Spatial Interference Cancellation for Multi-Antenna Mobile Ad Hoc Networks

Kaibin Huang, Jeffrey G. Andrews, Dongning Guo, Robert W. Heath, Jr., and Randall A. Berry (Submission date: July 11, 2008)

Abstract

Interference between nodes is a critical impairment in mobile ad hoc networks (MANETs). This paper studies the role of multiple antennas in mitigating such interference. Specifically, a network is studied in which receivers apply zero-forcing beamforming to cancel the strongest interferers. Assuming a network with Poisson distributed transmitters and independent Rayleigh fading channels, the transmission capacity is derived, which gives the maximum number of successful transmissions per unit area. Mathematical tools from stochastic geometry are applied to obtain the asymptotic transmission capacity scaling and characterize the impact of inaccurate channel state information (CSI). It is shown that, if each node cancels L interferers, the transmission capacity decreases as $\Theta(e^{\frac{1}{L+1}})$ as the outage probability ϵ vanishes. For fixed ϵ , as L grows, the transmission capacity increases as $\Theta(L^{1-\frac{2}{\alpha}})$ where α is the path-loss exponent. Moreover, CSI inaccuracy is shown to have no effect on the transmission capacity scaling as ϵ vanishes, provided that the CSI training sequence has an appropriate length, which we derived. Numerical results suggest that canceling merely one interferer by each node increases the transmission capacity by an order of magnitude or more, even when the CSI is imperfect.

I. INTRODUCTION

In a mobile ad hoc network (MANET), the mutual interference between nodes poses a fundamental limit on the throughput of peer-to-peer communication. This paper studies mitigating the effect of interference by provisioning nodes with multiple antennas. Specifically, each receiver uses zero-forcing beamforming to cancel the interference from the strongest interferers, whereas each transmitter simply chooses a random beam. This approach requires only limited local coordination and hence is suitable for a MANET.

This paper considers a simple network consisting of Poisson distributed transmitters and independent Rayleigh fading channels. We quantify the gains in network performance in terms of the *transmission capacity* (TC) as a function of the system parameters, including the amount and accuracy of channel state information (CSI) at each receiver. The TC is defined in [1] as the maximum density of successful transmissions so that a typical receiver satisfies an outage probability constraint for a target signal-to-interference-and-noise ratio (SINR). In other words, TC gives the average throughput per unit area.

K. Huang is with Yonsei University, S. Korea; J. G. Andrews and R. W. Heath, Jr. are with The University of Texas at Austin; D. Guo and R. A. Berry are with Northwestern University. Email: huangkb@yonsei.ac.kr, {rheath, jandrews}@ece.utexas.edu, {dguo, rberry}@eecs.northwestern.edu. This work was funded by the DARPA IT-MANET program under the grant W911NF-07-1-0028.

2

The derived TC scaling suggests that multiple antennas can significantly improve the performance of MANETs, even with inaccurate CSI.

A. Prior Work and Motivation

For Poisson distributed transmitters, TC was introduced in [1] for single-antenna MANETs assuming fixed transmission power and an ALOHA-like medium access control (MAC) layer. TC has also been used to study opportunistic transmissions [2], distributed scheduling [3], coverage [4], network irregularity [5], bandwidth partitioning [6], successive interference cancellation [7], and multi-antenna transmission [8] in MANETs [9]. This paper differs from [8] in that the antennas are employed for interference cancellation instead of interference averaging through diversity techniques. After the publication of preliminary results from the current investigation [10], [11], related work has been reported on the TC of multi-antenna MANETs that use spatial multiplexing and space-time block coding [12], space division multiple access [13], or optimally allocate the spatial degrees of freedom for interference cancellation and link enhancement such as spatial multiplexing [14] and array gain [15].

As discussed in [2], the TC is related to the more widely known *transport capacity* introduced in [16]. Transport capacity is typically studied in terms of the scaling of a network's total throughput-distance product as a function of the network size, while TC gives the number of single-hop transmissions possible in a specific area and in terms of the actual design parameters. Furthermore, most work on the transport capacity assumes perfect scheduling and zero-outage, while we focus on a random access model with an outage requirement.

Besides spatial interference cancellation, there are several alternative approaches for mitigating interference in MANETs. For example, the *interference alignment* approach in [17] achieves the optimal number of degrees of freedom in a high signal-to-noise ratio (SNR) setting. This approach appears daunting in practice because it requires jointly designed precoders and perfect CSI of interference channels. In contrast, the current approach only requires each receiver to obtain CSI from nearby interferers and no coordination of transmit precoders. Another method for interference management, used in many practical MAC protocols, is to create an interferer-free area – a *guard zone* – around each receiving node through carrier sensing. As shown in [3], optimizing the guard-zone size leads to significant TC gain for a singleantenna MANET with respect to pure random access. The use of interference cancellation can be viewed as creating an effective guard zone without requiring that other nearby transmitters be suppressed.

Several papers have addressed other aspects of multi-antenna MANETs. For example, beamforming or directional antennas have been integrated with the MAC protocols for MANETs to achieve higher network spatial reuse or energy efficiency [18]–[27]. In addition, multi-antenna techniques have been applied to enable efficient routing in MANETs [28]. Directional antennas have been studied for suppressing interference in MANETs by spatial filtering [23]–[27]. Directional antennas, however, are only suitable for environments with sparse scattering and low angular spread. In contrast, beamforming is applicable for both sparse and rich scattering, and is hence adopted in this paper as well as in [19]–[21] for spatial interference cancellation. Most prior work focuses on designing MAC protocols and rely on simulations for throughput evaluation [18]–[27], [29]. In [30], [31], the use of directional antennas is shown to increase the linear scaling factor of network transport capacity, which, however, is too coarse for quantifying the

network throughput. In view of prior work, there still lacks theoretic characterization of the relationship between the TC of MANETs and spatial interference cancellation.

B. Contributions and Organization

Our main contributions are summarized as follows.

- Assuming Poisson distributed transmitters and spatially independently and identically distributed (i.i.d.) Rayleigh fading channels, bounds on the probability of signal-to-interference ratio (SIR) outage are derived with either perfect or imperfect CSI. These bounds are found to be reasonably tight and lead to bounds on TC.
- 2) Let L denote the number of canceled interferers per receiver. Irrespective of whether the CSI is perfect, as the outage probability ϵ vanishes, the asymptotic TC is shown to vanish as $\epsilon^{\frac{1}{L+1}}$, which decays slower for larger L.
- 3) For fixed ε, as L → ∞, the TC is shown to increase as L^{1-²/_α} where α is the path-loss exponent. Hence spatial interference cancellation is more effective for larger α. If instead of canceling interference, the antennas are used to maximize the array gain, the TC scales as L^{2/_α} as shown in [8].¹
- 4) The required training sequence length for CSI estimation is derived for constraining the increase in outage probability and loss in data rate caused by imperfect interference cancellation. Finally, the required training sequence length is obtained for the optimal TC scaling and shown to be proportional to $\log \frac{1}{\epsilon}$.

Simulation results show that canceling a few (two to four) interferers per node is sufficient for harvesting most of the available TC gain. A capacity gain of more than an order of magnitude can be achieved by canceling only one interferer at each node, even with imperfect CSI. Moreover, a moderate length of the CSI training sequence is observed to be sufficient.

The remainder of this paper is organized as follows. Section II describes the network and wireless channel models. The SIR outage probability and TC are analyzed for perfect and imperfect CSI in Sections III and IV, respectively. Numerical results are presented in Section V.

II. MATHEMATICAL MODELS AND METRICS

A. Network Model

The locations of potential transmitting nodes in a MANET are modeled as a 2-D Poisson point process with density λ_o following the common approach in the literature (see e.g., [32]). Time is slotted and in each time-slot potential transmitting nodes follow a random access protocol, namely that they transmit independently with a fixed probability P_t . Let T denote the coordinate of a transmitting node. Given the random access protocol, the set $\hat{\Phi} = \{T\}$ is also a homogeneous Poisson point process but with smaller density $\lambda = P_t \lambda_o$ [33]. Each transmitting node is associated with a receiving node located at a

¹Since this work was submitted, recent results have demonstrated that linear TC scaling with the number of antennas per node can be achieved by proportionally allocating the spatial degrees of freedom at each receiver to obtain array gain and to cancel interferers, which can be interpreted as the product of above two sub-linear scalings [15].

unit distance. Relaxing the assumption that a transmitter and a receiver are separated by a fixed distance affects the TC scaling only by a multiplicative factor (see e.g, [2]).

Consider a typical receiving node located at the origin, denoted as R_0 , and let T_0 be the corresponding transmitter. This constraint on R_0 and T_0 does not compromise generality since the transmitting node process is translation invariant. Furthermore, according to Slivnyak's theorem [34, Section 4.4], the remaining transmitting nodes, namely $\Phi = \Phi \setminus \{T_0\}$, remain a homogeneous Poisson point process with the same density λ .

The MANET is assumed to be interference limited and thus noise is neglected for simplicity.² Consequently, the reliability of data packets received by R_0 is determined by the SIR. Moreover, we assume that each data link has a single stream, and communications between nodes are perfectly synchronized at symbol boundaries. All transmitting nodes are assumed to use unit transmission power.

B. Channel Model

Let every node be equipped with N antennas, so that the link between each transmitter and receiver can be modeled as an $N \times N$ multiple-input-multiple-output (MIMO) channel. We assume narrowband channels with frequency-flat block fading. Moreover, each MIMO channel consists of path-loss and small-scale fading components. Specifically, the channel from $T \in \Phi$ to R_0 is $r_T^{-\alpha/2}\mathbf{G}_T$, where $\alpha > 2$ is the path-loss exponent, and \mathbf{G}_T is an $N \times N$ matrix of i.i.d. unit circularly symmetric complex Gaussian elements (we subsequently denote the distribution as $\mathcal{CN}(0,1)$). Beamforming is applied at each transmitter and receiver, where \mathbf{f}_0 , \mathbf{f}_T and \mathbf{v}_0 denote the beamforming vectors at T_0 , $T \in \Phi$ and R_0 , respectively. Then, measured at R_0 , the received power from T_0 is $W = |\mathbf{v}_0^{\dagger}\mathbf{G}_0\mathbf{f}_0|^2$, and the interference power from transmitter T is $I_T = r_T^{-\alpha}|\mathbf{v}_0^{\dagger}\mathbf{G}_T\mathbf{f}_T|^2$, where \dagger represents the Hermitian transpose matrix operation.

C. Transmission Capacity

Correct decoding of received data packets requires the SIR to exceed a threshold θ , which is identical for all receivers. In other words, the information rate for each link is equal to $\log_2(1 + \theta)$ assuming Gaussian signaling. Note that, with everything else the same, the outage probability increases with the transmitter density λ . The TC under the outage constraint is thus given by

$$C(\epsilon) = (1 - \epsilon) \log_2(1 + \theta)\lambda \tag{1}$$

where λ is such that the outage probability P_{out} is ϵ , i.e.,

$$P_{\mathsf{out}} = \Pr(\mathsf{SIR} < \theta) \tag{2}$$

$$= \epsilon.$$
 (3)

The metric $C(\epsilon)$ quantifies the spatial reuse efficiency of a single-hop MANET; the product $C(\epsilon)d$ is related to the transport capacity of a multi-hop MANET [2].³

²Addressing the effect of noise requires straightforward but tedious modifications of the current analysis.

³A new metric called *random access transport capacity* has been proposed for multi-hop MANETs in a recent work [35], which generalizes TC to an end-to-end scenario under several additional assumptions.

D. Zero-Forcing Beamforming

From the perspective of R_0 , the interference channel from an interferer T is in effect a channel vector $\mathbf{h}_T = r_T^{-\alpha/2} \mathbf{G}_T \mathbf{f}_T$. For convenience, we refer to the channel norm $J_T = ||\mathbf{h}_T||^2$ as the precancellation interference power for T. The receiver R_0 equipped with N antennas cancels L interferers where $L \leq N - 1$. Let $\mathcal{T} \subset \Phi$ comprise the L strongest interferers to be canceled by R_0 . Spatial interference cancellation at R_0 is realized by choosing the beamformer \mathbf{v}_0 to be orthogonal to the beams of the L strongest interferers, i.e., $\mathbf{v}_0^{\dagger} \mathbf{h}_T = 0$ for every $T \in \mathcal{T}$.

Conditioned on interference cancellation, R_0 applies the remaining (N-L) spatial degrees of freedom to enhance the received signal power by maximum ratio combining. To be specific, \mathbf{v}_0 solves the following optimization problem

maximize:
$$|\mathbf{v}_0^{\dagger}\mathbf{h}_0|^2$$

subject to: $\|\mathbf{v}_0^{\dagger}\mathbf{h}_T\|^2 = 0, \forall T \in \mathcal{T}$ (4)
 $\|\mathbf{v}_0\|^2 = 1.$

Note that \mathbf{v}_0 is uniquely determined by $\{\mathbf{h}_T \mid T \in \mathcal{T}\}$ and \mathbf{h}_0 with probability 1.

We let transmitter T_0 apply a fixed transmit beamformer \mathbf{f}_0 in lieu of an adaptive one (such as to perform maximum ratio transmission [36]) to prevent the transmitter and receiver from chasing each other's beam, so as to preserve network stability. Since the fading channels are isotropic, \mathbf{f}_0 is chosen to be the all-one vector normalized by $1/\sqrt{L}$. With such beamforming, multiple transmit antennas contribute no additional array gain.

E. The Effective SIR Model: Perfect CSI

We first characterize the SIR at R_0 assuming perfect CSI and hence perfect cancellation of the L strongest interferers. Since \mathbf{G}_0 is isotropic, $\mathbf{G}_0\mathbf{f}_0$ is an i.i.d. $\mathcal{CN}(0,1)$ vector. It follows from [15, Lemma 1] that the random variable W has the following chi-square distribution with (N - L) complex degrees of freedom,

$$f_W(w) = \frac{w^{N-L-1}}{\Gamma(N-L)}e^{-w}, \quad w \ge 0$$
 (5)

where Γ denotes the gamma function. The factor (N - L) specifies the array gain per link [37]. The effective interference model resulting from perfect interference cancellation is illustrated in Fig. 1(a). The SIR at R_0 is given as

$$\mathsf{SIR} = \frac{W}{\sum_{T \in \Phi \setminus \mathcal{T}} I_T}.$$
(6)

F. The Effective SIR Model: Imperfect CSI

Without the assumption of perfect and readily available CSI, receivers estimate the signal strength of their interferers and then identify the strongest ones by using the random training signature sequences inserted into transmitted signals [38]. Subsequently, each receiver requests their strongest interferers to transmit training sequences and uses them to estimate the corresponding interference channels. Let M denote the length of each training sequence. The L training sequences form an $L \times M$ matrix $\sqrt{M}\mathbf{Q}$ where \mathbf{Q} consists of orthonormal row vectors [39].



Fig. 1. Effective interference model for a typical receiver canceling two strongest interferences with (a) perfect CSI or (b) imperfect CSI. The distance in the figures decreases with increasing received power. The data and interference links are plotted using solid and dashed lines, respectively.

Define V as an $N \times L$ matrix comprising the vectors $\{\mathbf{h}_T \mid T \in \mathcal{T}\}$ as columns. The signal Y received by R_0 during the training phase is expressed as an $N \times M$ matrix:

$$\mathbf{Y} = \sqrt{M} \mathbf{V} \mathbf{Q} + \sum_{T \in \Phi \setminus \mathcal{T}} \mathbf{h}_T \mathbf{x}_T$$
(7)

where the factor \sqrt{M} normalizes the average power of each training sequence to be one, and the $1 \times M$ row vector \mathbf{x}_T contains $\mathcal{CN}(0,1)$ data symbols transmitted by node T. The summation term in (7) represents interference to the CSI estimation at R_0 . The CSI is estimated using the least-squares method to yield for every $T \in \mathcal{T}$:

$$\hat{\mathbf{h}}_T = \frac{1}{\sqrt{M}} \mathbf{Y} \mathbf{q}_T^{\dagger}$$
(8)

$$= \mathbf{h}_{T} + \frac{1}{\sqrt{M}} \sum_{T' \in \Phi \setminus \mathcal{T}} \mathbf{h}_{T'} \tilde{x}_{T',T}$$
(9)

where \mathbf{q}_T is the training sequence sent by node T and $\tilde{x}_{T',T} = \mathbf{x}_T \mathbf{q}_T^{\dagger}$ has the distribution $\mathcal{CN}(0,1)$.⁴

The SIR is derived as follows. The estimated CSI is applied for computing the beamformer \mathbf{v}_0 used at R_0 . Using (8) and under the zero-forcing constraint: $\mathbf{v}_0^{\dagger} \hat{\mathbf{h}}_T = 0, \forall T \in \mathcal{T}$, the residual interference at R_0 after beamforming, denoted as I_R , can be written as

$$I_{R} = \sum_{T \in \mathcal{T}} \mathbf{v}_{0}^{\dagger} \mathbf{h}_{T} x_{T}$$
$$= -\frac{1}{\sqrt{M}} \sum_{T \in \mathcal{T}} \sum_{T' \in \Phi \setminus \mathcal{T}} \mathbf{v}_{0}^{\dagger} \mathbf{h}_{T'} \tilde{x}_{T',T} x_{T}.$$
(10)

⁴The alternative minimum mean-square-error (MMSE) estimator requires knowledge of the covariance of the aggregate interference from the weak transmitters, which is difficult to measure accurately due to the presence of the strong interference.

As illustrated in Fig. 1(b), CSI estimation errors result in additional interference with respect to the case of perfect CSI. For the present case, the SIR in (6) is modified as

$$\widetilde{\mathsf{SIR}} = \frac{W}{\sigma_R^2 + \sum_{T \in \Phi \setminus \mathcal{T}} I_T} \tag{11}$$

where σ_R^2 denotes the variance of I_R in (10) conditioned on fixed Φ and channels.

III. TRANSMISSION CAPACITY WITH PERFECT CSI

This section focuses on the analysis of the outage probability and TC assuming prefect CSI. In particular, we derive the TC scaling with respect to the number of canceled interferers per node and the outage probability.

A. Point Processes and Auxiliary Results

Since the L strongest interferers are canceled, we refer to the (L+1)-st strongest interferer (in terms of pre-cancellation power) as the *primary interferer* and denote it as T_P , whose pre-cancellation interference power is $J_P = ||\mathbf{h}_{T_P}||^2$. The interferers with smaller pre-cancellation interference power are referred to as the *secondary interferers*. There are two reasons for separating the interferers. First, considering the primary interferer alone yields a lower bound on the outage probability to be derived in the sequel. Second, as we show shortly, the secondary interferers conditioned on J_P form a Poisson point process.

Recall that Φ stands for the homogeneous Poisson point process of all transmitters but T_0 . Define a marked point process [33] $\Psi = \{(T, J_T\} \mid T \in \Phi\}$ where the mark of node T is its pre-cancellation interference power J_T . Different interference channels are independent and hence J_T depends on T but not other points in Φ . Let μ and μ^* denote the mean measures of Φ and Ψ , respectively. By applying Marking Theorem, Ψ is shown to be a Poisson process on the product space $\mathbb{R}^2 \times \mathbb{R}^+$ with mean measure μ^* given as [33]

$$\mu^*(\mathcal{B}) = \iint_{(t,u)\in\mathcal{B}} \mu(dt)p(t,du)$$
(12)

where \mathcal{B} is a measurable subset of $\mathbb{R}^2 \times \mathbb{R}^+$ and $p(T, \cdot)$ represents the distribution of J_T conditioned on T. Let $\mathcal{G} = \mathbb{R}^2 \times (g, \infty)$. Note that $\Psi \cap \mathcal{G}$ is the set of interference whose pre-cancellation interference power is larger than g. From (12), $\mu^*(\mathcal{G})$ can be obtained as

$$\mu^*(\mathcal{G}) = \lambda \int_{t \in \mathbb{R}^2} \int_g^\infty p(t, du) dt$$
(13)

$$= 2\pi\lambda \int_0^\infty \int_g^\infty rp(r,du)dr$$
(14)

$$= 2\pi\lambda \int_0^\infty r \Pr(J_T > g \mid |T| = r) dr$$
(15)

where (13) follows from the homogeneity of Φ , and (14) uses polar coordinates and the fact that J_T depends only on the distance $r_T = |T|$ from the origin.

Let the point process consisting of all secondary interferers conditioned on J_P be denoted by

$$\Pi(J_P) = \{ (T, J_T) \mid T \in \Phi, 0 \le J_T < J_P \}.$$
(16)

The distribution of $\Pi(J_P)$ and J_P are characterized by the following lemmas, which are proved in Appendices A and B, respectively.

Lemma 1. Conditioned on $J_P = g$, $\Pi(J_P)$ is a Poisson point process on $\mathbb{R}^2 \times [0,g)$. Its mean measure is given by

$$\mu^*(\mathcal{B} \mid J_P = g) = \lambda \iint_{(t,u) \in \mathcal{B}} \mu(dt) p(t, du)$$
(17)

where \mathcal{B} is a measurable subset of $\mathbb{R}^2 \times [0,g)$ and $p(T,\cdot)$ is the distribution function of J_T conditioned on node T.

Lemma 2. The primary pre-cancellation interference power J_P has the following cumulative distribution function

$$\Pr(J_P \le g) = \sum_{k=0}^{L} \frac{\left(\nu\lambda g^{-\frac{2}{\alpha}}\right)^k}{\Gamma(k+1)} e^{-\nu\lambda g^{-\frac{2}{\alpha}}}$$
(18)

and the probability density function

$$f_P(g) = \frac{2(\nu\lambda)^{L+1}}{\alpha\Gamma(L+1)} g^{-\frac{2(L+1)}{\alpha} - 1} e^{-\nu\lambda g^{-\frac{2}{\alpha}}}$$
(19)

where $\nu = \frac{\pi \Gamma(N + \frac{2}{\alpha})}{\Gamma(N)}$.

After perfect interference cancellation, the interference power of a remaining interferer $T \in \Phi \setminus \mathcal{T}$ is $I_T = |\mathbf{v}_0^{\dagger} \mathbf{h}_T|^2$. Define $\tilde{\mathbf{h}}_T = \mathbf{h}_T / ||\mathbf{h}_T||$ and $\delta_T = |\mathbf{v}_0^{\dagger} \tilde{\mathbf{h}}_T|^2$. Then I_T and J_T are related by $I_T = J_T \delta_T$. The distributions of the random variables $\{\delta_T\}$ are specified in the following lemma.

Lemma 3. The set of random variables $\{\delta_T \mid T \in \Phi \setminus \mathcal{T}\}$ follow i.i.d. beta(1, N - 1) distributions as specified by the following probability density function

$$f_{\delta}(x) = (N-1)(1-x)^{N-2}, \quad x \in (0,1).$$
 (20)

Furthermore, δ_T is independent with J_T .

Proof: See Appendix C.

It follows that the primary interference power I_P is given as $I_P = J_P \delta_P$ where δ_P is a beta(1, N - 1) random variable. Given that δ_P is random, node T_P may not contribute the largest interference power despite its dominance over other uncanceled interference in terms of pre-cancellation interference power. Next, the total secondary interference power can be written as $I_S = \sum_{T \in \Phi \setminus (T \cup \{T_P\})} J_T \delta_T$. This summation is a type of *shot noise* [33] whose probability density function has no known closed-form expression except for some simple cases [2], [32]. Nevertheless, the conditional first and second moments of I_S can be obtained as shown in the following lemma.

Lemma 4. After perfect spatial interference cancellation, the secondary interference power IB conditioned on the primary pre-cancellation interference power $J_P = g$ has the following mean and variance

$$\mathbb{E}[I\beta | J_P = g] = \frac{2\nu\lambda}{N(\alpha - 2)}g^{1 - \frac{2}{\alpha}}$$
(21)

$$\operatorname{var}(I\mathfrak{B} | J_P = g) = \frac{2\nu\lambda}{N(N+1)(\alpha-1)}g^{2-\frac{2}{\alpha}}.$$
(22)

Proof: See Appendix D.

B. Bounds on Outage Probability

From (6), the outage probability $P_{\rm out}$ can be written as

$$P_{\text{out}} = \Pr(I\mathfrak{B} + I_P > W\theta^{-1}).$$
(23)

It follows that P_{out} can be lower bounded as

$$P_{\mathsf{out}} \ge \Pr(I_P > W\theta^{-1}) \tag{24}$$

by considering only the primary interferer, which is tight if the primary interfererence is dominant. Using $I_P = J_P \delta_P$ and (23), an upper bound on P_{out} is obtained as

$$P_{\mathsf{out}} \leq \Pr(I\mathfrak{B} + J_P > W\theta^{-1}) \tag{25}$$

$$= \Pr(J_P > W\theta^{-1}) + \Pr(I\mathfrak{B} > W\theta^{-1} - J_P \mid J_P \le W\theta^{-1}) \Pr(J_P \le W\theta^{-1})$$
(26)

where (25) holds since $\delta_P \leq 1$. The above upper bound can be further bounded by applying the following Chebyshev's inequality:

$$\Pr(I_{S} \ge a \mid J_{P} = g) \le \min\left\{\frac{\operatorname{var}(I_{S} \mid J_{P} = g)}{\{a - \mathbb{E}\left[I_{S} \mid J_{P} = g\right]\}^{2}}, 1\right\}, \quad \forall \ a > \mathbb{E}\left[I_{S} \mid J_{P} = g\right].$$
(27)

Based on (24), (26) and (27), bounds on the outage probability are derived as shown in the following lemma.

Lemma 5. For perfect spatial interference cancellation, the outage probability satisfies $P_{out}^{\ell}(\lambda) \leq P_{out} \leq P_{out}^{u}(\lambda)$ where:

1) The lower bound is

$$P_{\mathsf{out}}^{\ell}(\lambda) = 1 - \sum_{k=0}^{L} \frac{\left(\nu\lambda\theta^{\frac{2}{\alpha}}\right)^{k}}{\Gamma(k+1)} \mathbb{E}\left[\left(\frac{W}{\delta_{P}}\right)^{-\frac{2k}{\alpha}} e^{-\nu\lambda\theta^{\frac{2}{\alpha}}\left(\frac{W}{\delta_{P}}\right)^{-\frac{2}{\alpha}}}\right].$$
 (28)

2) Define the subsets \mathcal{D}_1 and \mathcal{D}_2 of the product space $\mathbb{R}^+ \times \mathbb{R}^+$ as

$$\mathcal{D}_1 = \left\{ (w,g) \mid w\theta^{-1} - \mathbb{E}[I\mathfrak{B} \mid J_P = g] \le g \le w\theta^{-1} \right\}$$
(29)

$$\mathcal{D}_2 = \left\{ (w,g) \mid g < w\theta^{-1} - \mathbb{E}[I\mathfrak{B} \mid J_P = g] \right\}.$$
(30)

The upper bound is

$$P_{\mathsf{out}}^u(\lambda) = \Lambda_1 + \Lambda_2 + \Lambda_3 \tag{31}$$

where

$$\Lambda_1 = 1 - \sum_{k=0}^{L} \frac{\left(\nu\lambda\theta^{\frac{2}{\alpha}}\right)^k}{\Gamma(k+1)} \mathbb{E}\left[W^{-\frac{2k}{\alpha}} e^{-\nu\lambda\theta^{\frac{2}{\alpha}}W^{-\frac{2}{\alpha}}}\right]$$
(32)

$$\Lambda_2 = \iint_{(w,g)\in\mathcal{D}_1} f_W(w) f_P(g) dw dg \tag{33}$$

$$\Lambda_{3} = \iint_{(w,g)\in\mathcal{D}_{2}} \min\left\{\frac{\operatorname{var}(I\mathfrak{B} \mid J_{P} = g)}{\left\{w\theta^{-1} - g - \mathbb{E}\left[I\mathfrak{B} \mid J_{P} = g\right]\right\}^{2}}, 1\right\} f_{W}(w)f_{P}(g)dwdg$$
(34)

with $\mathbb{E}(I_S | J_P = g)$ and $var(I_S | J_P = g)$ given in Lemma 4.

Proof: See Appendix E.

C. Asymptotic Transmission Capacity

Using the upper and lower bounds described in Lemma 5, the TC scaling is analyzed for a large number of canceled interferers per node $(L \to \infty)$ or varnishing outage probability $(\epsilon \to 0)$ as follows.

Increasing the number of antennas at each receiver allows more interferers to be canceled, leading to higher TC. The TC scaling as $L \to \infty$ is given in the following theorem.

Theorem 1. With perfect CSI and fixed array gain (N - L), if the number of canceled interferers per node L is sufficiently large, the transmission capacity is bounded as

$$\frac{1}{\pi} \left[\frac{\epsilon \left(\alpha - 2 \right)}{2\theta \mathbb{E}[W^{-1}]} \right]^{\frac{2}{\alpha}} \le \frac{C(L)}{(1 - \epsilon) \log_2(1 + \theta) L^{1 - \frac{2}{\alpha}}} \le \frac{2}{\pi} \left[\frac{\mathbb{E}[W]}{\theta(1 - \epsilon)} \right]^{\frac{2}{\alpha}}.$$
(35)

Proof: See Appendix F.

The relationship in (35) shows that TC grows in the order of $L^{1-2/\alpha}$, where the growth is faster for steeper path loss. Intuitively, if interference decays more quickly with distance, canceling the strongest interference more significantly.

Small target outage probability results in a network of sparse transmitting nodes (i.e., $\lambda \to 0$). For such a sparse network, the relationship between the outage probability and node density is given in the following lemma.

Lemma 6. With perfect CSI and $\lambda \rightarrow 0$, the outage probability is bounded as follows.

1) If $L + 1 \leq \alpha$, for sufficiently small λ ,

$$\kappa_1 \le \frac{\epsilon}{\lambda^{L+1}} \le \kappa_2 \tag{36}$$

where

$$\kappa_1 = \frac{E\left[\delta_P^{\frac{2}{\alpha}(L+1)}\right]E\left[W^{-\frac{2}{\alpha}(L+1)}\right]\left(\nu\theta^{\frac{2}{\alpha}}\right)^{L+1}}{\Gamma(L+2)}$$
(37)

$$\kappa_2 = \frac{2^{\frac{2}{\alpha}(L+1)+1} \mathbb{E}\left[W^{-\frac{2}{\alpha}(L+1)}\right] (\nu \theta^{\frac{2}{\alpha}})^{L+1}}{\Gamma(L+2)}.$$
(38)

2) If $L + 1 > \alpha$, for sufficiently small λ ,

$$\kappa_1 \le \frac{\epsilon}{\lambda^{L+1}} \quad and \quad \frac{\epsilon}{\lambda^{\alpha}} \le \kappa_3$$
(39)

where

$$\kappa_3 = \frac{8\theta^2 \nu^{\alpha} \Gamma(L-\alpha+2) \mathbb{E}[W^{-2}]}{N(N+1)(\alpha-1) \Gamma(L+1)}.$$
(40)

Proof: See Appendix G.

Note that the ratio $\frac{\kappa_2}{\kappa_1}$ decreases as L becomes smaller. This suggests that the asymptotic bounds are tighter for smaller values of L.

Using Lemma 6 and the TC definition in (1), we have the following TC scaling.

Theorem 2. With perfect CSI and small target outage probability $\epsilon \to 0$, the TC is bounded as follows. 1) If $L + 1 \le \alpha$, for sufficiently small ϵ ,

$$\kappa_2^{-\frac{1}{L+1}} \le \frac{C(\epsilon)}{\log(1+\theta)\epsilon^{\frac{1}{L+1}}} \le \kappa_1^{-\frac{1}{L+1}}$$
(41)

where κ_1 and κ_2 are given in Lemma 6.

2) If $L + 1 > \alpha$, for sufficiently small ϵ ,

$$\kappa_3^{-\frac{1}{\alpha}} \le \frac{C(\epsilon)}{\log_2(1+\theta)\epsilon^{\frac{1}{\alpha}}} \quad and \quad \frac{C(\epsilon)}{\log_2(1+\theta)\epsilon^{\frac{1}{L+1}}} \le \kappa_1^{-\frac{1}{L+1}}$$
(42)

where κ_3 is given in Lemma 6.

The above theorem shows that as ϵ decreases, $C(\epsilon)$ follows a power law. For $L+1 > \alpha$, only bounds on the exponent are known. The derivation of the exact scaling for $L+1 > \alpha$ requires a tighter upper bound on outage probability than that based on Chebyshev's inequality in (27). This may require analyzing the distribution function of the secondary interference power, which, however, has no known closed-form expression for the present case.

For $L + 1 \leq \alpha$, the exponent of the TC power law is 1/(L + 1). This power law indicates that L determines the sensitivity of TC to a change in the outage constraint. To facilitate our discussion, rewrite the scaling in Theorem 2 as $C(\epsilon) \cong \alpha \epsilon^{\frac{1}{L+1}}$ where " \cong " represents asymptotic equivalence for $\epsilon \to 0$. The sensitivity of TC towards changes of the outage constraint decreases inversely with the number of canceled interferers. Reducing the outage probability by two orders of magnitude decreases the TC by 10, 3.2, and 1.8-fold in the case of 1, 3 and 7 canceled interferers per node, respectively. Last, from simulation results in Section V, the TC scaling in Theorem 2 is observed to also hold for outage probabilities of practical interest ($\epsilon \leq 0.1$).

According to Theorem 2, the decay rate of the TC with varnishing outage probability can be slowed down by employing more antennas for interference cancellation at the cost of increasing CSI estimation overhead, which can be considered as overhead for local coordination among nearby nodes. However, to guarantee nonzero network capacity for zero outage probability, perhaps the better choice is to rely on centralized scheduling as in [16]. Nevertheless, such scheduling requires global coordination and potentially incurs much higher overhead than combining the random access protocol and spatial interference cancellation. Thus the current setup balances network performance and overhead.

IV. TRANSMISSION CAPACITY WITH IMPERFECT CSI

This section addresses the effect of imperfect CSI. First, consider the scenario where the network remains unchanged except that the CSI is imperfect and the users have to relax their quality-of-service (QoS) requirements, namely to tolerate higher outage probability represented by \tilde{P}_{out} and to lower the date rate from $\log_2(1+\theta)$ to $\log_2(1+\tilde{\theta})$, where $\tilde{\theta}$ denote the corresponding SIR threshold. Thus $\tilde{P}_{out} = \Pr(\widetilde{SIR} \leq \tilde{\theta})$ where \widetilde{SIR} is given in (11). The corresponding TC is

$$\tilde{C} = (1 - \tilde{P}_{\text{out}}) \log_2(1 + \tilde{\theta})\lambda.$$
(43)

It is interesting to investigate the required training sequence length under a constraint on the QoS degradation. To this end, define

$$\Delta P = \tilde{P}_{\mathsf{out}}(\tilde{\theta}) - P(\theta) \tag{44}$$

$$\Delta B = \log_2(1+\theta) - \log_2(1+\theta) \tag{45}$$

$$\Delta C = C - \tilde{C}.\tag{46}$$

We consider the following constraints on the QoS degradation: $\Delta P \leq \vartheta_p$ and $\Delta B \leq \vartheta_b$ with $\vartheta_p, \vartheta_b \geq 0$. The training sequence length that satisfies these constraints and the corresponding TC loss, specified by the ratio $\frac{\Delta C}{C}$, are shown in the following theorem.

Theorem 3. To satisfy the constraints $\Delta P \leq \vartheta_p$ and $\Delta B \leq \vartheta_b$, it is sufficient to choose the trainingsequence length as

$$M = \max\left(\left\lceil \frac{\log L - \log \vartheta_p}{\omega(2^{\vartheta_b} - 1)} \right\rceil, L\right)$$
(47)

with $\omega = [\Gamma(L+1)]^{-\frac{1}{L}}$. Moreover, the resultant TC loss normalized by C is bounded as

$$\frac{\Delta C}{C} \le \frac{\vartheta_p}{1-\epsilon} + \frac{\vartheta_b}{\log_2(1+\theta)}.$$
(48)

Proof: See Appendix H.

Note that (47) takes into account that $M \ge L$ (see Section II-F) and M is an integer. For stringent constraints $\vartheta_p \to 0$ and $\vartheta_b \to 0$, it can be observed from (47) that the training sequence length is approximately proportional to $\log \frac{1}{\vartheta_p}$ and $\frac{1}{\vartheta_b}$. Also, the upper-bound in (48) suggests that the normalized capacity loss is more sensitive to the variation of ϑ_p if ϵ is large and ϑ_b if $\log_2(1+\theta)$ is small.

Next, for small outage probability, the required training sequence length is derived for achieving the same TC scaling as for perfect CSI given identical QoS requirements. These results are shown in the following theorem.

Theorem 4. For $\epsilon \to 0$ and $\Delta B \xrightarrow{\epsilon} 0$, $\Delta P \xrightarrow{\epsilon} 0$,⁵, the following scaling of the training sequence length

$$\lim_{\epsilon \to 0} \frac{M}{\frac{1+\varrho}{\omega}\epsilon^{-\varrho}} = 1 \tag{49}$$

with an arbitrary $\rho > 0$ is sufficient for achieving the transmission-capacity scaling for perfect CSI as given in Theorem 2.

⁵The notation $A \xrightarrow{\epsilon} B$ represents the convergence $A \to B$ as $\epsilon \to 0$.



Fig. 2. Outage probability for different transmitting node densities and perfect CSI.

Proof: See Appendix I.

The above theorem shows that to achieve the optimal asymptotic TC, M increases slowly (sub-linearly with an arbitrary positive exponent) with $\frac{1}{\epsilon}$, indicating small CSI estimation overhead. The reason is that CSI inaccuracy rises mainly from weak interferers and thus its effect is moderate.

Suppose the CSI estimation process repeats for every channel coherence time t_c (in symbols). The overhead of CSI estimation can be regarded as the TC decrease by a factor of M/t_c . Simulation results in Fig. 6 reveal that the capacity gain from interference cancellation results mostly from canceling only a few strongest interferers and thus L can be kept small; furthermore, short training sequences (small M) are sufficient for approaching the TC achieved with perfect CSI. Thus, the overhead is insignificant if mobility is low (large t_c).

V. SIMULATION AND DISCUSSION

In this section, the bounds on outage probability and TC are evaluated using Monte Carlo simulation. The procedure for simulating a MANET follows that in [40]. The simulated ad hoc network lies on a two-dimensional disk and contains a number of transmitter-receiver pairs, which is a Poisson random variable with the mean equal to 200. The disk area is adjusted according to the node density. The typical receiver is placed at the center of the disk. We set the required SIR as $\theta = 3$ or 4.8 dB, the link array gain (N - L) = 2, and the path-loss exponent as $\alpha = 4$ unless specified otherwise.

A. Bounds on Outage Probability

For perfect CSI, the bounds on outage probability from Lemma 5 and simulated values are compared in Fig. 2. It can be observed that the outage probability is approximately proportional to λ^{L+1} . The bounds for L = 1 are tighter than those for L = 3. Moreover, the bounds on outage probability converge to the exact values as the transmitting node density λ decreases. These two observations can be explained



Fig. 3. Compare outage probability for perfect and imperfect CSI given different transmitting node densities. The trainingsequence length is $M = \{3, 5, 11\}$ and the number of canceled interferers per node is L = 3.

by the dominance of the primary interference over the secondary one as L or λ decreases, where the secondary interference causes the looseness of the bounds on outage probability.

In Fig. 3, given identical data rates ($\Delta B = 0$), the outage probability for imperfect CSI is observed to rapidly converge to the perfect-CSI counterpart as M increases. In particular, for M = 11, CSI inaccuracy increases outage probability by less than two-fold.

B. Scaling of Transmission Capacity

In Fig. 4, asymptotic bounds on TC in Theorem 2 are compared with the exact values obtained by simulation for perfect CSI and the range of target outage probability $\epsilon \in [10^{-5}, 10^{-1}]$. The corresponding curves are identified using the legends "asymptotic upper bound", "asymptotic lower bound", and "simulation". Different combinations of (L, α) are separated according to the cases of $L+1 \leq \alpha$ and $L+1 > \alpha$, corresponding to Fig. 4(a) and Fig. 4(b), respectively. As observed from Fig. 4(a), for $L+1 \leq \alpha$, the asymptotic upper bound on TC is tight even in the non-asymptotic range e.g., $\epsilon \in [0.01, 0.1]$. The tightness of this bound is due to the dominance of primary interference when L is small. Moreover, Fig. 4(b) shows that for $L+1 > \alpha$ the slopes of the "simulation" curves converge to those of the corresponding "asymptotic upper bound" curves as the target outage probability decreases. The above observations suggest that for both $L+1 \leq \alpha$ and $L+1 > \alpha$, the TC scaling for small target outage probability follows the power law with the same exponent $\frac{1}{L+1}$.

C. Transmission Capacity vs. Size of Antenna Array

In Fig. 5, the transmission capacity is plotted for an increasing number of canceled interferers per node assuming perfect CSI. Furthermore, different target outage probabilities, namely $\epsilon = \{10^{-1}, 10^{-2}, 10^{-3}\}$, are considered. As observed from Fig. 5, the cancellation of a few interferers by each node leads to a TC gain of an order of magnitude or more with respect to the case of no cancellation. For example, for $\epsilon = 10^{-2}$, canceling two interferers per node provides a 25-time TC gain. The cancellation of



(a) $L+1 \leq \alpha$



Fig. 4. Comparison between asymptotic bounds on TC and the exact values obtained by simulation for perfect CSI and the cases of (a) $L + 1 \le \alpha$ and (b) $L + 1 > \alpha$.

more interferers has a diminishing effect on the network capacity since it becomes limited by secondary interference. It is also observed that the outage constraint affects TC more significantly for smaller L.

The effect of imperfect CSI on TC is shown in Fig. 6, where TC is plotted for increasing L. The TC loss due to CSI estimation errors is observed to reduce as M increases. Such a loss is relatively small even for a moderate value of M. For instance, the TC reduction is 25% for M = 11 and L = 7. Next, even for small M (i.e., M = 3), a TC gain of more than an order of magnitude can be achieved by interference cancellation. This confirms the practicality of interference cancellation.

ACKNOWLEDGEMENT

The authors thank Nihar Jindal, Rahul Vaze, and Gustavo de Veciana for helpful discussions. In particular, Dr. Jindal suggested the derivation of Theorem 1.



Fig. 5. Transmission capacity by simulation for different node densities and perfect CSI. The target outage probability is $\epsilon = \{10^{-1}, 10^{-2}, 10^{-3}\}.$



Fig. 6. Transmission capacity by simulation for different node densities and imperfect CSI. The target outage probability is $\epsilon = 10^{-2}$.

APPENDIX

A. Proof for Lemma 1

Consider two disjoint measurable subsets \mathcal{B} and \mathcal{C} of $\mathbb{R}^2 \times [0,g)$. Let Ξ be the counting function such that $\Xi(\mathcal{B})$ gives the number of elements in \mathcal{B} . We first show that $\Xi(\Pi(g) \cap \mathcal{B})$ conditioned on $J_P = g$ is a Poisson random variable as follows. Define the sets $\mathcal{G}_{\tau} := \mathbb{R}^2 \times (g - \tau, \infty)$ and $\mathcal{D}_{\tau} = \mathbb{R}^2 \times [g - \tau, g + \tau]$

where $\tau > 0$. Then

$$\Pr(\Xi(\Pi(g) \cap \mathcal{B}) = n \mid J_P = g) = \lim_{\tau \to 0} \Pr(\Xi(\Phi \cap \mathcal{B}) = n \mid \Xi(\Phi \cap \mathcal{D}_{\tau}) = 1, \Xi(\Phi \cap \mathcal{G}_{\tau}) = L).$$
(50)

By letting $\tau \to 0$, the probability measures of $\mathcal{B} \cap \mathcal{D}_{\tau}$ and $\mathcal{B} \cap \mathcal{G}_{\tau}$ can be made arbitrarily small. Since Φ is Poisson distributed, $\Xi(\Phi \cap \mathcal{B})$ becomes independent with $\Xi(\Phi \cap \mathcal{D}_{\tau})$ and $\Xi(\Phi \cap \mathcal{G}_{\tau})$ in the limit of $\tau \to 0$. It follows that (50) can be rewritten as

$$\Pr(\Xi(\Pi(g) \cap \mathcal{B}) = n \mid J_P = g) = \Pr(\Xi(\Phi \cap \mathcal{B}) = n).$$
(51)

In other words, $\Xi(\Pi(g) \cap \mathcal{B})$ conditioned on $J_P = g$ follows the Poisson distribution with the same parameter as $\Xi(\Phi \cap \mathcal{B})$.

Next, we prove the independence between $\Xi(\Pi(g) \cap \mathcal{B})$ and $\Xi(\Pi(g) \cap \mathcal{C})$ given $J_P = g$. Their joint distribution function is written as

$$\Pr(\Xi(\Pi(g) \cap \mathcal{B}) = m, \Xi(\Pi(g) \cap \mathcal{C}) = n \mid J_P = g)$$

=
$$\lim_{\tau \to 0} \Pr(\Xi(\Phi \cap \mathcal{B}) = m, \Xi(\Phi \cap \mathcal{C}) = n \mid \Xi(\Phi \cap \mathcal{D}_{\tau}) = 1, \Xi(\Phi \cap \mathcal{G}_{\tau}) = L).$$
(52)

Following a similar argument as for getting (51), we can obtain from (52) that

$$\Pr(\Xi(\Pi(g) \cap \mathcal{B}) = m, \Xi(\Pi(g) \cap \mathcal{C}) = n \mid J_P = g) = \Pr(\Xi(\Phi \cap \mathcal{B}) = m, \Xi(\Phi \cap \mathcal{C}) = n)$$
$$= \Pr(\Xi(\Phi \cap \mathcal{B}) = m) \Pr(\Xi(\Phi \cap \mathcal{C}) = n) (53)$$

where the last equality is due to a property of the Poisson process Φ given that $\mathcal{B} \cap \mathcal{C} = \emptyset$. The independence between $\Xi(\Pi(g) \cap \mathcal{B})$ and $\Xi(\Pi(g) \cap \mathcal{C})$ conditioned on $J_P = g$ follows from (51) and (53).

Combining above results proves that given $J_P = g$, $\Pi(g)$ is Poisson distributed on the space $\mathbb{R}^2 \times [0, g)$; furthermore, $\Pi(g)$ has the mean measure as given in (12), which leads to (17). This completes the proof.

B. Proof for Lemma 2

The distribution of J_T can be analyzed by writing $J_T = r_T^{-\alpha} \rho_T$ where $\rho_T = \|\mathbf{G}_T \mathbf{f}_T\|^2$. Since \mathbf{G}_T is an i.i.d. $\mathcal{CN}(0,1)$ matrix and \mathbf{f}_T fixed, $\mathbf{G}_T \mathbf{f}_T$ is an i.i.d. $\mathcal{CN}(0,1)$ vector. Hence ρ_T has the chi-square distribution with N complex degrees of freedom. It follows that

$$\Pr(J_T > g \mid r_T = r) = \int_{r^{\alpha}g}^{\infty} \frac{u^{N-1}}{\Gamma(N)} e^{-u} du.$$
(54)

Substituting (54) into (15) gives

$$\mu^{*}(\mathcal{G}) = 2\pi\lambda \int_{0}^{\infty} \int_{r^{\alpha}g}^{\infty} \frac{ru^{N-1}}{\Gamma(N)} e^{-u} du dr$$

$$= 2\pi\lambda \int_{0}^{\infty} \int_{0}^{\left(\frac{u}{g}\right)^{\frac{1}{\alpha}}} \frac{ru^{N-1}}{\Gamma(N)} e^{-u} dr du$$

$$= \frac{\pi\lambda g^{-\frac{2}{\alpha}}}{\Gamma(N)} \int_{0}^{\infty} u^{N+\frac{2}{\alpha}-1} e^{-u} du$$

$$= \nu\lambda g^{-\frac{2}{\alpha}}$$
(55)

where ν is defined in the lemma statement. Then the distribution function of J_P can be written as

$$\Pr(J_P \le g) = \Pr(\Xi(\Psi \cap \mathcal{G}) \le L).$$
(56)

Since $\Xi(\Psi \cap \mathcal{G})$ is a Poisson random variable with the mean $\mu^*(\mathcal{G})$ given in (55), the desired cumulative distribution function in (18) follows from (56). Differentiating this function gives the probability density function of J_P as

$$f_P(g) = \frac{2}{\alpha} \sum_{k=0}^{L} \frac{(\nu\lambda)^{k+1}}{k!} g^{-\frac{2(k+1)}{\alpha} - 1} e^{-\nu\lambda g^{-\frac{2}{\alpha}}} - \frac{2}{\alpha} \sum_{k=1}^{L} \frac{(\nu\lambda)^k}{(k-1)!} g^{-\frac{2k}{\alpha} - 1} e^{-\nu\lambda g^{-\frac{2}{\alpha}}} \\ = \frac{2}{\alpha} e^{-\nu\lambda g^{-\frac{2}{\alpha}}} \left\{ \sum_{k=0}^{L} \frac{(\nu\lambda)^{k+1}}{k!} g^{-\frac{2(k+1)}{\alpha} - 1} - \frac{2}{\alpha} \sum_{k=0}^{L-1} \frac{(\nu\lambda)^{k+1}}{k!} g^{-\frac{2(k+1)}{\alpha} - 1} \right\}.$$

The desired result in (19) follows from the last equation.

C. Proof for Lemma 3

Consider an arbitrary interferer $T \in \Phi$ before interference cancellation. We can write the effective channel vector as $\mathbf{h}_T = J_T \tilde{\mathbf{h}}_T$. The isotropicity of \mathbf{h}_T has two consequences: J_T and $\tilde{\mathbf{h}}_T$ are independent and $\tilde{\mathbf{h}}_T$ is also isotropic. Recall that J_T is the criterion for selecting interference to cancel. Thus, the independence between J_T and $\tilde{\mathbf{h}}_T$ implies that the isotropicity of \mathbf{h}_T is unaffected by interference cancellation if node T is uncanceled.

Next, it can be observed from (4) that \mathbf{v}_0 is a linear function of the vectors \mathbf{h}_0 and $\{\mathbf{h}_T \mid T \in \mathcal{T}\}$, which are i.i.d. and isotropic. As a result, \mathbf{v}_0 is also isotropic as well as independent with other normalized channel vectors $\{\tilde{\mathbf{h}}_T \mid T \in \Phi \setminus \mathcal{T}\}$. Hence for an uncanceled interferer $T \in \Phi \setminus \mathcal{T}$, $\delta_T = |\mathbf{v}_0^{\dagger} \tilde{\mathbf{h}}_T|^2$ represents the product of two independent isotropic unit-norm random vectors \mathbf{v}_0 and $\tilde{\mathbf{h}}_T$, which is shown in [41] to have the beta(1, N - 1) distribution in (20). Moreover, the independence between δ_T and $\delta_{T'}$ for $T \neq T'$ follows from the independence between $\tilde{\mathbf{h}}_T$ and $\tilde{\mathbf{h}}_{T'}$. This proves the first claim in the lemma statement.

Last, given $T \in \Phi \setminus \mathcal{T}$, since both \mathbf{v}_0 and $\tilde{\mathbf{h}}_T$ are independent with J_T as mentioned above, the independence of δ_T with J_T is immediate. This completes the proof.

D. Proof of Lemma 4

Define the sum pre-cancellation secondary interference power as $J\mathfrak{B} := \sum_{T \in \Phi \setminus (\mathcal{T} \cup \{T_P\})} J_T$. Using Lemma 1, the application of Campbell's Theorem gives that [34]

$$\mathbb{E}[JB \mid J_P = g] = \lambda \int_{\mathbb{R}^2} \int_0^g up(|x|, du) du dx$$

$$= 2\pi \lambda \int_0^\infty \int_0^g rup(r, du) du dr$$

$$= 2\pi \lambda \int_0^\infty \int_0^{r^\alpha g} r^{1-\alpha} \frac{\rho^N}{\Gamma(N)} e^{-\rho} d\rho dr$$

$$= 2\pi \lambda \int_0^\infty \int_{\left(\frac{\rho}{g}\right)^{\frac{1}{\alpha}}}^\infty r^{1-\alpha} \frac{\rho^N}{\Gamma(N)} e^{-\rho} dr d\rho$$

$$= \frac{2\pi \lambda}{\alpha - 2} g^{1-\frac{2}{\alpha}} \int_0^\infty \frac{\rho^{N+\frac{2}{\alpha}-1}}{\Gamma(N)} e^{-\rho} d\rho$$

$$= \frac{2\nu \lambda}{\alpha - 2} g^{1-\frac{2}{\alpha}}.$$
(57)

Next, the expectation of $I\beta$ conditioned on $J_P = g$ is given as

$$\mathbb{E}[I\mathfrak{B} \mid J_P = g] = \mathbb{E}\left[\sum_{T \in \Phi \setminus (\mathcal{T} \cup \{T_P\})} J_T \delta_T \mid J_P = g\right]$$

$$= \mathbb{E}\left[\sum_{T \in \Phi \setminus (\mathcal{T} \cup \{T_P\})} J_T \mathbb{E}(\delta_T) \mid J_P = g\right]$$

$$= \frac{1}{N} \mathbb{E}[J\mathfrak{B} \mid J_P = g]$$
(58)

where (58) holds since δ_T is independent with J_P according to Lemma 3, and (59) uses $\mathbb{E}[\delta_T] = \frac{1}{N}$ derived using the distribution function in (20). Substituting (57) into (59) gives the desired result in (21).

Like (57), the conditional variance of $J\beta$ is obtained by applying Campbell's Theorem as follows

$$\operatorname{var}(J\mathfrak{B} \mid J_P = g) = 2\pi\lambda \int_0^\infty \int_{\left(\frac{\rho}{g}\right)^{\frac{1}{\alpha}}}^\infty r \left(r^{-\alpha}\rho\right)^2 \frac{\rho^{N-1}}{\Gamma(N)} e^{-\rho} d\rho dr$$
$$= \frac{\pi\lambda}{\alpha - 1} g^{2-\frac{2}{\alpha}} \int_0^\infty \rho^{N+\frac{2}{\alpha}-1} e^{-\rho} d\rho$$
$$= \frac{\nu\lambda}{\alpha - 1} g^{2-\frac{2}{\alpha}}.$$
(60)

Since δ_T is independent with J_P , the conditional variance of $I\beta$ is obtained by applying modified Campbell's Theorem [33, p77]

$$\operatorname{var}(I\mathfrak{B} | J_P = g) = \mathbb{E}[\delta_T^2]\operatorname{var}(J\mathfrak{B} | J_P = g).$$
(61)

The second moment $\mathbb{E}[\delta_T^2]$ can be obtained using the distribution function in (20) as

$$\mathbb{E}[\delta_T^2] = (N-1)B(3, N-1) \\ = \frac{2}{N(N+1)}$$
(62)

where B denotes the beta function, and (62) applies the formula $B(x, y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}$ from [42, 8.384]. Combining (60), (61) and (62) gives the desired result in (22). This completes the proof.

E. Proof for Lemma 5

We can rewrite (24) as

$$P_{\mathsf{out}} \ge 1 - \Pr\left(J_P \le \frac{W\theta^{-1}}{\delta_P}\right)$$

Substituting (18) into the above equation gives the lower bound on P_{out} as shown in (28). Similarly, we can obtain the first term of the P_{out} upper bound in (26) as

$$\Pr(J_P > W\theta^{-1}) = \Lambda_1 \tag{63}$$

where Λ_1 is defined (32). Given $\mathcal{D}_1 \cup \mathcal{D}_2 = \{(w,g) \mid g < w\theta^{-1}\}$, the second term of the P_{out} upper bound in (26) can be expanded and then upper bounded as

$$\Pr(I\mathfrak{B} > W\theta^{-1} - J_P \mid J_P \leq W\theta^{-1}) \Pr(J_P \leq W\theta^{-1})$$

$$\leq \Pr((W, J_P) \in \mathcal{D}_1) + \Pr(I\mathfrak{B} > W\theta^{-1} - J_P \mid (W, J_P) \in \mathcal{D}_2) \Pr((W, J_P) \in \mathcal{D}_2)$$

$$\leq \Lambda_2 + \Lambda_3$$
(64)

where Λ_2 and Λ_3 are defined in the lemma statement, and (64) applies Chebyshev's inequality in (27). Combining (26), (63) and (64) gives the desired P_{out} upper bound in (31). This completes the proof.

F. Proof of Theorem 1

Given the distribution of J_P in (19), and applying Campbell's Theorem, we obtain that

$$\mathbb{E}[I_P] = \mathbb{E}[J_P]\mathbb{E}[\delta_P]$$

$$= \frac{2(\nu\lambda)^{L+1}}{N\alpha\Gamma(L+1)} \int_0^\infty g^{-\frac{2(L+1)}{\alpha}} e^{-\nu\lambda g^{-\frac{2}{\alpha}}} dg$$

$$= \frac{\Gamma\left(L+1-\frac{\alpha}{2}\right)(\nu\lambda)^{\frac{\alpha}{2}}}{N\Gamma(L+1)}.$$
(65)

Moreover, from (19) and (21),

$$\mathbb{E}[I_S] = \mathbb{E}[\mathbb{E}[I_S | J_P]]$$

$$= \frac{4(\nu\lambda)^{L+2}}{N\alpha(\alpha-2)\Gamma(L+1)} \int_0^\infty g^{-\frac{2(L+2)}{\alpha}} e^{-\nu\lambda g^{-\frac{2}{\alpha}}} dg \qquad (66)$$

$$= \frac{2(\nu\lambda)^{\frac{1}{2}}\Gamma\left(L-\frac{\alpha}{2}+2\right)}{N(\alpha-2)\Gamma(L+1)}$$
(67)

where (66) uses (22). Using (65) and (67), the total interference power I_{Σ} has the following expectation:

$$\mathbb{E}[I_{\Sigma}] = \frac{2(\nu\lambda)^{\frac{\alpha}{2}}}{\alpha - 2} \times \frac{L\Gamma\left(L - \frac{\alpha}{2} + 1\right)}{N\Gamma(L + 1)}.$$
(68)

Using Markov's inequality

$$P_{\mathsf{out}} \le \mathbb{E}\left\{\frac{\mathbb{E}[I_{\Sigma}]}{W\theta^{-1}}\right\}.$$
(69)

Since $P_{out} = \epsilon$, it follows from (68) and (69) that

$$\lambda \ge \frac{1}{\nu} \left\{ \frac{N\epsilon \left(\alpha - 2\right)}{2L\theta \mathbb{E}[W^{-1}]} \times \frac{\Gamma(L+1)}{\Gamma\left(L - \frac{\alpha}{2} + 1\right)} \right\}^{\frac{2}{\alpha}}.$$
(70)

Next, a lower bound on λ can be derived using the method in [15] where the success probability $(1-P_{out})$ is upper bounded using Markov's inequality. To this end, let \mathcal{U} denote the set of (L+1) strongest uncanceled interferers in terms of pre-cancellation interference power. Thus, the weakest interferer in \mathcal{U} corresponds to $\hat{J} = \min_{T \in \mathcal{U}} J_T$. Using above definitions,

$$1 - P_{\text{out}} \leq \Pr\left(\frac{W}{\sum_{T \in \mathcal{U}} J_T \delta_T} \geq \theta\right)$$

$$\leq \Pr\left(\frac{W}{\hat{J}_T \sum_{T \in \mathcal{U}} \delta_T} \geq \theta\right)$$

$$\leq \theta^{-1} \mathbb{E}[W] \mathbb{E}[\hat{J}^{-1}] \mathbb{E}\left[\frac{1}{\sum_{T \in \mathcal{U}} \delta_T}\right]$$
(71)

where (71) uses Markov's inequality. Note that the probability density function of \hat{J} is given by (19) with L replaced with 2L. Thus, similar to (65), it can be obtained that

$$\mathbb{E}[\hat{J}^{-1}] = \frac{\Gamma(2L+1+\frac{\alpha}{2})}{\Gamma(2L+1)(\nu\lambda)^{\frac{\alpha}{2}}}.$$
(72)

Substituting the above equation into (71) gives

$$\lambda \leq \frac{1}{\nu} \left\{ \frac{\mathbb{E}[W]}{\theta(1-\epsilon)} \times \frac{\Gamma(2L+1+\frac{\alpha}{2})}{\Gamma(2L+1)} \times \mathbb{E}\left[\frac{1}{\sum_{T \in \mathcal{U}} \delta_T}\right] \right\}^{\frac{2}{\alpha}}.$$
(73)

The scaling of λ can be derived using (70) and (73) and applying Kershaw's inequality [43]:

$$\left(x + \frac{s}{2}\right)^{1-s} < \frac{\Gamma(x+1)}{\Gamma(x+s)} < \left(x - \frac{1}{2} + \sqrt{s + \frac{1}{4}}\right)^{1-s}, \quad x > 0, \ 0 < s < 1.$$
(74)

Specifically, given $\Gamma(1+x) = x\Gamma(x)$, we can write

$$\frac{\Gamma(L+1)}{\Gamma\left(L-\frac{\alpha}{2}+1\right)} = \frac{\Gamma\left(L-\lceil\frac{\alpha}{2}\rceil+2\right)}{\Gamma\left(L+1-\lceil\frac{\alpha}{2}\rceil+\Delta\alpha\right)} \prod_{n=0}^{\lceil\frac{\alpha}{2}\rceil-2} (L-n)$$
(75)

where $\Delta \alpha = \lceil \frac{\alpha}{2} \rceil - \frac{\alpha}{2}$ and hence $0 \le \Delta \alpha < 1$. Using Kershaw's inequality, for $L > \lceil \frac{\alpha}{2} \rceil$

$$\left(L - \left\lceil \frac{\alpha}{2} \right\rceil + 1 + \frac{\Delta\alpha}{2} \right)^{1 - \Delta\alpha} < \frac{\Gamma\left(L - \left\lceil \frac{\alpha}{2} \right\rceil + 2\right)}{\Gamma\left(L + 1 - \left\lceil \frac{\alpha}{2} \right\rceil + \Delta\alpha\right)} < \left(L - \left\lceil \frac{\alpha}{2} \right\rceil + \frac{1}{2} + \sqrt{\Delta\alpha + \frac{1}{4}} \right)^{1 - \Delta\alpha}.$$
(76)

It follows from (75) and (76) that

$$\lim_{L \to \infty} \frac{\Gamma(L+1)}{L^{\frac{\alpha}{2}} \Gamma\left(L - \frac{\alpha}{2} + 1\right)} = 1.$$
(77)

Similarly, we can show that ν as defined in Lemma 2 scales as:

$$\lim_{L \to \infty} \frac{\nu}{\pi L^{\frac{2}{\alpha}}} = 1.$$
(78)

Combining (70), (77) and (78) gives

$$\liminf_{L \to \infty} \frac{\lambda}{L^{1-\frac{2}{\alpha}}} \ge \frac{1}{\pi} \left\{ \frac{\epsilon \left(\alpha - 2\right)}{2\theta \mathbb{E}[W^{-1}]} \right\}^{\frac{2}{\alpha}}.$$
(79)

Again, the application of Kershaw's inequality yields

$$\lim_{L \to \infty} \frac{\Gamma(2L+1+\frac{\alpha}{2})}{(2L)^{\frac{\alpha}{2}} \Gamma(2L+1)} = 1.$$
(80)

Moreover, it follows from the strong law of larger numbers that $\lim_{L\to\infty} \frac{\sum_{T\in\mathcal{U}} \delta_T}{L+1} = \mathbb{E}[\delta_T] = \frac{1}{L}$. Since (N-L) is fixed,

$$\lim_{L \to \infty} \mathbb{E}\left[\frac{1}{\sum_{T \in \mathcal{U}} \delta_T}\right] = 1.$$
(81)

Substituting (80) and (81) into (73) gives

$$\limsup_{L \to \infty} \frac{\lambda}{L^{1-\frac{2}{\alpha}}} \le \frac{2}{\pi} \left\{ \frac{\mathbb{E}[W]}{\theta(1-\epsilon)} \right\}^{\frac{2}{\alpha}}.$$
(82)

The desired result follows from (79) and (82) as well as the TC definition.

G. Proof of Lemma 6

For $\lambda \to 0$, we can obtain from (28) that

$$P_{\mathsf{out}}^{\ell}(\lambda) = \sum_{k=L+1}^{\infty} \frac{\left(\nu\lambda\theta^{\frac{2}{\alpha}}\right)^{k}}{\Gamma(k+1)} \mathbb{E}\left[\left(\frac{W}{\delta_{P}}\right)^{-\frac{2k}{\alpha}} e^{-\nu\lambda\theta^{\frac{2}{\alpha}}\left(\frac{W}{\delta_{P}}\right)^{-\frac{2}{\alpha}}}\right]$$
$$= \mathbb{E}\left[\left(\frac{W}{\delta_{P}}\right)^{-\frac{2}{\alpha}(L+1)}\right] \frac{\left(\nu\lambda\theta^{\frac{2}{\alpha}}\right)^{L+1}}{\Gamma(L+2)} + O(\lambda^{L+2})$$
$$= \kappa_{1}\lambda^{L+1} + O(\lambda^{L+2})$$
(83)

where (83) uses the distributions of W and δ_P in (5) and Lemma 3, respectively, and κ_1 is defined in the lemma statement. The first inequalities in (36) and (39) follow from the last equation.

Next, we prove the second inequalities in (36) and (39) as follows. Similar to (84), for $\lambda \to 0$, Λ_1 in (32) is obtained as

$$\Lambda_1 = \frac{E\left[W^{-\frac{2}{\alpha}(L+1)}\right] \left(\nu\theta^{\frac{2}{\alpha}}\right)^{L+1}}{\Gamma(L+2)} \lambda^{L+1} + O(\lambda^{L+2}).$$
(85)

The asymptotic expression for Λ_2 in (33) is derived as

$$\Lambda_2(\lambda) = \frac{2(\nu\lambda)^{L+1}}{\alpha\Gamma(L+1)} \int_0^\infty \int_{w\theta^{-1}+O(\lambda)}^{w\theta^{-1}} g^{-\frac{2(L+1)}{\alpha}-1} e^{-\nu\lambda g^{-\frac{2}{\alpha}}} dg f_W(w) dw$$
(86)

$$= \frac{2(\nu\lambda)^{L+1}}{\alpha\Gamma(L+1)} \int_0^\infty \left[\left(w\theta^{-1} \right)^{-\frac{2(L+1)}{\alpha}-1} + O(\lambda) \right] \times O(\lambda) f_W(w) dw$$

$$= O(\lambda^{L+2}). \tag{87}$$

where (86) uses (21). To derive the asymptotic expression for Λ_3 in (34), it is split into two terms as $\Lambda_3 = \Lambda_{3,1} + \Lambda_{3,2}$ where

$$\Lambda_{3,i} = \iint_{(w,g)\in\mathcal{D}_{2,i}} \min\left\{\frac{\operatorname{var}(I \mid J_P = g)}{\{w\theta^{-1} - g - \mathbb{E}\left[I \mid J_P = g\right]\}^2}, 1\right\} f_W(w) f_P(g) dwdg$$
(88)

with

$$\mathcal{D}_{2,1} = \left\{ (w,g) \mid \frac{w\theta^{-1}}{2} \le g + \mathbb{E}[I\mathfrak{B} \mid J_P = g] < w\theta^{-1} \right\}$$
$$\mathcal{D}_{2,2} = \left\{ (w,g) \mid 0 \le g + \mathbb{E}[I\mathfrak{B} \mid J_P = g] < \frac{w\theta^{-1}}{2} \right\}.$$
(89)

For $\lambda \to 0$, $\Lambda_{3,1}$ is obtained as

$$\Lambda_{3,1} \leq \iint_{(w,g)\in\mathcal{D}_{2,1}} f_W(w) f_P(g) dw dg$$

$$= \frac{(\nu\lambda)^{L+1}}{\Gamma(L+1)} \int_0^\infty \int_{(w\theta^{-1})^{-\frac{2}{\alpha}}+O(\lambda)}^{\left(\frac{w\theta^{-1}}{2}\right)^{-\frac{2}{\alpha}}+O(\lambda)} g^L dg f_W(w) dw + O(\lambda^{L+2})$$

$$= \frac{\left[2^{\frac{2(L+1)}{\alpha}}-1\right] \mathbb{E}\left[W^{-\frac{2(L+1)}{\alpha}}\right] \left(\theta^{\frac{2}{\alpha}}\nu\right)^{L+1}}{\Gamma(L+2)} \lambda^{L+1} + O(\lambda^{L+2}).$$
(90)

Next, $\Lambda_{3,2}$ defined in (88) is upper bounded as

$$\Lambda_{3,2} \leq \iint_{(w,g)\in\mathcal{D}_{2,2}} \frac{\operatorname{var}(I\mathbb{B}\mid J_P = g)}{(w\theta^{-1} - g - \mathbb{E}[I\mathbb{B}\mid J_P = g])^2} f_P(g) f_W(w) dw dg$$

$$\leq \int_0^\infty \int_0^{\frac{w\theta^{-1}}{2} + O(\lambda)} \frac{4\operatorname{var}(I\mathbb{B}\mid J_P = g)}{(w\theta^{-1})^2} f_P(g) dg f_W(w) dw.$$
(91)

where (91) holds since $g + \mathbb{E}[I\beta \mid J_P = g] < \frac{w\theta^{-1}}{2}$ according to (89). To simplify notation, define $\eta = \frac{8\theta^2}{N(N+1)(\alpha-1)\Gamma(L+1)}$. Substituting the distribution functions in (19) and (22) into (91) gives

$$\Lambda_{3,2} \leq \frac{2}{\alpha} \eta(\nu\lambda)^{L+2} \int_{0}^{\infty} w^{-2} \int_{0}^{\frac{w\theta^{-1}}{2} + O(\lambda)} g^{-\frac{2}{\alpha}(L+2)+1} e^{-\nu\lambda g^{-\frac{2}{\alpha}}} dg f_{W}(w) dw$$

= $\eta(\nu\lambda)^{\alpha} \int_{0}^{\infty} w^{-2} \int_{\left(\frac{w\theta^{-1}}{2}\right)^{-\frac{2}{\alpha}} \nu\lambda + O(\lambda^{2})}^{\infty} g^{L-\alpha+1} e^{-g} dg f_{W}(w) dw.$ (92)

For $L - \alpha + 1 < 0$, we obtain using (92) that

$$\Lambda_{3,2} \leq \eta(\nu\lambda)^{\alpha} \int_{0}^{\infty} w^{-2} \left[\left(\frac{w\theta^{-1}}{2} \right)^{-\frac{2}{\alpha}} \nu\lambda \right]^{L-\alpha+1} f_{W}(w) dw + O(\lambda^{L+2}) \\ = \frac{2^{\frac{2}{\alpha}(L+1)+1}(L+1)}{N(N+1)(\alpha-1)} \times \frac{\mathbb{E} \left[W^{-\frac{2}{\alpha}(L+1)} \right] (\theta^{\frac{2}{\alpha}} \nu)^{L+1}}{\Gamma(L+2)} \lambda^{L+1} + O(\lambda^{L+2}).$$
(93)

Since $P_{\mathsf{out}}^u \leq \Lambda_1 + \Lambda_2 + \Lambda_{3,1} + \Lambda_{3,2}$,

$$\limsup_{\lambda \to 0} \frac{P_{\mathsf{out}}^u}{\lambda^{L+1}} \le \liminf_{\lambda \to 0} \frac{\Lambda_1}{\lambda^{L+1}} + \liminf_{\lambda \to 0} \frac{\Lambda_2}{\lambda^{L+1}} + \liminf_{\lambda \to 0} \frac{\Lambda_{3,1}}{\lambda^{L+1}} + \liminf_{\lambda \to 0} \frac{\Lambda_{3,2}}{\lambda^{L+1}}.$$
 (94)

The substitution of (85), (87), (90) and (93) into (94) gives that

$$\limsup_{\lambda \to 0} \frac{P_{\mathsf{out}}^{u}}{\lambda^{L+1}} \leq \left[1 + \frac{2(L+1)}{N(N+1)(\alpha-1)} \right] \frac{2^{\frac{2}{\alpha}(L+1)} \mathbb{E} \left[W^{-\frac{2}{\alpha}(L+1)} \right] (\theta^{\frac{2}{\alpha}} \nu)^{L+1}}{\Gamma(L+2)} \\ \leq 2 \times \frac{2^{\frac{2}{\alpha}(L+1)} \mathbb{E} \left[W^{-\frac{2}{\alpha}(L+1)} \right] (\theta^{\frac{2}{\alpha}} \nu)^{L+1}}{\Gamma(L+2)} \tag{95}$$

where the last inequality holds since $N \ge L + 1$ and $\alpha > 2$. The second inequality in (36) follows. For $L - \alpha + 1 \ge 0$, $\Gamma(L - \alpha + 2)$ is finite and hence we obtain from (92) that

$$\Lambda_{3,2} \le \kappa_3 \lambda^\alpha + O(\lambda^{\alpha+1}) \tag{96}$$

-

where κ_3 is defined in the lemma statement. Substituting (85), (87), (90) and (96) into (94) gives the second equality in (39). This completes the proof.

H. Proof of Theorem 3

Using (11) and by definition, the outage probability for imperfect CSI is given as

$$\tilde{P}_{\mathsf{out}}(\tilde{\theta}) = \mathbb{E}\left[\Pr\left(\left.W\tilde{\theta}^{-1} \le \sigma_R^2 + I_{\Sigma}\right| \Phi, \{\mathbf{h}_T\}\right)\right].$$
(97)

Conditioned on Φ and $\{\mathbf{h}_T\}$, the randomness of the residual interference I_R in (10) depends only on the data symbols $\{x_T\}$ that follow i.i.d. $\mathcal{CN}(0,1)$ distributions. Thus, the conditional variance σ_R^2 of I_R can be written as

$$\sigma_{R}^{2} = \frac{1}{M} \sum_{T \in \mathcal{T}} \left| \sum_{T' \in \Phi \setminus \mathcal{T}} r_{T'}^{-\alpha/2} \mathbf{v}_{0}^{\dagger} \mathbf{G}_{T'} \tilde{x}_{T',T} \right|^{2}$$
$$= \frac{1}{M} \sum_{T \in \mathcal{T}} \left| \sum_{T' \in \Phi \setminus \mathcal{T}} I_{T'} \frac{\mathbf{v}_{0}^{\dagger} \mathbf{G}_{T'} \tilde{x}_{T',T}}{|\mathbf{v}_{0}^{\dagger} \mathbf{G}_{T'}|} \right|^{2}.$$
(98)

Let ~ represent equivalence in distribution. Since $\left\{\frac{\mathbf{v}_{0}^{\dagger}\mathbf{G}_{T'}\tilde{x}_{T',T}}{|\mathbf{v}_{0}^{\dagger}\mathbf{G}_{T'}|}\right\}$ consists of i.i.d. $\mathcal{CN}(0,1)$ elements, we obtain from (98) that conditioned on Φ and $\{\mathbf{h}_{T}\}$,

$$\sigma_R^2 \sim \frac{1}{M} \sum_{T' \in \Phi \setminus \mathcal{T}} I_{T'} \sum_{T \in \mathcal{T}} z_T$$

$$\sim \frac{\zeta}{M} I_{\Sigma}$$
(99)

where $\{z_T\}$ are i.i.d. exponential random variables with unit mean and ζ is a chi-square random variable having L complex degrees of freedom. By substituting (99) into (97), we obtain that

$$\tilde{P}_{\mathsf{out}}(\tilde{\theta}) = \Pr\left(W\tilde{\theta}^{-1} \le I_{\Sigma}\left(1 + \frac{\zeta}{M}\right)\right).$$
(100)

Given Z > 0, the above expression can be expanded as

$$\tilde{P}_{\mathsf{out}}(\tilde{\theta}) = \Pr\left(W\tilde{\theta}^{-1} \le I_{\Sigma}\left(1 + \frac{\zeta}{M}\right) \middle| \zeta \le Z\right) \Pr(\zeta \le Z) +$$

$$\Pr\left(W\tilde{\theta}^{-1} \le I_{\Sigma}\left(1 + \frac{\zeta}{M}\right) \middle| \zeta > Z\right) \Pr(\zeta > Z)$$

$$\le \Pr\left(W\tilde{\theta}^{-1} \le I_{\Sigma}\left(1 + \frac{Z}{M}\right)\right) + \Pr(\zeta > Z).$$
(101)
(102)

By setting $\theta = \left(1 + \frac{Z}{M}\right)\tilde{\theta}$, the inequality in (102) reduces to

$$\tilde{P}_{\mathsf{out}}(\tilde{\theta}) \le P_{\mathsf{out}}(\theta) + \Pr(\zeta > Z)$$

It follows that

$$\Delta P \leq \Pr(\zeta > Z)$$

$$\leq 1 - (1 - e^{-\omega Z})^{L}$$
(103)

$$\leq Le^{-\omega Z}$$
 (104)

where ω is defined in the theorem statement, (103) applies Alzer's inequalities for the incomplete Gamma function [44], and (104) uses Bernoulli's inequality. Moreover, given $\theta = \left(1 + \frac{Z}{M}\right)\tilde{\theta}$, the rate loss is

bounded as

$$\Delta B = \log_2(1+\theta) - \log_2\left(1 + \frac{\theta}{1+\frac{Z}{M}}\right)$$

$$\leq \log_2(1+\theta) - \log_2\left(\frac{1+\theta}{1+\frac{Z}{M}}\right)$$

$$= \log_2\left(1 + \frac{Z}{M}\right).$$
(105)

From (104) and (105), to satisfy the constraints $\Delta P \leq \vartheta_p$ and $\Delta B \leq \vartheta_b$, it is sufficient that

$$Le^{-\omega Z} = \vartheta_p$$

$$\log_2\left(1 + \frac{Z}{M}\right) = \vartheta_b.$$
(106)

Solving the above equations gives the training sequence length in (47).

Finally, from (1) and (43), the capacity loss defined in (46) is upper bounded as

$$\begin{split} \Delta C &= (1-\epsilon)\lambda \log_2(1+\theta) - (1-\tilde{P}_{\mathsf{out}})\lambda \log_2(1+\tilde{\theta}) \\ &= \lambda \left[(1-\epsilon) \log_2(1+\theta) - (1-\tilde{P}_{\mathsf{out}}) \log_2(1+\theta) + (1-\tilde{P}_{\mathsf{out}}) \log_2(1+\theta) \right] \\ &= \lambda \left[\Delta P \log_2(1+\theta) + (1-\tilde{P}_{\mathsf{out}}) \Delta B \right] \\ &= C \left[\frac{\Delta P}{1-\epsilon} + \frac{1-\tilde{P}_{\mathsf{out}}}{1-\epsilon} \times \frac{\Delta B}{\log_2(1+\theta)} \right] \\ &\leq C \left[\frac{\Delta P}{1-\epsilon} + \frac{\Delta B}{\log_2(1+\theta)} \right]. \end{split}$$

The desired result in (48) follows from the above inequality, completing the proof.

I. Proof of Theorem 4

We prove in the sequel that the training sequence length stated in the theorem achieves the TC scaling in (41) with $\Delta P \xrightarrow{\epsilon} 0$ and $\Delta B \xrightarrow{\epsilon} 0$. The parallel proof concerning the other capacity scaling in (42) is similar and omitted for brevity. Recall that λ is fixed regardless of whether CSI is perfect. Thus, it follows from (41) that for sufficiently small ϵ ,

$$\kappa_2^{-\frac{1}{L+1}} \le \frac{\tilde{C}(\epsilon)}{\log(1+\tilde{\theta})\epsilon^{\frac{1}{L+1}}} \le \kappa_1^{-\frac{1}{L+1}}.$$
(107)

As in Appendix H, we set $\theta = (1 + \frac{Z}{M}) \tilde{\theta}$ with Z > 0 and thus (104) holds. Furthermore, we choose Z and M such that $Le^{-\omega Z} = \epsilon^{1+\varrho}$ and $\frac{Z}{M} = \epsilon^{\varrho} \log \frac{1}{\epsilon}$ with $\varrho > 0$, which yields $\Delta P \xrightarrow{\epsilon} 0$ as a result of

(104) and has two other consequences:

$$M = \frac{1+\varrho}{\omega}\epsilon^{-\rho} + \frac{\log L}{\omega} \times \frac{\epsilon^{-\varrho}}{\log \frac{1}{\epsilon}}$$
(108)

$$= \frac{1+\varrho}{\omega}\epsilon^{-\rho} + o\left(\epsilon^{-\varrho}\right) \tag{109}$$

$$\lim_{\epsilon \to 0} \frac{\theta}{\tilde{\theta}} = \lim_{\epsilon \to 0} \left(1 + \frac{Z}{M} \right)$$
$$= \lim_{\epsilon \to 0} \left(1 + \epsilon^{\varrho} \log \frac{1}{\epsilon} \right)$$
$$= 1.$$
(110)

It follows from (110) that $\Delta B \xrightarrow{\epsilon} 0$ and $\lim_{\epsilon \to 0} \frac{\log_2(1+\theta)}{\log_2(1+\tilde{\theta})} = 1$. Combining the last inequality and (107) gives that for sufficiently small ϵ ,

$$\kappa_2^{-\frac{1}{L+1}} \le \frac{\tilde{C}(\epsilon)}{\log(1+\theta)\epsilon^{\frac{1}{L+1}}} \le \kappa_1^{-\frac{1}{L+1}}$$
(111)

Therefore, the TC scaling for imperfect CSI is identical to the perfect-CSI counterpart in (41). Furthermore, the scaling of M in (49) follows from (108). Since above results hold for an arbitrary $\rho > 0$, the proof is complete.

REFERENCES

- [1] S. P. Weber, X. Yang, J. G. Andrews, and G. de Veciana, "Transmission capacity of wireless ad hoc networks with outage constraints," *IEEE Trans. on Inform. Theory*, vol. 51, pp. 4091–4102, Dec. 2005.
- [2] S. P. Weber, J. G. Andrews, and N. Jindal, "The effect of fading, channel inversion, and threshold scheduling on ad hoc networks," *IEEE Trans. on Inform. Theory*, vol. 53, pp. 4127–4149, Nov. 2007.
- [3] A. Hasan and J. G. Andrews, "The guard zone in wireless ad hoc networks," *IEEE Trans. on Wireless Communications*, vol. 6, pp. 897–906, Mar. 2007.
- [4] R. K. Ganti and M. Haenggi, "Regularity, interference, and capacity of large ad hoc networks," in *Proc., IEEE Asilomar*, Oct. 2006.
- [5] J. Venkataraman, M. Haenggi, and O. Collins, "Shot noise models for the dual problems of cooperative coverage and outage in random networks," in *Proc., Allerton Conf. on Comm., Control, and Computing*, Sept. 2006.
- [6] N. Jindal, J. G. Andrews, and S. P. Weber, "Bandwidth partitioning in decentralized wireless networks," *IEEE Trans. on Wireless Communications*, vol. 7, pp. 5408–5419, Jul. 2008.
- [7] S. P. Weber, J. G. Andrews, X. Yang, and G. de Veciana, "Transmission capacity of wireless ad hoc networks with successive interference cancelation," *IEEE Trans. on Inform. Theory*, vol. 53, pp. 2799–2814, Aug. 2007.
- [8] A. M. Hunter, J. G. Andrews, and S. P. Weber, "Transmission capacity of ad hoc networks with spatial diversity," *IEEE Trans. on Wireless Communications*, vol. 7, pp. 5058–5071, Dec. 2008.
- [9] M. Haenggi, J. G. Andrews, F. Baccelli, O. Dousse, and M. Franceschetti, "Stochastic geometry and random graphs for the analysis and design of wireless networks," *IEEE Journal on Selected Areas in Communications*, vol. 27, pp. 1029–1046, Jul. 2009.
- [10] K. Huang, J. G. Andrews, R. W. Heath, Jr., D. Guo, and R. A. Berry, "Spatial interference cancelation for mobile ad hoc networks: Imperfect CSI," in *Proc., IEEE Asilomar*, pp. 131–135, Oct. 2008.
- [11] K. Huang, J. G. Andrews, R. W. Heath, Jr., D. Guo, and R. A. Berry, "Spatial interference cancelation for mobile ad hoc networks: Perfect CSI," in *Proc.*, *IEEE Globecom*, pp. 1–5, Nov. 2008.
- [12] R. H. Y. Louie, M. R. McKay, and I. B. Collings, "Spatial multiplexing and diversity communications in ad hoc networks," *to appear in IEEE Trans. on Inform. Theory.*

- [13] M. Kountouris and J. G. Andrews, "Transmission capacity scaling of SDMA in wireless ad hoc networks," in Proc., Information Theory Workshop, pp. 534–538, Oct. 2009.
- [14] R. Vaze and R. W. Heath Jr., "Optimal use of multiple antennas in ad-hoc networks : Transmission capacity perspective," *submitted to IEEE Trans. on Inform. Theory.*
- [15] N. Jindal, J. G. Andrews, and S. Weber, "Multi-antenna communication in ad hoc networks: achieving MIMO gains with SIMO transmission," *submitted to IEEE Trans. on Communications*.
- [16] P. Gupta and P. R. Kumar, "The capacity of wireless networks," *IEEE Trans. on Inform. Theory*, vol. 46, pp. 388–404, Mar. 2000.
- [17] V. R. Cadambe and S. A. Jafar, "Interference alignment and the degrees of freedom for the K user interference channel," *IEEE Trans. on Inform. Theory*, vol. 54, pp. 3425–3441, Aug. 2008.
- [18] M. Zorzi, J. Zeidler, A. Anderson, B. Rao, J. Proakis, A. L. Swindlehurst, M. Jensen, and S. Krishnamurthy, "Cross-layer issues in MAC protocol design for MIMO ad hoc networks," *IEEE Wireless Communications Magazine*, vol. 13, pp. 62–76, Aug. 2006.
- [19] B. Hamdaoui and K. G. Shin, "Characterization and analysis of multi-hop wireless MIMO network throughput," in Proc., Intl. Symp. on Mobile Ad Hoc Networking and Computing, pp. 120–129, 2007.
- [20] J. C. Mundarath, P. Ramanathan, and B. D. V. Veen, "A cross layer scheme for adaptive antenna array based wireless ad hoc networks in multipath environments," *Wireless Networks*, vol. 13, no. 5, pp. 597–615, 2007.
- [21] J.-S. Park, A. Nandan, M. Gerla, and H. Lee, "SPACE-MAC: enabling spatial reuse using MIMO channel-aware MAC," in Proc., IEEE Intl. Conf. on Communications, vol. 5, pp. 3642–3646, May 2005.
- [22] M. Z. Siam, M. Krunz, A. Muqattash, and S. Cui, "Adaptive multi-antenna power control in wireless networks," in *Proc.*, *Intl. Conf. on Wireless Comm. and Mobile Computing*, pp. 875–880, Jul. 2006.
- [23] R. Ramanathan, J. Redi, C. Santivanez, D. Wiggins, and S. Polit, "Ad hoc networking with directional antennas: a complete system solution," *IEEE Journal on Selected Areas in Communications*, vol. 23, pp. 496–506, Mar. 2005.
- [24] A. Deopura and A. Ganz, "Provisioning link layer proportional service differentiation in wireless networks with smart antennas," Wireless Networks, vol. 13, Mar. 2007.
- [25] K. Sundaresan and R. Sivakumar, "A unified MAC layer framework for ad-hoc networks with smart antennas," *IEEE Trans.* on Networking, vol. 15, pp. 546–559, Mar. 2007.
- [26] H. Singh and S. Singh, "Smart-ALOHA for multi-hop wireless networks," *Mobile Networks and Applications*, vol. 10, pp. 651–662, May 2005.
- [27] R. Ramanathan, "On the performance of ad hoc networks with beamforming antennas," in *Proc., Intl. Symp. on Mobile Ad Hoc Networking and Computing*, pp. 95–105, Oct. 2001.
- [28] Y. Wu, L. Zhang, Y. Wu, and Z. Niu, "Interest dissemination with directional antennas for wireless sensor networks with mobile sinks," in *Proc., Intl. Conf. on Embedded Networked Sensor Systems*, pp. 99–111, Nov. 2006.
- [29] S. Kumar, V. S. Raghavan, and J. Deng, "Medium access control protocols for ad hoc wireless networks: A survey," Ad Hoc Networks, vol. 4, pp. 326–358, May 2006.
- [30] J. Zhang and S. C. Liew, "Capacity improvement of wireless ad hoc networks with directional antennae," SIGMOBILE Mobile Computing Comm. Review, vol. 10, Apr. 2006.
- [31] S. Yi, Y. Pei, S. Kalyanaraman, and B. Azimi-Sadjadi, "How is the capacity of ad hoc networks improved with directional antennas?," *Wireless Networks*, vol. 13, no. 5, pp. 635–648, 2007.
- [32] F. Baccelli, B. Blaszczyszyn, and P. Muhlethaler, "An ALOHA protocol for multihop mobile wireless networks," *IEEE Trans. on Inform. Theory*, vol. 52, pp. 421–36, Feb. 2006.
- [33] J. F. C. Kingman, Poisson processes. Oxford University Press, 1993.
- [34] D. Stoyan, W. S. Kendall, and J. Mecke, Stochastic Gemoetry and its Applications. Wiley, 2nd ed., 1995.
- [35] J. G. Andrews, S. Weber, M. Kountouris, and M. Haenggi, "Random access transport capacity," *IEEE Trans. on Wireless Communications*, vol. 9, pp. 2101–2111, Jun. 2010.
- [36] T. K. Y. Lo, "Maximum ratio transmission," IEEE Trans. on Communications, vol. 47, pp. 1458–1461, Oct. 1999.
- [37] A. Paulraj, R. Nabar, and D. Gore, Introduction to Space-Time Wireless Communications. Cambridge University Press, 2003.

- [38] S. Verdu, Multiuser Detection. Cambridge, UK: Cambridge, 1998.
- [39] T. L. Marzetta and B. M. Hochwald, "Fast transfer of channel state information in wireless systems," *IEEE Trans. on Sig. Proc.*, vol. 54, pp. 1268–1278, Apr. 2006.
- [40] S. P. Weber and M. Kam, "Computational complexity of outage probability simulations in mobile ad-hoc networks," in Proc., Conf. on Information Sciences and Systems, Mar. 2005.
- [41] C. K. Au-Yeung and D. J. Love, "On the performance of random vector quantization limited feedback beamforming in a MISO system," *IEEE Trans. on Wireless Communications*, vol. 6, pp. 458–462, Feb. 2007.
- [42] I. Gradshteyn and I. Ryzhik, Table of integrals, series, and products. Academic Press, 7 ed., 2007.
- [43] D. Kershaw, "Some extensions of W. Gautschi's inequalities for the Gamma function," *Math. of Computation*, vol. 41, pp. 607–611, Oct. 1983.
- [44] H. Alzer, "On some inequalities for the incomplete Gamma function," *Mathematics of Computation*, vol. 66, pp. 771–778, Apr. 2005.