

# MIMO Interference Alignment in Random Access Networks

Behrang Nosrat-Makouei, Radha Krishna Ganti, Jeffrey G. Andrews, and Robert W. Heath, Jr.,

## Abstract

In this paper, we analyze a multiple-input multiple-output (MIMO) interference channel where nodes are randomly distributed on a plane as a spatial Poisson cluster point process. Each cluster uses interference alignment (IA) to suppress intra-cluster interference but unlike most work on IA, we do not neglect inter-cluster interference. We also connect the accuracy of channel state information to the distance between the nodes, i.e. the quality of CSI degrades with increasing distance. Accounting for the training and feedback overhead, we derive the transmission capacity of this MIMO IA ad hoc network and then compare it to open-loop (interference-blind) spatial multiplexing. Finally, we present exemplary system setups where spatial multiplexing outperforms IA due to the imperfect channel state information or the non-aligned inter-cluster interference.

## I. INTRODUCTION

Interference alignment (IA) achieves the degrees of freedom in the  $K$ -user MIMO interference channel [2]. IA confines the interference to a subspace at each receiver such that an interference-free subspace becomes available for the desired signal transmission. Except for *blind* IA techniques [3], [4], which usually attain a lower multiplexing gain, IA requires cooperation between the transmitting/receiving nodes using global channel state information (CSI) [5] or some form of channel reciprocity. Although the impact of inaccurate CSI on the performance of IA has been studied before [6], existence of non-cooperating nodes casting non-aligned interference has been mostly ignored.

### A. Background

In large networks, such as mobile ad hoc and cellular networks, IA will be used independently in separate clusters and the nearby nodes that are not coordinating with any one cluster will cause non-aligned interference at the receivers. There are three main reasons.

- 1) Number of antennas at each node is a limiting factor. It is shown in [2] that the number of nodes that can cooperate through MIMO IA is limited by the number of antennas at each node.
- 2) Overhead practically limits the cluster size. The overhead of IA grows super linearly with the number of users [7], and hence it is likely that small groups of nodes will coordinate to perform IA.

Behrang Nosrat-Makouei, Jeffrey G. Andrews, and Robert W. Heath, Jr., are with the Department of Electrical and Computer Engineering, The University of Texas at Austin, Austin, TX 78712 USA (e-mail: behrang.n.m@mail.utexas.edu; jandrews@ece.utexas.edu; rheath@ece.utexas.edu). Radha Krishna Ganti is with the Department of Electrical Engineering, Indian Institute of Technology Madras, Chennai, India (e-mail: rganti@ee.iitm.ac.in).

This work was supported by the DARPA IT-MANET program, Grant W911NF-07-1-0028, and by the Army Research Labs, Grant W911NF-10-1-0420. This work was presented in part at Asilomar Conf. on Signals, Syst. & Comput., Pacific Grove, CA, Nov. 2011 [1].

- 3) More cooperation is not always better. Recently, [8] showed that because of inherent channel uncertainty, there is a moderate cluster size above which spectral efficiency at best saturates, and in many practical scenarios (e.g. when pilots are used for channel estimation), actually decreases if more nodes join the cluster to cooperate.

In this case, single cluster analysis (e.g. DoF studies) does not capture the impact of interference from the other nodes in the network and can lead to unrealistic cooperation gains which would not be attainable if the *inter-cluster* interference was accounted for [8], [9]. Most work on the performance of IA, however, is confined to single-cluster performance analysis (see [6], [10]–[12] and references therein).

When dealing with large networks, a relevant metric of the system performance is the transmission capacity [13], defined as the number of successful transmission per unit area, subject to a constraint on outage probability. The transmission capacity, in contrast to other network-wide system performance metrics such as transport capacity, generally leads to closed-form expressions or tight bounds providing insight into the network design parameters [14]. To the best of our knowledge, little prior work on the network-wide performance of MIMO IA systems exists. In [15], the spatial distribution of nodes is taken into account for deriving the point-to-point outage probability but the accuracy of the acquired CSI is ignored. By assuming perfect CSI, however, as discussed in [6, Section V], the authors in [15] effectively favor IA over other transmission techniques which either do not require CSI at the transmitters or are less sensitive to CSI imperfections. Therefore, the goal of this paper is take into account both the node distribution and the CSI uncertainty to better understand the performance of MIMO IA in large decentralized networks.

### B. Contributions

In this paper, we find the transmission capacity of a large ad hoc network where nodes are partitioned into separate clusters each cooperating through IA. We assume a four-stage transmission protocol. In the first stage, with a finite length *training period*, imperfect CSI for the cross links is obtained through MMSE channel estimation. In the second stage, the estimated CSI is fed back to the other nodes in the cluster during the *feedback period*. In the third stage, the IA transmit/receive filters are computed. In the last stage, using the rest of the finite-length channel block, the nodes communicate data using a cluster-wise slotted Aloha-like channel access protocol where at random, all nodes in a cluster either transmit simultaneously or turn off their transmission. MIMO IA, as discussed in this paper, requires synchronization and coordination in each cluster and therefore a coordinated decision to transmit or not is reasonable.

Toward deriving the transmission capacity of this network, we first derive the exact point-to-point outage probability at a typical receiver. Then, assuming fixed feedback overhead, we solve for the optimum training period locally maximizing each cluster's goodput. Next, we derive the transmission capacity and compare it to a network with the same topology where only a single transmit/receiver pair in each cluster utilizes spatial multiplexing at each time instant.

Our results indicate that the transmission technique of choice is a function of the node density, the mobility of the nodes, the transmit power, and the characteristics of the underlying communications medium. For example, in dense networks with high transmit power, spatial multiplexing (SM) over an orthogonal channel access strategy such as time division multiple access (TDMA) can outperform IA due to lower inter-cluster interference. Also, the signal-to-noise-ratio (SNR) switching point between IA and TDMA+SM decreases with increasing density and mobility. Our initial work in [1] only deals with point-to-point outage probability. This paper

elaborates on the claims of [1], solves for the optimum training period, finds the corresponding transmission capacity of the network, provides easier to compute bounds in several important cases, and presents new simulation results.

### C. Organization and Notation

The remainder of the paper is organized as follows. In Section II we present the system model. In Section III we analyze the performance of intra-cluster MIMO IA through quantifying point-to-point outage probability and the transmission capacity. In Section IV we derive parallel performance metrics for the same network utilizing spatial multiplexing. We present numerical results in Section V followed by concluding remarks in Section VI. Also, Tables I and II present the notation used and the important symbols defined in this manuscript.

## II. SYSTEM MODEL

The spatial locations of the potential transmitters,  $\Phi$ , are modeled as a planar Neyman-Scott cluster point process [16]. In this process, the cluster centers are modeled by a parent homogeneous Poisson point process (PPP)  $\Phi_p$  of density  $\tilde{\lambda}_p$ . Each parent point  $x \in \Phi_p$  forms the center of a cluster around which  $K$  daughter points are uniformly distributed in a circle of radius  $R$ . The resulting process<sup>1</sup> is a stationary point process of density  $K\tilde{\lambda}_p$ . We assume clusters randomly access the channel with probability  $P_A$  effectively reducing the density of this PPP to  $\lambda_p = P_A\tilde{\lambda}_p$ . The receiver of a transmitter at  $x$  is denoted by  $\hat{x}$  and is assumed to be randomly located at distance  $D_r$  from its transmitter forming an  $N \times N$  MIMO link. The receivers are not part of the point process  $\Phi$ . An instance of the nodes' location is shown in Fig. 1.

In this paper, a typical transmitter (a transmitter chosen at random) is considered and its performance is analyzed. This transmitter is typical in the sense that the performance of IA in this node is a representative of the average performance of IA in the network [16], [17]. Since the underlying point process is stationary, without loss of generality, it can be assumed that the typical transmitter is at the origin<sup>2</sup>. Denote the cluster to which the transmitter at the origin belongs by  $\Psi_o$ . The received signal at receiver  $\hat{x}$ ,  $x \in \Psi_o$ , is

$$\mathbf{y}_{\hat{x}} = \sum_{z \in \Psi_o} \sqrt{g_{\hat{x}z}} \mathbf{H}_{\hat{x}z} \mathbf{F}_z \tilde{\mathbf{s}}_z + \mathbf{J}_c + \mathbf{u}_{\hat{x}}, \quad (1)$$

where  $\mathbf{J}_c = \sum_{z \in \Phi/\Psi_o} \sqrt{g_{\hat{x}z}} \mathbf{H}_{\hat{x}z} \mathbf{F}_z \tilde{\mathbf{s}}_z$  is the inter-cluster interference,  $g_{\hat{x}z}$  and  $\mathbf{H}_{\hat{x}z}$  represent the pathloss and the matrix of channel coefficients between the transmitter  $z$  and the receiver  $\hat{x}$ ,  $\mathbf{F}_z$  is the precoder at transmitter  $z$  with the transmitted signal  $\tilde{\mathbf{s}}_z$  such that  $\mathbb{E}\{\tilde{\mathbf{s}}_z^* \tilde{\mathbf{s}}_z\} = P$ , and  $\mathbf{u}_{\hat{x}} \sim \mathcal{CN}(\mathbf{0}, N_o \mathbf{I})$  is the additive white Gaussian noise. In this paper, it is assumed that  $\mathbf{F}_z^* \mathbf{F}_z = \mathbf{I}$ , because of tractability and the observation that the gain attained otherwise, such as with the MMSE algorithm in [19] or the Max-SINR algorithm in [18], is limited and confined to the low SNR regime where the inter-cluster interference is generally not dominant. In every cluster, channel state information is estimated at the receivers as in [20] and conveyed to all other nodes of the cluster using an error-free instantaneous feedback link. We propose to model the uncertainty in the MIMO channels using a Gauss-Markov model of the form [21], [22]

$$\mathbf{H}_{\hat{x}z} = \sqrt{1 - \beta_{\hat{x}z}^2} \mathbf{H}_{\hat{x}z}^w + \beta_{\hat{x}z} \mathbf{E}_{\hat{x}z} \quad x, z \in \Psi_o, \quad (2)$$

<sup>1</sup>The parent points  $\Phi_p$  will not be a part of the final point process.

<sup>2</sup>The whole point process can be translated so the randomly picked node is located at the origin.

where  $\mathbf{H}_{\hat{x}z}^w$  is the estimated channel,  $\mathbf{E}_{\hat{x}z}$  represents the estimation error with i.i.d. terms distributed as  $\mathcal{CN}(0, 1)$ , and  $\beta_{\hat{x}z}^2$  is the normalized variance of the estimation error. It is assumed that the channel is quasi-static block-fading such that  $\mathbf{H}$  is constant for a block duration of length  $T$  and then changes independently. Training, feedback, and data transmission are assumed to be all orthogonal in time, in the same coherence time or frame  $T$  [23]. Hence,  $\beta_{\hat{x}z}$  is set to be related to the average received SNR at each link,  $\gamma_{\hat{x}z}$ , as [20, Section II.B]

$$\beta_{\hat{x}z}^2 = \frac{1}{1 + \frac{T_t}{N} \gamma_{\hat{x}z}} = \frac{1}{1 + T_t \frac{\gamma_o g_{\hat{x}z}}{N}}, \quad (3)$$

where  $\gamma_o = \frac{P}{N_o}$  and  $T_t \geq KN$  is the number of channel instances spent for training  $\mathbf{H}_{\hat{x}z}$  [10]. For analytical tractability, it is also assumed that  $\mathbf{H}^w$  is used to construct the precoders/equalizers and nodes effectively ignore the imperfection in CSI in their design.

### III. INTRA-CLUSTER INTERFERENCE ALIGNMENT

At each cluster, a  $K$ -user system of IA is feasible if there exists a set of matrices  $\mathcal{W} = \{\mathbf{W}_{\hat{z}} | z \in \Psi_o\}$  such that, given the received signal of (1), the following constraints are met [5]:

$$\begin{cases} \text{rank}(\mathbf{W}_{\hat{x}} \mathbf{H}_{\hat{x}x} \mathbf{F}_x) = N_s & \forall x, z \in \Psi_o, \\ \mathbf{W}_{\hat{x}} \mathbf{H}_{\hat{x}z} \mathbf{F}_z = \mathbf{0} & \forall z \neq x \end{cases} \quad (4)$$

where  $\mathbf{W}_{\hat{x}}$  is the combining filter used at receiver  $\hat{x}$  and  $N_s$  is the number of interference-free streams each transmitter can send to its receiver. The linear equalizer presented in [6] and the projection matrix presented in [19, Section III.A] are examples of a possible receive filter in (4). It is assumed that the IA precoders are designed using the alternating minimization algorithm in [19, Section III.A] such that  $\mathbf{F}_x$  is independent of  $\mathbf{H}_{\hat{x}x}$  for all  $x \in \Psi$ . Also, it is assumed that the set of  $\{N, N_s, K\}$  constitutes a feasible IA system, which for the MIMO interference channel requires that  $2N - (K + 1)N_s \geq 0$ .

#### A. Characterizing the SINR

From (4), interference at receiver  $\hat{x}$  is confined to an  $N - N_s$  dimensional subspace. Let  $\{\{\cdot\}\}$  represent horizontal concatenation of the elements in  $\{\cdot\}$ . Then, as IA precoders/equalizers are constructed using  $\mathbf{H}^w$  as given by (2), the  $N \times (K-1)N_s$  matrix of  $\mathbf{J}_{\hat{x}} = [\{\mathbf{H}_{\hat{x}z}^w \mathbf{F}_z | z \neq x, z \in \Psi_o\}]$  spans an  $N - N_s$  dimensional subspace. Let the singular value decomposition of  $\mathbf{J}_{\hat{x}}$  be  $\mathbf{U}_{\mathbf{J}_{\hat{x}}} \Sigma_{\mathbf{J}_{\hat{x}}} \mathbf{V}_{\mathbf{J}_{\hat{x}}}^*$  and let the rows of  $\mathbf{W}_{\hat{x}}$  be the columns of  $\mathbf{U}_{\mathbf{J}_{\hat{x}}}$  corresponding to zero singular values in  $\Sigma_{\mathbf{J}_{\hat{x}}}$ . As  $\mathbf{W}_{\hat{x}}$  is independent of  $\mathbf{H}_{\hat{x}x}^w \mathbf{F}_x$ , it satisfies the conditions in (4) and is a valid zero-forcing (ZF) equalizer for IA. Using this ZF receiver, the post-processing SINR of the  $n$ th stream at receiver  $\hat{x}$  is

$$\gamma_{\hat{x},n}^{\text{IA}} = \frac{g_{\hat{x}x} (1 - \beta_{\hat{x}x}^2) \tilde{\mathbf{h}}_{\hat{x}x}^* \tilde{\mathbf{h}}_{\hat{x}x}}{\underbrace{\frac{N_s}{\gamma_o} + \underbrace{g_{\hat{x}x} \beta_{\hat{x}x}^2 \tilde{\mathbf{e}}_{\hat{x}x}^* \tilde{\mathbf{e}}_{\hat{x}x}}_{I_s}} + \underbrace{\sum_{z \in \Psi_o/x} g_{\hat{x}z} \beta_{\hat{x}z}^2 \tilde{\mathbf{e}}_{\hat{x}z}^* \tilde{\mathbf{e}}_{\hat{x}z}}_{I_e} + \underbrace{\sum_{z \in \Phi/\Psi_o} g_{\hat{x}z} \tilde{\mathbf{h}}_{\hat{x}z}^* \tilde{\mathbf{h}}_{\hat{x}z}}_{I_i}}, \quad (5)$$

where for all  $z \in \Phi$ ,  $\tilde{\mathbf{h}}_{\hat{x}z} = (\mathbf{e}_n^* \mathbf{W}_{\hat{x}} \mathbf{H}_{\hat{x}z}^w \mathbf{F}_z)^*$ ,  $\tilde{\mathbf{e}}_{\hat{x}z} = (\mathbf{e}_n^* \mathbf{W}_{\hat{x}} \mathbf{E}_{\hat{x}z} \mathbf{F}_z)^*$ , and  $\mathbf{e}_n$  is the  $n$ th column of an  $N_s \times N_s$  identity matrix. Note that  $I_s$  represents the residual error from the direct link and, as the distance between the transmitter and receiver is constant, its pathloss (and the error variance) are not random variables and so it is separated from  $I_e$  to emphasize this point. Let the entries

of  $\mathbf{H}_{\hat{x}z}$  and  $\mathbf{E}_{\hat{x}z}$  be i.i.d. Gaussian terms. As  $\mathbf{W}_{\hat{x}}$  and  $\mathbf{F}_z$  are unitary matrices independent of  $\mathbf{H}_{\hat{x}z}$  and  $\mathbf{E}_{\hat{x}z}$ , due to the doubly unitarily invariance of the Gaussian distribution,  $\tilde{\mathbf{h}}_{\hat{x}z}$  and  $\tilde{\mathbf{e}}_{\hat{x}z}$  will be column vectors of length  $N_s$  with i.i.d. Gaussian terms and independent of each other (as  $\mathbf{H}_{\hat{x}z}$  and  $\mathbf{E}_{\hat{x}z}$  are independent of each other). Note that (5) is in fact independent of the stream index  $n$ .

### B. Probability of Outage

In (5), since  $\tilde{\mathbf{h}}_{\hat{x}z}^* \tilde{\mathbf{h}}_{\hat{x}z}$  and  $\tilde{\mathbf{e}}_{\hat{x}z}^* \tilde{\mathbf{e}}_{\hat{x}z}$  are i.i.d.  $\Gamma(N_s, 1)$  random variables, we denote them both by  $h_{\hat{x}z}$  for notational simplicity. Denote the typical transmitter at the origin by  $o$ . Considering a transmitter at the origin is equivalent to conditioning on the existence of a point at the origin. Since every point belongs to some cluster, conditioning on the existence of a point at the origin equals the presence of a cluster with a daughter point at the origin. Since the parent point process is a PPP, an additional cluster with a daughter point at the origin can be added to it without changing the statistics of the other points of the process. Succinctly, the Palm probability of a Neyman-Scott cluster process is  $\mathbb{P}^o = \mathbb{P} * \Psi_o$  where  $*$  denotes superposition [16]. This implies that assuming a point of the cluster process at the origin equals the original point process  $\Phi$  plus an additional cluster which has a point at the origin. Also this additional cluster at the origin  $\Psi_o$  is independent of the original process  $\Phi$ . Since the tagged transmitter at origin does not contribute to the interference at the receiver, it is convenient to use the reduced Palm probability denoted by  $\mathbb{P}^{!o}$  instead of Palm probability. Reduced Palm probability is similar to Palm probability, except that the point at the origin is not considered in the computation of the probability and hence  $\mathbb{P}^{!o} = \mathbb{P} * \{\Psi_o \setminus \{o\}\}$ . From (5), the probability of success is therefore  $\mathbb{P}_s^{\text{IA}}(\theta) = \mathbb{P}^{!o}(\gamma_{\hat{o},n}^{\text{IA}} > \theta)$ , where  $\theta$  is the SINR threshold.

**Theorem 1.** *For the system model described in Section II, the success probability when each cluster uses IA is*

$$\mathbb{P}_s^{\text{IA}}(\theta) = \sum_{k=0}^{N_s-1} \frac{(-\eta)^k}{k!} \frac{d^k}{ds^k} e^{-s \frac{N_s}{\gamma_o}} \left( \frac{\frac{T_t \gamma_o}{N} + D_r^\alpha}{s + \frac{T_t \gamma_o}{N} + D_r^\alpha} \right)^{N_s} \mathcal{L}_{\mathbf{I}_e}^{!o}(s) \mathcal{L}_{\mathbf{I}_i}(s) \Big|_{s=\eta}, \quad (6)$$

where  $\mathcal{L}_{\mathbf{I}_e}^{!o}(s)$  and  $\mathcal{L}_{\mathbf{I}_i}(s)$  are Laplace transforms of intra and inter-cluster interference given by

$$\mathcal{L}_{\mathbf{I}_e}^{!o}(s) = \frac{1}{\pi R^2} \int_{B(o,R)} \left[ \frac{1}{\pi R^2} \int_{B(o,R)} \left( \frac{\frac{T_t \gamma_o}{N} + \|x-y-\hat{o}\|^\alpha}{s + \frac{T_t \gamma_o}{N} + \|x-y-\hat{o}\|^\alpha} \right)^{N_s} dx \right]^{K-1} dy, \quad (7)$$

$$\mathcal{L}_{\mathbf{I}_i}(s) = \exp \left( -\lambda_p \int_{\mathbb{R}^2} 1 - \left[ \frac{1}{\pi R^2} \int_{B(o,R)} \left( \frac{\|x-y\|^\alpha}{s + \|x-y\|^\alpha} \right)^{N_s} dx \right]^K dy \right), \quad (8)$$

where  $\eta = \frac{\theta}{g_{\hat{o}o}(1-\beta_{\hat{o}o}^2)}$ .

*Proof:* See Appendix A. ■

Note that the integrals in (7) and (8) can be computed by switching to polar coordinates and exchanging the order of differentiation and integration<sup>3</sup>. Although the probability of successful transmission given by (6) is in closed-form, numerically computing it, especially for  $N_s >$

<sup>3</sup>The integral in (7) is always finite since the domain of integration is a bounded set. The integral in (8) can be shown to be finite using the fact that  $\|x-y\| \geq \|y\| - R$  for the inner integral and  $1 - x^m \leq \exp(-x^m)$ ,  $m \geq 1$  for the outer integral.

1 where differentiation is required, is not trivial. Lemma 1 provides bounds on the Laplace transforms in (7) and (8) that are easily computable. Also, Lemma 2 can be used to avoid the differentiation in (6) when  $N_s > 1$ .

**Lemma 1.** For  $K > 2$

$$\mathcal{L}_{\mathbf{I}_i}(s) \leq \exp\left(-\lambda_p s^{2/\alpha} \frac{\Gamma(KN_s + 2/\alpha)\Gamma(1 - 2/\alpha)}{\Gamma(KN_s)}\right) \quad (9)$$

$$\mathcal{L}_{\mathbf{I}_e}^{\text{lo}}(s) \leq \frac{1}{2\pi} \int_{\theta=0}^{2\pi} \int_{x=0}^{2R} \frac{h_r(x)}{(1 + sf(\sqrt{x^2 + D_r^2} - 2xD_r \cos(\theta)))^{N_s(K-1)}} dx d\theta, \quad (10)$$

where

$$h_r(x) = \begin{cases} \frac{x}{R^2} I_{1-\frac{x^2}{4R^2}}(3/2, 1/2) & x \leq 2R \\ 0 & \text{Otherwise} \end{cases} \quad (11)$$

and  $I_y(a, b)$  is the regularized incomplete beta function.

*Proof:* See Appendix B. ■

**Lemma 2.** If  $\eta > (N_s - 1)/e$ , the success probability is bounded by

$$P_s^{\text{IA}} \leq \sum_{k=0}^{N_s-1} \frac{\eta^k}{k!} e^{-(\eta-k/e)\frac{N_s}{\gamma_0}} \mathbb{E} e^{-(\eta-k/e)g_{\text{oo}}\beta_{\text{oo}}^2 h_{\text{oo}}} \mathcal{L}_{\mathbf{I}_e}^{\text{lo}}(\eta - k/e) \mathcal{L}_{\mathbf{I}_i}(\eta - k/e). \quad (12)$$

*Proof:* See Appendix C. ■

For  $N_s > 1$ , unlike (6), (12) does not have a differentiation operator. Note that the condition on  $N_s$  in Lemma 2 is equivalent to  $N_s < 1 + \frac{e\theta D_r^\alpha}{1-\beta_{\text{oo}}^2}$  which is true for practical  $N_s$  and typical operating regime where  $\theta \gg 1$ ,  $D_r > 1$ , and  $\beta_{\text{oo}}^2 \ll 1$ . Also, results from Lemma 1 can be used to further simplify the expression in Lemma 2.

### C. Optimizing the Training Period

For a given block fading of length  $T$ , the transmitters spend  $T_t \geq KN$  channel instances for training the links. We also assume a perfect analog feedback link where the receivers send the trained channels over a period of  $T_f = K^2N$  channel instances to the transmitters. In practice, the transmitters select  $T_t$  to optimize some performance criteria. In this paper, we assume transmitters use the goodput at each cluster defined as

$$\begin{aligned} \hat{T}_t &= \arg \max_{T_t} \frac{T - K^2N - T_t}{T} KN_s P_s^{\text{IA}}(\theta) \log_2(1 + \theta) \\ \text{s.t. } & KN \leq T_t \leq T - K^2N, \end{aligned} \quad (13)$$

where  $\frac{T-K^2N-T_t}{T}$  accounts for the transmission opportunities lost due to overhead,  $KN_s$  is the total number of streams in each cluster, and  $P_s^{\text{IA}}(\theta) \log_2(1 + \theta)$  is the rate multiplied by the times the connection exists, i.e. SINR passes the threshold  $\theta$ . Note that in (13),  $P_s^{\text{IA}}(\theta)$  is implicitly a function of the training period  $T_t$ .

For a given node mobility and hence a Doppler frequency  $f_d \approx \frac{1}{T}$ , the training period can be written as a fraction of the total block length  $T$ , i.e.  $T_t = \delta T = \delta \frac{1}{f_d}$ . Therefore, the optimization problem of (13) can be rewritten as

$$\begin{aligned} \hat{\delta} &= \arg \max_{\delta} (1 - \delta - f_d K^2 N) K N_s P_s^{\text{IA}}(\theta) \log_2(1 + \theta) \\ \text{s.t. } & f_d K N \leq \delta \leq (1 - f_d K^2 N), \end{aligned} \quad (14)$$

where

$$P_s^{\text{IA}} = \sum_{k=0}^{N_s-1} \frac{(-\eta)^k}{k!} \frac{d^k}{ds^k} e^{-s \frac{N_s}{\gamma_o}} \left( \frac{\frac{\delta \gamma_o}{N} + f_d D_r^\alpha}{\frac{\delta \gamma_o}{N} + f_d(s + D_r^\alpha)} \right)^{N_s} \mathcal{L}_{I_e}^{\text{!o}}(s, f_d) \mathcal{L}_{I_i}(s) \Big|_{s=\eta},$$

$$\eta = \frac{\theta D_r^\alpha (\gamma_o \delta + N D_r^\alpha f_d)}{\gamma_o \delta}, \mathcal{L}_{I_e}^{\text{!o}}(s, f_d) = \frac{1}{\pi R^2} \int_{B(o,R)} \left[ \frac{1}{\pi R^2} \int_{B(o,R)} \frac{\frac{\delta \gamma_o}{N} + f_d \|x-y-\hat{o}\|^\alpha}{\frac{\delta \gamma_o}{N} + f_d (s + \|x-y-\hat{o}\|^\alpha)} dx \right]^{K-1} dy, \text{ and } \mathcal{L}_{I_i}(s) \text{ is given}$$

by (8).

With a convex relaxation on  $T_t$  (and therefore  $\delta$ ) to change its domain to the real numbers, the optimization problem of (14) is convex and solvable with any of the numerical optimization algorithms [24]. Note that, although complicated, the derivative of the objective function in (14) w.r.t  $\delta$  is computable and evaluating the objective function or its derivatives for any set of values is possible. Next we provide approximate closed form solutions for some common cases.

1) *Single stream from each transmitter:* When  $N_s = 1$ ,  $P_s^{\text{IA}}$  in (14) simplifies to

$$P_s^{\text{IA}}(N_s = 1) = e^{-\eta \frac{N_s}{\gamma_o}} \frac{\frac{\delta \gamma_o}{N} + f_d D_r^\alpha}{\frac{\delta \gamma_o}{N} + f_d (\eta + D_r^\alpha)} \mathcal{L}_{I_e}^{\text{!o}}(\eta, f_d) \mathcal{L}_{I_i}(\eta). \quad (15)$$

The highly non-linear dependency of (15) on  $\delta$  can be converted into a polynomial one following the Taylor expansion approximation method proposed in [25]. Let  $g(\delta, f_d)$  be the objective function in (14). Rewrite  $g(\delta, f_d)$  as its Taylor series around  $f_d = 0$  (infinite block length and hence perfect training) keeping all the other variables constant

$$\begin{aligned} \hat{\delta} &\approx \arg \max_{\delta} g(\delta, f_d = 0) + \frac{\partial g}{\partial f_d} \Big|_{f_d=0} f_d + \frac{1}{2} \frac{\partial^2 g}{\partial f_d^2} \Big|_{f_d=0} f_d^2 \\ \text{s.t. } & f_d K N \leq \delta \leq (1 - f_d K^2 N). \end{aligned} \quad (16)$$

Simplifying the terms in (16) and removing the constant scaling coefficients from the optimization problem yields

$$\begin{aligned} \hat{\delta}_1 &\approx \arg \max_{\delta} -\delta + C_1 \frac{1}{\delta} + C_2 \frac{1}{\delta^2} \\ \text{s.t. } & f_d K N \leq \delta \leq (1 - f_d K^2 N), \end{aligned} \quad (17)$$

where  $C_1$  and  $C_2$  are given in Appendix D. Let  $\delta_1$  be the relevant root of the first derivative of (17). The optimum training period will be given by

$$\hat{T}_{t,1} = \min(\max(KN, [\delta_1 T]), T - K^2 N). \quad (18)$$

When  $\frac{T_t}{N} g_{\delta o} \gamma_o = \frac{T_t}{N} \frac{\gamma_o}{D_r^\alpha} \gg 1$ , which is true for high SNR,  $\beta_{\delta o} \approx 0$  which implies that  $\eta \approx \theta D_r^\alpha$ . In this case,  $\mathcal{L}_{I_i}(\eta)$  is a strictly positive function independent of  $T_t$  and (17) simplifies to

$$\begin{aligned} \hat{\delta}_2 &\approx \arg \max_{\delta} -\delta + D_1 \frac{1}{\delta} + D_2 \frac{1}{\delta^2} \\ \text{s.t. } & f_d K N \leq \delta \leq (1 - f_d K^2 N), \end{aligned} \quad (19)$$

where  $D_1$  and  $D_2$  are given in Appendix E. Similar to  $\hat{T}_{t,1}$ , let  $\delta_2$  be the relevant root of the first derivative of (19). The optimum training period will be given by

$$\hat{T}_{t,2} = \min \left( \max (KN, [\delta_2 T]), T - K^2 N \right). \quad (20)$$

2) *Greater than one stream from each transmitter:* For  $N_s > 1$ , the same approach taken to derive (18) and (20) can be used to approximately solve for the optimum training period. Alternatively, bounds given in (9), (10), and Lemma 2 can be used to derive simpler (and less accurate) results. We believe the involved expressions in this case do not provide additional insight into the problem at this point and are not presented here.

#### D. Transmission Capacity

Let  $q(\lambda_p) = 1 - P_s^{\text{IA}}(\theta) = \epsilon$ . Accounting for overhead, the normalized transmission capacity is

$$C(\epsilon) = \frac{T - K^2 N - \hat{T}_t}{T} q^{-1}(\epsilon) K (1 - \epsilon). \quad (21)$$

With the probability of successful transmission as given by (6), the exact expression of  $q^{-1}(\epsilon)$  for general  $N_s$  is not analytically tractable. Next we give its exact expression for the case of  $N_s = 1$  and provide a bound for general  $N_s > 1$ .

1) *Single stream from each transmitter:* Using (6)

$$q(\lambda_p) = 1 - e^{-\hat{\eta} \frac{N_s}{\gamma_o}} \left( \frac{\frac{\hat{T}_t \gamma_o}{N} + D_r^\alpha}{s + \frac{\hat{T}_t \gamma_o}{N} + D_r^\alpha} \right) \mathcal{L}_{I_e}^{\text{lo}}(\hat{\eta}) \mathcal{L}_{I_i}(\hat{\eta}) = \epsilon, \quad (22)$$

where  $\hat{\eta} = \eta|_{T_t = \hat{T}_t}$ . Substituting (7) and (8) into (22) gives

$$q^{-1}(\epsilon) = \lambda_p^\epsilon = \frac{\log_e \left( \frac{e^{-\hat{\eta} \frac{N_s}{\gamma_o}}}{1 - \epsilon} \frac{\frac{\hat{T}_t \gamma_o}{N} + D_r^\alpha}{\frac{\hat{T}_t \gamma_o}{N} + \hat{\eta} + D_r^\alpha} \frac{1}{\pi R^2} \int_{B(o,R)} \left[ \frac{1}{\pi R^2} \int_{B(o,R)} \left( \frac{\frac{\hat{T}_t \gamma_o}{N} + \|x-y-\delta\|^\alpha}{\hat{\eta} + \frac{\hat{T}_t \gamma_o}{N} + \|x-y-\delta\|^\alpha} \right)^{N_s} dx \right]^{K-1} dy \right)}{\int_{\mathbb{R}^2} 1 - \left[ \frac{1}{\pi R^2} \int_{B(o,R)} \left( \frac{\|x-y\|^\alpha}{\hat{\eta} + \|x-y\|^\alpha} \right)^{N_s} dx \right]^K dy}, \quad (23)$$

where  $\hat{T}_t$  is the optimum training period obtained earlier. Using (23), (21) can be computed.

2) *Greater than one stream from each transmitter:* When  $\hat{\eta} > (N_s - 1)/e$ , using Lemma 2

$$q(\lambda_p) \geq 1 - \sum_{k=0}^{N_s-1} \frac{\hat{\eta}^k}{k!} e^{-(\hat{\eta}-k/e) \frac{N_s}{\gamma_o}} \mathbb{E} e^{-(\hat{\eta}-k/e) g_{\delta o} \beta_{\delta o}^2 h_{\delta o}} \mathcal{L}_{I_e}^{\text{lo}}(\hat{\eta} - k/e) \mathcal{L}_{I_i}(\hat{\eta} - k/e). \quad (24)$$

Note that the expression for the optimum training when  $N_s > 1$  was not presented in Section III-C2. Now, if  $\hat{\eta} \gg k/e$  for all  $k \in \{0, \dots, N_s - 1\}$ ,  $\hat{\eta} - k/e \approx \hat{\eta}$  and hence (24) equals

$$\begin{aligned} q(\lambda_p) &\geq 1 - \left( \sum_{k=0}^{N_s-1} \frac{\hat{\eta}^k}{k!} \right) e^{-\hat{\eta} \frac{N_s}{\gamma_o}} \mathbb{E} e^{-\hat{\eta} g_{\delta o} \beta_{\delta o}^2 h_{\delta o}} \mathcal{L}_{I_e}^{\text{lo}}(\hat{\eta}) \mathcal{L}_{I_i}(\hat{\eta}) \\ \Rightarrow \mathcal{L}_{I_i}(\hat{\eta}) &\geq \frac{1 - \epsilon}{\left( \sum_{k=0}^{N_s-1} \frac{\hat{\eta}^k}{k!} \right) e^{-\hat{\eta} \frac{N_s}{\gamma_o}} \mathbb{E} e^{-\hat{\eta} g_{\delta o} \beta_{\delta o}^2 h_{\delta o}} \mathcal{L}_{I_e}^{\text{lo}}(\hat{\eta})} \end{aligned}$$

$$\Rightarrow \lambda_p^\epsilon \leq \frac{\log_e \left( \frac{1}{1-\epsilon} \left( \sum_{k=0}^{N_s-1} \frac{\hat{\eta}^k}{k!} \right) e^{-\hat{\eta} \frac{N_s}{\gamma_o}} \mathbb{E} e^{-\hat{\eta} g_{\hat{o}o} \beta_{\hat{o}o}^2 h_{\hat{o}o}} \mathcal{L}_{\mathbf{I}_e}^{\text{!o}}(\hat{\eta}) \right)}{\int_{\mathbb{R}^2} \left[ 1 - \frac{1}{\pi R^2} \int_{B(o,R)} \left( \frac{\|x-y\|^\alpha}{\hat{\eta} + \|x-y\|^\alpha} \right)^{N_s} dx \right]^K dy}. \quad (25)$$

To simplify the integrations in (25), one could also use the bound given by (10) for  $\mathcal{L}_{\mathbf{I}_e}^{\text{!o}}(\hat{\eta})$  or use a similar approach taken to obtain (9) to find an appropriate bound for the denominator of (25).

#### IV. SPATIAL MULTIPLEXING

An alternative strategy to IA is spatial multiplexing with an orthogonal multiple access technique, e.g. TDMA. We analyze open-loop spatial multiplexing where at each cluster only a single transmit/receiver pair communicate  $N$  streams without precoding. In short, we refer to TDMA+SM as SM. For a fair comparison with IA, we assume a similar CSI imperfection as in (2). Therefore, the signal at a typical receiver can be written as

$$\mathbf{y}_{\hat{x}} = g_{\hat{x}\hat{x}} \left( \sqrt{(1-\beta_{\hat{x}\hat{x}}^2)} g_{\hat{x}\hat{x}} \mathbf{H}_{\hat{x}\hat{x}}^w + \beta_{\hat{x}\hat{x}} \mathbf{E}_{\hat{x}\hat{x}} \right) \tilde{\mathbf{s}}_x + \mathbf{J}_c + \mathbf{u}_{\hat{x}},$$

where  $\mathbf{J}_c$  is defined in (1) with the difference that each cluster only has a single transmitter. In this case, the point process of the transmitters simplifies to the Poisson point process of the parent points. When  $\beta_{\hat{x}\hat{x}} = 0$ , after a zero-forcing receiver based on  $\mathbf{H}_{\hat{x}\hat{x}}$ , the SINR of the  $n$ th stream at a typical receiver can be written as

$$\gamma_{\hat{x},n}^{\text{SM}} = \frac{g_{\hat{x}\hat{x}}}{\mathbf{e}_n^* (\mathbf{H}_{\hat{x}\hat{x}}^w)^{-1} \left( \frac{N}{\gamma_o} \mathbf{I}_N + \sum_{z \in \Phi/\Psi_o} g_{\hat{x}z} \mathbf{H}_{\hat{x}z} \mathbf{H}_{\hat{x}z}^* \right) ((\mathbf{H}_{\hat{x}\hat{x}}^w)^{-1})^* \mathbf{e}_n}, \quad (26)$$

where we have assumed the interference values from different transmitters are independent. The probability of success is given by [26]

$$P_s^{\text{SM}}(\theta) = \mathbb{P}^{\text{!o}}(\gamma_{\hat{x},n}^{\text{SM}} > \theta) = \exp \left( -\lambda_p \theta^{\frac{2}{\alpha}} D_r^2 \mathcal{J} - \theta D_r^\alpha \frac{N}{\gamma_o} \right), \quad (27)$$

where  $\mathcal{J} = \frac{\pi \Gamma(N + \frac{2}{\alpha}) \Gamma(1 - \frac{2}{\alpha})}{\Gamma(N)}$ . With imperfect CSI, we assume the receivers compute their zero-forcing receivers based on the estimated channel values and ignore the estimation error. Then, the denominator of (26) will have an additional term of  $g_{\hat{x}\hat{x}} \beta_{\hat{x}\hat{x}}^2 \mathbf{E}_{\hat{x}\hat{x}} \mathbf{E}_{\hat{x}\hat{x}}^*$  inside the parenthesis. As the channel estimation error is independent of the estimated channel and the transmitter/receiver distance is fixed to  $D_r$ , this additional term is effectively increasing the noise spectral density from  $N_o$  to  $N_o + P g_{\hat{x}\hat{x}} \beta_{\hat{x}\hat{x}}^2$ . As the numerator of (26) also changes to  $g_{\hat{x}\hat{x}} (1 - \beta_{\hat{x}\hat{x}}^2)$ , in case of imperfect CSI as in (2), the probability of success at a typical receiver for any of the streams changes from (27) to

$$P_s^{\text{SM}}(\theta) = \mathbb{P}^{\text{!o}}(\gamma_{\hat{x},n}^{\text{SM}} > \theta) = \exp \left( -\lambda_p \tilde{\theta}^{\frac{2}{\alpha}} D_r^2 \mathcal{J} - \tilde{\theta} D_r^\alpha \frac{N \tilde{N}_o}{P} \right), \quad (28)$$

where  $\tilde{\theta} = \frac{\theta}{1 - \beta_{\hat{x}\hat{x}}^2}$  and  $\tilde{N}_o = N_o + P g_{\hat{x}\hat{x}} \beta_{\hat{x}\hat{x}}^2$ .

### A. Optimum Training Period

Similar to the IA case, each node selects a training period that optimizes its own (its own cluster's) goodput. Unlike IA, however, here the minimum training period for each transmit/receive pair is  $N$  and no resources are spent for feedback. Hence

$$\begin{aligned} \hat{T}_t^{\text{SM}} &= \underset{T_t}{\operatorname{argmax}} \quad \frac{T - T_t}{T} NP_s^{\text{SM}}(\theta) \log_2(1 + \theta) \\ \text{s.t.} \quad & N \leq T_t \leq T. \end{aligned} \quad (29)$$

It is possible to show that the optimization problem of (29), after a convex relaxation on  $T_t$ , is also a convex problem and any of the numerical optimization algorithms [24] can be used to solve for the optimum  $\hat{T}_t^{\text{SM}}$ . Nevertheless, similar to the optimization problem of (14), we first replace the objective function of (29) with its second order Taylor expansion around  $f_d = 0$  (infinite block length and hence perfect training) holding all the other variables constant, and then obtain an approximate closed-form solution for  $\hat{T}_t^{\text{SM}}$

$$\begin{aligned} \hat{\delta}^{\text{SM}} &= \underset{\delta^{\text{SM}}}{\operatorname{argmax}} \quad (1 - \delta) NP_s^{\text{SM}}(\theta) \log_2(1 + \theta) \\ \text{s.t.} \quad & N \leq T_t \leq T. \end{aligned} \quad (30)$$

Removing the constant coefficients and simplifying the terms, the new optimization problem derived from the 2nd order Taylor series expansion of the objective function in (30) is

$$\begin{aligned} \hat{\delta}^{\text{SM}} &\approx \underset{\delta}{\operatorname{argmax}} \quad -\delta + B_1 \frac{1}{\delta} + B_2 \frac{1}{\delta^2} \\ \text{s.t.} \quad & f_d N \leq \delta \leq 1, \end{aligned} \quad (31)$$

where  $B_1$  and  $B_2$  are given in Appendix F. Similar to the IA case, let  $\delta_3$  be the relevant root of the first derivative of (31). The optimum training period will be given by

$$\hat{T}_t^{\text{SM}} = \min(\max(N, [\delta_3 T]), T). \quad (32)$$

### B. Transmission Capacity

For a given block length  $T$ , accounting for overhead, the normalized transmission capacity is

$$c(\epsilon) = \frac{T - \hat{T}_t^{\text{SM}}}{T} N \lambda_p^\epsilon (1 - \epsilon). \quad (33)$$

Using (28),  $\lambda_p^\epsilon$  is found as

$$\lambda_p^\epsilon = \frac{N}{\tilde{\theta}^{2/\alpha} D_r^2 \mathcal{J}} \left( \log \left( \frac{1}{1 - \epsilon} \right) - \frac{\tilde{\theta} D_r^\alpha N \tilde{N}_o}{P} \right), \quad (34)$$

where in computing  $\tilde{\theta}$  and  $\tilde{N}_o$ , the optimum training period,  $\hat{T}_t^{\text{SM}}$ , is used.

## V. NUMERICAL EXAMPLES AND DISCUSSION

In this section, we present numerical examples demonstrating both the accuracy of the developed theory for analyzing large IA networks and the effectiveness of such analysis in discovering operating regimes where IA outperforms other transmission schemes. For simplicity, we focus on the case of  $K = 3$ ,  $N = 2$ , and  $N_s = 1$ . To further reduce the number of involved variables, we set  $R$  to  $D_c = 0.5\lambda_p^{0.5}$  (average distance between the cluster centers) and set  $D_r$  to  $\frac{R}{5}$ . In this way, by increasing  $D_c$ , all the nodes move away from each other. We set  $N_o$  to 1 and unless otherwise stated, it is also assumed that all the clusters are transmitting simultaneously. In the forthcoming discussions, perfect channel estimation corresponds to very large values of  $T_t$  where  $\beta$  is assumed to be 0 and the worst channel estimation corresponds to assigning the least number of channel instances required for training the cross links in an IA cluster, i.e.  $T_t = KN$ . Note that for perfect CSI, (38) and (7) are both equal to 1.

**Probability of successful transmission with perfect CSI** Assume perfect channel state information and  $\gamma_o = 30$  dB.  $P_s^{\text{IA}}$  and  $P_s^{\text{SM}}$  given by (6) and (27) as a function of the SINR threshold for three values of the average distance between the cluster centers is shown in Fig. 2. As can be seen, for very dense networks, SM outperforms IA which can be explained by the reduced inter-cluster interference due to only a single transmit/receiver pair being active at each cluster. In addition, for moderately dense networks, IA outperforms SM only at high  $\theta$  which highlights the nonlinear nature of such comparisons. For very low node density, IA has better performance than SM when the increased distance between transmit/receive pairs does not enforce a zero probability of successful transmission.

**Probability of successful transmission with imperfect CSI** Now assume the worst channel estimation error variance for IA with the same  $\gamma_o = 30$  dB. The questions of interest are i) how much the performance of IA will be affected by introducing the imperfect CSI and ii) how does the relative performance between IA and SM change in this scenario. For the same values of  $D_c$  and  $\theta$  as in Fig. 2, the probabilities of successful transmission for IA and SM when the training period is lowered to  $KN = 6$  is shown in Fig. 3.  $P_s^{\text{IA}}$  and  $P_s^{\text{SM}}$  remain relatively unchanged for dense networks but are reduced for moderate and high node densities. Compared to the perfect CSI case, SM now outperforms IA for a larger range of  $\theta$  for moderate node densities. Also, IA has lost most of its advantages at moderate node densities (from maximum of 0.1 to 0.06) and the gap between IA and SM has decreased for the widely dispersed network. As a reference, the error of (6) when  $T_t = 6$  compared with numerical results for the curves shown in Fig. 3, in the worst case, is less than 0.0035.

**Optimum training period** Again, assume  $\gamma_o = 30$  dB. The optimum training period as obtained using (18) and (32) for IA and SM together with the corresponding optimum values found through numerically optimizing (13) and (29) for  $\theta = 20$  dB and two values of cluster radius (and hence average cluster center distance) as a function of total frame length,  $T$ , is shown in Fig. 4. As can be seen, (18) and (32) can be used to accurately find the optimum training period for a wide range of node density and total frame length. As expected, IA requires more training than SM but surprisingly, the ratio of the optimum training periods between the two transmission techniques seems to be a constant value independent of the total block length,  $T$ , which requires further analysis not in the scope of this work. Also, as the Taylor expansions used to derive (4) and (18) are at  $T = \infty$ , small inaccuracies at small frame lengths is expected. Moreover, the numerical errors in numerically calculating the integrals of the coefficients in (18) become harder to confine as  $R$  (and hence the integration ranges) increases which explains the deviation of the results obtained using (18) from the true optimum points for  $R = 5$ .

**Transmission capacity** Now fix the SINR threshold in (23) and (34) to  $\theta = 17$  dB and fix the maximum tolerable outage probability to  $\epsilon = 0.1$ . The maximum cluster density,  $\lambda_p^\epsilon$ , as given by (23) and (34) and the corresponding transmission capacities given by (22) and (33) for two values of cluster radius,  $R$ , and two values of the total block length,  $T$ , as a function of transmit SNR,  $\gamma_o$ , are shown in Fig. 5 and Fig. 6. Note that for  $R = 5$  only  $T = 200$  curves are shown as  $T = 1000$  curves were not conveying any new observations. Also, for both IA and SM, optimum training periods are used to compute the transmission capacity. As seen in Fig. 5, except for the high SNR values, increasing  $T$  allows for a better channel estimate at the receivers and consequently each receiver can tolerate more inter-cluster interference and hence a large cluster density. Also as the radius of each cluster increases, the channel estimation errors increase and the maximum tolerable cluster density decreases. Although by increasing  $T$ , the crossing point between IA and SM for  $\lambda_p^\epsilon$  shifts to a lower SNR value, the corresponding crossing point for the transmission capacity shifts to higher SNR values which highlights the tradeoff between a higher obtainable throughput by increasing  $K$  in each cluster and the accompanying higher accumulated non-aligned interference at each receiver due to the imperfect CSI estimation.

## VI. CONCLUSION

In this paper, we obtained an expression for the probability of successful transmission in a clustered MIMO IA network when the communication takes place over a Rayleigh fading quasi-static block-fading channel of a finite length and the impact of channel estimation on the accuracy of the obtained CSI is taken into account. We formulated an optimization of the training period length and explicitly solved it for the case of a single stream from each transmitter. We then presented the exact transmission capacity for the same case of a single stream from each transmitter and also provided an upper bound for the general case. Through simulations, we showed that probability of successful transmission and transmission capacity of TDMA+SM can be higher than IA in dense networks or when the mobility of the nodes is high.

Our analyses in this paper were based on a generic IA filter design approach where the only goal of the transmitters is to align the interference without optimizing the direct links. Any such variations of the IA algorithm can potentially increase the gains of IA but quantifying such transmission techniques are out of the scope of this work. Moreover, we assumed a perfect analog feedback link for IA where only the minimum required channel resources are used to relay the exact estimated channel values at the receivers to the rest of the nodes in each cluster. By using the results of [10], it is possible to include the analog feedback analysis but doing so would result in more involved expressions not benefiting the current scope of this work. Other optimization problems, such as optimizing the SINR threshold for the maximum tolerable outage probability, can be readily defined using the results of this paper.

## APPENDIX A PROOF OF THEOREM 1

Since  $h_{\hat{o}o} \sim \Gamma(N_s, 1)$ , its complementary cumulative distribution function is  $F(x) = \frac{\Gamma(N_s, x)}{\Gamma(N_s)} = e^{-x} \sum_{k=0}^{N_s-1} \frac{x^k}{k!}$ . Hence,

$$P_s^{\text{IA}}(\theta) = \mathbb{P}^{\text{lo}} \left( \frac{g_{\hat{o}o}(1 - \beta_{\hat{o}o}^2)h_{\hat{o}o}}{\frac{N_s}{\gamma_o} + I_s + I_e + I_i} > \theta \right) = \mathbb{P}^{\text{lo}} \left( h_{\hat{o}o} > \eta \left( \frac{N_s}{\gamma_o} + I_s + I_e + I_i \right) \right)$$

$$\stackrel{(a)}{=} \sum_{k=0}^{N_s-1} \frac{\eta^k}{k!} \mathbb{E}^{!o} \left[ \left( \frac{N_s}{\gamma_o} + I_s + I_e + I_i \right)^k e^{-\eta \left( \frac{N_s}{\gamma_o} + I_s + I_e + I_i \right)} \right] \quad (35)$$

$$\stackrel{(b)}{=} \sum_{k=0}^{N_s-1} \frac{(-\eta)^k}{k!} \frac{d^k}{ds^k} \mathbb{E}^{!o} e^{-s \left( \frac{N_s}{\gamma_o} + I_s + I_e + I_i \right)} \Big|_{s=\eta}, \quad (36)$$

where (a) follows from distribution of  $h_{\hat{o}o}$  and (b) from the derivative property of the Laplace transform. Since  $I_s$ ,  $I_e$ , and  $I_i$  in (5) are all independent, (36) equals

$$\begin{aligned} P_s^{\text{IA}}(\theta) &= \sum_{k=0}^{N_s-1} \frac{(-\eta)^k}{k!} \frac{d^k}{ds^k} e^{-s \frac{N_s}{\gamma_o}} \mathbb{E} e^{-s I_s} \mathbb{E}^{!o} e^{-s I_e} \mathbb{E}^{!o} e^{-s I_i} \Big|_{s=\eta} \\ &= \sum_{k=0}^{N_s-1} \frac{(-\eta)^k}{k!} \frac{d^k}{ds^k} e^{-s \frac{N_s}{\gamma_o}} \mathbb{E} e^{-s g_{\hat{o}o} \beta_{\hat{o}o}^2 h_{\hat{o}o}} \mathcal{L}_{I_e}^{!o}(s) \mathcal{L}_{I_i}(s) \Big|_{s=\eta}, \end{aligned} \quad (37)$$

where  $\mathcal{L}_{I_e}^{!o}(s)$  is the Laplace transform of the intra-cluster interference w.r.t the reduced Palm measure and  $\mathcal{L}_{I_i}(s)$  is the Laplace transform of the inter-cluster interference. Since  $h_{\hat{o}o} \sim \Gamma(N_s, 1)$ , using (3),

$$\mathbb{E} e^{-s g_{\hat{o}o} \beta_{\hat{o}o}^2 h_{\hat{o}o}} = \left( \frac{\frac{T_t \gamma_o}{N} + D_r^\alpha}{s + \frac{T_t \gamma_o}{N} + D_r^\alpha} \right)^{N_s}. \quad (38)$$

We now evaluate  $\mathcal{L}_{I_e}^{!o}(s)$

$$\begin{aligned} \mathcal{L}_{I_e}^{!o}(s) &= \mathbb{E}^{!o} \left[ e^{-s \sum_{z \in \Psi_o} g_{\hat{o}z} \beta_{\hat{o}z}^2 h_{\hat{o}z}} \right] = \mathbb{E}^{!o} \left[ \prod_{z \in \Psi_o} e^{-s g_{\hat{o}z} \beta_{\hat{o}z}^2 h_{\hat{o}z}} \right] \\ &\stackrel{(a)}{=} \mathbb{E}^{!o} \left[ \prod_{z \in \Psi_o} \left( \frac{1}{1 + s g_{\hat{o}z} \beta_{\hat{o}z}^2} \right)^{N_s} \right] \\ &= \mathbb{E}^{!o} \left[ \prod_{z \in \Psi_o} \left( \frac{\frac{T_t \gamma_o}{N} + \|z - \hat{o}\|^\alpha}{s + \frac{T_t \gamma_o}{N} + \|z - \hat{o}\|^\alpha} \right)^{N_s} \right], \end{aligned} \quad (39)$$

where (a) follows from the Laplace transform of  $h_{\hat{o}z}$  and in the last step  $\|x\|^{-\alpha}$  was substituted for the path loss. Observe that (39) is the probability generating functional (PGF) of the representative cluster  $\Psi_o$  w.r.t its reduced Palm probability. We now use the following result from [17, Lemma 1] which we state for completeness. For any function  $f(x) : \mathbb{R}^2 \rightarrow \mathbb{R}^+$ ,

$$\mathbb{E}^{!o} \prod_{x \in \Psi_o} f(x) = \frac{1}{\pi R^2} \int_{A=B(o,R)} \left( \frac{1}{\pi R^2} \int_{A=B(o,R)} f(x-y) dx \right)^{K-1} dy. \quad (40)$$

Using (40), (39) equals

$$\mathcal{L}_{I_e}^{!o}(s) = \frac{1}{\pi R^2} \int_{B(o,R)} \left[ \frac{1}{\pi R^2} \int_{B(o,R)} \left( \frac{\frac{T_t \gamma_o}{N} + \|x-y-\hat{o}\|^\alpha}{s + \frac{T_t \gamma_o}{N} + \|x-y-\hat{o}\|^\alpha} \right)^{N_s} dx \right]^{K-1} dy. \quad (41)$$

We now focus on  $\mathcal{L}_{\mathbf{I}_i}(s)$

$$\begin{aligned}\mathcal{L}_{\mathbf{I}_i}(s) &= \mathbb{E}^{!o} \left[ e^{-s \sum_{x \in \Phi} g_{\hat{o}x} h_{\hat{o}x}} \right] = \mathbb{E}^{!o} \left[ \prod_{x \in \Phi} \left( \frac{1}{1 + s \|x - \hat{o}\|^{-\alpha}} \right)^{N_s} \right] \\ &\stackrel{(a)}{=} \exp \left( -\lambda_p \int_{\mathbb{R}^2} 1 - \left[ \frac{1}{\pi R^2} \int_{B(o,R)} \left( \frac{\|x-y\|^\alpha}{s + \|x-y\|^\alpha} \right)^{N_s} dx \right]^K dy \right),\end{aligned}\quad (42)$$

where (a) follows from (43) which characterizes the PGF of a Poisson cluster process [16]

$$\mathbb{E} \prod_{x \in \Phi} f(x) = \exp \left( -\lambda_p \int_{\mathbb{R}^2} 1 - \left( \frac{1}{\pi R^2} \int_{B(o,R)} f(x-y) dx \right)^K dy \right). \quad (43)$$

Note that in (41),  $x$  and  $y$  vary within a cluster, while in (42),  $y$  varies over the whole plane. Also, in (41), the interference is only from the other  $K-1$  transmitters in the cluster and hence the exponent of  $K-1$ , while in (42), the interference is from all the  $K$  transmitters at each cluster and hence the exponent of  $K$ .

## APPENDIX B PROOF OF LEMMA 1

Define the following PDF

$$f_p(x) = \begin{cases} \frac{1}{\pi R^2} & x \in B(o, R) \\ 0 & \text{Otherwise} \end{cases}. \quad (44)$$

Now, using (44)

$$\begin{aligned}\mathcal{L}_{\mathbf{I}_i}(s) &= \exp \left( -\lambda_p \int_{\mathbb{R}^2} 1 - \left[ \frac{1}{\pi R^2} \int_{B(o,R)} \mathbb{E}_h [e^{-sh\|x-y\|^{-\alpha}}] dx \right]^K dy \right) \\ &= \exp \left( -\lambda_p \int_{\mathbb{R}^2} 1 - \left[ \int_{\mathbb{R}^2} \mathbb{E}_h [e^{-sh\|x-y\|^{-\alpha}}] f_p(x) dx \right]^K dy \right).\end{aligned}\quad (45)$$

Therefore, the inner integral in (45) can be seen as the expectation of  $\mathbb{E}_h [e^{-sh\|x-y\|^{-\alpha}}]$  w.r.t  $f_p(x)$ . Using Jensen's inequality, as  $(\mathbb{E}_x [z])^K \leq \mathbb{E}_x [z^K]$  for  $K > 2$ , it follows from (45) that

$$\begin{aligned}\mathcal{L}_{\mathbf{I}_i}(s) &\leq \exp \left( -\lambda_p \int_{\mathbb{R}^2} 1 - \int_{\mathbb{R}^2} f_p(x) \mathbb{E}_h [e^{-sh\|x-y\|^{-\alpha}}]^K dx dy \right) \\ &\stackrel{(a)}{=} \exp \left( -\lambda_p \int_{\mathbb{R}^2} \int_{\mathbb{R}^2} \left( 1 - \left( \frac{\|x-y\|^\alpha}{s + \|x-y\|^\alpha} \right)^{KN_s} \right) f_p(x) dx dy \right) \\ &= \exp \left( -\lambda_p \int_{\mathbb{R}^2} f_p(x) \int_{\mathbb{R}^2} \left( 1 - \left( \frac{\|y'\|^\alpha}{s + \|y'\|^\alpha} \right)^{KN_s} \right) dy' dx \right) \\ &= \exp \left( -\lambda_p s^{2/\alpha} \frac{\Gamma(KN_s + 2/\alpha) \Gamma(1 - 2/\alpha)}{\Gamma(KN_s)} \int_{\mathbb{R}^2} f_p(x) dx \right) \\ &= \exp \left( -\lambda_p s^{2/\alpha} \frac{\Gamma(KN_s + 2/\alpha) \Gamma(1 - 2/\alpha)}{\Gamma(KN_s)} \right),\end{aligned}$$

where (a) follows from  $h \sim \Gamma(N_s, 1)$ . Similarly for  $\mathcal{L}_{I_e}^{!o}$

$$\begin{aligned} \mathcal{L}_{I_e}^{!o}(s) &= \frac{1}{\pi R^2} \int_{B(o,R)} \left[ \frac{1}{\pi R^2} \int_{B(o,R)} \mathbb{E}_h[e^{-shf(\|x-y-\hat{\delta}\|)}] dx \right]^{K-1} dy \\ &\stackrel{(a)}{\leq} \mathbb{E}_{x,y} \left[ \left( \mathbb{E}_h[e^{-shf(\|x-y-\hat{\delta}\|)}] \right)^{K-1} \right], \end{aligned} \quad (46)$$

where  $f(x) = \frac{1}{\|x\|^\alpha + \frac{T_t \gamma_o}{N}}$ ,  $x$  and  $y$  are two points independently and uniformly distributed in  $B(o, R)$ , and (a) follows from the Jensen's inequality. Evaluating the expectation w.r.t  $h$ , (46) equals

$$\mathcal{L}_{I_e}^{!o}(s) \leq \mathbb{E} \left[ \frac{1}{(1 + sf(\|x - y - \hat{\delta}\|))^{N_s(K-1)}} \right]. \quad (47)$$

Let  $(r, \theta)$  be the polar representation of  $x - y$ , where  $x$  and  $y$  are two points independently and uniformly distributed on  $B(o, r)$ . Then  $\theta \sim U(0, 2\pi)$  and from [27] the PDF of  $r$  is given by (11). Note that (11) equals  $I_{1-\frac{x^2}{4R^2}}(3/2, 1/2) = \frac{2}{\pi} \left( \arcsin(\sqrt{1-x^2/4R^2}) - \frac{x}{2R} \sqrt{1-x^2/4R^2} \right)$ . Without loss of generality, we can assume that  $\hat{\delta} = (D_r, 0)$ . Then (47) simplifies to (10).

#### APPENDIX C PROOF OF LEMMA 2

From (35) we have

$$\begin{aligned} P_s^{\text{IA}} &= \sum_{k=0}^{N_s-1} \frac{\eta^k}{k!} \mathbb{E}^{!o} \left[ \left( \frac{N_s}{\gamma_o} + I_s + I_e + I_i \right)^k e^{-\eta \left( \frac{N_s}{\gamma_o} + I_s + I_e + I_i \right)} \right] \\ &= \sum_{k=0}^{N_s-1} \frac{\eta^k}{k!} \mathbb{E}^{!o} \left[ \left( \frac{N_s}{\gamma_o} + I_s + I_e + I_i \right)^k e^{-\left( \frac{k}{e} + \eta - \frac{k}{e} \right) \left( \frac{N_s}{\gamma_o} + I_s + I_e + I_i \right)} \right] \\ &\stackrel{(a)}{\leq} \sum_{k=0}^{N_s-1} \frac{\eta^k}{k!} \mathbb{E}^{!o} \left[ e^{-(\eta - \frac{k}{e}) \left( \frac{N_s}{\gamma_o} + I_s + I_e + I_i \right)} \right] \\ &= \sum_{k=0}^{N_s-1} \frac{\eta^k}{k!} e^{-(\eta - k/e) \frac{N_s}{\gamma_o}} \mathbb{E} e^{-(\eta - k/e) I_s} \mathbb{E}^{!o} e^{-(\eta - k/e) I_e} \mathbb{E} e^{-(\eta - k/e) I_i}, \end{aligned} \quad (48)$$

where (a) follows from the observation  $x^k e^{-\frac{k}{e}x} \leq 1$  for  $x > 0$  and the proof follows by comparing (48) to (37) and making the necessary change of variables.

#### APPENDIX D COEFFICIENTS OF (17)

$$\begin{aligned} C_1 &= \frac{\theta D_r^\alpha f_d N}{2\gamma_o^4} \left( -2\theta D_r^{2\alpha} f_d K N \gamma_o (A_4 K \lambda_p \gamma_o + 1) \right. \\ &\quad \left. + 2\gamma_o^2 (-A_1 f_d (K-1)N + K \gamma_o (f_d K^2 N - 1)) \right. \\ &\quad \left. - D_r^\alpha \gamma_o^2 (2 + 2A_4 K \lambda_p \gamma_o + f_d N (2 + K(K+1)(\theta-2) - 2A_4 K^3 \lambda_p \gamma_o)) \right) \end{aligned}$$

$$\begin{aligned}
& -\theta D_r^{3\alpha} f_d N (1 + K \lambda_p \gamma_o (2A_4 + (A_5 + A_4(2 + A_4 K \lambda_p)) \gamma_o))), \\
C_2 = & \frac{N^2 \theta D_r^\alpha f_d^2}{2\gamma_o^4} \left( 2\gamma_o^2 D_r^\alpha + \theta D_r^{3\alpha} - 2A_1 \gamma_o^2 - 2K \gamma_o^2 D_r^\alpha + 2A_1 K \gamma_o^2 \right. \\
& + K^2 \gamma_o^2 \theta D_r^\alpha + 2K \gamma_o \theta D_r^{2\alpha} + K \gamma_o^2 \theta D_r^\alpha \\
& \left. + K \lambda_p \theta D_r^{3\alpha} (A_4^2 K \gamma_o^2 \lambda_p + 2A_4 K \gamma_o^2 D_r^{-\alpha} + 2A_4 \gamma_o + 2A_3 \gamma_o^2 + A_5 \gamma_o^2) \right),
\end{aligned}$$

where

$$\begin{aligned}
A_1 &= \frac{1}{\pi R^2} \int_{B(o,R)} \frac{1}{\pi R^2} \int_{B(o,R)} \|x-y-\hat{o}\|^\alpha dx dy, \quad A_2 = \|x-y\|^\alpha + \theta D_r^\alpha, \\
A_3 &= \int_{\mathbb{R}^2} \left[ \frac{1}{\pi R^2} \int_{B(o,R)} \frac{\|x-y\|^\alpha}{(A_2)^3} dx \right] \left[ \frac{1}{\pi R^2} \int_{B(o,R)} \frac{\|x-y\|^\alpha}{A_2} dx \right]^{K-1} dy, \\
A_4 &= \int_{\mathbb{R}^2} \left[ \frac{1}{\pi R^2} \int_{B(o,R)} \frac{\|x-y\|^\alpha}{(A_2)^2} dx \right] \left[ \frac{1}{\pi R^2} \int_{B(o,R)} \frac{\|x-y\|^\alpha}{A_2} dx \right]^{K-1} dy, \\
A_5 &= (K-1) \int_{\mathbb{R}^2} \left[ \frac{1}{\pi R^2} \int_{B(o,R)} \frac{\|x-y\|^\alpha}{(A_2)^2} dx \right]^2 \left[ \frac{1}{\pi R^2} \int_{B(o,R)} \frac{\|x-y\|^\alpha}{A_2} dx \right]^{K-2} dy.
\end{aligned}$$

#### APPENDIX E

##### COEFFICIENTS OF (19)

$$\begin{aligned}
D_1 &= \frac{f_d \theta D_r^\alpha}{\frac{\gamma_o}{N}} \left( -K + \frac{f_d}{\frac{\gamma_o}{N}} \left( \frac{K^3 N \gamma_o}{N} - D_r^\alpha (\theta + 1) - \frac{\theta D_r^\alpha}{2} (K-1)(K+2) - (K-1)A_1 \right) \right), \\
D_2 &= \frac{f_d^2 \theta D_r^\alpha}{\left(\frac{\gamma_o}{N}\right)^2} \left( D_r^\alpha (\theta + 1) + \frac{\theta D_r^\alpha}{2} (K-1)(K+2) + (K-1)A_1 \right),
\end{aligned}$$

where  $A_1$  is given in Appendix D.

#### APPENDIX F

##### COEFFICIENTS OF (31)

$$\begin{aligned}
B_1 &= -B_2 - \frac{D_r^\alpha f_d}{\gamma_o^2 \alpha} \left( N \alpha \theta D_r^\alpha + N \gamma_o \alpha \theta + 2\mathcal{J} \gamma_o \theta^{\frac{2}{\alpha}} D_r^2 \lambda_p \right), \\
B_2 &= \frac{D_r^{2\alpha} f_d^2}{2\gamma_o^4 \alpha^2} \left( \left( \alpha N \theta (\gamma_o + D_r^\alpha) + 2\gamma_o \mathcal{J} \theta^{\frac{2}{\alpha}} D_r^2 \lambda_p \right)^2 - 2\gamma_o^2 \alpha \mathcal{J} \theta^{\frac{2}{\alpha}} D_r^2 \lambda_p (2/\alpha - 1) \right),
\end{aligned}$$

where  $\mathcal{J}$  is defined in (27).

## REFERENCES

- [1] B. Nosrat-Makouei, K. G. Ganti, J. G. Andrews, and R. W. Heath, Jr., "MIMO interference alignment in random access networks," in *Proc. 45th Asilomar Conference on Signals, Systems and Computers*, Pacific Grove, CA, Nov. 2011.
- [2] T. Gou and S. A. Jafar, "Degrees of freedom of the  $K$  user  $M \times N$  MIMO interference channel," *IEEE Transactions on Information Theory*, vol. 56, no. 12, pp. 6040–6057, Dec. 2010.
- [3] S. Jafar, "Blind interference alignment," *IEEE Journal of Selected Topics in Signal Processing*, vol. 6, no. 3, pp. 216–227, June 2012.
- [4] T. Gou, C. Wang, and S. A. Jafar, "Aiming perfectly in the dark - blind interference alignment through staggered antenna switching," *IEEE Transactions on Signal Processing*, vol. 59, no. 6, pp. 2734–2744, June 2011.
- [5] V. R. Cadambe and S. A. Jafar, "Interference alignment and degrees of freedom of the  $K$ -user interference channel," *IEEE Transactions on Information Theory*, vol. 54, no. 8, pp. 3425–3441, Aug. 2008.
- [6] B. Nosrat-Makouei, J. G. Andrews, and R. W. Heath, Jr., "MIMO interference alignment over correlated channels with imperfect CSI," *IEEE Transactions on Signal Processing*, vol. 59, no. 6, pp. 2783–2794, June 2011.
- [7] S. W. Peters and R. W. Heath, Jr., "User partitioning for less overhead in MIMO interference channels," *IEEE Transactions on Wireless Communications*, vol. 11, no. 2, pp. 592–603, Feb. 2012.
- [8] A. Lozano, R. W. Heath, Jr., and J. G. Andrews, "Fundamental limits of cooperation," submitted to *IEEE Transactions on Information Theory*, 2012. [Online]. Available: <http://arxiv.org/pdf/1204.0011v1.pdf>
- [9] F. Baccelli and B. Blaszczyszyn, *Stochastic Geometry and Wireless Networks, Part II: Applications*, ser. Foundations and trends in networking. Now Publishers, 2009.
- [10] O. El Ayach and R. W. Heath, Jr., "Interference alignment with analog channel state feedback," *IEEE Transactions on Wireless Communications*, vol. 11, no. 2, pp. 626–636, Feb. 2012.
- [11] B. Nosrat-Makouei, J. G. Andrews, and R. W. Heath, Jr., "User arrival in MIMO interference alignment networks," *IEEE Transactions on Wireless Communications*, vol. 11, no. 2, pp. 842–851, Feb. 2012.
- [12] S. Gollakota, S. D. Perli, and D. Katabi, "Interference alignment and cancellation," *ACM SIGCOMM Computer Communication Review*, vol. 39, no. 4, pp. 159–170, Aug. 2009.
- [13] S. P. Weber, X. Yang, J. G. Andrews, and G. de Veciana, "Transmission capacity of wireless ad hoc networks with outage constraints," *IEEE Transactions on Information Theory*, vol. 51, no. 12, pp. 4091–4102, Dec. 2005.
- [14] S. Weber, J. G. Andrews, and N. Jindal, "An overview of the transmission capacity of wireless networks," *IEEE Transactions on Communications*, vol. 58, no. 12, pp. 3593–3604, Dec. 2010.
- [15] R. Tresch, G. Alfano, and M. Guillaud, "Interference alignment in clustered ad hoc networks: high reliability regime and per-cluster aloha," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing*, Prague, Czech Republic, May 2011, pp. 3348–3351.
- [16] D. Stoyan, W. S. Kendall, and J. Mecke, *Stochastic Geometry and Its Applications*, ser. Wiley-interscience paperback. John Wiley & Sons, 2008.
- [17] R. K. Ganti and M. Haenggi, "Interference and outage in clustered wireless ad hoc networks," *IEEE Transactions on Information Theory*, vol. 55, no. 9, pp. 4067–4086, Sept. 2009.
- [18] K. Gomadam, V. R. Cadambe, and S. A. Jafar, "Approaching the capacity of wireless networks through distributed interference alignment," in *Proc. IEEE Global Telecommunications Conference*, New Orleans, LA, Nov. 2008, pp. 1–6.
- [19] S. W. Peters and R. W. Heath, Jr., "Cooperative algorithms for MIMO interference channels," *IEEE Transactions on Vehicular Technology*, vol. 60, no. 1, pp. 206–218, Jan. 2011.
- [20] G. Caire, N. Jindal, M. Kobayashi, and N. Ravindran, "Multiuser MIMO achievable rates with downlink training and channel state feedback," *IEEE Transactions on Information Theory*, vol. 56, no. 6, pp. 2845–2866, June 2010.
- [21] C. Wang, E. K. S. Au, R. D. Murch, W. H. Mow, R. S. Cheng, and V. Lau, "On the performance of the MIMO zero-forcing receiver in the presence of channel estimation error," *IEEE Transactions on Wireless Communications*, vol. 6, no. 3, pp. 805–810, Mar. 2007.
- [22] L. Musavian, M. R. Nakhai, M. Dohler, and A. H. Aghvami, "Effect of channel uncertainty on the mutual information of MIMO fading channels," *IEEE Transactions on Vehicular Technology*, vol. 56, no. 5, pp. 2798–2806, Sept. 2007.
- [23] M. Kobayashi, N. Jindal, and G. Caire, "Training and feedback optimization for multiuser MIMO downlink," *IEEE Transactions on Communications*, vol. 59, no. 8, pp. 2228–2240, Aug. 2011.
- [24] S. P. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge Univ Pr, 2004.
- [25] N. Jindal and A. Lozano, "A unified treatment of optimum pilot overhead in multipath fading channels," *IEEE Transactions on Communications*, vol. 58, no. 10, pp. 2939–2948, Oct. 2010.
- [26] R. H. Y. Louie, M. R. McKay, and I. B. Collings, "Open-loop spatial multiplexing and diversity communications in ad hoc networks," *IEEE Transactions on Information Theory*, vol. 57, no. 1, pp. 317–344, Jan. 2011.
- [27] J. Hammersley, "The distribution of distance in a hypersphere," *The Annals of Mathematical Statistics*, vol. 21, no. 3, pp. 447–452, 1950.

TABLE I  
NOTATION USED IN THIS PAPER

Symbol	Explanation
$\mathbf{a}$ & $\mathbf{A}$	$\mathbf{a}$ is a vector and $\mathbf{A}$ is a matrix
$\mathbf{A}^*$ , $\mathbf{A}^T$ , & $\mathbf{A}^{-1}$	Conjugate transpose, transpose, and inverse of $\mathbf{A}$
$\mathbf{A}_{n \times m}$	$\mathbf{A}$ is a matrix of dimension $n \times m$
$\mathbf{A}(n, m)$	Element on the $n$ th row and $m$ th column of $\mathbf{A}$
$\text{rank}(\mathbf{A})$	Rank of $\mathbf{A}$
$[a]$	Closest integer to $a$
$\mathbf{I}_N$ & $\mathbf{0}_{a \times b}$	$N \times N$ identity matrix and $a \times b$ matrix of all zeros
$B(o, R)$	The disk of radius $R$ centered on the origin
$\mathbb{P}^{lo}$	The reduced Palm probability measure
$\mathbf{A}_{i,j}$ vs. $\mathbf{A}$	When $\mathbf{A}_{i,j}$ correspond to $i$ and $j$ nodes, $\mathbf{A}$ is a generic random variable i.i.d. to the specific ones

TABLE II  
IMPORTANT SYMBOLS DEFINED IN THIS PAPER

$P_A$	Channel access probability of each cluster
$\Phi_p$	Spatial locations of the cluster centers with density $\tilde{\lambda}_p$
$K$	Number of transmit/receive pairs in each cluster
$N$	Number of antennas at each node
$N_s$	Number of streams from each transmitter
$\Phi$	The spatial locations of the potential transmitters with density $K\tilde{\lambda}_p$
$\Psi_o$	The cluster to which the transmitter at the origin belongs by
$\lambda_p$	Density of the active transmitters (equals $P_A\tilde{\lambda}_p$ )
$D_r$	Distance of a receiver from its corresponding transmitter
$D_c$	Average distance between the cluster centers
$R$	Radius of each cluster
$\beta_{\hat{x}z}^2$ & $g_{\hat{x}z}$	Variance of the channel estimation error and pathloss for the link between transmitter $z$ and receiver $\hat{x}$
$T$ , $f_d$ , $T_t$ , & $T_f$	Length of the constant channel block, Doppler frequency ( $f_d \approx \frac{1}{T}$ ), training period, and feedback period
$\theta$	SINR threshold for successful transmission

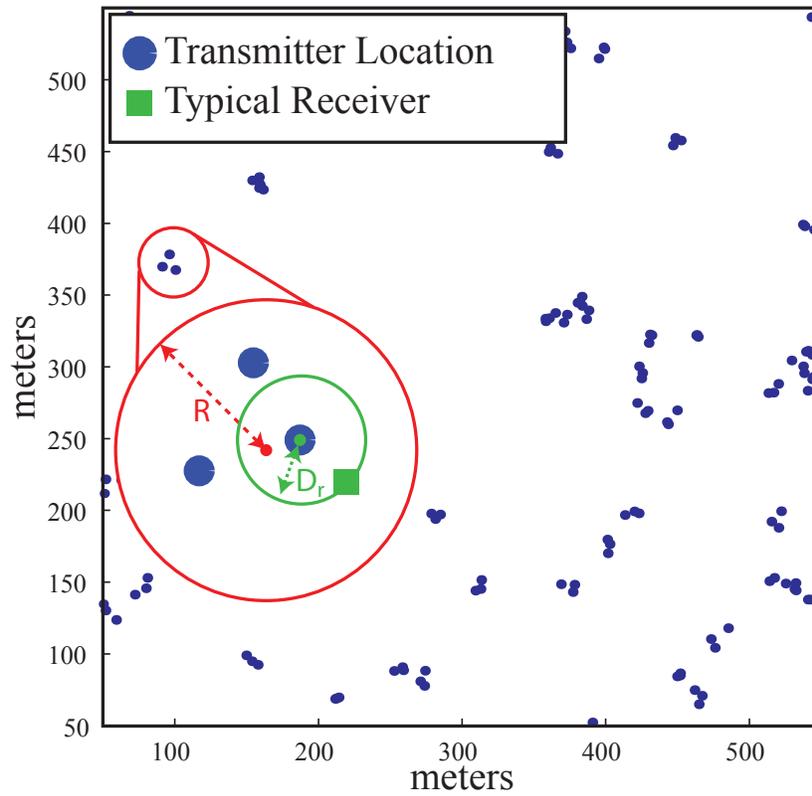


Fig. 1. An instance of the transmitter's distribution when the number of transmitters per cluster is  $K = 3$ .

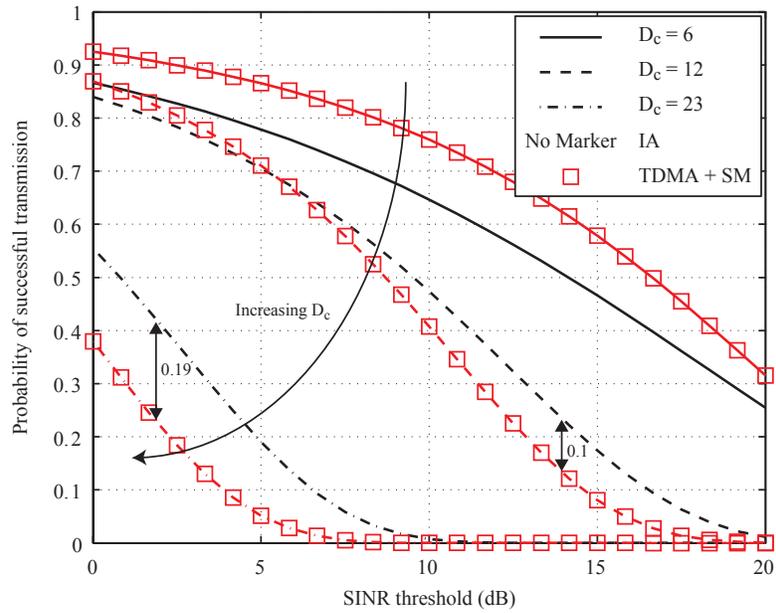


Fig. 2. The probability of successful transmission for IA and TDMA+SM as a function of the SINR threshold for some values of the average distance between the cluster centers  $D_c$  with perfect CSI.

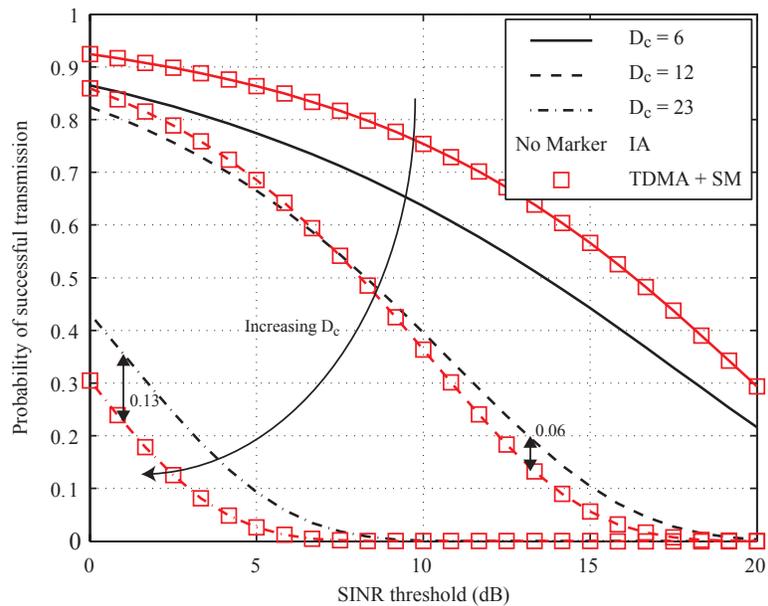


Fig. 3.  $P_s^{\text{IA}}$  and  $P_s^{\text{SM}}$  as a function of the SINR threshold for some values of the average distance between the cluster centers  $D_c$  with the smallest training period of  $T_t = 6$ .

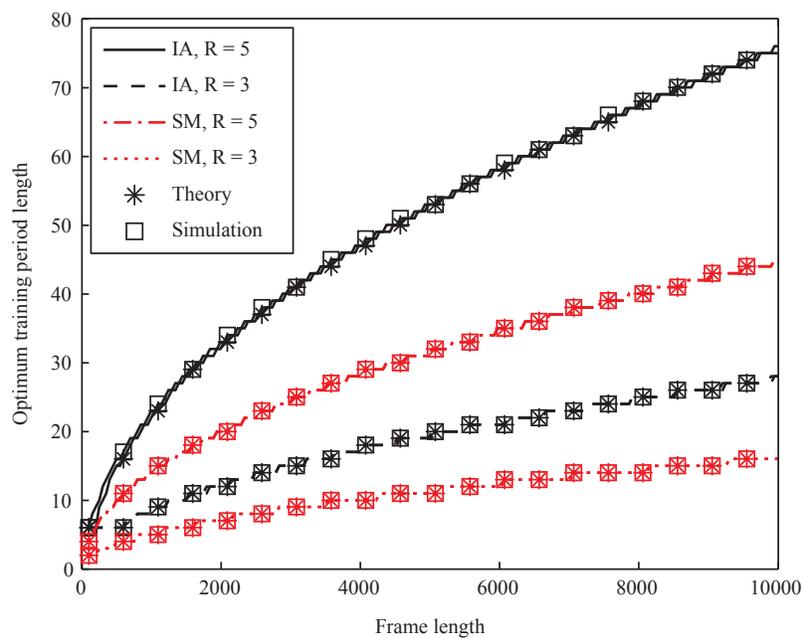


Fig. 4. The optimum training period as obtained using (18) and (32) for IA and SM together with the corresponding optimum values found through numerically optimizing (13) and (29) at two values of cluster radius (and hence average cluster center distance) for  $\theta = 20$  dB and  $\gamma_o = 30$  dB as a function of total frame length,  $T$ .

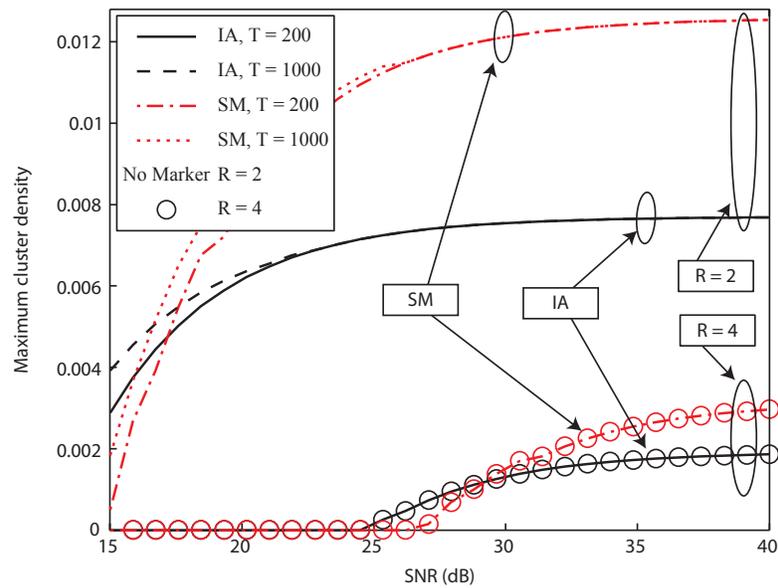


Fig. 5. The maximum cluster density,  $\lambda_P^\epsilon$ , as given by (23) and (34) for two values of cluster radius,  $R$ , and two values of the total block length,  $T$ , at the SINR threshold of  $\theta = 17$  dB and the maximum tolerable outage probability of  $\epsilon = 0.1$  as a function of transmit SNR,  $\gamma_o$ .

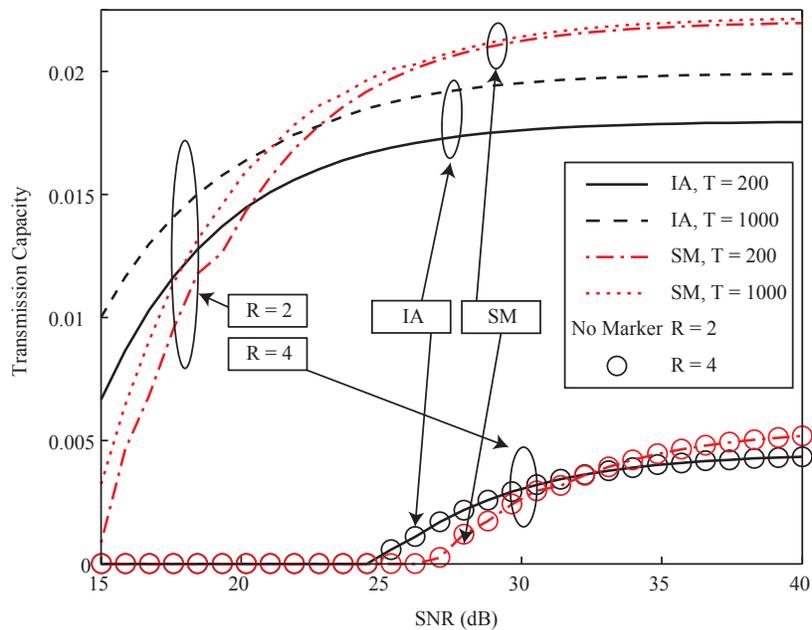


Fig. 6. The transmission capacity,  $c(\epsilon)$ , as given by (22) and (33) for two values of cluster radius,  $R$ , and two values of the total block length,  $T$ , at the SINR threshold of  $\theta = 17$  dB and the maximum tolerable outage probability of  $\epsilon = 0.1$  as a function of transmit SNR,  $\gamma_o$ . In each case, the optimum training periods as given by (18) and (32) were used to compute the transmission capacities.