

Statistical Analysis of Array Gain for Cooperative MISO Transmitters without CSI

Sebastià Galmés, *Member, IEEE*, and Ozgur B. Akan, *Senior Member, IEEE*

Abstract—Virtual Multiple-Input Single-Output (MISO) is recently proposed to extend the benefits of transmitter space diversity to networks in which the deployment of antenna arrays on individual nodes is infeasible from a practical point of view. Ad-hoc and sensor networks are examples of these type of networks. In these scenarios, nodes equipped with single antenna can cooperatively transmit to emulate an antenna array. However, cooperative transmissions require knowledge of the channel state either at the transmitter side or the receiver side in order to achieve full performance gains. Several solutions are proposed in the literature under these assumptions, but at the expense of increased overhead and energy consumption. In this paper, the array gain at the receiver from the non-coherent combining of the signals from multiple transmitters is analyzed in statistical sense, under the assumption that the channel knowledge is unavailable. The transmitters are assumed to be randomly spread over a circular region. More specifically, exact or very accurate closed-form expressions for the expectation and variance of the array gain are obtained, and then a complete statistical distribution is postulated and validated by means of heuristic procedures, goodness-of-fit tests and specialized software. The results obtained in this paper can be especially useful for the implementation of two-tiered wide area sensor networks.

Index Terms—Multiple Input Single Output (MISO), wireless sensor networks, received power, signal-to-noise ratio (SNR), path-loss and shadowing model, expectation, variance, squared coefficient of variation, statistical distribution.

I. INTRODUCTION

SPACE diversity techniques have gained considerable importance in wireless systems due to their benefits in terms of spectral efficiency and performance. In essence, space diversity techniques exploit the availability of independent signal paths due to multiple antennas at the transmitter and/or receiver sides of the communication channel. Among the performance benefits, the use of multiple independently fading signal paths leads to a considerable reduction in the average number of outage events. This is because the probability that these paths experience deep fades simultaneously (outage probability) becomes very small. Also, a substantial improvement in the slope of the average error probability as a function of the average signal-to-noise ratio (SNR) is achieved. These enhancements are captured through the concept of *diversity gain*. Another performance measure is the *array gain*, which

is equal to the SNR increase that is obtained from transmitter and/or receiver diversity compared to the SNR that would be obtained with just a single antenna at both sides of the communication channel. Both array and diversity gains are associated with the simultaneous transmission of the same data stream over the independent paths. However, in the case that both the transmitter and receiver sides have multiple antennas, an alternative form of exploitation is obtained by decomposing the resulting Multiple-Input Multiple-Output (MIMO) channel into a number of parallel channels over which independent data streams can be multiplexed. In this case, the *multiplexing gain* measures the increase in data rate and spectral efficiency with regard to a system based on single antennas at both sides. [1] provides a detailed analysis of space diversity and MIMO systems.

Despite the above benefits, direct application of space diversity techniques to some systems, like ad-hoc and sensor networks, becomes infeasible from both the practical and economical points of view. This is due to the fact that these techniques require complex transceiver circuitry and large amount of signal processing power, which exceed the capabilities of low-powered nodes. Moreover, their limited size does not allow for accommodating antenna arrays. Hence, the implementation of space diversity in wireless ad-hoc and sensor networks relies on the ability of nodes to cooperate (*cooperative diversity*). Essentially, in these scenarios a virtual antenna array can be made of the single-antenna nodes via local information exchanges.

This paper deals with the special case of Cooperative MISO (C-MISO), by considering a circular cluster of arbitrarily large number of randomly-deployed nodes transmitting to a remote (target) destination. The focus is on the array gain, and thus the form of cooperation assumed in this paper is based on pure message replication, which means that the transmitters send the same replica of a message at a time, but with the attenuation and phase differences caused by the diversity of paths. It is well-known that the maximization of the array gain can only be achieved by coherently combining the signals at the target receiver, fact that requires knowledge of channel state at the transmitter side. In this way, every node can conveniently weight the signal to be transmitted so that all received signals have about the same power and are superposed in phase (synchronized). However, again this requirement is not suitable for systems like ad-hoc and sensor networks, which rely on low-powered low-cost devices. This leads to the main goal of this paper, which is to demonstrate that the absence of channel state information (CSI) as well as the randomness of locations, do not preclude from obtaining a significant array gain from a set of cooperative

Manuscript received November 8, 2012; revised April 10 and October 16, 2013; accepted November 28, 2013. The associate editor coordinating the review of this paper and approving it for publication was H. Yousefi'zadeh.

S. Galmés is with the Dept. of Mathematics and Computer Science, University of the Balearic Islands, Spain (e-mail: sebastia.galmes@uib.es).

O. B. Akan is with the Next-Generation and Wireless Communications Laboratory (NWCL), Dept. of Electrical and Electronics Engineering, Koc University, Istanbul 34450, Turkey (e-mail: akan@ku.edu.tr).

Digital Object Identifier 10.1109/TWC.2014.011714.121745

transmitters. In other words, in this paper it is assumed that signals from transmitters randomly deployed within a cluster, are directly added up without any preprocessing or preconditioning. To the best of our knowledge, this is the first work attempting to analyze the effects on array gain of the non-coherent combination of signals from multiple randomly-deployed transmitters over a circular cluster. Moreover, since transmitter locations are random, and signals are not previously weighted or compensated, the resulting array gain is also random. Thus, more specifically, the theoretical goal of this paper is to provide a complete statistical characterization of the array gain under the aforementioned conditions. In particular, exact or very accurate closed-form expressions are obtained for the expectation and variance of the array gain. Then, a complete characterization of its statistical distribution is first postulated and then validated by means of specialized software and goodness-of-fit tests.

From a more practical point of view, the results obtained in this paper can be applied to the class of sensor networks called time-driven. In time-driven sensor networks, communication is triggered by nodes, which periodically report sensed data to the base station. Typically, these networks are deployed in a structured manner, either by locating nodes at strategic sampling positions, or according to some regular pattern. Sensed data are typically routed through multi-hop predetermined paths that form a data-gathering tree rooted at the base station. Because the selection of strategic locations is driven by application requirements or because the monitored variables exhibit low spatial variability, the resulting deployment may be sparse, with large inter-node distances that exceed the transmission range of single nodes. Thus, the implementation of such Wide Area Sensor Networks (WASN) may be enhanced by deploying clusters instead of single nodes at selected locations and using the cooperative scheme proposed in this paper. Note that sensor nodes constitute a relevant example of low-cost transmitters with limited capabilities, for which knowledge of channel state would not only be impractical, but also energy-consuming: the dynamics of the wireless channel would require periodic refreshment of channel state information. So, the improved array gain that can be obtained from the proposed scheme would contribute to increase the transmission range for a given transmission power, or to reduce transmission power for a given transmission range. This, in turn, would allow for the implementation of WASN with no need to rely on other access technologies, like xG.

The rest of this paper is organized as follows. Section II describes the related work. In Section III, an aggregate propagation model for cooperative transmitters is developed. In Section IV, exact or accurate closed-form expressions for the mean, variance and squared coefficient of variation of the array gain are obtained. In Section V, simulation results validating the theoretical expressions are provided. In Section VI the statistical distribution of the array gain is characterized and proposed solutions are validated for the special case of no shadowing. Then, Section VII deals with the implications of the results on cluster dimensioning for large-scale two-tiered sensor networks. Finally, in Section VIII, the main conclusions and further research issues are drawn.

II. RELATED WORK

One of the pioneering works in the application of MIMO techniques to sensor networks is [2], which focuses on the analysis of energy efficiency and delay of cooperative MIMO. The analysis includes the local energy costs necessary for joint packet transmission and reception, and considers an arbitrary number of nodes at both the transmitter and the receiver sides. However, for simplicity, results are only provided for Alamouti schemes with at most two transmit and two receive antennas. A subsequent work is [3], which considers some additional overheads as well as the impact of the channel path loss parameter in the evaluation of energy consumption. As in [2], the analysis covers multiple cooperation nodes at both transmitter and receiver sides, but again results are only provided for 2×1 and 2×2 Alamouti-based schemes. Another interesting contribution is [4], which analyzes the increase in spectral efficiency that can be achieved with channel state information at the transmitter (CSIT), an assumption that was precluded in previous works based on Alamouti schemes and space-time block codes (STBC). This assumption is also adopted in [5], which analyzes the performance of collaborative beamforming using the theory of random arrays. In this case, CSIT takes the form of phase knowledge, which can be achieved via synchronization of each transmitter node to a beacon sent from the destination node (closed-loop scenario) or via acquisition of (relative) transmitter location information from the beacon of a nearby reference point (open-loop scenario). In any case, nodes within a cluster can collaboratively transmit in such a way that their signals are coherently added in the target direction. Given the random deployment of nodes, the achievable performance is analyzed in statistical sense, and as a result the probability distribution of the far-field beampattern of a randomly generated phased array is obtained. This result can also be viewed as an extension of the theory of linear random arrays [6] to random arrays on a disk. The work presented in [7] also deals with collaborative beamforming, but the analysis focuses on the energy overhead necessary to adjust the phases of the transmitter signals so that they constructively interfere at the receiver. More specifically, the paper investigates the net energy benefits of collaborative beamforming as a function of the number of participant nodes and the size of data to be transmitted.

Basically, the works described so far focus on the physical layer aspects of cooperative MISO. However, other works address the design of upper layer protocols that allow the implementation of cooperative MISO exploiting its benefits. For instance, the work presented in [8] considers a sensor network following a two-tiered architecture, which is constructed by grouping nodes in collaborative sets or clusters attending to proximity. Its main focus is on the development of a scheduled protocol that governs inter-cluster and intra-cluster information exchanges, via the so-called inter-cluster and intra-cluster slots, respectively. It assumes MIMO transmissions based on the extension of Alamouti schemes to STBC for more than two transmit antennas, and the corresponding analysis provides results in terms of probability of outage, diversity gain and energy efficiency. [9] proposes a new cooperative low power listening (LPL) MAC protocol, which combines redundant

control packets and one of two cooperative schemes different from STBC: optimal beamforming (BF) and spatial multiplexing (SM). As packet redundancy ensures higher reliability (in terms of retransmission probability), the cooperative schemes allow for energy savings. Both retransmission probability and energy savings are characterized via analytical models, and it is concluded that the new protocol with optimal BF and the number of cooperative transmitters set to 2 represents the most promising solution in terms of energy consumption. An additional set of works go beyond the MAC layer. Particularly, the authors of [10] propose a cross-layer framework for exploiting virtual MISO links in ad hoc networks. The approach spans the physical, MAC and routing layers and provides a significant improvement of the end to end performance in terms of throughput and delay, as well as robustness to mobility and interference-induced link failures. In this context, another contribution proposes Cooperative Source Routing protocol to convert physical layer gain into network level performance improvement [11].

These precedent works rely on the assumption of state channel knowledge at least at the receiver, which inevitably introduces extra signaling and energy consumption. Given the importance of reducing these overheads in wireless sensor networks, this paper considers the case with no form of channel knowledge. Hence, the main focus is on the statistical distribution of the array gain that results from the non-coherent aggregation of signals from multiple, randomly-deployed transmitters. More specifically, the contributions of this paper can be listed as follows:

- The statistical distribution of the array gain for cooperative MISO with no CSI is derived, based on the theoretical considerations and multiple sets of simulation results. Exact or accurate closed-form expressions for the mean and variance of the array gain are obtained.
- Unlike previous works, the intra-cluster distances between multiple transmitters, which lead to time-shifts between the corresponding signals, are not neglected, and their influence on the amplitude and phase of the received signal is analyzed.
- A criterion for cluster dimensioning is provided on the basis of statistical guarantees for the array gain. This can be especially useful for large-scale sensor networks, as the enhanced array gain will allow for larger distances between sampling locations.

As we know, this is the first work attempting to derive statistical information about the non-coherent aggregation of signals from arbitrarily large number of randomly-located transmitters.

III. AGGREGATE PATH-LOSS MODEL

In this section, an aggregate path-loss and shadowing model for a set of randomly-deployed cooperative transmitters is constructed. The objective is to obtain an expression for the power of the received signal resulting from superposition of multiple individual transmissions at a generic target destination. However, before addressing this issue, the main features of the system under consideration are set up and a statistical characterization of the distance to the destination is provided.

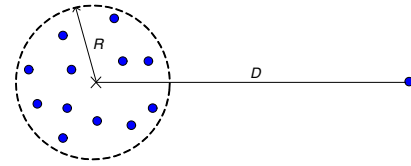


Fig. 1. System under analysis: a group of sensor nodes randomly scattered over a circular region of radius R centered at distance D of a target receiver.

A. System Model

The system under consideration is a circular cluster of radius R with N randomly-deployed transmitters and a target, located at a distance D from the center of the cluster (see Fig. 1). The nodes within the cluster cooperatively transmit by emulating the antenna array of a real MISO scheme. The target may correspond to the cluster head of the next-hop cluster or to the base station. All transmitter nodes and the receiver are assumed to be static. The details of the intra-cluster protocol that governs local information exchanges, as well as possible cluster head rotation are out of the scope of this paper. Adopted assumptions follow:

- As the paper focuses on the array gain, small-scale propagation effects, namely fading, will be ignored. This is in agreement with [1], where the array gain is defined as the increase in average signal-to-noise ratio due to cooperation in absence of fading, which means that only large-scale propagation effects, namely path-loss and shadowing are considered.
- In the construction of the aggregate path-loss and shadowing model, it will be assumed that inter-cluster distances are substantially larger than intra-cluster ones. This is typical for the sparse two-tiered topologies arising in wide area sensor networks. According to Fig. 1, $D \gg R$. Moreover, R cannot be too large in order to limit the transmission costs of intra-cluster communications and to avoid excessive ISI (Inter-symbol Interference) at the receiver. Typically, this means that R is in the order of a few tens of meters, whereas D may achieve several hectometers.
- Once nodes within a cluster receive a packet broadcast by the cluster head, they start to transmit simultaneously. The only phase effects considered in the analysis are caused by the diversity of distances between each intra-cluster node and the target. This assumption can be adopted without losing the generality of analysis, as any de-synchronization among nodes can essentially be mapped into the distance diversity already considered in the present work.

B. Characterization of Distance Diversity

The next step in the analysis of the array gain is the statistical characterization of the distance between a randomly deployed node and the target destination. Fig. 2 shows the reference scheme. Note that, without loss of generality, the origin of the coordinate system (\mathbf{O}) is located at the center of the cluster and the target destination (\mathbf{T}) on the x -axis. Then, the distance d between any point (x, y) within the circular

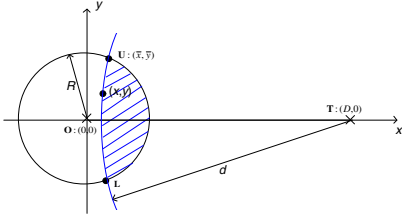


Fig. 2. Reference scheme for the characterization of distance diversity.

cluster and the target destination is given by:

$$d = d(x, y) = \sqrt{(x - D)^2 + y^2} \quad (1)$$

Note that d is a random variable that ranges from $D - R$ to $D + R$. In turn, it depends on the two random variables x and y , whose joint density function $f(x, y)$ obeys the following:

$$f(x, y) = \begin{cases} \frac{1}{\pi R^2} & , x^2 + y^2 \leq R^2 \\ 0 & , x^2 + y^2 > R^2 \end{cases} \quad (2)$$

Accordingly, the cumulative distribution function of d can be expressed as follows:

$$F(d) = \iint_{S(d)} f(x, y) dx dy = \frac{1}{\pi R^2} \iint_{S(d)} dx dy \quad (3)$$

Here, $S(d)$ denotes the shadowed region in Fig. 2, that is, the region of points in the cluster that are located at a distance d or less from the target destination. Obviously, the cumulative distribution is the ratio between such a shadowed region and the area of the circular cluster. Also, note that the shadowed region results from the intersection of two circles: the circle of radius R centered at \mathbf{O} and the circle of radius d centered at \mathbf{T} . Accordingly, a reference point in the analysis is \mathbf{U} , which is the upper cross point of the two corresponding circumferences in Fig. 2. Its coordinates (\bar{x}, \bar{y}) can be obtained using the well-known circumference equations:

$$\bar{x} = \bar{x}(d) = \frac{R^2 + D^2 - d^2}{2D} \quad (4a)$$

$$\bar{y} = \bar{y}(d) = \frac{1}{2D} \sqrt{d^2(2D^2 + 2R^2 - d^2) - (D^2 - R^2)^2} \quad (4b)$$

Correspondingly, the coordinates of the lower cross point \mathbf{L} in Fig. 2 are $(\bar{x}, -\bar{y})$. In addition, the line segment connecting \mathbf{U} and \mathbf{L} can be used to define two reference triangles, \mathbf{OUL} and \mathbf{TUL} , whose areas can be formulated in terms of \bar{x} and \bar{y} . These areas, along with the areas of the circular sectors \mathbf{OUL} and \mathbf{TUL} , can be conveniently combined to obtain an exact result for the cumulative distribution function (the details of the geometry-based and other standard calculations have been omitted, as they are not central to the purposes of the paper):

$$F(d) = \frac{1}{\pi R^2} \left(R^2 \arctan \frac{\bar{y}}{\bar{x}} + d^2 \arctan \frac{\bar{y}}{D - \bar{x}} - D\bar{y} \right), \quad D - R \leq d \leq D + R \quad (5)$$

Then, by taking the first derivative of (5) and performing numerous manipulations (again omitted), the following exact

result for the probability density function $f(d)$ can be obtained:

$$f(d) = \frac{2d}{\pi R^2} \arctan \left(\frac{\bar{y}}{D - \bar{x}} \right), \quad D - R \leq d \leq D + R \quad (6)$$

Recall that \bar{x} and \bar{y} are given by (4a) and (4b), respectively. For illustration purposes, expression (6) is plotted in Fig. 3 for different combinations of R and D . Initially, the radius is increased from 10 to 30 m, while maintaining the ratio D/R . Then, the radius is doubled without varying the distance to the target. All the curves are slightly left-skewed because for $d = D$, the area of the shadowed region in Fig. 2 is always less than half the circle centered at \mathbf{O} . This is more noticeable for relatively small values of D/R , such as the case of $R = 60m$ and $D = 300m$. The following lemma provides a general expression for the moments of the distance to the target.

Lemma 1. The k -th moment of the distance to the target destination results from multiplying the distance between the center of the cluster and the target destination (D), raised to the k -th power, by a factor that only depends on the ratio R/D .

Proof. The k -th moment of the distance to the target destination is defined as follows:

$$E[d^k] = \int_{D-R}^{D+R} d^k f(d) dd \quad (7)$$

Then, recalling expressions (6), (4a) and (4b), introducing the change of variable $t = d/R$ in the evaluation of the integral in (7), and defining $\alpha = R/D$, we can derive the following:

$$E[d^k] = D^k I_k(\alpha), \quad \text{where} \quad (8a)$$

$$I_k(\alpha) = \int_{1/\alpha-1}^{1/\alpha+1} \alpha^k t^{k-1} g(t, \alpha) dt \quad (8b)$$

$$g(t, \alpha) = \frac{2t^2}{\pi} \arctan \left(\frac{\sqrt{-\alpha^4 t^4 + 2\alpha^2(\alpha^2 + 1)t^2 - (\alpha^4 - 2\alpha^2 + 1)}}{\alpha^2 t^2 - \alpha^2 + 1} \right) \quad (8c)$$

□

Corollary. From this lemma, we can derive the following expressions for expectation, second moment, squared coefficient of variation and skewness of the distance to the target destination:

$$E[d] = \mu(\alpha) = D I_1(\alpha) \quad (9)$$

$$E[d^2] = D^2 I_2(\alpha) = D^2 + \frac{R^2}{2} \quad (10)$$

$$SCV[d] = \rho(\alpha) = \frac{E[d^2]}{E[d]^2} - 1 = \frac{D^2 + \frac{R^2}{2}}{D^2 I_1^2(\alpha)} - 1 = \frac{1 + \frac{\alpha^2}{2}}{I_1^2(\alpha)} - 1 \quad (11)$$

$$SK[d] = v(\alpha) = \frac{E[(d - \mu(\alpha))^3]}{E[d - \mu(\alpha)]^3} = \frac{I_3(\alpha) - 3(1 + \alpha^2/2)I_1(\alpha) + 2I_1^3(\alpha)}{(1 + \alpha^2/2 - I_1^2(\alpha))^{3/2}} \quad (12)$$

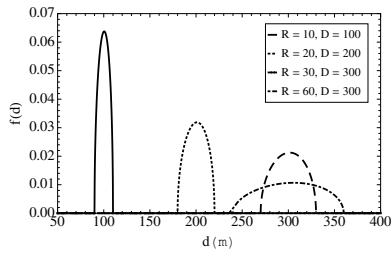


Fig. 3. Probability density function of the distance-to-the-target for different combinations of R and D (both in meters).

Note that due to its definition, α is in the interval $[0, 1]$. Accordingly, $I_1(\alpha) = E[d]/D$ always takes values very close to 1 (between 1 and 1.13 approximately). Similarly, the squared coefficient of variation takes very small values (around 17% at maximum), which indicates low relative variability with respect to the mean. This fact could apparently diminish the relevance of analyzing the statistics of the array gain as a result of the random deployment over a circle. However, the squared coefficient of variation experiences a significant amplification, when the analysis of variability is extended to the array gain. Finally, the skewness of the distance to the receiver, which exclusively depends on α as in the case of the squared coefficient of variation is always slightly negative. This means that the corresponding probability density function is slightly left-skewed.

C. Aggregate Power

Among available path-loss and shadowing models for wireless communications, one of the most commonly used in system design and analysis is the combined log-distance path-loss and log-normal shadowing model [1] [12]. Thus, the purpose of the subsequent analysis is to obtain a model for the aggregate power at the receiver (target destination), given that each communication pair is governed by the combined model. To achieve this goal, the received signal from a single transmitter is formulated first, and then the aggregate model is constructed.

The ratio of transmit-to-receive power in dB for the combined log-distance path-loss and log-normal shadowing model, with adapted notation, can be expressed as [1] [12]:

$$\frac{P_t}{P_r}(dB) = \bar{L}(d_0) + 10\eta \log\left(\frac{d}{d_0}\right) + \Psi \quad (13)$$

In this expression, d_0 is the (close-in) reference distance, η is the path-loss exponent, $\bar{L}(d_0)$ is the average path loss at a distance d_0 from the transmitter, and Ψ is a zero-mean Gaussian distributed random variable (in dB) with standard deviation σ_Ψ (also in dB). This random variable represents the variations due to shadowing about the distance-dependent component of (13). Note that, as stated in [1] the Gaussian distribution can yield sufficiently negative values so as to make $P_r > P_t$, which is physically impossible. Thus, in order to capture the underlying physical model, a truncated version is used such that Ψ always larger than some negative value. On the other hand, in order to avoid the heavy tail of the lognormal distribution and make simulations more efficient,

an upper bound has also been adopted in this paper, meaning that Ψ is no larger than some positive (high) value.

In general, the parameters involved in (13) are determined empirically [1] [12]. For instance, the reference distance is evaluated from measurements close to the transmitter, although in the far-field region of the transmitting antenna. Regarding the path-loss exponent, it depends on the propagation environment, but it usually ranges from values slightly below the free-space propagation scenario ($\eta = 2$) to larger values (up to 6 or more) when obstructions are present. Also, empirical studies in outdoor scenarios have shown that the standard deviation σ_Ψ typically varies between 4dB and 13dB. Finally, the evaluation of $\bar{L}(d_0)$ is based on either close-in measurements or on a free space assumption from the transmitter to the reference distance. Then, following the notation in [1] for the mathematical expression of received signals, the signal generated at the receiver by any single transmitter, under the combined path loss and shadowing model, is as follows:

$$r_1(t) = Re \left\{ \left(\frac{d_0^\eta}{10^{\bar{L}(d_0)/10}} \frac{1}{d^\eta 10^{\Psi/10}} \right)^{\frac{1}{2}} u(t) e^{-j\frac{2\pi}{\lambda}d} e^{j2\pi f_c t} \right\} \quad (14)$$

This expression assumes that d is large compared with d_0 (long-range segment) and that the transmission is narrowband relative to the delay spread among all multi-path components. Specifically, $u(t)$ is the complex envelope or equivalent low-pass representation of the transmitted signal (thus including any conformance pulse), and f_c and λ are the carrier frequency and carrier wavelength respectively.

Next, considering N randomly-deployed cooperative transmitters, the aggregate signal at the receiver can be formulated by assuming that each communication pair obeys the model given by (14). To this end, let us define $d(i)$ as the distance between any transmitter i and the target destination, where i varies between 1 and N . Then, the received signal can be written as follows:

$$r_N(t) = Re \left\{ \left(\sum_{i=1}^N \left(\frac{d_0^\eta}{10^{\bar{L}(d_0)/10}} \frac{1}{d(i)^\eta 10^{\Psi(i)/10}} \right)^{\frac{1}{2}} u(t - \tau(i)) e^{-j\frac{2\pi}{\lambda}d(i)} \right) e^{j2\pi f_c t} \right\} \quad (15)$$

In this expression, $\Psi(i)$ represents the attenuation in dB due to shadowing in the path between the i^{th} transmitter and the receiver, and $\tau(i)$ denotes the delay spread relative to some reference, which can be a hypothetical ray traveling from the origin of coordinates \mathbf{O} to the target destination (see Fig. 2). Accordingly, this delay can vary between $-R/c$ and R/c , where c is the speed of electromagnetic waves. However, assuming again narrowband transmission, $\tau(i) \ll B_u^{-1}$, where B_u is the bandwidth of the complex envelope, and $u(t - \tau(i)) \approx u(t)$. This leads to the simplification of the expression in (15):

$$r_N(t) = Re \left\{ \left[\sum_{i=1}^N \left(\frac{d_0^\eta}{10^{\bar{L}(d_0)/10}} \frac{1}{d(i)^\eta 10^{\Psi(i)/10}} \right)^{\frac{1}{2}} \right] u(t) e^{j2\pi f_c t} \right\}$$

$$e^{-j\frac{2\pi}{\lambda}d(i)} \Big] u(t)e^{j2\pi f_c t} \Big\} \quad (16)$$

Therefore, the aggregate power at the receiver is as follows:

$$P_N = P_t \frac{d_0^\eta}{10^{\bar{L}(d_0)/10}} \left| \sum_{i=1}^N \frac{e^{-j\frac{2\pi}{\lambda}d(i)}}{d(i)^{\eta/2} 10^{\Psi(i)/20}} \right|^2 \quad (17)$$

An alternative for (17) can be obtained using Euler's formula and developing squared modulus:

$$P_N = P_t \frac{d_0^\eta}{10^{\bar{L}(d_0)/10}} \left(\left(\sum_{i=1}^N \frac{\cos\left(\frac{2\pi}{\lambda}d(i)\right)}{d(i)^{\eta/2} 10^{\Psi(i)/20}} \right)^2 + \left(\sum_{i=1}^N \frac{\sin\left(\frac{2\pi}{\lambda}d(i)\right)}{d(i)^{\eta/2} 10^{\Psi(i)/20}} \right)^2 \right) \quad (18)$$

Next, in Section IV, the array gain is introduced and characterized statistically. For the sake of mathematical completeness, the analysis in this section will cover the full range of parameter α , although the long-range approximation makes it only valid for α sufficiently smaller than 1.

IV. STATISTICAL CHARACTERIZATION OF THE ARRAY GAIN

For a given propagation model (combining path loss and shadowing in its more general form), the array gain of a MISO system is defined as the ratio of the power received at the target with cooperative transmission, and the power that would be obtained with a single transmitter only. For simplicity, such transmitter can be virtually located at the origin (see Fig. 2). Also, in order to use an unbiased reference, we can assume that the shadowing component corresponding to this virtual transmitter takes the average value, that is, 0dB. Then, the received power can be obtained by the expressions (17) or (18) for $N = 1$, $d(1) = D$ and $\Psi(1) = 0$:

$$P_1 \Big|_{\substack{d(1) = D \\ \Psi(1) = 0}} = P_t \frac{d_0^\eta}{10^{\bar{L}(d_0)/10} D^\eta} \quad (19)$$

Accordingly, the array gain G can be formulated as:

$$G = \frac{P_N}{P_1 \Big|_{\substack{d(1) = D \\ \Psi(1) = 0}}} = D^\eta \left| \sum_{i=1}^N \frac{e^{-j\frac{2\pi}{\lambda}d(i)}}{d(i)^{\eta/2} \cdot 10^{\Psi(i)/20}} \right|^2 \quad (20)$$

Here, expression (17) for P_N has been used. Again, if the Euler's formula is applied, G can be alternatively expressed as follows:

$$G = D^\eta \cdot \left(\left(\sum_{i=1}^N \frac{\cos\left(\frac{2\pi}{\lambda}d(i)\right)}{d(i)^{\eta/2} \cdot 10^{\Psi(i)/20}} \right)^2 + \left(\sum_{i=1}^N \frac{\sin\left(\frac{2\pi}{\lambda}d(i)\right)}{d(i)^{\eta/2} \cdot 10^{\Psi(i)/20}} \right)^2 \right) \quad (21)$$

Note that the array gain does not depend on technical specifications about the transceivers, but only on the carrier

frequency (or carrier wavelength), the statistics of the shadowing random variable, the number of intra-cluster transmitters and the geometric parameters of the system, i.e., the distance between cluster and target destination, and cluster radius. Note also that expression (21) defines the array gain as a random variable that depends on two sets of N random variables, $d(i)$ and $\Psi(i)$. However, whereas the set of distances are mutually independent (since sensor nodes are assumed to be randomly deployed) and identically distributed according to (5) or (6), the set of shadowing components are in general correlated Gaussian random variables, because obstructing objects typically interfere several signal paths. Nevertheless, in order to preserve the analytical tractability, independence will be assumed from now on (this assumption will not affect the results obtained in this paper because, as a first step approach, this paper will only focus on the effects of distance diversity in absence of shadowing).

A. Average Array Gain

In order to obtain average array gain, let us introduce the following auxiliary random variable:

$$Y(\varphi) = \frac{\cos\left(\frac{2\pi}{\lambda}d + \varphi\right)}{d^{\eta/2} \cdot 10^{\Psi/20}} \quad (22)$$

Note that $Y(\varphi)$ is a function of the random variables d and Ψ , and the adjustment parameter φ , which is either 0 to yield a pure cosine or $-\pi/2$ to yield a pure sine. Note that expression (21) essentially defines the array gain as a result of sampling the random variable given by (22). Under the assumption that such a sampling is independent, the two sums in (21) can be viewed as the aggregation of N independent samples of $Y(\varphi)$ for $\varphi = 0$ and $\varphi = -\pi/2$, respectively. More specifically, if $\bar{Y}(N, \varphi)$ denotes the statistical sample mean constructed from N independent samples of $Y(\varphi)$, (21) can be rewritten in a more compact way:

$$G = D^\eta \cdot N^2 (\bar{Y}^2(N, 0) + \bar{Y}^2(N, -\pi/2)) \quad (23)$$

Therefore, the expected array gain is as follows:

$$E[G] = D^\eta \cdot N^2 (E[\bar{Y}^2(N, 0)] + E[\bar{Y}^2(N, -\pi/2)]) \quad (24)$$

From statistics, it is well known that a sample mean of size N is an unbiased estimator of the (true) population mean, and that its variance is N times lower than the (true) population variance. Hence, the following equalities hold:

$$\begin{aligned} E[\bar{Y}^2(N, 0)] &= \text{Var}[\bar{Y}(N, 0)] + E^2[\bar{Y}(N, 0)] \\ &= \frac{\text{Var}[Y(0)]}{N} + E^2[Y(0)] \end{aligned} \quad (25a)$$

$$\begin{aligned} E[\bar{Y}^2(N, -\pi/2)] &= \text{Var}[\bar{Y}(N, -\pi/2)] + E^2[\bar{Y}(N, -\pi/2)] \\ &= \frac{\text{Var}[Y(-\pi/2)]}{N} + E^2[Y(-\pi/2)] \end{aligned} \quad (25b)$$

Next, by introducing (25a) and (25b) into (24) and performing some calculations, we can derive the following exact expression for the expected array gain:

$$\begin{aligned} E[G] &= D^\eta \left(N \cdot E \left[\frac{1}{d^\eta \cdot 10^{\Psi/10}} \right] + \right. \\ &\quad \left. N(N-1) (E^2[Y(0)] + E^2[Y(-\pi/2)]) \right) \end{aligned} \quad (26)$$

Since Ψ does not depend on distance (this dependence is captured by the path-loss component in (13), which explains that the average of the shadowing component is assumed to be zero), expectations in (26) can be factorized. Specifically, by introducing $X(\varphi)$ such that $Y(\varphi) = X(\varphi)/10^{\Psi/20}$, expression (26) can be rewritten in this way:

$$E[G] = D^\eta \left(N \cdot E[1/d^\eta] \cdot E[1/10^{\Psi/10}] + N(N-1) (E^2[X(0)] + E^2[X(-\pi/2)]) \cdot E^2[1/10^{\Psi/20}] \right) \quad (27)$$

The expectations corresponding to the shadowing component cannot be expressed in closed-form, but they can be numerically evaluated once the variance of the Gaussian distribution is specified. Regarding the evaluation of $E[1/d^\eta]$ and $E[X(\varphi)]$, only R and D need to be specified, because these parameters completely determine the probability density function of the target distance, i.e., (6). A detailed analysis of (27) leads to the derivation of the following lemma for the expected array gain:

Lemma 2. Under the combined path-loss and shadowing model, the expected array gain only depends on the number of transmitters, the path-loss exponent, the variance of the random attenuation due to shadowing and the ratios α and $\beta = R/\lambda$.

Proof. Recalling the definition of $g(t, \alpha)$ in (8c), it can be shown that:

$$E[1/d^\eta] = \frac{1}{D^\eta} \cdot \int_{\frac{1}{\alpha}-1}^{\frac{1}{\alpha}+1} \alpha^{-\eta} \cdot t^{-\eta-1} \cdot g(t, \alpha) \cdot dt = D^{-\eta} \cdot I_{-\eta}(\alpha) \quad (28a)$$

$$E \left[\cos \left(\frac{2\pi}{\lambda} \cdot d + \varphi \right) / d^{\eta/2} \right] = \frac{1}{D^{\eta/2}} \cdot \int_{\frac{1}{\alpha}-1}^{\frac{1}{\alpha}+1} \alpha^{-\eta/2} \cdot t^{-\eta/2-1} \cdot \cos(2\pi\beta \cdot t + \varphi) \cdot g(t, \alpha) \cdot dt = D^{-\eta/2} \cdot \tilde{I}_{-\eta/2}(\alpha, \beta, \varphi) \quad (28b)$$

Here, $I_{-\eta}(\alpha)$ and $\tilde{I}_{-\eta/2}(\alpha, \beta, \varphi)$ have been introduced consistently with (8b). Then, the expected array gain given by (27) can be alternatively expressed as:

$$E[G] = N \cdot I_{-\eta}(\alpha) \cdot E[1/10^{\Psi/10}] + N(N-1) \cdot \left(\tilde{I}_{-\eta/2}^2(\alpha, \beta, 0) + \tilde{I}_{-\eta/2}^2(\alpha, \beta, -\pi/2) \right) \cdot E^2[1/10^{\Psi/20}] \quad (29)$$

Therefore, the expected array gain exclusively depends on N , η , σ_Ψ , α and β . \square

In particular, for the scenarios considered in this paper, where $R \gg \lambda$, the expected array gain can be well approximated by the first term in (27) or (29):

$$E[G] \cong N \cdot D^\eta \cdot E[1/d^\eta] \cdot E[1/10^{\Psi/10}] = N \cdot I_{-\eta}(\alpha) \cdot E[1/10^{\Psi/10}] \quad (30)$$

This approximation arises from the observation that, for $R \gg \lambda$, the following inequality holds:

$$E^2 \left[\frac{\cos \left(\frac{2\pi}{\lambda} \cdot d + \varphi \right)}{d^{\eta/2}} \right] \ll E \left[\frac{1}{d^\eta} \right], \quad \varphi = 0, -\pi/2 \quad (31)$$

To illustrate the validity of this statement, let us first consider the ratio between the right and left-hand sides of (31). Note that it only depends on α , β , ϕ and η :

$$\frac{E \left[\frac{1}{d^\eta} \right]}{E^2 \left[\frac{\cos \left(\frac{2\pi}{\lambda} \cdot d + \varphi \right)}{d^{\eta/2}} \right]} = \frac{1/D^\eta \cdot I_{-\eta}(\alpha)}{\left(\frac{1}{D^{\eta/2}} \cdot \tilde{I}_{-\eta/2}(\alpha, \beta, \varphi) \right)^2} = \frac{I_{-\eta}(\alpha)}{\tilde{I}_{-\eta/2}^2(\alpha, \beta, \varphi)} \quad (32)$$

This ratio was evaluated for η between 1 and 6, α between 0 and 1 and β sufficiently large ($\beta \geq 100$), leading to very high results in all cases. This validated approximation (30) for usual working conditions, that is, carrier frequencies in wireless transmission bands, like ISM, and cluster radius on the order of several tens of meters). It can also be shown that:

$$\lim_{\alpha \rightarrow 0} I_{-\eta}(\alpha) = 1 \quad (33a)$$

$$\lim_{\alpha \rightarrow 1} I_{-\eta}(\alpha) = \infty \quad (33b)$$

Correspondingly:

$$\lim_{\alpha \rightarrow 0} E[G] = N \cdot E[1/10^{\Psi/10}] \quad (34a)$$

$$\lim_{\alpha \rightarrow 1} E[G] = \infty \quad (34b)$$

This latter result, though mathematically correct, is unrealistic because the analysis performed in this paper is only valid for α sufficiently smaller than 1 (long-range segment approximation). It is also particularly interesting to notice that, for any η , $I_{-\eta}(\alpha) \geq 1$ and monotonically increasing with α , and for a given α , it increases as η increases. The latter might be surprising, because it implies that the expected array gain given by (30) increases with the path-loss exponent. However, what it really means is that the benefits of the array of transmitters are more significant under bad scenarios in terms of path loss.

Expression (30) contrasts with the array gain of other schemes, as shown in Table I, which compares the following cases by assuming that every transmitter (node) has limited energetic resources: SISO (Single Input Single Output), that is, one transmitter with a single antenna and one receiver, with a battery B at the transmitter; SISO with a battery N times larger; conventional MISO, that is, a single transmitter with N antennas and channel state information at the transmitter side (CSI); virtual or cooperative MISO (C-MISO) with CSI; and finally, C-MISO without CSI, which is the case analyzed in this paper. The comparison includes other magnitudes, such as the energetic resources at node and system levels, the energy wasted to achieve a prescribed signal-to-noise ratio at the receiver, both at node and system level, and the node lifetime. For the two latter magnitudes, the energy wasted in the SISO schemes, namely E , is taken for reference. Note that

TABLE I
COMPARISON BETWEEN DIFFERENT SCHEMES. NL: NODE LEVEL, SL: SYSTEM LEVEL

Scheme	Energy avail., NL	Energy avail., SL	Array gain	Energy wasted, NL	Energy wasted, SL	Node lifetime
SISO	$N \cdot B$	$N \cdot B$	1	E	E	$N \cdot B/E$
SISO	B	B	1	E	E	B/E
MISO + CSI	B	B	N	E/N	E/N	$N \cdot B/E$
C-MISO + CSI	B	$N \cdot B$	N^2	E/N^2	E/N	$N^2 \cdot B/E$
C-MISO / no CSI	B	$N \cdot B$	Random G , with average given by: $E[G] \cong N \cdot I_{-\eta}(\alpha) \cdot$ $E[1/10^{\Psi/10}]$	E/G	$N \cdot E/G$	$G \cdot B/E$

TABLE II

EXACT AND APPROXIMATE VALUES OF THE EXPECTED ARRAY GAIN, FOR DIFFERENT COMBINATIONS OF N , α , η AND σ_{Ψ} . $\sigma_{\Psi} = 0$ MEANS NO SHADOWING.

N	$1/\alpha$	η	σ_{Ψ} (dB)	$E[G]$ (exact.)	$E[G]$ (approx.)
10	5	2	0	10.2055	10.2055
100	5	2	0	102.057	102.055
10	100	2	0	10.0005	10.0005
100	100	2	0	100.007	100.005
10	5	6	0	12.0093	12.0092
100	5	6	0	120.096	120.092
10	100	6	0	10.0045	10.0045
100	100	6	0	100.047	100.045
10	5	2	13	9.00254	9.00253
100	5	2	13	90.0264	90.0253
10	100	2	13	8.8217	8.8217
100	100	2	13	88.218	88.217
10	5	6	13	10.5937	10.5936
100	5	6	13	105.938	105.936
10	100	6	13	8.82523	8.82522
100	100	6	13	88.2533	88.2522

the proposed scheme is the only one that yields a random array gain, which potentially may lead to lifetime values relatively close to the case of C-MISO with CSI.

To further support the previous results, both the exact and approximate expressions of the expected array gain have been evaluated for different combinations of N , α , η and σ_{Ψ} . In order to maximize the scope of the comparison, extreme values have been assigned to all these parameters. The ratio β has been set to 30, which may correspond to a carrier frequency in the ISM band, such as $f_c = 433$ MHz, and a cluster radius around 20m. In case of shadowing, the lower and upper truncation values have been respectively set to -7dB and 70dB, which represents a very pessimistic scenario. Results are shown in Table II. As it can be noticed, there is practically no difference between exact and approximate results. Also note that, even under bad conditions, the resulting array gain is relatively high.

B. Analysis of Variance

Central to the variance analysis is obtaining an expression for the second moment. By taking the square of (20) and its expectation, the following preliminary formula can be obtained:

$$E[G^2] = E \left[D^{2\eta} \left| \sum_{i=1}^N \frac{e^{-j\frac{2\pi}{\lambda} \cdot d(i)}}{d(i)^{\eta/2} \cdot 10^{\Psi(i)/20}} \right|^4 \right] \quad (35)$$

Again, if Euler's formula is applied, an equivalent expression is obtained:

$$E[G^2] = E \left[D^{2\eta} \left(\left(\sum_{i=1}^N \frac{\cos\left(\frac{2\pi}{\lambda} \cdot d(i)\right)}{d(i)^{\eta/2} \cdot 10^{\Psi(i)/20}} \right)^2 + \left(\sum_{i=1}^N \frac{\sin\left(\frac{2\pi}{\lambda} \cdot d(i)\right)}{d(i)^{\eta/2} \cdot 10^{\Psi(i)/20}} \right)^2 \right) \right] \quad (36)$$

By applying similar approximations to (31) based on $R \gg \lambda$ and the independence assumption, and recalling (28a), the following result can be achieved:

$$E[G^2] \cong N \cdot I_{-2\eta}(\alpha) \cdot E \left[\frac{1}{10^{\Psi/5}} \right] + 2N(N-1) \cdot I_{-\eta}^2(\alpha) \cdot E^2 \left[\frac{1}{10^{\Psi/10}} \right] \quad (37)$$

Finally, since $Var[G] = E[G^2] - E^2[G]$, where the mean and second moments are given by (30) and (37) respectively, the variance of the array gain can be finally approximated as:

$$Var[G] \cong N \cdot I_{-2\eta}(\alpha) \cdot E \left[\frac{1}{10^{\Psi/5}} \right] + N(N-2) \cdot I_{-\eta}^2(\alpha) \cdot E^2 \left[\frac{1}{10^{\Psi/10}} \right] \quad (38)$$

Then, the squared coefficient of variation is given by:

$$\rho[G] = \frac{Var[G]}{E^2[G]} \cong \frac{1}{N} \cdot \frac{I_{-2\eta}(\alpha)}{I_{-\eta}^2(\alpha)} \cdot \frac{E \left[\frac{1}{10^{\Psi/5}} \right]}{E^2 \left[\frac{1}{10^{\Psi/10}} \right]} + \frac{N(N-2)}{N^2} \quad (39)$$

These results reveal that the variability of the array gain (variance, squared coefficient of variation) only depends on the relation R/D , the cluster size N and the parameters of the propagation model (assuming $R \gg \lambda$). Since $I_{-2\eta}(\alpha)$, $I_{-\eta}(\alpha)$ and the shadowing-related terms are independent of N , it can be easily shown that the squared coefficient of variation tends to 1 as N increases: $\lim_{N \rightarrow \infty} \rho[G] = 1$. However, depending on the ratio α and the variance of the shadowing component, converging to 1 can be from left or right. In fact, a detailed analysis of (39) leads to the following statements:

- If $h_{-\eta}(\alpha) = I_{-2\eta}(\alpha)/I_{-\eta}^2(\alpha)$, then $h_{-\eta}(\alpha) \geq 1$ for $\alpha \geq 0$, with $\frac{\partial h_{-\eta}(\alpha)}{\partial \alpha} > 0 \forall \alpha, \forall \eta$.
- For a given α , $h_{-\eta}(\alpha)$ increases as η increases.
- The ratio $E[10^{-\Psi/5}]/E^2[10^{-\Psi/10}]$ is greater than 1 for $\sigma_{\Psi} > 0$, and is equal to 1 for $\sigma_{\Psi} = 0$.

From these results, and given the fact that α is strictly greater than 0 (for a finite D), it can be concluded that $\rho[G]$ is

always greater than $1 - 1/N$. Moreover, there exists $\alpha = \alpha_{ref}$ such that $\rho[G] = 1$, with $\rho[G] < 1$ for $\alpha < \alpha_{ref}$ and $\rho[G] > 1$ for $\alpha > \alpha_{ref}$, also independently of N :

$$\rho[G] = 1 \Rightarrow \frac{I_{-2\eta}(\alpha_{ref})}{I_{-2\eta}^2(\alpha_{ref})} \cdot \frac{E [10^{-\Psi/5}]}{E^2 [10^{-\Psi/10}]} = 2 \quad (40)$$

Obviously, the solution to this equation depends on the path loss exponent and the shadowing component. Note also that, as stated in Section III-B, although the distance to the receiver exhibits low variability, this may be significantly amplified when evaluated for the array gain.

V. SIMULATION RESULTS

In order to validate the theoretical results obtained previously, several simulation tests were performed by using the *Mathematica* package. The carrier frequency and cluster radius were respectively set to 433 MHz and 20 m. For the propagation environment, a pessimistic scenario was selected by fixing $\eta = 5$, $\sigma_\Psi = 10$ dB and a range from -5 dB to 100 dB for the shadow fading component. All the results were obtained after 2000 or 5000 simulation runs (*i.e.* 2000 or 5000 different random deployments), leading to 95% confidence intervals less than 10% for both the expectation and the variance of the array gain. In particular, for the sample variance, the confidence interval based on the chi-square distribution is used [16]. The comparison covers cluster sizes varying from 10 to 100 in steps of 10, and four representative values of α , namely 0.20, 0.40, 0.60 and 0.80 (although large values for α are unrealistic, they were considered just for mathematical validation). This is shown in Fig. 4. As it can be noticed, simulation results match almost perfectly the theoretical values. In the case of the expected array gain – Fig. 4(a), a linear increase is observed in such a way that the larger α results in steeper straight line. Similarly, Fig. 4(b) shows the results corresponding to the squared coefficient of variation. Note that a log-plot is used, since it was observed that the sensitiveness of this coefficient with respect to both N and α increases with the value of the latter.

VI. STATISTICAL DISTRIBUTION OF THE ARRAY GAIN

The next step in the analysis is the adjustment of a statistical distribution for the array gain. As this paper constitutes a first-step approach to analyzing the benefits of cooperation in absence of CSI, the subsequent discussion focuses on the special case of no shadowing ($\sigma_\Psi = 0$) in order to capture the specific effects of distance diversity on that distribution. Moreover, given the dependence of the type of distribution on whether $\rho[G]$ is smaller or larger than 1, a detailed evaluation of expression (39) was performed. In particular, it was observed that the evolution of the ratio $E [10^{-\Psi/5}] / E^2 [10^{-\Psi/10}]$ in terms of the standard deviation and truncation values of the shadowing component is complex, generally involving local maxima and minima. Obviously, this requires a deeper analysis that would exceed the purposes of this paper. On the other hand, the path-loss exponent is set to 4, which has been commonly adopted from empirical observations [13] [14] [15] [16]. In this case, the solution to (40) is $\alpha_{ref} = 0.41856$ and we will focus on the case $\alpha < \alpha_{ref}$, as it fully captures

TABLE III

MAIN FEATURES OF THE DISTRIBUTION TO BE FITTED. IN THIS TABLE, $f(G)$ STANDS FOR THE THEORETICAL DISTRIBUTION OF THE ARRAY GAIN.

Features	Theoretical results
Range	$(0, G_{\max})$, with $G_{\max} = N^2/(1-\alpha)^4$
Limiting values	$\lim_{G \rightarrow 0} f(G) = 0$; $\lim_{G \rightarrow G_{\max}} f(G) = 0$
Mean	$E[G] \cong N I_{-4}(\alpha)$
Squared coefficient of variation	$\rho[G] = \frac{1}{N} \frac{I_{-8}(\alpha)}{I_{-4}^2(\alpha)} + \frac{N(N-2)}{N^2} < 1$ for $\alpha < \alpha_{ref}$

the validity region of the analysis performed in this paper (D much larger than R). Moreover, it also corresponds to the wide area deployments.

In general, using some theoretical a priori information is recommended in order to postulate plausible distributions. From the results obtained in Section IV, particularized for $\sigma_\Psi = 0$, Table III summarizes some features and metrics of the desired distribution. In particular, the first and second features described in the table state that the distribution to be fitted is bounded by two extreme unreachable values. On one hand, the lower endpoint is 0 as the array gain can be extremely small but always strictly positive, as there is no way to cancel (20). On the other hand, the maximum value of the array gain corresponds to the hypothetical case in which all the transmitters are located at the rightmost position within the circle depicted in Fig. 2, that is, at distance $D - R$ from the target destination. In this case, it can be shown that (20) yields the maximum gain $G_{\max} = N^2/(1-\alpha)^4$.

In addition to a priori theoretical results, histograms with different data sets can provide insights on the shape and the family of distributions to be considered. For instance, Fig. 5(a) shows the histogram obtained from sampling the array gain when $\alpha = 0.1$ and $N = 30$ by ExpertFit [18] (described later) from 1000 samples generated by Mathematica under the same settings of Section V. Similarly, Fig. 5(b) shows the histogram for $\alpha = 0.4$ and $N = 70$. In fact, these histograms were obtained for all the combinations of α and N . At a first glance, they seem to correspond to an exponential behavior. However, the exponential distribution is consistent neither with the limiting value at the lower endpoint formulated in Table III, nor with a squared coefficient of variation that should be strictly less than 1. However, the histograms shown in Fig. 5 may correspond to distributions that have a mode close to the lower endpoint, thus exhibiting a sharp peak at this mode. The most common is the Beta distribution, which is bounded, and the truncated version of the Gamma, Lognormal and Weibull distributions, which are non-negative and unbounded (the reason to consider the truncated version is to fulfill the finite range specification in Table III). Table IV shows the parameter fitting for each of these distributions, based on the true mean and squared coefficient of variation of the array gain.

In general, the behavior of the distributions shown in Table IV at the lower endpoint depends on their shape parameters (except the Lognormal distribution, whose density always yields 0 at $G = 0$ regardless of its parameters). This allows us to discard the Beta distribution, since its first shape parameter

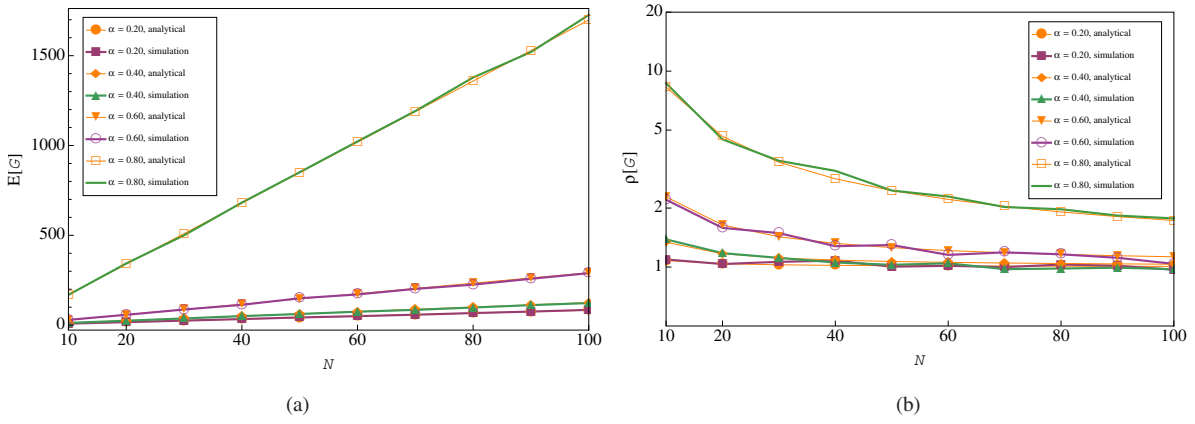


Fig. 4. Evolution of the expected array gain (a) and the squared coefficient of variation of the array gain (b) in terms of the cluster size, for different α values.

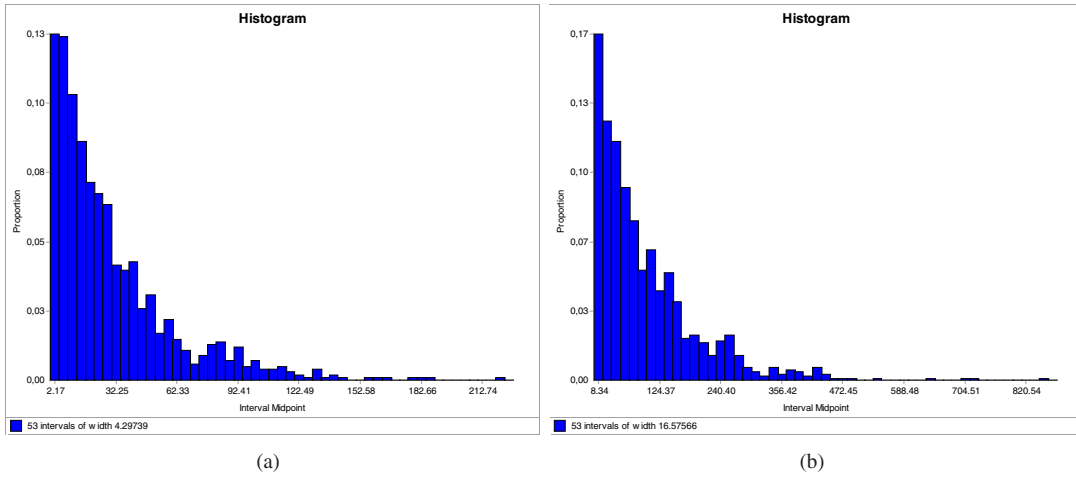


Fig. 5. Histogram generated by ExpertFit from 1000 samples of the array gain for $\alpha = 0.1$ $N = 30$ (a), $\alpha = 0.4$ $N = 70$ (b).

(sh_1) usually results in a value less than 1 for $\rho[G]$ close to 1 (see sh_1 in Table IV), which in turn yields a limiting value at the lower endpoint equal to infinity instead of 0 (see [19]). Whereas, the shape parameters obtained for both the Gamma and Weibull distributions are always above 1 for $\rho[G] < 1$. From Table IV, this statement is obviously true in the case of the Gamma distribution. In the case of the Weibull distribution, the following statement holds:

$$\exists x < 1 : \frac{1}{x} \frac{\Gamma(2x)}{\Gamma^2(x)} < 1 \quad (41)$$

(Note that the solution for the shape parameter would be $sh = 1/x$).

Next, this preliminary analysis is completed by a deeper study based on the ExpertFit software (version 8). This is a powerful statistical package that allows for automatically determining the distributions, including their parameters, that best fit to a data set. It also includes other features for definite assessment, like graphical plots and goodness-of-fit tests (Chi-square, Kolmogorov-Smirnoff and Anderson-Darling). The analysis that follows is divided into two parts. Firstly, the automatic fitting feature of ExpertFit is used to obtain the best ranked distributions. In the second part, the Anderson-Darling test, which is recognized to be the most powerful, is applied.

TABLE IV
PARAMETER FITTING OF VARIOUS PLAUSIBLE DISTRIBUTIONS BASED ON THE KNOWN VALUES FOR THE EXPECTATION AND SQUARED COEFFICIENT OF VARIATION OF THE ARRAY GAIN. IN THIS TABLE, sh AND sc STAND RESPECTIVELY FOR THE SHAPE AND SCALE PARAMETERS.

Distribution	Parameter fitting
Beta (a, b, sh_1, sh_2)	$a = 0 ; b = G_{\max}$ $sh_1 = \frac{b - E[G](1 + \rho[G])}{b \rho[G]}$ $sh_2 = \frac{(b - E[G])(b - E[G](1 + \rho[G]))}{b E[G] \rho[G]}$
LN(sh, sc)	$sh = \sqrt{\text{Ln}(1 + \rho[G])}$ $sc = \text{Ln}(E[G]) - sh^2/2$
Gamma(sh, sc)	$sh = 1/\rho[G]$ $sc = E[G] \rho[G]$
Weibull(sh, sc)	Find $sh : 2 sh \frac{\Gamma(2/sh)}{\Gamma^2(1/sh)} - 1 = \rho[G]$ $sc = E[G] \frac{sh}{\Gamma(1/sh)}$

A. Automatic fitting

ExpertFit ranks the distributions (with estimates of their parameters) in terms of their capability to fit a given data set by means of a relative score between 0 and 100. In addition, it provides an absolute evaluation that qualifies the fitting capability of every distribution as good, indeterminate or bad. To summarize the results, Table V shows the average score obtained for the a priori most theoretically-

TABLE V
AVERAGE SCORE AND COEFFICIENT OF VARIATION THAT RESULT FROM
APPLYING THE AUTOMATIC FITTING FEATURE OF EXPERTFIT.

Distribution	Average score	Coefficient of variation (%)
Gamma	87.66	6.74
Weibull	87.04	9.99
Pearson type VI	79.63	14.03
Lognormal	38.55	10.68

feasible distributions, along with the corresponding coefficient of variation (ratio between standard deviation and average) expressed as a percentage. For each distribution, these measures were obtained over 40 different data sets, resulting from the combination of $\alpha = \{0.1, 0.2, 0.3, 0.4\}$ and $N \in [10, 100]$. In turn, each data set consisted of 1000 samples of the array gain. As noticed, the initially postulated Lognormal distribution, gets a very low score with ExpertFit despite its theoretical feasibility (being the absolute qualification equal to "bad" for all data sets). The reason is that in all tests, the peak of such distribution overestimates the empirical density values significantly. However, a new distribution comes into play, namely the Pearson type VI (PT6) distribution, which achieves a high evaluation, although not the highest one. In addition, the PT6 distribution introduces an extra difficulty in the process of parameter fitting, as it is defined by three parameters instead of two (one scale and two shape parameters). This is not the case for both the Gamma and Weibull distributions, which are based on two parameters (see Table IV) and have received the best scores. Although the Gamma distribution achieves an overall evaluation that is slightly superior in terms of average and coefficient of variation, the similarity with the Weibull distribution means that applying any quantitative method (like cluster dimensioning in the next section) based on either statistical characterization would lead to no significant difference. Nevertheless, in order to obtain a higher discrimination between these characterizations, another experiment was performed using data sets of 5000 samples. However, still no significant changes were detected. Thus, for a definite assessment, the Anderson-Darling test was applied by focusing on the Gamma and Weibull distributions (next subsection).

To complete this first experimental stage, Tables VI-IX compare the parameters obtained theoretically and those generated by ExpertFit from the application of the MLE (Maximum Likelihood Estimator) method (as part of the automatic fitting process). As it can be noticed, theoretical and fitted parameters are in very good agreement in all cases.

Finally, the residual area that results from truncating the two distributions at $G = G_{max}$, was also evaluated for the 40 data sets. In all cases the results were extremely low or practically 0. This is due to the fact that G_{max} is usually very large and both the Gamma and Weibull distributions are light-tailed. Hence, their natural (unbounded) versions suffice to represent the statistical behavior of the array gain, with no need to rely on the truncated versions.

B. Anderson-Darling test

In the last experiment, we apply the Anderson-Darling test to the 40 data sets generated previously. As Fig. 6 reveals, the null hypothesis cannot be rejected for any distribution at a level of significance of 10%. Particularly, Fig. 6(a) refers to the Gamma distribution. As noticed, all test statistics are below the critical value. The same happens in Fig. 6(b) for the Weibull distribution. These results lead us to the conclusion that there is no difference between using a Gamma distribution or a Weibull distribution to characterize the statistical behavior of the array gain, at least from a practical point of view. The next section, which considers the dimensioning of transmitter clusters on the basis of the two characterizations, will highlight this fact.

VII. CLUSTER DIMENSIONING

The results obtained in this paper suggest that cluster dimensioning cannot be exclusively based on the expected array gain, if at least some statistical guarantee on array gain is to be provided. This is due to the variability around the expected value, which in turn is caused by the random deployment. Fortunately, the knowledge of the statistical distribution that characterizes the random gain simplifies the problem to finding the number of transmitters such that the array gain is above some threshold with high probability. This condition can be expressed as

$$\text{Find } N: \text{prob}[G \geq G_q] \geq q \quad (42)$$

Note that G_q is nothing but the percentile $(1 - q)100$ of the distribution of the array gain.

The results obtained in the previous section state that both the Gamma and Weibull distributions can accurately reproduce the random nature of the array gain. Fig. 7(a)-8(b) provide a definite assessment of this statement. In each figure, the cluster size in terms of the required array gain is plotted for different statistical guarantees (q), considering both distributions for a specific value of α . As it is observed, in all cases there is practically no difference between the two distributions in terms of the dimensioning of clusters. Note also that, as stated before, the required cluster sizes are significantly higher than those predicted from the expected array gain.

VIII. CONCLUSIONS

This paper has addressed the scenario of multiple randomly-deployed cooperative transmitters forming a cluster and sending information to a remote receiver. For this MISO scheme, no form of channel state information has been assumed, and the expectation and squared coefficient of variation of the corresponding array gain have been completely characterized by adopting a combined path-loss and shadowing model between every transmitter and the receiver. Specifically, exact or very accurate closed-form expressions have been obtained in terms of the cluster size, cluster radio, target distance, path-loss exponent and shadowing component. As this paper represents a first-step approach to the analysis of cooperation in absence of CSI, the characterization of the statistical distribution has focused on the simpler case of no shadowing. Under these conditions, it has been shown that the array gain

TABLE VI
COMPARISON BETWEEN THEORETICAL AND EXPERTFIT PARAMETERS FOR THE GAMMA DISTRIBUTION AND DIFFERENT CLUSTER SIZES $\alpha = 0.1$ (LEFT), $\alpha = 0.2$ (RIGHT).

N	Shape (theoretical)	Shape (ExpertFit)	Scale (theoretical)	Scale (ExpertFit)	N	Shape (theoretical)	Shape (ExpertFit)	Scale (theoretical)	Scale (ExpertFit)
10	1.1061	1.07792	9.22431	9.54244	10	1.09022	1.10667	9.95275	10.02504
20	1.05038	1.00388	19.4274	20.58234	20	1.04316	1.04541	20.8035	21.07654
30	1.03303	0.99779	29.6304	30.16485	30	1.02837	0.99762	31.6542	32.52350
40	1.02457	1.01392	39.8334	42.19290	40	1.02112	1.03130	42.5049	41.07919
50	1.01956	1.00645	50.0365	51.82210	50	1.01683	0.99667	53.3556	53.37703
60	1.01625	0.94071	60.2395	66.66348	60	1.01398	0.98469	64.2063	65.41380
70	1.01389	1.01825	70.4426	72.75257	70	1.01196	0.98282	75.0569	77.95617
80	1.01214	0.94314	80.6456	85.74653	80	1.01045	0.93820	85.9076	91.59060
90	1.01077	1.01574	90.8487	94.64508	90	1.00928	0.96505	96.7583	98.26546
100	1.00968	0.97094	101.052	107.97315	100	1.00834	0.96819	107.609	108.84919

TABLE VII
COMPARISON BETWEEN THEORETICAL AND EXPERTFIT PARAMETERS FOR THE GAMMA DISTRIBUTION AND DIFFERENT CLUSTER SIZES, $\alpha = 0.3$ (LEFT), $\alpha = 0.4$ (RIGHT).

N	Shape (theoretical)	Shape (ExpertFit)	Scale (theoretical)	Scale (ExpertFit)	N	Shape (theoretical)	Shape (ExpertFit)	Scale (theoretical)	Scale (ExpertFit)
10	1.06065	0.97253	11.3854	12.26707	10	1.01185	1.03631	14.0063	13.65093
20	1.02943	1.00555	23.4612	23.53351	20	1.00589	1.01657	28.1786	28.29677
30	1.01943	0.99594	35.537	36.45009	30	1.00392	1.00213	42.351	42.37957
40	1.0145	0.95838	47.6129	51.18716	40	1.00294	0.90603	56.5233	65.05896
50	1.01157	1.00903	59.6887	60.88556	50	1.00235	0.99165	70.6957	71.10901
60	1.00962	0.94070	71.7646	80.25972	60	1.00196	0.99698	84.868	88.00897
70	1.00823	0.95887	83.8404	89.50208	70	1.00168	0.95812	99.0404	101.88272
80	1.0072	0.99849	95.9163	99.03940	80	1.00147	0.95747	113.213	121.72880
90	1.00639	1.00982	107.992	113.72055	90	1.0013	0.95868	127.385	136.27682
100	1.00575	1.04231	120.068	115.40265	100	1.00117	1.02230	141.557	141.75506

TABLE VIII
COMPARISON BETWEEN THEORETICAL AND EXPERTFIT PARAMETERS FOR THE WEIBULL DISTRIBUTION AND DIFFERENT CLUSTER SIZES, $\alpha = 0.1$ (LEFT), $\alpha = 0.2$ (RIGHT).

N	Shape (theoretical)	Shape (ExpertFit)	Scale (theoretical)	Scale (ExpertFit)	N	Shape (theoretical)	Shape (ExpertFit)	Scale (theoretical)	Scale (ExpertFit)
10	1.05209	1.05519	10.4106	10.50414	10	1.04441	1.07551	11.041	11.40540
20	1.02497	1.00443	20.6134	20.70057	20	1.02142	1.04380	21.8915	22.40344
30	1.01642	0.99576	30.8163	30.04421	30	1.01411	1.00341	32.7421	32.49151
40	1.01223	1.00456	41.0192	42.86301	40	1.01052	1.01384	43.5927	42.57974
50	1.00975	1.00575	51.2222	52.28134	50	1.00839	0.98980	54.4434	52.94177
60	1.0081	0.96298	61.4252	61.69044	60	1.00697	0.99377	65.2941	64.24266
70	1.00693	1.00784	71.6283	74.32524	70	1.00597	0.99220	76.1447	76.37044
80	1.00605	0.96197	81.8313	79.48739	80	1.00522	0.95908	86.9954	84.35207
90	1.00538	1.01076	92.0343	96.53009	90	1.00463	0.97517	97.8461	93.80479
100	1.00483	0.98286	102.237	104.07136	100	1.00417	0.98055	108.697	104.49645

TABLE IX
COMPARISON BETWEEN THEORETICAL AND EXPERTFIT PARAMETERS FOR THE WEIBULL DISTRIBUTION AND DIFFERENT CLUSTER SIZES, $\alpha = 0.3$ (LEFT), $\alpha = 0.4$ (RIGHT).

N	Shape (theoretical)	Shape (ExpertFit)	Scale (theoretical)	Scale (ExpertFit)	N	Shape (theoretical)	Shape (ExpertFit)	Scale (theoretical)	Scale (ExpertFit)
10	1.03	0.98621	12.2221	11.85974	10	1.00591	1.02736	14.2074	14.30214
20	1.01464	1.00587	24.2978	23.72201	20	1.00294	1.01633	28.3798	28.95675
30	1.00968	0.99479	36.3736	36.22082	30	1.00196	1.00136	42.5521	42.49403
40	1.00723	0.97584	48.4494	48.54072	40	1.00147	0.94781	56.7244	57.57335
50	1.00577	1.02015	60.5252	62.00884	50	1.00117	0.99652	70.8967	70.40549
60	1.0048	0.97127	72.601	74.56717	60	1.00098	1.00249	85.069	87.83425
70	1.00411	0.97551	84.6768	84.89487	70	1.00084	0.97466	99.2414	96.53379
80	1.00359	1.01037	96.7526	99.30510	80	1.00073	0.97719	113.414	115.41633
90	1.00319	1.00762	108.828	115.19873	90	1.00065	0.98223	127.586	129.69618
100	1.00287	1.02833	120.904	121.64751	100	1.00059	1.01567	141.758	145.85020

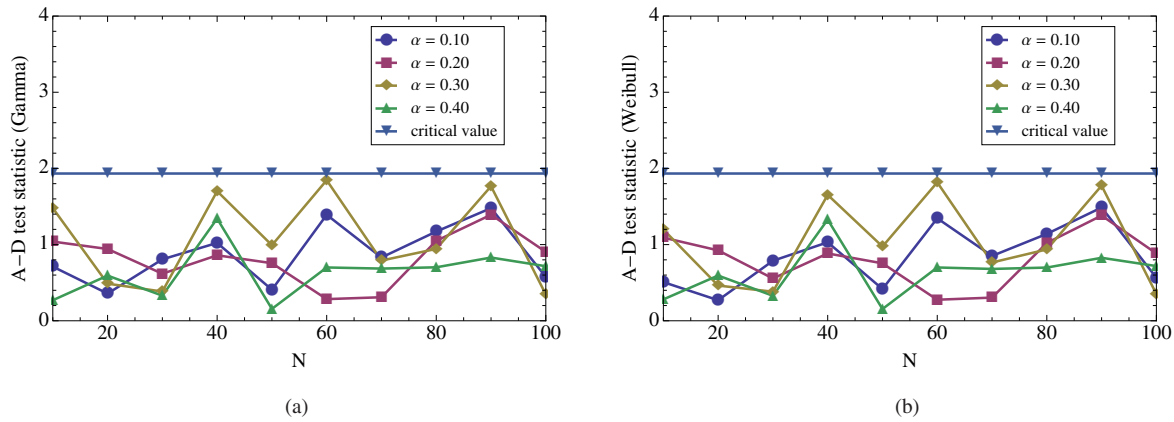


Fig. 6. Anderson-Darling test statistics under the hypothesis null that the array gain obeys a Gamma distribution (a) and Weibull distribution (b). The critical value is for a 10% level of significance.

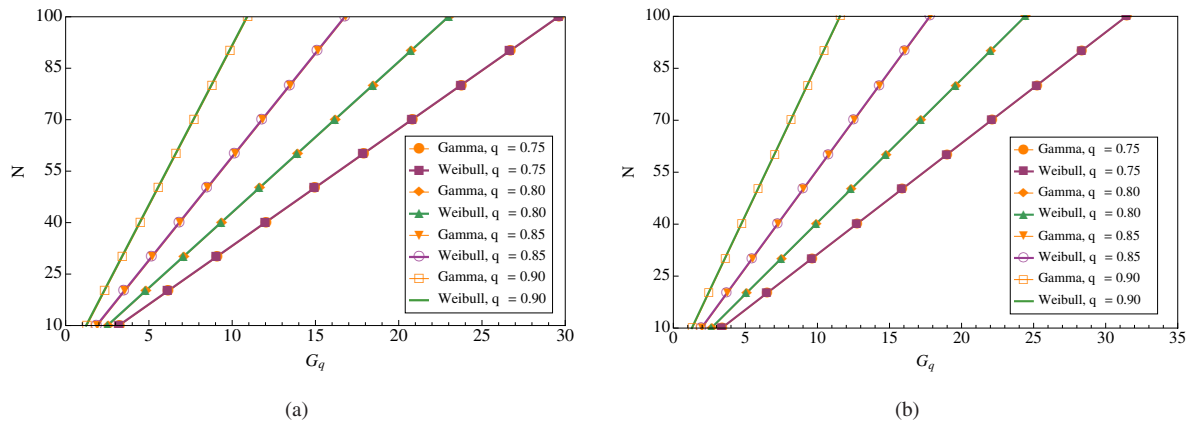


Fig. 7. Cluster size in terms of the required array gain for different statistical guarantees. Both the Gamma and Weibull distributions are considered, with $\alpha = 0.1$ (a) and $\alpha = 0.2$ (b).

is well described by a Gamma or Weibull distribution, whose parameters can be easily calculated from the mean and squared coefficient of variation previously obtained.

Further research can focus on the characterization of the statistical distribution of the array gain under shadowing conditions. This can also include the natural auto-correlation of the shadowing component, which has been ignored to preserve the analytical tractability, as well as additional random independent time shifts among all transmitters. In fact, we expect that the benefits of cooperation in these less favorable scenarios are even greater. On the other hand, some preliminary results have shown that, given a bad deployment (a deployment for which the resulting array gain is small), a small variation along the polar angle of the target location within the receiving cluster, can lead to very high gains and thus practically 100% statistical guarantees. So, this phenomenon also deserves further analysis. Finally, the consideration of cluster shapes other than circular may contribute to obtain even better distributions for the array gain.

ACKNOWLEDGEMENTS

This work is funded in part by the Spanish Ministry of Science and Innovation under contract TIN2010-16345. Also, the authors would like to express their gratitude to Murat Kocaoglu, Ph. D. student at Koc University, Istanbul, for his contribution to this final version of the paper.

REFERENCES

- [1] A. Goldsmith, *Wireless Communications*. Cambridge University Press, 2005.
- [2] S. Cui, A. J. Goldsmith, and A. Bahai, "Energy-efficiency of MIMO and cooperative MIMO techniques in sensor networks," *IEEE J. Sel. Areas Commun.*, vol. 22, no. 6, pp. 1089–1098, Aug. 2004.
- [3] S. Jayaweera, "Virtual MIMO-based cooperative communication for energy-constrained wireless sensor networks," *IEEE Trans. Wireless Commun.*, vol. 5, no. 5, pp. 984–989, May 2006.
- [4] S. J. Kim, R. E. Cagley, and R. A. Iltis, "Spectrally efficient communication for wireless sensor networks using a cooperative MIMO technique," *Wireless Netw.*, vol. 13, no. 3, pp. 397–407, June 2007.
- [5] H. Ochiai, P. Mitran, H. V. Poor, and V. Tarokh, "Collaborative beamforming for distributed wireless ad hoc sensor networks," *IEEE Trans. Signal Process.*, vol. 53, no. 11, pp. 4110–4124, Nov. 2005.
- [6] Y. T. Lo, "A mathematical theory of antenna arrays with randomly spaced elements," *IRE Trans. Antennas Propag.*, vol. AP-12, pp. 257–268, May 1964.
- [7] J. Feng, Y.-H. Lu, B. Jung, and D. Peroulis, "Energy efficient collaborative beamforming in wireless sensor networks," in *Proc. 2009 IEEE International Symp. Circuits, Syst.*
- [8] A. D. Coso, S. Savazzi, U. Spagnolini, and C. Ibars, "Virtual MIMO channels in cooperative multi-hop wireless sensor networks," in *Proc. 2006 Conf. Inf. Sciences, Syst.*
- [9] M. R. Ahmad, E. Dutkiewicz, and X. Huang, "Performance evaluation of MAC protocols for cooperative MIMO transmissions in sensor networks," in *Proc. 2008 ACM International Symp. Performance Evaluation Wireless Ad Hoc, Sensor, Ubiquitous Netw.*
- [10] G. Jakllari, S. V. Krishnamurthy, M. Faloutsos, P. V. Krishnamurthy, and O. Ercetin, "A cross-layer framework for exploiting virtual MISO links in mobile ad hoc networks," *IEEE Trans. Mobile Comput.*, vol. 6, no. 6, pp. 579–594, June 2007.

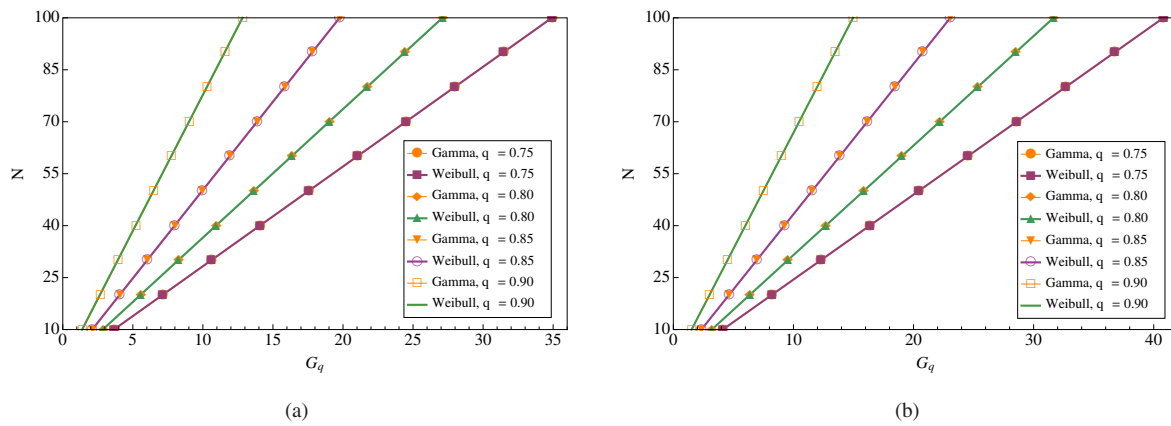


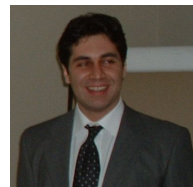
Fig. 8. Cluster size in terms of the required array gain for different statistical guarantees. Both the Gamma and Weibull distributions are considered, with $\alpha = 0.3$ (a) and $\alpha = 0.4$ (b).

- [11] Y. Guan, Y. Xiao, C.-C. Shen, and L. Cimini, "CSR: cooperative source routing using virtual MISO in wireless ad hoc networks," in *Proc. 2011 IEEE WCNC*.
- [12] T. S. Rappaport, *Wireless Communications - Principles and Practice*, 2nd ed. Prentice-Hall, 2002.
- [13] W. B. Heinzelman, A. P. Chandrakasan, and H. Balakrishnan, "An application-specific protocol architecture for wireless microsensor networks," *IEEE Trans. Wireless Commun.*, vol. 1, no. 4, pp. 660–670, Oct. 2002.
- [14] M. A. Labrador and P. M. Wightman, *Topology Control in WSN*. Springer Science + Business Media B. V., 2009.
- [15] O. Younis and S. Fahmy, "Distributed clustering in ad-hoc sensor networks: a hybrid, energy-efficient approach," in *Proc. 2004 IEEE INFOCOM*.
- [16] T. Stoyanova, F. Kerasiotis, A. Prayati, and G. Papadopoulos, "Evaluation of impact factors on RSS accuracy for localization and tracking applications," in *Proc. 2007 ACM MOBIWAC*.
- [17] D. C. Montgomery and G. C. Runger, *Applied Statistics and Probability for Engineers*. Wiley, 2007.
- [18] ExpertFit. Available: <http://www.averill-law.com/distribution-fitting/>
- [19] A. M. Law and W. D. Kelton, *Simulation Modeling and Analysis*, 3rd ed. McGraw-Hill Higher Education, 2000.



and cognitive radios.

Sebastià Galmés (M'10) received the M.Sc. degree in Electrical Engineering from Polytechnic University of Catalonia, Barcelona, Spain, in 1989, and the Ph.D. degree in Computer Science from University of the Balearic Islands, Palma de Mallorca, Spain, in 1999, where he is currently Associate Professor in the Department of Mathematics and Computer Science and Head of the Performance Engineering Group. His current research interests focus on wireless communications, wireless sensor networks, nanoscale and molecular communications



Ozgur B. Akan (M'00-SM'07) received the Ph.D. degree in electrical and computer engineering from the Broadband and Wireless Networking Laboratory, School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, in May 2004. He is currently a full Professor with the Department of Electrical and Electronics Engineering, Koc University, Associate Director for the Graduate School of Science and Engineering, and the Director of Next-generation and Wireless Communications Laboratory (NWCL). His current research interests

are in nanoscale, molecular and quantum communications, next-generation wireless communications, signal processing for communications, information theory. Prof. Akan is an Associate Editor for *IEEE TRANSACTIONS ON COMMUNICATIONS*, *IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY*, Editor for the *International Journal of Communication Systems* (Wiley), Editor for the *European Transactions on Telecommunications*, and Editor for *Nano Communication Networks Journal* (Elsevier). He is an IEEE Senior Member (Communications Society, Information Theory Society, Signal Processing Society), and a member of ACM. He is appointed as IEEE Communications Society (ComSoc) Distinguished Lecturer effective January 2011.