}\left(\ldots, \frac{\sigma_v^2}{\sigma_{n_l}^4 \lambda_l}, \ldots \right) \right\} \\
&\text{subject to} && \sum_{l=1}^L \lambda_l \frac{1}{\gamma_l}[\mathbf{C}_\theta + \mathbf{C}_n]_{ll} \leq \bar{P}, \quad \text{(P5)}
\end{aligned}
$$

where $\mathbf{Q} = (\mathbf{C}_\theta^{-1} + \mathbf{C}_n^{-1})^{-1}$, $\mathbf{C}_n = \text{diag}(\sigma_{n_1}^2, \ldots, \sigma_{n_L}^2)$ and we have discarded $\text{Tr}\{\mathbf{Q}\}$ from $\widetilde{\text{MSE}}$ in (16), as it does not depend on λ_l. We have the following result:

Theorem 3: The optimal solution to (P5) is given by

$$
\lambda_l^\star = \frac{\bar{P}[\mathbf{Q}^2]_{ll}^{1/2}}{\frac{\sigma_{n_l}^2}{\sqrt{\gamma_l}}[\mathbf{C}_\theta + \mathbf{C}_n]_{ll}^{1/2} \sum_m \frac{[\mathbf{Q}^2]_{mm}^{1/2} d_m^\beta [\mathbf{C}_\theta + \mathbf{C}_n]_{mm}^{1/2}}{\sigma_{n_m}^2}}. \tag{23}
$$

Proof: The proof is omitted since it is analogous to the proof of Theorem 1. $\square$

We have the following corollary regarding the MSE of the optimal power allocations in (23).

Corollary 1: As $N \to \infty$, the MSE floor incurred by using the optimal power allocation derived in (23) becomes

$$\lim_{N \to \infty} \text{MSE} = \text{Tr} \left\{ \left(\mathbf{C}_\theta^{-1} + \mathbf{C}_n^{-1} \right)^{-1} \right\}. \qquad (24)$$

Proof: Letting $N \to \infty$ in (8) using the power allocations (23) yields the result. □

Remark 5: As suggested by Corollary 1, the MSE cannot decay to zero as $N \to \infty$. This is indeed expected since using the favorable propagation condition, there exist N orthogonal channels between the sensors and the FC, and it is well-known (see, e.g., [11]) that the MSE for an orthogonal multiple access channel cannot decay to zero as $N \to \infty$.

D. Analysis of (P2): Uncorrelated Source

Now, we assume that the source vector $\boldsymbol{\theta}$ is uncorrelated, i.e., the source covariance matrix $\mathbf{C}_\theta = \text{diag}(\sigma_{\theta_1}^2, \ldots, \sigma_{\theta_L}^2)$, where we are able to derive closed-form solution. In this case, the optimal solution to the sensor power allocation can be found in closed-form as follows.

Theorem 4: Provided $\frac{\sqrt{\gamma_l} \sigma_{\theta_l}^2}{(\sigma_{\theta_l}^2 + \sigma_{n_l}^2)^{1/2}}$ is ordered decreasingly in $l \in \{1, \ldots, L\}, \exists$ a unique $M^\star$ such that

$$M^\star = \max \left\{ M \in \{1, \ldots, L\} : \frac{(\sigma_{\theta_l}^2 + \sigma_{n_l}^2)^{1/2}}{\sqrt{\gamma_l} \sigma_{\theta_l}^2} \right.$$

$$\left. < \frac{N\bar{P}/\sigma_v^2 + \sum_{m=1}^M d_m^\beta}{\sum_{m=1}^M \frac{\sigma_{\theta_m}^2 d_m^\beta}{(\sigma_{\theta_m}^2 + \sigma_{n_m}^2)^{1/2}}} \right\}. \qquad (25)$$

Then, the optimal λ_l is given by

$$\lambda_l^\star = \frac{1}{N} \left[\frac{\sigma_v^2 \sigma_{\theta_l}^2 \sum_{m=1}^{M^\star} d_m^{2\beta}}{(\sigma_{\theta_l}^2 + \frac{\sigma_{n_l}^2}{\sqrt{\gamma_l}})^{3/2} \sum_{m=1}^{M^\star} \frac{\sigma_{\theta_m}^2 d_m^\beta}{(\sigma_{\theta_m}^2 + \sigma_{n_m}^2)^{1/2}}} - \frac{\sigma_v^2}{\sigma_{\theta_l}^2 + \sigma_{n_l}^2} \right]$$

$$+ \frac{\bar{P}\sigma_{\theta_l}^2}{(\sigma_{\theta_l}^2 + \frac{\sigma_{n_l}^2}{\sqrt{\gamma_l}})^{3/2} \sum_{m=1}^{M^\star} \frac{\sigma_{\theta_m}^2 d_m^\beta}{(\sigma_{\theta_m}^2 + \sigma_{n_m}^2)^{1/2}}} \qquad (26)$$

for $l = 1, \ldots, M^\star$, and $\lambda_l^\star = 0$ for $l = M^\star + 1, \ldots, L$. Further, the amplitude of the optimal amplification gain for sensor l becomes $|\alpha_l^\star| = \sqrt{\lambda_l^\star/\gamma_l}$.

Proof: The proof is omitted since it is analogous to the proof of Theorem 2. □

Corollary 2: As $N \to \infty$, the asymptotic MSE floor incurred by using the optimal power allocation derived in (26) becomes

$$\lim_{N \to \infty} \text{MSE} = \sum_{l=1}^L \frac{1}{1/\sigma_{\theta_l}^2 + 1/\sigma_{n_l}^2}. \qquad (27)$$

Proof: First, it can be verified from (25) that as $N \to \infty, M^\star \to L$. Hence, $\forall l$, we obtain $\lambda_l > 0$. Then, we find the limit of objective function as $N \to \infty$, which yields (27). □

Remark 6: We note that the expression in (27) coincides with (24) when $\mathbf{C}_\theta = \text{diag}(\sigma_{\theta_1}^2, \ldots, \sigma_{\theta_L}^2)$. As can be observed from (24), as $N \to \infty$, the MSE floor does not depend on channel parameters, power allocation and noise at the FC, rather depends on source and sensor noise parameters.

V. EXTENSIONS TO CORRELATED NOISE AND/OR GENERIC CHANNEL

Up to this point, our analysis and design have been built upon the assumption of uncorrelated channels and FC noise. In massive MIMO, large arrays of antennas are deployed in a fixed volume which may increase spatial correlation among antennas and correlation among thermal noise components at the FC. Therefore, in this section, we include the assumption of correlated noise, i.e., non-diagonal $\mathbf{C}_v$, and correlated channels (the columns of the channel matrix $\mathbf{H}$ are correlated) into our analysis.

A. Correlated Noise at FC

First, we consider the case of correlated noise and state the following result. Let the covariance matrix of the noise at the FC be denoted by $\mathbf{C}_v$. Then, recalling that $\mathbf{H} = \mathbf{G}\boldsymbol{\Gamma}^{1/2}$, as $N \to \infty$, then using ([40], Theorem 3.4) and ([40], Theorem 3.7) we can show that

$$\mathbf{H}^H \mathbf{C}_v^{-1} \mathbf{H} \overset{a}{=} \text{Tr}\left\{\mathbf{C}_v^{-1}\right\} \boldsymbol{\Gamma}, \qquad (28)$$

where $\boldsymbol{\Gamma}$ is a diagonal matrix containing the attenuation coefficients γ_l (cf. Section II).

It is interesting to investigate analytically how noise correlation at the FC, and in particular $\mathbf{C}_v$, affects the performance. Let us first consider the power optimization problem subject to an MSE constraint. Given the MSE expression in (5), and applying the matrix inversion lemma on $\mathbf{C}_w$, we can then exploit (28) in order to find the asymptotic expression (in N) for the MSE. Next, we use a variable change $\widetilde{\lambda}_l \triangleq \text{Tr}\{\mathbf{C}_v^{-1}\}|\alpha_l|^2 \gamma_l$, which yields the following optimization problem for $\widetilde{\lambda}_l, \forall l \in \{1, \ldots, L\}$,

$$\underset{\{\widetilde{\lambda}_l \geq 0\}_{l=1}^L}{\text{minimize}} \quad \sum_{l=1}^L \frac{1}{\text{Tr}\{\mathbf{C}_v^{-1}\}} \frac{\widetilde{\lambda}_l}{\gamma_l}[\mathbf{C}_\theta + \mathbf{C}_n]_{ll}$$

$$\text{subject to} \quad \text{Tr}\left\{ \left(\mathbf{C}_\theta^{-1} + \widetilde{\boldsymbol{\Lambda}} - \widetilde{\boldsymbol{\Lambda}}(\mathbf{C}_n^{-1} + \widetilde{\boldsymbol{\Lambda}})^{-1}\widetilde{\boldsymbol{\Lambda}} \right)^{-1} \right\} \leq \bar{d}, \qquad (29)$$

where $\widetilde{\boldsymbol{\Lambda}} \triangleq \text{diag}(\widetilde{\lambda}_1 \ldots, \widetilde{\lambda}_L)$.

Note that the constant $\frac{1}{\text{Tr}\{\mathbf{C}_v^{-1}\}}$ can be pulled out of the optimization problem since it is independent of the variables λ_l. The solution to the resulting optimization problem is therefore independent of the covariance matrix $\mathbf{C}_v$ due to the same argument as used in Remark 3. This leads to the following important conclusion:

Remark 7: Through (29) one can compare the total optimal power associated with the correlated noise scenario and that associated with the uncorrelated noise scenario. For this purpose, if one assumes that in the correlated case the diagonal elements of the matrix remain the same but the off-diagonal elements become non-zero (this type of correlation structure is used for example in the exponential model [41]), then based on ([42], theorem 1.2), it can be shown that $\text{Tr}\{\mathbf{C}_v^{-1}\} > \sum_{i=1}^N 1/[\mathbf{C}_v]_{ii}$ as long as $\mathbf{C}_v$ is positive definite. Hence, by studying the objective function in (29), it can be observed that the total optimal power P_{tot} decreases in the presence of noise correlation at the FC. Similar analysis can be carried out to study the impact of noise

correlation on MSE. In this case, it is straightforward to see that, using the same change of variable to $\widetilde{\lambda}_l$, the available total power budget $\bar{P}$ will be transformed into $\mathrm{Tr}\{\mathbf{C}_v^{-1}\}\bar{P}$ which affects the optimal power allocations, and as a result, noise correlation would improve the optimal MSE as well.

B. Generic Channel Model

Now, we study a generic MIMO channel. We let $\mathbf{G} = [\mathbf{g}_1 \ \mathbf{g}_2 \ \dots \ \mathbf{g}_L]$, in the formulation introduced in (2), be a random matrix with columns $\mathbf{g}_l = \mathbf{U}_l \boldsymbol{\Sigma}_l^{1/2} \widetilde{\mathbf{g}}_l \in \mathbb{C}^{N \times 1}$, where, $\forall l$, $\mathbf{U}_l$ is a $N \times r$ (with $N \geq r$) unitary eigenvector matrix, and $\boldsymbol{\Sigma}_l$ is a $r \times r$ diagonal eigenvalue matrix including r non-zero eigenvalues. Further, the entries of $\widetilde{\mathbf{g}}_l \in \mathbb{C}^{r \times 1}$ are i.i.d. random variables drawn from $\mathcal{CN}(0, 1)$. Hence, we denote the covariance matrix associated with the lth column of $\mathbf{G}$ by $\mathbf{C}_l \triangleq \mathbb{E}[\mathbf{g}_l \mathbf{g}_l^H] = \mathbf{U}_l \boldsymbol{\Sigma}_l \mathbf{U}_l^H \in \mathbb{C}^{N \times N}$. This correlation model is used for example in [43], [44]. We also note that the transmit-side (sensor-side) channel correlation, or correlation among the rows of $\mathbf{G}$ is reasonably neglected due to the favorable propagation condition. That is to say, the transmit-side channels are decorrelated using reasonably large antenna arrays at the FC (see, e.g., [45], [46]) and sufficient spatial sensor separation.

We will make the assumption, as in [45], [47], that $\mathrm{Tr}\{\mathbf{C}_l\}$ scales with N. However, $\mathbf{C}_l$ can have any arbitrary structure, thus accommodating generic spatial propagation environment and arbitrary geometry. This assumption reflects that an increase in the number of antennas N corresponds to an increased number of degrees of freedom (i.e., richness of scattering), as a consequence of improved spatial resolution and array aperture. Clearly, the aforementioned condition implies that the rank of $\mathbf{C}_l$ should grow linearly with N. Of course, this rank scaling law includes both the cases of full rank ($r = N$) and rank-deficient channels ($r = \rho N$, where $0 < \rho < 1$), where the latter case might happen due to insufficient richness of scattering for large arrays [45].

In the case of a generic MIMO channel, we can show the following asymptotic result using the results in ([40], Theorem 3.4) and ([40], Theorem 3.7). Let $\mathbf{C}_l$ denote the channel covariance matrix between the lth sensor and the FC with N antennas. Then, provided $\mathrm{Tr}\{\mathbf{C}_l\}$ scales with N, as $N \to \infty$, it follows that

$$\mathbf{H}^H \mathbf{H} \overset{a}{=} \mathrm{diag}(\mathrm{Tr}\{\mathbf{C}_1\}, \mathrm{Tr}\{\mathbf{C}_2\}, \dots, \mathrm{Tr}\{\mathbf{C}_L\})\boldsymbol{\Gamma}. \quad (30)$$

Remark 8: In the correlated channel scenario, one can consider $\mathrm{diag}(\mathrm{Tr}\{\mathbf{C}_1\}, \mathrm{Tr}\{\mathbf{C}_2\}, \dots, \mathrm{Tr}\{\mathbf{C}_L\})\boldsymbol{\Gamma}$ as an asymptotic (in N) approximation for $\mathbf{H}^H \mathbf{H}$ in the MSE expression (5). Therefore, since $\mathbf{H}^H \mathbf{H}$ is diagonalized, similar analysis to the uncorrelated channel case can be carried out to derive optimal sensor power allocations. For instance, by comparing (30) with (7), the closed-form optimal power allocations in (17), (19), (23) and (26) are modified by multiplying the coefficients γ_l, by $\frac{1}{N}\mathrm{Tr}\{\mathbf{C}_l\}, \forall l$.

Let us now consider a special case, where $\forall l, \mathbf{C}_l = \mathbf{C}_g$, i.e., all channel covariance matrices are assumed to be equivalent. Using a change of variable $\widetilde{\lambda}_l \triangleq \mathrm{Tr}\{\mathbf{C}_g\}|\alpha_l|^2 \gamma_l / \sigma_v^2$ and

defining $\widetilde{\boldsymbol{\Lambda}} \triangleq \mathrm{diag}(\widetilde{\lambda}_1, \dots, \widetilde{\lambda}_L)$, carried out in (29), the optimization problem (P1) becomes

$$\underset{\{\widetilde{\lambda}_l \geq 0\}_{l=1}^L}{\text{minimize}} \quad \sum_{l=1}^L \frac{\sigma_v^2}{\mathrm{Tr}\{\mathbf{C}_g\}} \frac{\widetilde{\lambda}_l}{\gamma_l}[\mathbf{C}_\theta + \mathbf{C}_n]_{ll}$$

$$\text{subject to} \quad \mathrm{Tr}\left\{\left(\mathbf{C}_\theta^{-1} + \widetilde{\boldsymbol{\Lambda}} - \widetilde{\boldsymbol{\Lambda}}(\mathbf{C}_n^{-1} + \widetilde{\boldsymbol{\Lambda}})^{-1}\widetilde{\boldsymbol{\Lambda}}\right)^{-1}\right\} \leq \bar{d},$$

$$\quad (31)$$

which indicates that the optimization can be solved regardless of the covariance matrix $\mathbf{C}_g$, however, the value at the optimal point would be inversely proportional to $\mathrm{Tr}\{\mathbf{C}_g\}$.

C. Generic Channel and Noise

In the case that both channel and noise correlations are present, the above results can be constructed to form the following general result which can be followed by using ([40], Theorem 3.4) and ([40], Theorem 3.7). Let $\mathbf{C}_l$ denote the channel covariance matrix between the lth sensor and the FC with N antennas, and let $\mathbf{C}_v$ be the noise covariance matrix at the FC. Then, provided $\mathrm{Tr}\{\mathbf{C}_l\}$ scales with N, as $N \to \infty$, it follows that

$$\mathbf{H}^H \mathbf{C}_v^{-1} \mathbf{H} \overset{a}{=} \mathrm{diag}\left(\mathrm{Tr}\left\{\mathbf{C}_v^{-1}\mathbf{C}_1\right\}, \dots, \mathrm{Tr}\left\{\mathbf{C}_v^{-1}\mathbf{C}_L\right\}\right)\boldsymbol{\Gamma}. \quad (32)$$

D. Imperfect Channel Estimation

In practice, the knowledge of the state of the communication channel is very crucial in order to obtain promising performance gains of MIMO communication systems. For this purpose, channel gains need to be estimated, using, e.g., training sequences/pilot signals at the receiver or transmitter and fed back to where this information is needed. The channel gains are, therefore, subject to errors or imprecision, such as estimation error, noise or quantization error, which need to be considered throughout the design procedure.

Let us assume that the estimated channel matrix be modeled as $\widehat{\mathbf{H}} = \mathbf{H} + \mathbf{E}$, where $\mathbf{H}$ is the nominal channel matrix, and $\mathbf{E} = [\sqrt{\epsilon_1}\mathbf{e}_1^\top \ \sqrt{\epsilon_2}\mathbf{e}_2^\top \ \dots \ \sqrt{\epsilon_N}\mathbf{e}_N^\top]^\top \in \mathbb{C}^{N \times L}$ be the error matrix, where $\mathbf{e}_i \in \mathbb{C}^{1 \times L}$ $(i = 1, \dots, N)$ is a vector accounts for the error from the channels between the L sensors and the ith antenna at the FC. Furthermore, we assume that the error matrix $\mathbf{E}$ is independent of $\mathbf{H}$, and each element in $\mathbf{e}_i$ is drawn from $\mathcal{CN}(0, 1)$. Thus, $\epsilon_i \geq 0$ $(i = 1, \dots, N)$ is a constant that denotes the variance of the error vector $\mathbf{e}_i$.

In this scenario, we need to adjust the MSE in (6) into the case where channel estimation error exists in order to find the optimal power allocations. We have

$$\widehat{\mathbf{H}}^H \widehat{\mathbf{H}} = (\mathbf{H} + \mathbf{E})^H (\mathbf{H} + \mathbf{E}) \overset{a}{=} N\boldsymbol{\Gamma} + \left(\sum_{i=1}^N \epsilon_i\right)\mathbf{I}_L, \quad (33)$$

where the last asymptotic equality follows by the fact that $\mathbf{E}^H \mathbf{E} = \widehat{\mathbf{E}}^H \mathrm{diag}(\epsilon_1, \dots, \epsilon_N)\widehat{\mathbf{E}}$, where $\widehat{\mathbf{E}} \triangleq [\mathbf{e}_1^\top \ \dots \ \mathbf{e}_N^\top]^\top$, and from the favorable propagation condition, as $N \to \infty$.

Letting $\widehat{\lambda}_l \triangleq (\gamma_l + \frac{1}{N}\sum_{i=1}^N \epsilon_i)|\widehat{\alpha}_l|^2$, where $\widehat{\alpha}_l$ $(l \in \{1, \dots, L\})$ denotes the lth sensor amplification factor in the case of imperfect channel estimates, and defining

$\widehat{\mathbf{\Lambda}} \triangleq \mathrm{diag}(\widehat{\lambda}_1, \ldots, \widehat{\lambda}_L)$, the MSE in (6) can be asymptotically approximated as

$$\mathrm{MSE} \stackrel{a}{=} \mathrm{Tr}\Bigg\{\Big(\mathbf{C}_\theta^{-1} + N\sigma_v^{-2}\widehat{\mathbf{\Lambda}}$$
$$-N^2\sigma_v^{-4}\widehat{\mathbf{\Lambda}}\Big(\mathbf{C}_n^{-1} + N\sigma_v^{-2}\widehat{\mathbf{\Lambda}}\Big)^{-1}\widehat{\mathbf{\Lambda}}\Big)^{-1}\Bigg\}. \quad (34)$$

Moreover, the total power, considering channel estimation error, is stated as

$$P_{\mathrm{tot}} = \sum_{l=1}^{L} \frac{\widehat{\lambda}_l[\mathbf{C}_\theta + \mathbf{C}_n]_{ll}}{\gamma_l + \frac{1}{N}\sum_{i=1}^{N}\epsilon_i}. \quad (35)$$

Using the MSE and total power formulations in (34) and (35), respectively, the design procedure for optimal power allocation can be pursued in a similar manner to those in Section IV. We finalize this section with the following remark.

Remark 9: Note that when ϵ_i ($i = 1, \ldots, N$) is small, then $\gamma_l + \frac{1}{N}\sum_{i=1}^{N}\epsilon_i \approx \gamma_l$. Hence, the optimal power allocation in the case of imperfect channel estimation would be very close to that of the perfect channel estimation case. On the other hand, when ϵ_i is large, such that γ_l gets dominated by $\frac{1}{N}\sum_{i=1}^{N}\epsilon_i$, then the optimal $\widehat{\lambda}_l^\star$ incurred by minimizing the total power subject to MSE constraint would not depend on ϵ_i anymore (since $\frac{1}{N}\sum_{i=1}^{N}\epsilon_i$ is independent of the index l and can be pulled out of the objective function). As a result, the total *nominal* power consumed by sensors would not depend on ϵ_i, and the performance saturates as ϵ_i increases.

It is also possible to show that the optimal total power, under the MSE constraint, would decrease proportionally with $1/N$. First note that if ϵ_i ($i = 1, \ldots, N$) is a random variable with a finite mean, then, for sufficiently large N, $\frac{1}{N}\sum_{i=1}^{N}\epsilon_i \to \mathbb{E}[\epsilon_i]$. This shows that the denominator in (35) becomes a constant (i.e., independent of N), and as a result the optimal total power decreases proportionally with $1/N$. Analogously, if ϵ_i is deterministic and uniformly bounded (i.e., $0 \le \epsilon_i \le \epsilon_{max}$), then it clearly follows that $\frac{1}{N}\sum_{i=1}^{N}\epsilon_i = \mathcal{O}(1)$. This shows that the optimal total power decreases proportional with $1/N$ in this case also. Later, in Section VI, we compare the scaling laws (in $1/N$) for the two scenarios of perfect and imperfect channel estimation via numerical simulations.

VI. NUMERICAL SIMULATIONS

In this section, we quantify the performance of the proposed optimization methods. We basically evaluate the total power, by solving the power allocation problem (P1), or MSE, by solving the power allocation problem (P2).

A. Numerical Setups

In all simulation studies, we assume $L = 15$ number of sensors. We also assume a homogenous scenario, where $\forall l \in \{1, \ldots, L\}$, the source-to-sensor noise variances $\sigma_{n_l}^2 = 10^{-3}$ W. Further, $\sigma_v^2 = 10^{-6}$ W. Unless otherwise stated, we assume that the source is uncorrelated, and the variance of source entries is chosen the same and equal to $\sigma_{\theta_l}^2 = 1$. We also set the pathloss exponent to $2\beta = 2$ (a free-space propagation scenario), and sensor-to-FC distances d_l are uniformly distributed ranging from 20 to 70 meters.
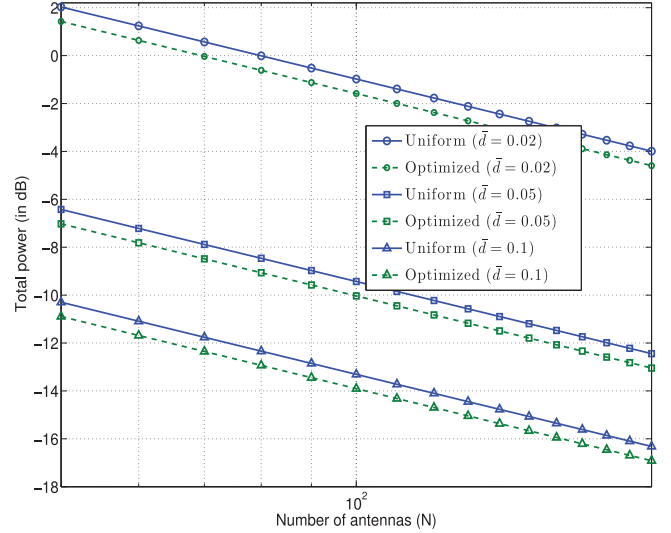

Fig. 3. Total power consumed by sensors as a function of number of antennas N for different values of distortion threshold $\bar{d}$ using optimized design and uniform power allocation.

B. Numerical Results

In Fig. 3, we plot total power consumed by sensors as a function of number of antennas N (varying from 50 to 200 at a step size of 10) using the optimized power allocation (19) and uniform power allocation, for varying distortion threshold $\bar{d}$. For the uniform power allocation, we assume that all sensor nodes consume equal power, i.e., $|\alpha_l|^2[\mathbf{C}_\theta + \mathbf{C}_n]_{ll}$ are the same $\forall l$. It can be observed that the total power in dB-log scale decays linearly in N for both methods. In all setups, the optimal power allocation outperforms the uniform power allocation.

Next, in Fig. 4, we plot the MSE as a function of number of antennas N using the optimized power allocation (26) and uniform power allocation. As observed, the optimal power allocation gives a lower MSE compared to the uniform power allocation. Also, as we provide a higher power to sensors (corresponding to higher $\bar{P}$), the resulting power allocation leads to a lower MSE. We note that as $N \to \infty$, all the curves, irrespective of optimality and power levels $\bar{P}$, approach to an MSE floor of 0.015 which is determined from (27).

Now, we study the correlated source case. We consider the exponential covariance matrix model [41] for the source, where each entry at row i and column j of the source covariance matrix $\mathbf{C}_\theta$ is chosen as $\rho_\theta^{|i-j|}$ in which $0 \le \rho_\theta \le 1$ is known as correlation coefficient. In Fig. 5, MSE is plotted as a function of number of antennas N for different values of the correlation coefficient ρ_θ and for fixed $\bar{P} = 0.1\,W$. The curves in Fig. 5 are associated with the exact analysis (by solving (P2) using CVX solver [39] or equivalently by solving the KKT conditions in (22)) shown by solid line, and the approximate analysis (23) shown by dashed line, respectively. We observe that the approximate solution is tight in all numerical setups. As can be also expected, higher source correlation leads to lower MSE. This is due to the fact that each source entry consists of information from other entries as a result of correlation which yields to a more accurate estimation. We note that each curve, associated with a different ρ_θ, approaches to an MSE floor determined by (24). However, since in this setting $\mathbf{C}_\theta^{-1}$ is dominated by $\mathbf{C}_n^{-1}$, the curves reach to a more or less the same MSE floor of 0.015 as $N \to \infty$.
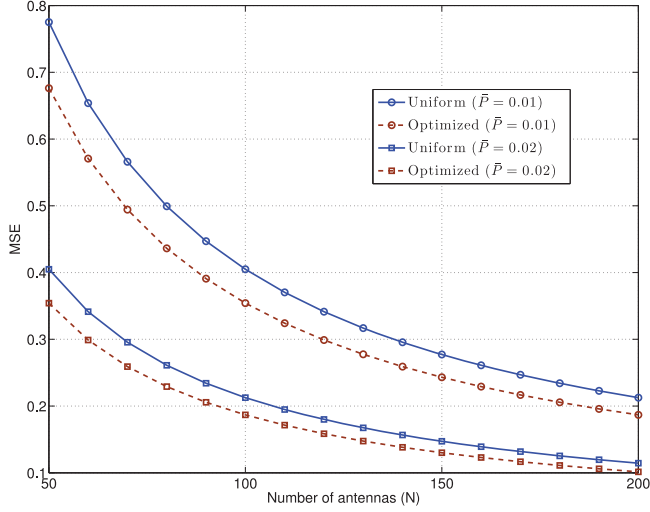
Fig. 4. MSE as a function of number of antennas N at the FC for different values of power threshold $\bar{P}$ using optimized design and uniform power allocation.
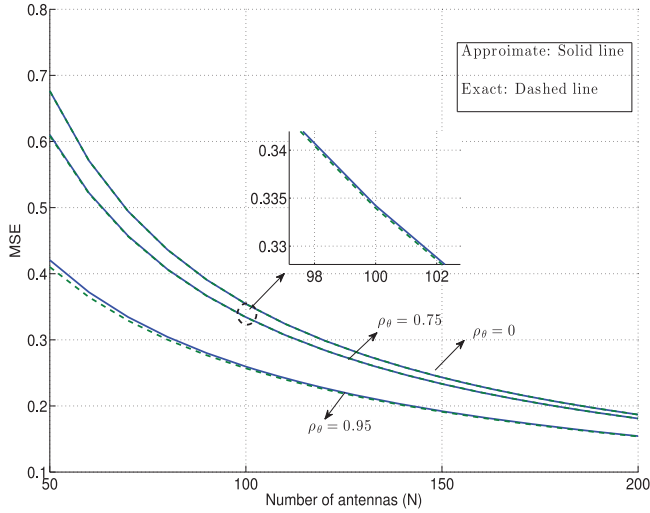

Fig. 5. MSE as a function of number of antennas N at the FC for different values of correlation parameter ρ_θ.
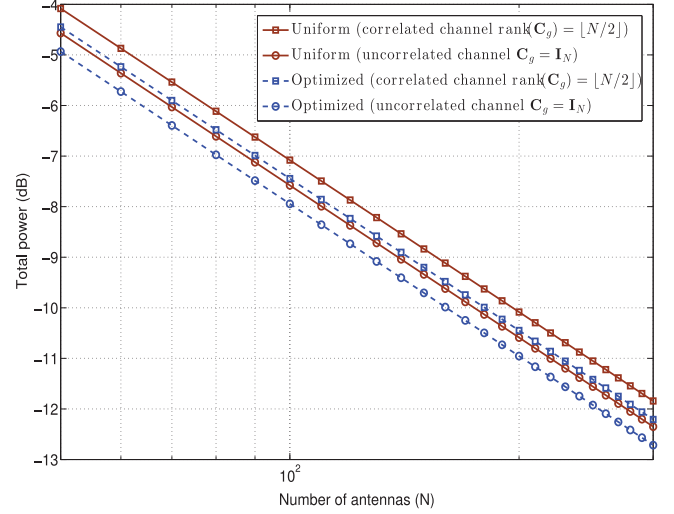

Fig. 6. Total power consumption as a function of number of antennas N at the FC for different channel correlation scenarios using optimized and uniform power allocations.
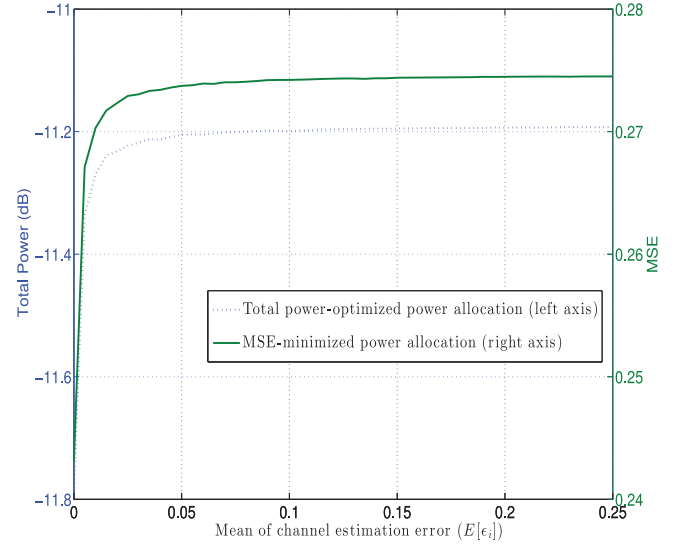

Fig. 7. MSE and total power consumption as a function of mean channel estimation error $\mathbb{E}[\epsilon_i]$.

In Fig. 6, we demonstrate the effect of channel correlation on total power as a function of N using optimized (by solving (31)) and uniform power allocations. We set $\bar{d} = 0.05$, and also assume that the channel covariance matrices are the same, i.e., $\mathbf{C}_l \triangleq \mathbf{C}_g, \forall l$. We study two cases (as discussed in Section V-B):

1) Correlated channel, where $r = \text{rank}(\mathbf{C}_g) = \lfloor N/2 \rfloor$. We generate the covariance matrix as $\mathbf{C}_g = \mathbf{U}\boldsymbol{\Sigma}\mathbf{U}^H$, where $\boldsymbol{\Sigma}$ is a $r \times r$ diagonal matrix whose diagonal elements are chosen as eigenvalues of the exponential correlation model with correlation coefficient 0.7, and the columns of $\mathbf{U}$ as r eigenvectors associated with the eigenvalues.

2) Uncorrelated channel, where $\mathbf{C}_g = \mathbf{I}_N$.

We observe from the curves in Fig. 6 that the channel correlation degrades the performance as discussed earlier in Section V-B for both power allocation methods. It can be easily verified that, in this setting, if $\text{rank}(\mathbf{C}_g) = \frac{1}{c}N$ for some constant $c > 1$, then the total power increases by a factor c (cf. (31)).

In the concept of MIMO channel capacity, it has been also shown in [48] that a correlated channel provides a smaller ca-

pacity due to less degrees of freedom with respect to uncorrelated channels. It should be also mentioned if the covariance matrix is full-rank, i.e., $r = N$, and generated under the exponential model, then the performance would be equivalent to the uncorrelated case since $\text{Tr}\{\mathbf{C}_g\} = N$.

In our final experiments, we evaluate the performance of the optimized design when channel estimation is erroneous. In order to study how the channel estimation error affects the performance, we solve the optimization problems corresponding to the design criteria (34) and (35), and then insert their solutions (including the error variance ϵ_i) into the original design criteria (8) and (9). In Fig. 7, using a double y-axis, we plot the effect of mean of error variance $\mathbb{E}[\epsilon_i]$ $(i = 1, \ldots, N)$ both on the total power and on the MSE. We randomly generate $\epsilon_i, \forall i$, according to a non-negative uniform distribution whose mean varies from 0 (no estimation error) to 0.25 with a step size 0.01. We also set $N = 150, \bar{P} = 0.01$ W, and $\bar{d} = 0.05$. As can be seen from Fig. 7, total power consumption and MSE are quite sensi-
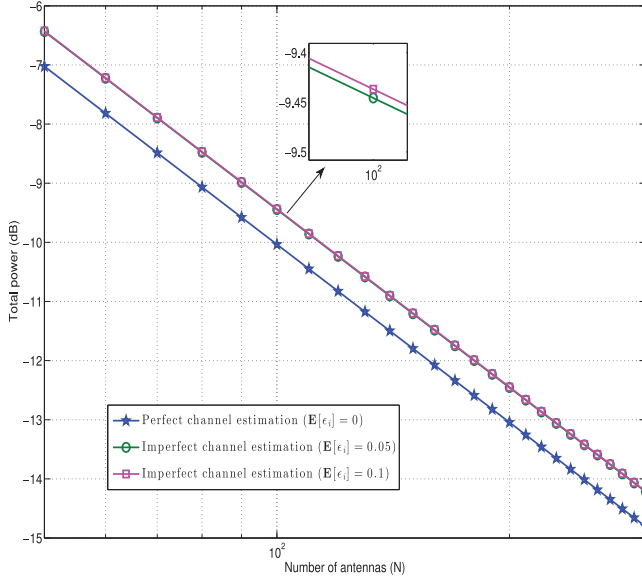
Fig. 8. Total power consumption as a function number of antennas N for perfect and imperfect channel estimation scenarios.

tive to the channel estimation error, and by increasing ϵ, power consumption increases and MSE degrades. Note that the curve corresponding to the total-power minimization problem (dotted line) finally saturates due to the reasons discussed in Remark 9.

Finally, in Fig. 8, we compare the scaling laws for the optimal total power consumption as N increases in three cases: when channel estimation is perfect, i.e., $\epsilon_i = 0$, and when channel estimation is imperfect, where we set $\mathbb{E}[\epsilon_i] = 0.05$ and $\mathbb{E}[\epsilon_i] = 0.1$. The remaining simulation setups are kept as the same as those of the previous study. As can be seen, in these cases, the total power (as discussed in Remark 9) decays proportional with $1/N$. However, in the imperfect channel estimation scenario, the total power consumption would increase by almost 0.5 dB compared to the perfect channel estimation scenario. The small gap between the two cases $\mathbb{E}[\epsilon_i] = 0.05$ and $\mathbb{E}[\epsilon_i] = 0.1$, as observed in Fig. 8, can be also explained from the saturation of the "total power" curves in Fig. 7.

VII. CONCLUSION

We have studied a decentralized multi-sensor estimation estimation problem, where L sensor nodes amplify their observations and forward them over fading MACs to a FC equipped with large arrays of antennas, denoted by $N \gg L$. Under this condition, we were able to obtain analytical results for the MSE under the favorable propagation condition. We have optimized the amplification factor at each sensor node with the objectives of minimizing total power consumption or MSE for source estimation, under a maximum distortion or total power constraint, respectively. Our analysis revealed that as N increases, the optimal total power consumption of the sensors decay with a factor proportional to $1/N$ in order to satisfy a targeted MSE. Numerical studies have demonstrated the efficiency of the proposed optimal power allocation compared to uniform power allocation. We have also investigated practical scenarios, where channel gains or additive channel noise are correlated, and further, where channel gains are subject to estimation error.

For instance, we analytically showed that under rank-deficient channel correlation the performance cannot be improved, and numerically quantified such performance degradation. In these cases, we have numerically illustrated that the MSE and power are adversely affected by increasing correlation and channel estimation error.

APPENDIX A
PROOF OF THEOREM 1

It can be easily shown that the objective and the constraint in (P3) are both convex, therefore, it can be solved by the standard KKT condition technique. Introducing the Lagrange multiplier $\mu \geq 0$, the Lagrangian can be written as

$$\mathcal{L}(\{\lambda_l\}_{l=1}^L, \mu) = \sum_{l=1}^L \lambda_l \frac{1}{\gamma_l} [\mathbf{C}_\theta + \mathbf{C}_n]_{ll}$$
$$+ \mu \left(\sum_{l=1}^L [\mathbf{Q}]_{ll} + \frac{1}{N} \sum_{l=1}^L \frac{\sigma_v^2}{\lambda_l \sigma_{n_l}^4} [\mathbf{Q}^2]_{ll} - \bar{d} \right).$$

Taking the derivative of the Lagrangian with respect to $\lambda_l > 0$, and letting it equal zero, we have

$$\lambda_l = \frac{\sqrt{\frac{\mu}{N}} \frac{\sigma_v}{\sigma_{n_l}^2} [\mathbf{Q}^2]_{ll}^{1/2}}{[\mathbf{C}_\theta + \mathbf{C}_n]_{ll}^{1/2} / \sqrt{\gamma_l}}. \qquad (36)$$

Then, plugging (36) back into the constraint in (P3), the Lagrange multiplier μ is determined as

$$\mu = \frac{1}{N} \left(\frac{\sum_l \frac{\sigma_v^2 \sqrt{\gamma_l}}{\sigma_{n_l}^2} [\mathbf{Q}^2]_{ll}^{1/2} [\mathbf{C}_\theta + \mathbf{C}_n]_{ll}^{1/2}}{\bar{d} - \sum_l [\mathbf{Q}]_{ll}} \right)^2. \qquad (37)$$

Finally, inserting (37) back into (36), the optimal $\lambda_l^\star$ is given by (17).

APPENDIX B
PROOF OF THEOREM 2

Using the KKT condition technique, and introducing the Lagrange multiplier $\mu \geq 0$, the Lagrangian (by neglecting the constraints $\lambda_l \geq 0$ for the moment) can be written as

$$\mathcal{L}(\{\lambda_l\}_{l=1}^L, \mu) = \sum_{l=1}^L \lambda_l \frac{1}{\gamma_l} \left(\sigma_{\theta_l}^2 + \sigma_{n_l}^2 \right)$$
$$+ \mu \left(\sum_{l=1}^L \frac{\sigma_{\theta_l}^2 \sigma_{n_l}^2 \lambda_l + \frac{\sigma_{\theta_l}^2 \sigma_v^2}{N}}{(\sigma_{\theta_l}^2 + \sigma_{n_l}^2)\lambda_l + \frac{\sigma_v^2}{N}} - \bar{d} \right).$$

Taking the first derivative of the Lagrangian $\mathcal{L}(\{\lambda_l\}_{l=1}^L, \mu)$ with respect to λ_l, and letting it equal zero, we obtain (also noting that $\lambda_l \geq 0$)

$$\lambda_l = \left[\frac{\sqrt{\frac{\mu}{N}} \sigma_{\theta_l}^2 \sigma_v \sqrt{\gamma_l}}{(\sigma_{\theta_l}^2 + \sigma_{n_l}^2)^{3/2}} - \frac{\sigma_v^2}{N(\sigma_{\theta_l}^2 + \sigma_{n_l}^2)} \right]^+. \qquad (38)$$

Now, we define $z_l \triangleq \frac{\sigma_{\theta_l}^2 \sqrt{\gamma_l}}{(\sigma_{\theta_l}^2 + \sigma_{n_l}^2)^{1/2}} - \frac{\sigma_v}{\sqrt{N\mu}}$ such that $\lambda_l = \frac{\sigma_v \sqrt{\mu/N}}{\sigma_{\theta_l}^2 + \sigma_{n_l}^2}[z_l]^+$. Hence, since $z_1 \geq z_2 \geq \cdots \geq z_L$ (based on the statement in Theorem 2), we assume that the first $M^\star$ values of $z_l, l = 1, \ldots, L$, are positive, and the remaining $L - M^\star$ values are zero, where $M^\star \leq L$. Due to the above assumption, $M^\star$ is unique, since the threshold $\frac{\sigma_v}{\sqrt{N\mu}}$ for comparing z_l only depends on N, σ_v and μ, which are constant for a given system. Plugging (38) back into the constraint in (P4), we determine μ, and inserting back into (38) yields the optimal values $\lambda_l^\star$ in (19). Now, in order to determine $M^\star$, we let it be equal to the maximum integer in the interval $[1, L]$ such that $z_l > 0$, which yields (18).

References

[1] A. Shirazinia, S. Dey, D. Ciuonzo, and P. Salvo Rossi, "Massive MIMO for decentralized estimation over coherent multiple access channels," in *Proc. IEEE Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, Jun. 2015, pp. 241–245.

[2] A. Ribeiro and G. Giannakis, "Bandwidth-constrained distributed estimation for wireless sensor networks—Part I: Gaussian case," *IEEE Trans. Signal Process.*, vol. 54, no. 3, pp. 1131–1143, Mar. 2006.

[3] I. Schizas, G. Giannakis, and Z.-Q. Luo, "Distributed estimation using reduced-dimensionality sensor observations," *IEEE Trans. Signal Process.*, vol. 55, no. 8, pp. 4284–4299, Aug. 2007.

[4] J.-J. Xiao and Z.-Q. Luo, "Decentralized estimation in an inhomogeneous sensing environment," *IEEE Trans. Inf. Theory*, vol. 51, no. 10, pp. 3564–3575, Oct. 2005.

[5] T. Wimalajeewa and S. Jayaweera, "Optimal power scheduling for correlated data fusion in wireless sensor networks via constrained PSO," *IEEE Trans. Wireless Commun.*, vol. 7, no. 9, pp. 3608–3618, Sept. 2008.

[6] M. 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.

[7] M. Banavar, A. 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.

[8] M. Goldenbaum and S. Stanczak, "On multiantenna sensor networks with interference: Energy consumption vs. robustness," in *Proc. Int. ITG Workshop Smart Antennas*, Mar. 2012, pp. 125–132.

[9] X. Zhang, H. Poor, and M. Chiang, "Optimal power allocation for distributed detection over MIMO channels in wireless sensor networks," *IEEE Trans. Signal Process.*, vol. 56, no. 9, pp. 4124–4140, Sep. 2008.

[10] M. Banavar, A. Smith, C. Tepedelenlioglu, and A. Spanias, "Distributed detection over fading MACs with multiple antennas at the fusion center," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Mar. 2010, pp. 2894–2897.

[11] J.-J. Xiao, S. Cui, Z.-Q. Luo, and A. Goldsmith, "Linear coherent decentralized estimation," *IEEE Trans. Signal Process.*, vol. 56, no. 2, pp. 757–770, Feb. 2008.

[12] A. Behbahani, A. Eltawil, and H. Jafarkhani, "Decentralized estimation under correlated noise," *IEEE Trans. Signal Process.*, vol. 62, no. 21, pp. 5603–5614, Nov. 2014.

[13] S. Cui, J.-J. Xiao, A. Goldsmith, Z.-Q. Luo, and H. Poor, "Estimation diversity and energy efficiency in distributed sensing," *IEEE Trans. Signal Process.*, vol. 55, no. 9, pp. 4683–4695, Sep. 2007.

[14] I. Bahceci and A. Khandani, "Linear estimation of correlated data in wireless sensor networks with optimum power allocation and analog modulation," *IEEE Trans. Commun.*, vol. 56, no. 7, pp. 1146–1156, Jul. 2008.

[15] 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.

[16] M. Gastpar and M. Vetterli, "Source-channel communication in sensor networks," in *Lect. Notes in Comput. Science*. New York, NY, USA: Springer, 2003, pp. 162–177.

[17] M. Gastpar, B. Rimoldi, and M. Vetterli, "To code, or not to code: Lossy source-channel communication revisited," *IEEE Trans. Inf. Theory*, vol. 49, no. 5, pp. 1147–1158, May 2003.

[18] A. Smith, M. Banavar, C. Tepedelenlioglu, and A. Spanias, "Distributed estimation over fading MACs with multiple antennas at the fusion center," in *Proc. Asilomar Conf. Signals, Syst., Comput.*, Nov. 2009, pp. 424–428.

[19] T. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590–3600, Nov. 2010.

[20] H. Q. Ngo, E. Larsson, and T. Marzetta, "Energy and spectral efficiency of very large multiuser MIMO systems," *IEEE Trans. Commun.*, vol. 61, no. 4, pp. 1436–1449, Apr. 2013.

[21] E. Larsson, O. Edfors, F. Tufvesson, and T. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 186–195, Feb. 2014.

[22] E. Björnson, M. Kountouris, and M. Débbah, "Massive MIMO and small cells: Improving energy efficiency by optimal soft-cell coordination," in *Proc. 20th Int. Conf. Telecommun. (ICT)*, May 2013, pp. 1–5.

[23] H. Q. Ngo, E. Larsson, and T. Marzetta, "Energy and spectral efficiency of very large multiuser MIMO systems," *IEEE Trans. Commun.*, vol. 61, no. 4, pp. 1436–1449, Apr. 2013.

[24] J. Vieira, S. Malkowsky, K. Nieman, Z. Miers, N. Kundargi, L. Liu, I. Wong, V. Owall, O. Edfors, and F. Tufvesson, "A flexible 100-antenna testbed for massive MIMO," in *Proc. IEEE Globecom Workshops*, Dec. 2014, pp. 287–293.

[25] E. Björnson, E. G. Larsson, and T. L. Marzetta, "Massive MIMO: 10 myths and one grand question," *IEEE Commun. Mag.*, 2015, submitted for publication.

[26] N. Shariati, E. Björnson, M. Bengtsson, and M. Debbah, "Low-complexity polynomial channel estimation in large-scale MIMO with arbitrary statistics," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 815–830, Oct. 2014.

[27] L. Lu, G. Li, A. Swindlehurst, A. Ashikhmin, and R. Zhang, "An overview of massive MIMO: Benefits and challenges," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 742–758, Oct. 2014.

[28] F. Jiang, J. Chen, and A. Swindlehurst, "Phase-only analog encoding for a multi-antenna fusion center," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Mar. 2012, pp. 2645–2648.

[29] F. Jiang, J. Chen, A. Swindlehurst, and J. Lopez-Salcedo, "Massive MIMO for wireless sensing with a coherent multiple access channel," *IEEE Trans. Signal Process.*, vol. 63, no. 12, pp. 3005–3017, Jun. 2015.

[30] D. Ciuonzo, P. Salvo Rossi, and S. Dey, "Massive MIMO channel-aware decision fusion," *IEEE Trans. Signal Process.*, vol. 63, no. 3, pp. 604–619, Feb. 2015.

[31] A. Savvides, C.-C. Han, and M. B. Strivastava, "Dynamic fine-grained localization in Ad-Hoc networks of sensors," in *Proc. Annu. ACM Int. Conf. Mobile Comput. Netw. (MobiCom)*, 2001, pp. 166–179.

[32] X. Cheng, A. Thaeler, G. Xue, and D. Chen, "TPS: A time-based positioning scheme for outdoor wireless sensor networks," in *Proc. IEEE INFOCOM*, Mar. 2004, vol. 4, pp. 2685–2696.

[33] H. Chen, G. Wang, Z. Wang, H. So, and H. Poor, "Non-line-of-sight node localization based on semi-definite programming in wireless sensor networks," *IEEE Trans. Wireless Commun.*, vol. 11, no. 1, pp. 108–116, Jan. 2012.

[34] U. Madhow, D. Brown, S. Dasgupta, and R. Mudumbai, "Distributed massive MIMO: Algorithms, architectures and concept systems," in *Proc. Inf. Theory Appl. Workshop (ITA)*, Feb. 2014, pp. 1–7.

[35] S. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1993.

[36] P. Smith, C. Neil, M. Shafi, and P. Dmochowski, "On the convergence of massive MIMO systems," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Jun. 2014, pp. 5191–5196.

[37] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[38] L. Vandenberghe and S. Boyd, "Semidefinite programming," *SIAM Rev.*, vol. 38, pp. 49–95, 1994.

[39] M. Grant and S. Boyd, CVX: Matlab software for disciplined convex programming. ver. 1.21, Apr. 2011 [Online]. Available: http://cvxr.com/

[40] R. Couillet and M. Debbah, *Random Matrix Methods for Wireless Communications*. Cambridge, U.K.: Cambridge Univ. Press, 2011.

[41] S. Loyka, "Channel capacity of MIMO architecture using the exponential correlation matrix," *IEEE Commun. Lett.*, vol. 5, no. 9, pp. 369–371, Sep. 2001.

[42] M. Fiedler, "Old and new about positive definite matrices," *J. Linear Algebra Appl.*, vol. 484, pp. 496–503, Nov. 2015.

[43] J. Hoydis, S. T. Brink, and M. Debbah, "Massive MIMO: How many antennas do we need?," in *Proc. Annu. Allerton Conf. Commun., Control, Comput.*, Sep. 2011, pp. 545–550.

[44] A. Adhikary, J. Nam, J.-Y. Ahn, and G. Caire, "Joint spatial division and multiplexing—The large-scale array regime," *IEEE Trans. Inf. Theory*, vol. 59, no. 10, pp. 6441–6463, Oct. 2013.

[45] E. Björnson, J. Hoydis, M. Kountouris, and M. Debbah, "Massive MIMO systems with non-ideal hardware: Energy efficiency, estimation, and capacity limits," *IEEE Trans. Inf. Theory*, vol. 60, no. 11, pp. 7112–7139, Nov. 2014.

[46] X. Gao, O. Edfors, F. Rusek, and F. Tufvesson, "Massive MIMO performance evaluation based on measured propagation data," *IEEE Trans. Wireless Commun.*, vol. 14, no. 7, pp. 3899–3911, 2015.

[47] F. Rusek, D. Persson, B. K. Lau, E. Larsson, T. Marzetta, O. Edfors, and F. Tufvesson, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Process. Mag.*, vol. 30, no. 1, pp. 40–60, Jan. 2013.

[48] V. Veeravalli, Y. Liang, and A. Sayeed, "Correlated MIMO wireless channels: Capacity, optimal signaling, and asymptotics," *IEEE Trans. Inf. Theory*, vol. 51, no. 6, pp. 2058–2072, Jun. 2005.

Amirpasha Shirazinia (S'08–M'14) received the B.Sc. and M.Sc. degrees in Electrical Engineering from Ferdowsi University of Mashhad and Amirkabir University of Technology, respectively, in 2005 and 2008. He received Ph.D. in Telecommunications from the Royal Institute of Technology (KTH), Stockholm, Sweden, in April 2014. Since May 2014, he has been working as a post-doctoral researcher with the Department of Engineering Sciences in Uppsala University, Uppsala, Sweden. His research interests include statistical and sparse signal processing, wireless communications, compressed sensing, massive MIMO and source and channel coding.

Subhrakanti Dey (M'96–SM'06) was born in India, in 1968. He received the B.Tech. and M.Tech. degrees from the Department of Electronics and Electrical Communication Engineering, Indian Institute of Technology, Kharagpur, in 1991 and 1993, respectively, and the Ph.D. degree from the Department of Systems Engineering, Research School of Information Sciences and Engineering, Australian National University, Canberra, in 1996.

He is currently a Professor with the Department of Engineering Sciences in Uppsala University, Sweden. Prior to this, he was a Professor with the Department of Electrical and Electronic Engineering, University of Melbourne, Parkville, Australia, from 2000 until early 2013. From September 1995 to September 1997, and September 1998 to February 2000, he was a Postdoctoral Research Fellow with the Department of Systems Engineering, Australian National University. From September 1997 to September 1998, he was a Postdoctoral Research Associate with the Institute for Systems Research, University of Maryland, College Park. His current research interests include networked control systems, wireless communications and networks, signal processing for sensor networks, and stochastic and adaptive signal processing and control.

Professor Dey currently serves on the Editorial Board of IEEE TRANSACTIONS ON SIGNAL PROCESSING and IEEE TRANSACTIONS ON CONTROL OF NETWORK SYSTEMS. He was also an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING during 2007–2010 and the IEEE TRANSACTIONS ON AUTOMATIC CONTROL during 2004–2007, and Associate Editor for *Elsevier Systems and Control Letters* during 2003–2013.

Domenico Ciuonzo (S'11–M'14) received the B.Sc. and M.Sc. (*summa cum laude*) degrees in computer engineering and the Ph.D., respectively, in 2007, 2009 and 2013, from Second University of Naples (SUN), Aversa, Italy. In 2011 he was involved in the Visiting Researcher Programme of NATO CMRE, La Spezia, Italy. In 2012 he was a visiting scholar at the ECE Department of the University of Connecticut, Storrs, U.S. In 2013–2014, he was a PostDoc at the Department of Industrial and Information Engineering of SUN, under the project "MICENEA", funded by MIUR within the program Futuro in Ricerca (FIRB) 2013. In 2015 he held a visiting appointment at Department of Electronics and Telecommunications, Norwegian University of Science and Technology, Trondheim, Norway. Since 2014 he is a PostDoc researcher at DIETI, University of Naples, "Federico II", Italy. His research interests are mainly in the areas of Data Fusion, Statistical Signal Processing, Target Tracking and Wireless Sensor Networks. His reviewing activity was recognized by IEEE COMMUNICATIONS LETTERS and the IEEE TRANSACTIONS ON COMMUNICATIONS, which nominated him "Exemplary Reviewer" in 2013 and 2014, respectively. Since 2014 he serves as Associate Editor for the IEEE AEROSPACE AND ELECTRONIC SYSTEMS MAGAZINE.

Pierluigi Salvo Rossi (SM'11) was born in Naples, Italy, on April 26, 1977. He received the "Laurea" degree in telecommunications engineering (*summa cum laude*) and the Ph.D. degree in computer engineering in 2002 and 2005, respectively, both from the University of Naples "Federico II", Naples, Italy. During his Ph.D. studies, he has been a researcher at the Inter-departmental Research Center for Signals Analysis and Synthesis (CIRASS), University of Naples "Federico II", Naples, Italy; at the Department of Information Engineering, Second University of Naples, Aversa (CE), Italy; at the Communications and Signal Processing Laboratory (CSPL), Department of Electrical and Computer Engineering, Drexel University, Philadelphia, PA, U.S.; and an adjunct professor at the Faculty of Engineering, Second University of Naples, Aversa (CE), Italy. From 2005 to 2008 he worked as a postdoc at the Department of Computer Science and Systems, University of Naples "Federico II", Naples, Italy; at the Department of Information Engineering, Second University of Naples, Aversa (CE), Italy; at the Department of Electronics and Telecommunications, Norwegian University of Science and Technology, Trondheim, Norway; and visited the Department of Electrical and Information Technology, Lund University, Lund, Sweden.

In 2008, he was appointed assistant professor (tenured in 2011) in telecommunications at the Department of Industrial and Information Engineering, Second University of Naples, Aversa (CE), Italy; and visited the Department of Electronics and Telecommunications, Norwegian University of Science and Technology, Trondheim, Norway and the Excellence Center for Wireless Sensor Networks (WISENET), Uppsala University, Uppsala, Sweden.

Currently, he is an associate professor in signal processing at the Department of Electronics and Telecommunications, Norwegian University of Science and Technology, Trondheim, Norway. He is an Associate Editor of the IEEE COMMUNICATIONS LETTERS and of the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS. His research interests fall within the areas of communications and signal processing.