Information Theoretic Upper Bound on the Capacity of Wireless Backhaul Networks

Harpreet S. Dhillon and Giuseppe Caire

***Abstract*—We derive an information theoretic upper bound on the capacity of a wireless backhaul network modeled as a classical *random extended network*, except that we assume the number of antennas at each base station (BS) also scales up as an arbitrary function of network size. The antenna scaling is justified because of the increasing maturity of higher transmission frequencies which enables us to pack large number of antennas in small form factors. The main technical arguments are based on the generalization of *geometric exponential stripping* technique of [1] to channel matrices with complex-valued channel gains. An important consequence of our result is a lower bound on the number of antennas per BS required for network scalability.**

I. Introduction

Scaling laws provide a useful way to characterize capacity of large wireless networks. Initiated by Gupta and Kumar's seminal work [2], this area received a lot of attention in the past decade, which significantly improved our understanding of large wireless networks. Perhaps the most widely known result is the *square-root* n law with which the capacity of a n node network scales [2]. While it was first believed that this may not be achievable for a random network setting, Franceschetti *et al.* [3] closed this gap with an achievable strategy using percolation theory. Another notable contribution was by Özgür *et al.* [4], which demonstrated that linear scaling with n was possible using hierarchical cooperation. An interesting dichotomy appeared when Franceschetti *et al.* [5] argued that linear scaling is not possible and that *square-root* n law is the fundamental limit. This debate was settled independently by Lee *et al.* [6] and Özgür *et al.* [7], where the main conclusion was that both the earlier results are correct and that they are applicable in different regimes.

A notable exception in this otherwise well-studied area is the capacity of multi-antenna wireless networks, especially where the number of antennas can also scale as some arbitrary function of total number of nodes n. This regime is quickly becoming relevant for wireless backhaul in the current cellular networks. Due to the increasing understanding of propagation at higher transmission frequencies and push from key industrial players to go for millimeter-wave communications, it is now becoming realistic to pack more and more antennas in small form factors [8], [9]. While there have been some recent attempts to understand practical considerations for the design of a multi-antenna backhaul network, e.g., [10], there are no known results on the performance limits. In this paper, we take a step in this direction and derive an information theoretic upper bound on its capacity by modeling this regime as a random extended network, where the number of antennas per node can also scale with the network size. Note that this antenna scaling assumption, although not mainstream, already appears in the literature in slightly different contexts. For instance, [11] studies the capacity of single-antenna ad hoc networks with infrastructure support, where the number of antennas per infrastructure BS increases with the network size.

The study of information theoretic bounds for the capacity scaling of wireless networks was initiated by Xie *et al.* in [12], where the main result was restricted to power-law pathloss exponent $\alpha > 6$. Two follow-up contributions of particular interest for this paper are [1], [13], where in [13] Lévêque *et al.* derived a useful bound valid for any $\alpha > 2$, and in [1] Franceschetti gave an alternate geometric proof for the bound of [13]. However, both [1], [13] ignored phase rotation in the channel gains. In this paper, we generalize the geometric arguments of [1] to more general channel models, and derive an information theoretic upper bound for multi-antenna random extended networks. An important consequence of this result is a lower bound on the number of antennas per node required to make the backhaul network scalable, i.e., to provide non-zero rate to each source-destination pair independent of n.

II. System Model

We consider a cellular network where the BS locations are sampled from a unit density homogeneous Poisson Point Process (PPP) $\Phi \subset \mathbb{R}^2$. Note that the results derived in this paper trivially generalize to any given finite BS density. For the scaling results, we consider the *random extended network model*, and focus our attention on a box B_n with size $\sqrt{n} \times \sqrt{n}$ [14]. The number of BSs lying in B_n is a Poisson distributed random variable with mean n. For the fixed BS density, we are interested in the capacity scaling of the network formed by BSs inside B_n as $n \to \infty$. It should be noted that as $n \to \infty$, B_n also grows, eventually encompassing all the points of Φ. We assume uniform traffic across B_n. The source-destination pairs are picked uniformly at random, such that each BS is the destination of exactly one source BS.

We assume BSs do not have access to wired backhaul. All the data is communicated over wireless backhaul links sharing the same spectrum. For the wireless backhaul links, each BS has $\Psi(n)$ antennas, where $\Psi(n)$ is a non-decreasing function of n. Scaling the number of antennas with the network size n may not come naturally to some readers, but it should be noted

The authors are with the Communication Sciences Institute (CSI), Department of Electrical Engineering, University of Southern California, Los Angeles, CA (email: {hdhillon; caire}@usc.edu).

This work was supported by NSF CCF-1161801 and a gift from Intel for research on 5G mm-wave wireless networks.

that it is not unrealistic, especially with the increasing maturity of higher transmission frequencies, e.g., at 28 and 38 GHz [8], which allows to pack more and more antennas in manageable form factors. For instance, a matchbook-sized prototype of 64 antenna array has already been demonstrated [9]. We further assume that the physical dimensions (size) of antenna array does not change with n. Denoting the distance between the k^{th} antenna of the transmitting BS to the i^{th} antenna of the receiving BS by d_{ik}, the baseband channel gain h_{ik} between these two antennas is

$$h_{ik} = \sqrt{l(d_{ik})} \exp\left(j\theta_{ik}\right), \tag{1}$$

where $l(d_{ik}) = \min\{1, d_{ik}^{-\alpha}\}$ is a bounded power-law pathloss function with exponent $\alpha > 2$, and θ_{ik} denotes phase rotation. For line-of-sight channels, $\theta_{ik} = \frac{2\pi d_{ik}}{\lambda}$, where λ is the transmission wavelength. As will be evident in the next section, our analysis holds for any given $\{\theta_{ik}\}$, $1 \leq i, k \leq \Psi(n)$, irrespective of their joint distribution. Each node is assumed to have a maximum power constraint of P watts.

Before concluding this section, we introduce the main metric of interest and the order notation. We denote the network throughput, i.e., total number of bits successfully transmitted per second by all nodes in B_n, by $T(n)$. The per source-destination rate is $R(n)$. The following probabilistic version of the ordering notation is used [3]. We write $f(n) = O(g(n))$ with high probability (w.h.p.) if $\exists$ a constant K independent of n such that $\lim_{n\to\infty} \mathbb{P}(f(n) \leq Kg(n)) = 1$. Similarly, $f(n) = \Omega(g(n))$ if $g(n) = O(f(n))$. In the same spirit, any general event $\mathcal{A}_n$ is said to occur w.h.p. if $\lim_{n\to\infty} \mathbb{P}(\mathcal{A}_n) = 1$. For notational simplicity, the bandwidth is assumed to be 1 Hz, the noise power spectral density to be 1 watts/Hz, and the information is measures in nats (1 nat = $\log_2(e)$ bits).

III. Information-Theoretic Upper Bound

In this section, we derive an information-theoretic upper bound on the network throughput $T(n)$. Note that for a bound to be information-theoretic, it has to solely depend upon the physical constraints of the system without making any assumptions about the transmission strategy. Before going into the technical details of the upper bound, we state the following useful result, which follows directly from the Chernoff bound for Poisson distribution. For the relevant Chernoff bound, please refer to [3, Appendix II] or [15, Theorem 5.4].

Lemma 1. *For a homogeneous PPP $\Phi \subset \mathbb{R}^d$ with density λ, let $N(\mathcal{A})$ be the number of points in any measurable set $\mathcal{A} \subset \mathbb{R}^d$. Denote the Lebesgue measure of $\mathcal{A}$ by $|\mathcal{A}|$. We have*

$$\lim_{|A|\to\infty} \mathbb{P}(N(\mathcal{A}) > 2\lambda|\mathcal{A}|) = 0. \tag{2}$$

The main idea behind the derivation of this information-theoretic upper bound is to use the information cut-set bound. We first partition the box B_n into two equal boxes, each with side lengths $\sqrt{n} \times \sqrt{n}/2$, as shown in Fig. 1. We will study the information flow across the common edge of the two boxes, i.e., this edge acts as a cut. By Lemma 1, it is easy to deduce that w.h.p. there are less than n points in each box. Since there are $O(n)$ source-destination pairs that need to transmit across this cut, the upper bound on the information flow across this cut also gives an upper bound (in order) for $T(n)$, from which the upper bound on source-destination rate $R(n)$ will directly follow. Note further that the information flow across the cut is upper bounded by the capacity of the effective multiple-input multiple-output (MIMO) channel, say C_n, with the BSs to the left of the cut operating as an effective transmitter and the BSs to the right as a receiver. Since, we are interested in the upper bound, we assume there are exactly n BSs on either side of the cut. The $n\Psi(n) \times n\Psi(n)$ effective channel matrix, denoted by $\mathbf{H}$, is given by

$$\mathbf{H} = \begin{bmatrix} \mathbf{H}_{11} & \mathbf{H}_{12} & \dots & \mathbf{H}_{1n} \\ \mathbf{H}_{21} & \mathbf{H}_{22} & \dots & \mathbf{H}_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{H}_{n1} & \mathbf{H}_{n2} & \dots & \mathbf{H}_{nn} \end{bmatrix}, \tag{3}$$

where $\mathbf{H}_{ik}$ is a $\Psi(n) \times \Psi(n)$ channel matrix from k^{th} BS from the left of the cut to the i^{th} BS to the right of the cut. Before deriving the actual bound, it is useful to understand a key property of this matrix, which is especially relevant for the line-of-sight (LoS) propagation. Recall that assuming array size and the transmission wavelength to be constants, the channel rank, more precisely degrees of freedom (DoF), decrease with the increasing transmitter-receiver separation. Beyond a certain separation, the channel rank collapses to 1. For a more formal statement of this result, interested readers should refer to [6], [7], [16]. As an aside, if we assume each $\mathbf{H}_{ik}$, $1 \leq i, k \leq n$ to be a single rank matrix, following upper bound on $\text{rank}(\mathbf{H})$ can be derived.

Lemma 2 (Rank of a block matrix with rank 1 blocks)**.** *Assuming each $\mathbf{H}_{ik}$, $1 \leq i, k \leq n$ to be a single rank matrix, the rank of $\mathbf{H}$ given by* (3) *is upper bounded by*

$$\text{rank}(\mathbf{H}) \leq \min\{n\Psi(n), n^2\}. \tag{4}$$

Denoting the basis vector of the column (equivalently row) space of $\mathbf{H}_{ik}$ by $\mathbf{v}_{ik}$, (4) *holds with equality if $\exists\{\mathbf{v}_{ik}\}$ such that $\dim\left(\text{span}(\{\mathbf{v}_{ik}\}_{k=1}^{n})\right) = \min\{n, \Psi(n)\}$, for all i.*

Proof. Since $\mathbf{H}_{ik}$ is a rank 1 matrix, it can be expressed as

$$\mathbf{H}_{ik} = \begin{bmatrix} \mu_1\mathbf{v}_{ik} & \mu_2\mathbf{v}_{ik} & \dots & \mu_m\mathbf{v}_{ik} \end{bmatrix}, \tag{5}$$

where $\mathbf{v}_{ik}$ is a $\Psi(n) \times 1$ basis vector for the column space of $\mathbf{H}_{ik}$, and $\{\mu_i\}$ is a set of scalars. It is easy to see that each column of $\mathbf{H}$ can be represented as a linear combination of the following vectors

$$\mathbf{u}_j = \begin{bmatrix} \mathbf{0} \\ \vdots \\ \mathbf{v}_{ik} \\ \vdots \\ \mathbf{0} \end{bmatrix} \quad 1 \leq j \leq n^2, \tag{6}$$

where $k = \lceil \frac{j}{n} \rceil$, $i = j - (k-1)n$, and $\mathbf{0}$ is a $\Psi(n) \times 1$ vector of all 0s. The non-zero entries appear only in the i^{th} block of $\mathbf{u}_j$. Thus, $\text{rank}(\mathbf{H}) \leq n \times n$. Also, the rank cannot be larger than the minimum dimension of the matrix, which implies $\text{rank}(\mathbf{H}) \leq n\Psi(n)$, from which the inequality follows.

The condition for the equality of (4) follows from the fact that if $\dim(\mathrm{span}(\{\mathbf{v}_{ik}\}_{k=1}^{n})) = \min\{\Psi(n), n\}$, for all i, implies $\dim\left(\mathrm{span}(\{\mathbf{u}_j\}_{j=1}^{n^2})\right) = \min\{\Psi(n)n, n^2\}$. $\square$

Remark 1. *The above result shows that if there is sufficient randomness in the system, especially in terms of the BS and antenna locations, and if* $\Psi(n) \leq n$, $\mathrm{rank}(\mathbf{H}) = n\Psi(n)$, *despite its constituting blocks to be all single rank matrices. First, this result highlights the importance of cooperation in LoS propagation. Second, since the capacity of a channel is directly linked to its rank, more precisely DoF, this result shows that the bounds derived in this paper are not highly dependent on the DoF of each individual link. This explains our choice of general propagation model in* (1).

After this slight detour, we now derive an upper bound on the MIMO channel capacity C_n. Note that the condition $\mathrm{rank}(\mathbf{H}_{ik}) = 1$ above was just an aside for illustrative purposes and is not assumed in this discussion. Denoting the transmit symbol covariance matrix by $\mathbf{Q}$, C_n is

$$\begin{aligned} C_n &= \max_{\substack{\mathbf{Q}\succ 0 \\ \mathrm{Tr}(\mathbf{Q})\leq nP}} \log\det\left(\mathbf{I} + \mathbf{H}\mathbf{Q}\mathbf{H}^*\right) \\ &\overset{(a)}{\leq} \log\det\left(\mathbf{I} + nP\mathbf{H}\mathbf{H}^*\right) \\ &\overset{(b)}{\leq} \sum_{i=1}^{n\Psi(n)} \log\left(1 + nP(\mathbf{H}\mathbf{H}^*)_{ii}\right) \end{aligned} \tag{7}$$

where $\mathbf{H}^*$ denotes the conjugate transpose of $\mathbf{H}$, (a) follows by relaxing the total power constraint, i.e., instead of sharing nP power across n transmitters, we assume each BS can transmit at power nP, and (b) follows from the Hadamard inequality specialized for positive semi-definite matrices [17]. For the ease of argument, we assume that the distances between any pair of transmit and receive antennas between k^{th} BS on the left (transmitter) and the i^{th} BS on the right (receiver) are the same and equal to r_{ik}. The main idea now is to derive tight bounds on $(\mathbf{H}\mathbf{H}^*)_{ii}$ using the geometric properties of the point process that determine the distances of BSs from the cut. The bound is based on the tools developed in [14, Theorem 5.4.4] and [1], where a similar bound is derived for a single antenna network. Note that the "mirroring argument" used in [1], [13] to establish equivalence between singular and eigenvalues of the channel matrix is not directly applicable in our case due to complex-valued channel gains. However, as discussed in this section, this only requires a few technical adjustments in the original proof of [1].

For notational simplicity, we order the BSs on both sides of the cut by their respective distances from the cut. Further, the distance of BS i from the cut is denoted by $\hat{r}_i$. To get tighter bounds for $(\mathbf{H}\mathbf{H}^*)_{ii}$, we will use the *exponential stripping* technique introduced in [1]. As shown in Fig. 1, both the boxes on the either side of the cut are partitioned into $\lfloor \log \frac{\sqrt{n}}{2} \rfloor + 1$ vertical strips S_i. For $1 \leq i \leq \lfloor \log \frac{\sqrt{n}}{2} \rfloor$, the minimum distance of the BSs lying in S_i from the cut is $\frac{\sqrt{n}}{2e^i}$, which will be used to upper bound $(\mathbf{H}\mathbf{H}^*)_{ii}$. For $i = \lfloor \log \frac{\sqrt{n}}{2} \rfloor + 1$, i.e., the vertical strip closest to the cut, we will simply upper bound the path-loss by 1. Denoting the number of BSs in S_i by $X(S_i)$, the following holds w.h.p. $\forall\, i$

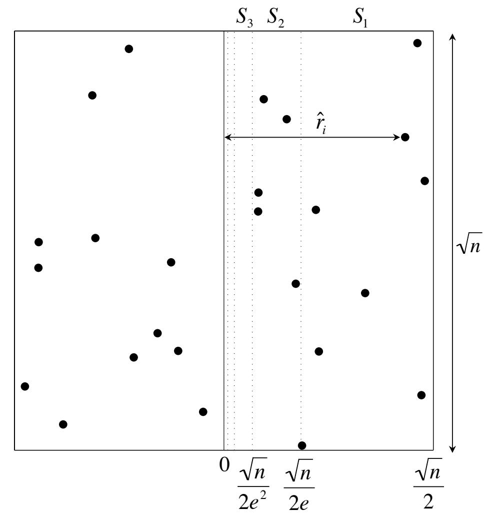

Fig. 1. The setup to derive information-theoretic upper bounds. The vertical strips are denoted by S_i, where $1 \leq i \leq \lfloor \log \frac{\sqrt{n}}{2} \rfloor + 1$.

$$X(S_i) \leq e\frac{n}{e^i}, \tag{8}$$

which directly follows from Lemma 1. From (7), we get

$$C_n \leq \Psi(n) \sum_{i=1}^{\log\frac{\sqrt{n}}{2}+1} X(S_i) \log\left(1 + nP(\mathbf{H}\mathbf{H}^*)_{ii}\right) \tag{9}$$

$$\leq e\Psi(n) \sum_{i=1}^{\log\frac{\sqrt{n}}{2}+1} \frac{n}{e^i} \log\left(1 + nP(\mathbf{H}\mathbf{H}^*)_{ii}\right), \tag{10}$$

where for notational simplicity, we expressed $\lfloor \log \frac{\sqrt{n}}{2} \rfloor$ simply as $\log \frac{\sqrt{n}}{2}$. Note that this does not compromise our results because eventually we are interested in the case $n \to \infty$. Recall that index i in $(\mathbf{H}\mathbf{H}^*)_{ii}$ denotes the vertical strip. This implicitly represents the diagonal values of $\mathbf{H}\mathbf{H}^*$ corresponding to the BSs located in S_i, which will all have the following common upper bound

$$\begin{aligned} (\mathbf{H}\mathbf{H}^*)_{ii} &= \Psi(n)\sum_{k=1}^{n} l(r_{ik}) \overset{(a)}{\leq} \Psi(n)\sum_{k=1}^{n} l(\hat{r}_i) = n\Psi(n) l(\hat{r}_i) \\ &\overset{(b)}{\leq} \begin{cases} n\Psi(n)\hat{r}_i^{-\alpha} & 1 \leq i \leq \log\frac{\sqrt{n}}{2} \\ n\Psi(n) & i = \log\frac{\sqrt{n}}{2} + 1 \end{cases} \\ &\overset{(c)}{\leq} \begin{cases} n^{1-\frac{\alpha}{2}}\Psi(n)2^{\alpha}e^{i\alpha} & 1 \leq i \leq \log\frac{\sqrt{n}}{2} \\ n\Psi(n) & i = \log\frac{\sqrt{n}}{2} + 1 \end{cases}, \end{aligned} \tag{11}$$

where (a) follows from the fact that $\hat{r}_i \leq r_{ik}$ and $l(\cdot)$ is a non-increasing function, (b) from the fact that for $1 \leq i \leq \log\frac{\sqrt{n}}{2}$, $\hat{r}_i$, which is the distance of the BS lying in S_i from the cut, is lower bounded by $\frac{\sqrt{n}}{2e^i} \geq 1$, which implies $l(\hat{r}_i) = \hat{r}_i^{-\alpha}$, and for $i = \log\frac{\sqrt{n}}{2} + 1$, we simply upper bound $l(\hat{r}_i)$ by 1, and (c) follows by lower bounding $\hat{r}_i$. Substituting (11) in (10)

$$C_n \leq e\Psi(n) \sum_{i=1}^{\log\frac{\sqrt{n}}{2}} \frac{n}{e^i} \log\left(1 + Pn^{2-\frac{\alpha}{2}}\Psi(n)2^{\alpha}e^{i\alpha}\right)$$

$$+ 2\Psi(n)\sqrt{n}\log\left(1+Pn^2\Psi(n)\right), \quad (12)$$

where the last term is $O(\Psi(n)\sqrt{n}\log(n^2\Psi(n)))$, which as we will see is not the bottleneck term. The main goal of the rest of this section is to find an upper bound on the summation of the first term. For notational simplicity, define $\kappa = \frac{\alpha}{2} - 2 - \log_n \Psi(n)$ and assume it to be positive, which implies that this derivation is applicable for $\alpha > 2\left(2+\log_n \Psi(n)\right)$. This condition will be required for the Taylor expansion of the log terms. We denote the summation term by $C_{\rm s}$, which is

$$\begin{aligned}
C_{\rm s} &= \sum_{i=1}^{\log\frac{\sqrt{n}}{2}} \frac{n}{e^i}\log\left(1+Pn^{2-\frac{\alpha}{2}}\Psi(n)2^\alpha e^{i\alpha}\right)\\
&= \sum_{i=1}^{\log\frac{\sqrt{n}}{2}} \frac{n}{e^i}\log\left(1+P2^\alpha\frac{e^{i\alpha}}{n^\kappa}\right)\\
&= \sum_{i=1}^{\frac{2\kappa}{\alpha}\log\frac{\sqrt{n}}{2}} \frac{n}{e^i}\log\left(1+P2^\alpha\frac{e^{i\alpha}}{n^\kappa}\right) + \\
&\sum_{i=\frac{2\kappa}{\alpha}\log\frac{\sqrt{n}}{2}+1}^{\log\frac{\sqrt{n}}{2}} \frac{n}{e^i}\log\left(1+P2^\alpha\frac{e^{i\alpha}}{n^\kappa}\right) = C_{\rm s_1}+C_{\rm s_2}. \quad (13)
\end{aligned}$$

Since the constant $P2^\alpha$ in the log terms of both $C_{\rm s_1}$ and $C_{\rm s_2}$ is independent of n and hence does not impact scaling of these terms, we will ignore it in the following discussion for notational simplicity. Using Taylor series expansion for log term, $C_{\rm s_1}$ can now be expressed as

$$\begin{aligned}
C_{\rm s_1} &= \sum_{i=1}^{\frac{2\kappa}{\alpha}\log\frac{\sqrt{n}}{2}} \frac{n}{e^i}\sum_{k=1}^{\infty}\frac{(-1)^{k+1}}{k}\frac{e^{ki\alpha}}{n^{k\kappa}}\\
&= n\sum_{k=1}^{\infty}\frac{(-1)^{k+1}}{k}\frac{1}{n^{k\kappa}}\sum_{i=1}^{\frac{2\kappa}{\alpha}\log\frac{\sqrt{n}}{2}} e^{i(k\alpha-1)}. \quad (14)
\end{aligned}$$

The summation with respect to i can be computed as

$$\begin{aligned}
\sum_{i=1}^{\frac{2\kappa}{\alpha}\log\frac{\sqrt{n}}{2}} e^{i(k\alpha-1)} &= \frac{e^{k\alpha-1}}{e^{k\alpha-1}-1}\left(\frac{n^{(k\alpha-1)\kappa/\alpha}}{2^{2(k\alpha-1)\kappa/\alpha}}-1\right)\\
&\stackrel{(a)}{\leq} m\left(\frac{n^{(k\alpha-1)\kappa/\alpha}}{2^{2(k\alpha-1)\kappa/\alpha}}-1\right), \quad (15)
\end{aligned}$$

where (a) follows by the fact that $\frac{e^{k\alpha-1}}{e^{k\alpha-1}-1}$ is uniformly upper bounded by some positive constant m. Substituting (15) back in (14), we get

$$\begin{aligned}
C_{\rm s_1} &\leq n\sum_{k=1}^{\infty}\frac{(-1)^{k+1}}{k}\frac{1}{n^{k\kappa}}m\left(\left(\frac{n}{4}\right)^{(k\alpha-1)\kappa/\alpha}-1\right)\\
&= C_{\rm s_{11}}+C_{\rm s_{12}}. \quad (16)
\end{aligned}$$

Ignoring again the constants, $C_{\rm s_{11}}$ can be upper bounded as

$$C_{\rm s_{11}} = n^{1-\frac{\kappa}{\alpha}}\sum_{k=1}^{\infty}\frac{(-1)^{k+1}}{k} = n^{1-\frac{\kappa}{\alpha}}\log 2 = O(n^{1-\frac{\kappa}{\alpha}}). \quad (17)$$

Similarly, $C_{\rm s_{12}}$ can be upper bounded as $C_{\rm s_{12}} =$

$$n\sum_{k=1}^{\infty}\frac{(-1)^{k+1}}{k}\frac{1}{n^{k\kappa}} = n\log(1+n^{-\kappa}) = O(n^{1-\kappa}). \quad (18)$$

Substituting (17) and (18) back in (16), we get

$$C_{\rm s_1} \leq O(n^{1-\frac{\kappa}{\alpha}}) + O(n^{1-\kappa}) = O(n^{1-\frac{\kappa}{\alpha}}), \quad (19)$$

where the last equality follows from the fact that $\alpha > 2$. We now turn our attention to $C_{\rm s_2}$ in (13), where we again ignore the constants. It can be expressed as

$$\begin{aligned}
C_{\rm s_2} &= \sum_{i=\frac{2\kappa}{\alpha}\log\frac{\sqrt{n}}{2}+1}^{\log\frac{\sqrt{n}}{2}} \frac{n}{e^i}\log\left(1+\frac{e^{i\alpha}}{n^\kappa}\right)\\
&= \sum\frac{n}{e^i}\log\left(e^{i\alpha}\cdot\frac{1}{n^\kappa}\cdot\left(1+\frac{n^\kappa}{e^{i\alpha}}\right)\right)\\
&= \sum\frac{n\alpha i}{e^i} - \sum\frac{n\kappa}{e^i}\log n + \sum\frac{n}{e^i}\log\left(1+\frac{n^\kappa}{e^{i\alpha}}\right)\\
&= C_{\rm s_{21}} - C_{\rm s_{22}} + C_{\rm s_{23}}, \quad (20)
\end{aligned}$$

where the limits of the summation for all the terms are the same as the first equation. We now look at the three terms separately starting with $C_{\rm s_{21}}$

$$\begin{aligned}
C_{\rm s_{21}} &= n\alpha\sum_{i=\frac{2\kappa}{\alpha}\log\frac{\sqrt{n}}{2}+1}^{\log\frac{\sqrt{n}}{2}}\frac{i}{e^i}\\
&\stackrel{(a)}{\leq} O(\sqrt{n}) + O\left(n^{1-\frac{\kappa}{\alpha}}\right) \stackrel{(b)}{=} O\left(n^{1-\frac{\kappa}{\alpha}}\right), \quad (21)
\end{aligned}$$

where (a) follows by computing the summation directly, and (b) from the fact that $\frac{\kappa}{\alpha} = \frac{1}{2} - \frac{1}{\alpha}\left(2+\log_n\Psi(n)\right) \leq \frac{1}{2}$. Similarly, the second term can be expressed as

$$\begin{aligned}
C_{\rm s_{22}} &= n\kappa\log(n)\sum_{i=\frac{2\kappa}{\alpha}\log\frac{\sqrt{n}}{2}+1}^{\log\frac{\sqrt{n}}{2}} e^{-i} \stackrel{(a)}{\leq} O(\sqrt{n}\log(n)) + \\
&O\left(n^{1-\frac{\kappa}{\alpha}}\log(n)\right) = O\left(n^{1-\frac{\kappa}{\alpha}}\log(n)\right), \quad (22)
\end{aligned}$$

where (a) again follows by computing the summation directly and the final result by the fact that $\kappa < \frac{\alpha}{2}$. Now we come to the final term $C_{\rm s_{23}}$, which can be expressed in terms of the Taylor series, $\log(1+x) = \sum_{k=1}^{\infty}\frac{(-1)^{k+1}}{k}x^k, |x|\leq 1$, as

$$\begin{aligned}
C_{\rm s_{23}} &= n\sum_{i=\frac{2\kappa}{\alpha}\log\frac{\sqrt{n}}{2}+1}^{\log\frac{\sqrt{n}}{2}}\frac{1}{e^i}\log\left(1+\frac{n^\kappa}{e^{i\alpha}}\right)\\
&= n\sum_{k=1}^{\infty}\frac{(-1)^{k+1}}{k}n^{k\kappa}\sum_{i=\frac{2\kappa}{\alpha}\log\frac{\sqrt{n}}{2}+1}^{\log\frac{\sqrt{n}}{2}}\frac{1}{e^{i(k\alpha+1)}}. \quad (23)
\end{aligned}$$

The summation with respect to i can be expressed as

$$\begin{aligned}
&\frac{e^{-k\alpha-1}}{1-e^{-k\alpha-1}}\frac{1}{e^{\frac{k\alpha+1}{\alpha}2\kappa\log\frac{\sqrt{n}}{2}}}\left[1-\frac{e^{k\alpha+1}}{e^{(k\alpha+1)\left(1-\frac{2\kappa}{\alpha}\right)\log\frac{\sqrt{n}}{2}}}\right]\\
&= \frac{e^{-k\alpha-1}}{1-e^{-k\alpha-1}}\left(\frac{4}{n}\right)^{k\kappa+\frac{\kappa}{\alpha}} - \frac{1}{1-e^{-k\alpha-1}}\left(\frac{4}{n}\right)^{\frac{\alpha k+1}{2}}. \quad (24)
\end{aligned}$$

Using the fact that both $\frac{e^{-k\alpha-1}}{1-e^{-k\alpha-1}}$ and $\frac{1}{1-e^{-k\alpha-1}}$ are upper bounded uniformly by positive constants, and ignoring the constants that do not impact scaling, $C_{s_{23}}$ can be upper bounded by the sum of the following two terms

$$C_{s_{231}} = n^{1-\frac{\kappa}{\alpha}} \sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k} = n^{1-\frac{\kappa}{\alpha}} \log 2 = O\left(n^{1-\frac{\kappa}{\alpha}}\right) \tag{25}$$

$$\begin{aligned} C_{s_{232}} &= \sqrt{n} \sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k} n^{(\kappa-\frac{\alpha}{2})k} \overset{(a)}{=} \sqrt{n} \log(1 + n^{\kappa-\frac{\alpha}{2}}) \\ &= O\left(\frac{\sqrt{n}}{n^{\frac{\alpha}{2}-\kappa}}\right) = O(\sqrt{n}), \end{aligned} \tag{26}$$

where (a) follows from the fact that since $\kappa - \frac{\alpha}{2} = -2 - \log_n \Psi(n) < 0$, $n^{k(\kappa-\frac{\alpha}{2})} < 1$, followed by using the appropriate Taylor series expansion. Combining these two results, we get the following upper bound on $C_{s_{23}}$

$$C_{s_{23}} \leq O\left(n^{1-\frac{\kappa}{\alpha}}\right) + O(\sqrt{n}) = O\left(n^{1-\frac{\kappa}{\alpha}}\right). \tag{27}$$

Substituting (21), (22) and (27) in (20), C_{s_2} is

$$\begin{aligned} C_{s_2} = \; & C_{s_{21}} - C_{s_{22}} + C_{s_{23}} \leq O\left(n^{1-\frac{\kappa}{\alpha}}\right) + O\left(n^{1-\frac{\kappa}{\alpha}} \log(n)\right) \\ & + O\left(n^{1-\frac{\kappa}{\alpha}}\right) = O\left(n^{1-\frac{\kappa}{\alpha}} \log(n)\right). \end{aligned} \tag{28}$$

Now combining (19) and (28), C_s can be upper bounded as

$$\begin{aligned} C_s &= C_{s_1} + C_{s_2} = O(n^{1-\frac{\kappa}{\alpha}}) + O\left(n^{1-\frac{\kappa}{\alpha}} \log(n)\right) \\ &= O\left(n^{1-\frac{\kappa}{\alpha}} \log(n)\right), \end{aligned} \tag{29}$$

which when substituted back to (12) completes the proof of the following main result of this paper.

Theorem 1. *For path-loss exponent $\alpha > 2\left(2 + \log_n \Psi(n)\right)$*

$$T(n) = O\left(\sqrt{n} n^{\frac{2}{\alpha}} \Psi(n)^{1+\frac{1}{\alpha}} \log n\right). \tag{30}$$

Remark 2 (Price for generality)**.** *Under a simplified channel model with real channel gains and $\Psi(n) = 1$, the information theoretic bound of [1], [13] is $T(n) = O(n^{\frac{1}{2}+\frac{1}{\alpha}} \log n)$. Putting $\Psi(n) = 1$ in Theorem 1, we get $T(n) = O(n^{\frac{1}{2}+\frac{2}{\alpha}} \log n)$, which although slightly weaker, is derived under more accurate propagation conditions.*

Remark 3 (Scalability)**.** *From Theorem 1, it is easy to check that for linear capacity scaling in n, we need $\Psi(n) = \Omega(\left[n^{\frac{1}{2}-\frac{2}{\alpha}} (\log n)^{-1}\right]^{\frac{\alpha}{1+\alpha}})$, which is meaningful for $\alpha > 4$. In particular, for high attenuation regime, we need to scale antennas almost as $\sqrt{n}$ to make our network scalable. In fact, in [18], $\Psi(n) = \sqrt{n}$ is shown to achieve scalability for any $\alpha > 2$ in LoS MIMO networks under slightly different BS placement model, showing that there exists a regime of BS physical sizes, inter-BS distances, and high frequency, where a scalable wireless backhaul can be effectively implemented with short hops, each achieving high MIMO multiplexing gain.*

IV. Conclusions

Wireless backhaul for current cellular networks is quickly becoming a necessity, especially in the context of urban small cell deployments. Two likely features of future backhaul networks are: (i) higher transmission frequencies, e.g., millimeter wave, and consequently, (ii) large number of transmit antennas. We modeled this network as a multi-antenna random extended network, where the number of antennas per BS can scale as some arbitrary function of the total number of BSs. Using geometric arguments, we derived an information theoretic upper bound on the capacity of this network. As a consequence, we also get a lower bound on the scaling of antennas required to make a backhaul network scalable.

Since the formal study of backhaul networks has just begun, there are numerous extensions possible for this work. For instance, it is important to include legacy BSs in the analysis that may additionally have fixed capacity wired backhaul. It is also important to study the performance of various practical multi-hop strategies in the context of capacity scaling.

References

[1] M. Franceschetti, "A note on Lévêque and Telatar's upper bound on the capacity of wireless *ad hoc* networks," *IEEE Trans. on Info. Theory*, vol. 53, no. 9, pp. 3207 – 3211, Sep. 2007.

[2] P. Gupta and P. R. Kumar, "The capacity of wireless networks," *IEEE Trans. on Info. Theory*, vol. 46, no. 2, pp. 388 – 404, Mar. 2000.

[3] M. Franceschetti, O. Dousse, D. N. C. Tse, and P. Thiran, "Closing the gap in the capacity of wireless networks via percolation theory," *IEEE Trans. on Info. Theory*, vol. 53, no. 3, pp. 1009 – 1018, Mar. 2007.

[4] A. Özgür, O. Lévêque, and D. N. C. Tse, "Hierarchical cooperation achieves optimal capacity scaling in *ad hoc* networks," *IEEE Trans. on Info. Theory*, vol. 53, no. 10, pp. 3549 – 3572, Oct. 2007.

[5] M. Franceschetti, D. Migliore, and P. Minero, "The capacity of wireless networks: Information-theoretic and physical limits," *IEEE Trans. on Info. Theory*, vol. 55, no. 8, pp. 3413 – 3424, Aug. 2009.

[6] S.-H. Lee and S.-Y. Chung, "Capacity scaling of wireless ad hoc networks: Shannon meets Maxwell," *IEEE Trans. on Info. Theory*, vol. 58, no. 3, pp. 1702 – 1715, Mar. 2012.

[7] A. Özgür, O. Lévêque, and D. N. C. Tse, "Spatial degrees of freedom of large distributed MIMO systems and wireless ad hoc networks," *IEEE Journal on Sel. Areas in Commun.*, vol. 31, no. 2, pp. 202 – 214, Feb. 2013.

[8] T. S. Rappaport, S. Sun, R. Mayzus, H. Zhao, Y. Azar, K. Wang, G. N. Wong, J. K. Schulz, M. Samimi, and F. Gutierrez, "Millimeter wave mobile communications for 5G cellular: It will work!" *IEEE Access*, vol. 1, pp. 335 – 349, 2013.

[9] A. Bleicher, "The 5G phone future: Samsung's millimeter-wave transceiver technology could enable ultrafast mobile broadband by 2020," *IEEE Spectrum*, vol. 50, no. 7, pp. 15 – 16, Jul. 2013.

[10] S. Hur, T. Kim, D. J. Love, J. V. Krogmeier, T. A. Thomas, and A. Ghosh, "Millimeter wave beamforming for wireless backhaul and access in small cell networks," *IEEE Trans. on Commun.*, vol. 61, no. 10, pp. 4391 – 4403, Oct. 2013.

[11] W.-Y. Shin, S.-W. Jeon, N. Devroye, M. H. Vu, S.-Y. Chung, Y. H. Lee, and V. Tarokh, "Improved capacity scaling in wireless networks with infrastructure," *IEEE Trans. on Info. Theory*, vol. 57, no. 8, pp. 5088 – 5102, Aug. 2011.

[12] L.-L. Xie and P. R. Kumar, "A network information theory for wireless communication: Scaling laws and optimal operation," *IEEE Trans. on Info. Theory*, vol. 50, no. 5, pp. 748 – 767, May 2004.

[13] O. Lévêque and I. E. Telatar, "Information-theoretic upper bounds on the capacity of large extended ad hoc wireless networks," *IEEE Trans. on Info. Theory*, vol. 51, no. 3, pp. 858 – 865, Mar. 2005.

[14] M. Franceschetti and R. Meester, *Random Networks for Communication: From Statistical Physics to Information Systems*. Cambridge University Press, 2007.

[15] M. Mitzenmacher and E. Upfal, *Probability and Computing: Randomized Algorithms and Probabilistic Analysis*. Cambridge University Press, 2005.

[16] M. Desgroseilliers, O. Lévêque, and E. Preissmann, "Spatial degrees of freedom of MIMO systems in line-of-sight environment," in *Proc., IEEE Intl. Symposium on Information Theory*, Istanbul, Turkey, Jul. 2013.

[17] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge University Press, 1990.

[18] H. S. Dhillon and G. Caire, "Scalability of line-of-sight massive MIMO mesh networks for wireless backhaul," in *Proc., IEEE Intl. Symposium on Information Theory*, Honolulu, HI, Jul. 2014.