

Scalability of Line-of-Sight Massive MIMO Mesh Networks for Wireless Backhaul

Harpreet S. Dhillon and Giuseppe Caire

Abstract—This paper considers an *extended wireless network with multi-antenna nodes in line-of-sight (LoS) propagation environment*. Assuming that the number of antennas at each node can be scaled as some arbitrary function of the number of nodes, we study the *scalability* of this network, i.e., its ability to deliver non-zero rate to each source-destination pair. Since the rank of the LoS multiple-input multiple-output (MIMO) channel starts collapsing with the increasing separation between the transmitter and the receiver, we consider two competing transmission strategies: (i) *long hop*: each source-destination pair minimizes the number of hops by sacrificing multiplexing gain and ideally achieving full power gain over each hop, and (ii) *short hop*: each source-destination pair communicates through a series of short hops each achieving full multiplexing gain. By characterizing the number of antennas required to achieve scalability in both the cases, we show that the antenna requirement is significantly less for the short hop case. These results have key applications in the design of wireless backhaul for cellular networks, where the possibility of having massive MIMO links is becoming a reality due to the increasing maturity of higher transmission frequencies, e.g., 28 and 38 GHz.

I. INTRODUCTION

Wireless backhaul is becoming a key component of modern cellular networks, especially in the context of dense urban deployment of small cells, where it may not always be possible to provide conventional wired backhaul to all the cell sites [1]. Besides, the increasing understanding of propagation at higher transmission frequencies provides an additional impetus by making it possible to pack more antennas in manageable form factors and hence establish high speed backhaul links among base stations (BSs) [2]. As a result, the wireless backhaul networks will likely operate in a slightly different regime compared to the more familiar networks of the past, i.e., there will likely be a gradual migration towards higher transmission frequencies, which will facilitate packing more and more antennas at the BSs. This regime where the transmission frequency and the number of antennas per BS scale with the network size is not very well studied in the literature, perhaps due to lack of motivation in the conventional single-antenna networks, and is the main focus of this work.

The study of asymptotic information rate of single-antenna large wireless networks received a lot of attention in the past decade. This was initiated by Gupta and Kumar's seminal work [3], which along with the refining results of Franceschetti *et al.* [4] established a well-known capacity-scaling result

of $\Theta(\sqrt{n})$ for single-antenna wireless networks, where n is the total number of nodes in the network. Using hierarchical cooperation, Özgür *et al.* [5] demonstrated that it is possible to do better and that a linear capacity scaling with n is achievable. On the contrary, Franceschetti *et al.* [6] argued that this is not possible and that $\Theta(\sqrt{n})$ holds independent of the fading and power attenuation models. This contradiction was settled independently in Lee *et al.* [7] and Özgür *et al.* [8], which showed that while both the results are correct, they are applicable in different operational regimes depending upon the spatial degrees-of-freedom (DoF) available with the MIMO channel. As a part of the analysis, the achievability results for spatial DoF for LoS MIMO channel were also established in [7], [8]. This led to an increased interest in understanding the MIMO channel matrix under LoS propagation model. Two very recent noteworthy contributions in this direction are by Desgroseilliers *et al.* [9], [10]. In [9], a matching upper bound (within logarithmic factors) is derived for the spatial DoF of LoS MIMO channels by approximating the LoS MIMO channel matrix with another random matrix that is easier to manipulate. Under the same approximation, [10] derives the number of significant singular values and an upper bound on the largest singular value of the LoS MIMO channel matrix.

In terms of wireless backhaul design, the focus so far has mostly been on system design under practical considerations, e.g., Hur *et al.* [11] studied millimeter wave beamforming for wireless backhaul and the effect of wind induced pole movement on beam alignment. On the other hand, the fundamental questions about *optimal* strategies and performance limits are not quite resolved yet. For instance, if a BS wants to communicate data to another, possibly distant, BS, it will likely have to transmit over multiple hops. Since in LoS propagation, which is the main focus of this paper, the MIMO channel starts becoming rank deficient with the increasing transmitter-receiver separation, the multiplexing gain achieved over each hop is highly dependent upon the hop length. This naturally leads to the following two competing strategies: (i) *long hop*, where each source-destination pair minimizes the number of hops by sacrificing multiplexing gain to form thin beams and ideally achieving full power gain over each hop, and (ii) *short hop*, where each source-destination pair communicates through a series of short hops, each achieving full multiplexing gain. The main technical contribution of this paper is the characterization of the number of antennas as a function of number of BSs n required by each strategy to make the network scalable, i.e., provide non-zero rate to each source-destination pair. As a consequence of this result, it is shown that the short hop strategy is significantly better in terms

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.

of the antenna requirement as the network size grows.

II. SYSTEM MODEL

We consider a LoS MIMO mesh network, where the locations of the nodes, e.g., BSs in a backhaul network, are modeled by a perturbed square lattice as shown in Fig. 1. The distance between the closest lattice points is assumed to be a constant c . Note that our analysis remains unchanged as long as each small $c \times c$ square in Fig. 1 contains exactly one point. The random perturbation allows for the fact that it may not always be possible to place BSs on an ideal lattice due to practical restrictions, such as the non availability or high leasing cost of certain cell sites. Further details about how the lattice is perturbed are not needed in this paper. Note that the analysis trivially generalizes to the case where each small square contains constant number of nodes.

For the scaling results, we consider the *extended network model*, where we focus our attention on the box B_n with size $c\sqrt{n} \times c\sqrt{n}$ containing n nodes. Fig. 1 is an example of B_n with $n = 64$. Note that although this model has traditionally been used to study *ad hoc* networks, it is also applicable to cellular backhaul because of the possibility of forming *ad hoc* multi-hop backhaul links between source-destination nodes (BSs). We are mainly interested in understanding the asymptotic capacity scaling of the network formed by the nodes inside B_n as $n \rightarrow \infty$. Note that as $n \rightarrow \infty$, the box B_n also grows, hence the name “extended network”. The source-destination pairs in B_n are picked uniformly at random, such that each node is a destination of exactly one source. It is worth noting that there is another popular model to study scaling results, namely *dense network model*, where the node density tends to infinity, while keeping the size of the box constant. Interested readers should refer to [5], [12] for a discussion on the technical differences in the two models. Since the focus of this paper is on wireless backhaul, it makes more sense to consider extended network model, which is representative of a large coverage area with a fixed BS density.

We assume that each node has $\Psi(n)$ antennas, where $\Psi: \mathbb{N}^+ \rightarrow \mathbb{N}^+$ is a monotonically non-decreasing function of n . As discussed in the sequel, we will be particularly interested in the case where $\Psi(n) = \sqrt{n}$. Scaling the number of antennas with the size of network may not seem natural to some readers, but it is worth noting that this is not unrealistic in the current networks due to the increasing maturity of millimeter wave communications, e.g., at 28 and 38 GHz [13]. Recall that as the carrier frequency increases, the physical size of the antennas decreases, hence making it possible to pack more antennas in a manageable form factor. For instance, a working prototype with a matchbook-sized array of 64 antennas has already been demonstrated by Samsung [14], which for $\Psi(n) = \sqrt{n}$, corresponds to the network size of $n = 64^2 = 4096$ in our setup. This is already an “asymptotic” regime for an urban cellular network, e.g., assuming 1 BS per 100×100 m², it already covers about 6.4×6.4 km² area.

We assume that the antennas at each node are uniformly distributed in squares of side \sqrt{a} , which is assumed to be much smaller than the distance between any two communicating

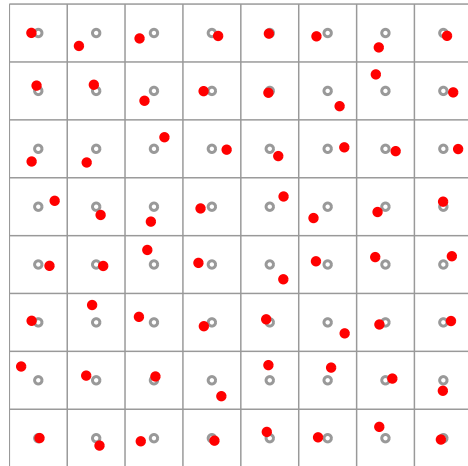


Fig. 1. The locations of the BSs modeled as a perturbed square lattice. The grid and the BS locations are denoted by hollow and filled circles, respectively.

nodes. For $\Psi(n) \rightarrow \infty$, the achievable spatial DoF for a LoS MIMO channel are [7], [8]

$$\begin{cases} \min \left\{ \Psi(n), \frac{\sqrt{a}}{\lambda} \right\}, & \text{when } d \in [1, \sqrt{a}] \\ \min \left\{ \Psi(n), \frac{a}{\lambda d} \right\}, & \text{when } d \in (\sqrt{a}, \frac{a}{\lambda}] \\ 1, & \text{when } d \in (\frac{a}{\lambda}, \infty), \end{cases} \quad (1)$$

where λ is the transmission wavelength and d is the transmitter-receiver separation. A matching upper bound (within logarithmic factors) on the DoF is derived recently in [9] under an element-by-element approximation of the LoS MIMO channel matrix by another random matrix. As discussed in the sequel, one particular case of interest is when a LoS MIMO link can achieve $\text{DoF} = \Psi(n)$ for all $d \leq d_{\max}$, for some constant d_{\max} independent of n . From (1), it is clear that for $d_{\max} \leq \sqrt{a}$ this can be achieved by: (i) fixing λ and scaling $\sqrt{a} \propto \Psi(n)$, or (ii) fixing a and scaling $\lambda \propto \Psi(n)^{-1}$. In a realistic urban backhaul network, d_{\max} would correspond to the separation between closest BSs or small cells and would be of the order of 100s of meters. Therefore, for reasonable form factors of the BSs, $d_{\max} \leq \sqrt{a}$ will never hold in practice. Now for the more realistic case of $d_{\max} > \sqrt{a}$, $\text{DoF} = \Psi(n)$ can be achieved by: (i) fixing λ and scaling $a \propto \Psi(n)$, or (ii) fixing a and scaling $\lambda \propto \Psi(n)^{-1}$. While it is not possible to keep growing the size of BS with $\Psi(n)$, it is indeed possible to assume $\lambda \propto \Psi(n)^{-1}$, which is consistent with the ongoing migration towards higher transmission frequencies [13]. The following remark on the relevance of these assumptions is in order.

Remark 1 (Urban small-cell network). Let us check if the above regime is consistent with real-world deployments. Consider an urban wireless backhaul network with $d_{\max} = 100$ m. Also consider $\Psi(n) = 64$, which we know is possible with the current technology [14]. To achieve full DoF at a carrier frequency of 30GHz, i.e., $\lambda = 1$ cm, we need $a = \Psi(n)\lambda d_{\max} = 64\text{m}^2$, which means a square array of side length 8m. This is clearly a good ball-park number.

The assumption that $\lambda \propto \Psi(n)^{-1}$ requires a careful treatment of link budget and pathloss model, which we do next.

Recall that in the LoS environment the channel gain between two single-antenna nodes with separation $d \gg \lambda$ is [7]

$$h = \sqrt{G} \frac{\lambda}{4\pi d} \exp\left(-j \frac{2\pi}{\lambda} d\right), \quad (2)$$

where G is the product of the antenna gains at the transmitter and the receiver. For $\lambda \propto \Psi(n)^{-1}$ and the proportionality constant of 1 (choice of constant does not matter), we get

$$h = \frac{\sqrt{G}}{4\pi d \Psi(n)} \exp(-j2\pi d \Psi(n)). \quad (3)$$

For a single-antenna transmission, if we assume transmit power and G to be constants independent of n , the received power at a fixed distance d goes down as $\Psi(n)^{-2}$. For a given receiver sensitivity, i.e., minimum received signal-to-noise ratio (SNR) required for successful communication, this would mean that two nodes with a finite distance between them can no longer communicate beyond a certain $\Psi(n)$. The more reasonable assumption is to scale up G or the transmit power $\propto \Psi(n)^2$ such that the received power at a distance d remains constant independent of n . In terms of the link budget, this is equivalent to assuming fixed transmit power and

$$h = d^{-1} \exp(-j2\pi d \Psi(n)). \quad (4)$$

Note that although these scaling assumptions are not mainstream, there have been instances, e.g., [7], where the scaling of transmission frequency and antenna gain with n have been considered, and [15], where the number of antennas at the infrastructure nodes are scaled by the network size in a hybrid wireless network. Besides, it should be clear that these are not any more unreasonable than familiar assumptions such as $n \rightarrow \infty$ or $\text{SNR} \rightarrow \infty$. In this work, we adopt a slightly modified version of the equivalent model (4) along with fixed transmit power P . In particular, the gain between the k^{th} transmit antenna and the i^{th} receive antenna is

$$h_{ik} = \min\{d_{ik}^{-\frac{\alpha}{2}}, 1\} \exp(-j2\pi d_{ik} \Psi(n)), \quad (5)$$

where the pathloss exponent $\alpha > 2$ captures larger attenuation that may result from multiple propagation paths [8], and the $\min\{d_{ik}^{-\frac{\alpha}{2}}, 1\}$ avoids singularity at the origin. Note that the constant $\frac{\sqrt{G}}{4\pi}$ is irrelevant for the scaling results and is hence ignored in (5). For notational simplicity, we assume amplitude term $\min\{d_{ik}^{-\frac{\alpha}{2}}, 1\}$ in (5) to be the same and equal to $\min\{d^{-\frac{\alpha}{2}}, 1\}$ for all the transmit-receive antenna pairs for a given link due to the assumption that the BS form factor is much smaller than the transmitter-receiver separation d . The differences in the path lengths can however not be ignored in the exponential term because they get amplified by $\Psi(n)$. The channel matrix between two nodes is denoted by \mathbf{H} with entries given by (5), and its conjugate transpose by \mathbf{H}^\dagger . It will be enriched, e.g., to include node indices, when required.

Our main focus will be on the throughput achievable simultaneously by each source-destination pair, which we denote by $R(n)$. Without loss of generality, the bandwidth will be assumed to be 1 Hz, which is shared by all the nodes, and the noise power spectral density N_0 to be 1 watt/Hz. We use the following probabilistic version of the ordering

notation as defined in [4]. We write $f(n) = O(g(n))$ with high probability (w.h.p.) if \exists a constant K independent of n such that $\lim_{n \rightarrow \infty} \mathbb{P}(f(n) \leq K g(n)) = 1$. Similarly, $f(n) = \Omega(g(n))$ if $g(n) = O(f(n))$.

III. LONG HOP: BEAMFORMING

In this section, we consider a transmission strategy where each node transmits a single stream of data at a constant rate to a farthest possible node in the direction of its destination in order to minimize the number of hops. This strategy is motivated by the fact that in the massive-MIMO regime, it is possible for a node to form thin beams and hence concentrate transmission energy in the direction of a far-away node without creating excessive interference to its nearby nodes. Our main goal is to find an upper bound on $R(n)$ as a function of $\Psi(n)$ and α . Recall that for a transmitter-receiver separation of $d > 1$, the maximum rate achievable for a single stream is

$$R(d) \stackrel{(a)}{=} \log(1 + P \lambda_{\max}(\mathbf{H}\mathbf{H}^\dagger)) \quad (6)$$

$$\stackrel{(b)}{\leq} \log(1 + P d^{-\alpha} \Psi(n)^2), \quad (7)$$

where $\lambda_{\max}(\mathbf{H}\mathbf{H}^\dagger)$ in (a) is the maximum Eigenvalue of $\mathbf{H}\mathbf{H}^\dagger$ and should not be confused with transmission wavelength λ , and (b) follows from the fact that $\lambda_{\max}(\mathbf{H}\mathbf{H}^\dagger) \leq \text{Tr}(\mathbf{H}\mathbf{H}^\dagger) = \Psi(n)^2$. To minimize the number of hops for a given source-destination pair, our goal is to maximize distance d for each hop keeping the transmission rate (7) constant. Assuming the minimum target received power to be P_0 , we have

$$P d^{-\alpha} \Psi(n)^2 \geq P_0 \Rightarrow d \leq \left(\frac{P}{P_0}\right)^{\frac{1}{\alpha}} \Psi(n)^{\frac{2}{\alpha}} = d_c. \quad (8)$$

Given d_c , we need a lower bound on source-destination separation in order to lower bound the number of hops needed. The lower bound on the distance is derived in the following Lemma. A simple proof is given in Appendix A.

Lemma 1 (Lower bound on source-destination separation). *The source-destination separation of a randomly chosen pair in B_n is $\Omega(n^{\frac{1}{2}-\epsilon})$ w.h.p., where $\epsilon > 0$.*

Using this result, the lower bound on the number of hops required for a randomly chosen source-destination pair N_h is

$$N_h = \Omega\left(\frac{n^{\frac{1}{2}-\epsilon}}{d_c}\right). \quad (9)$$

Since there are $\frac{n}{2}$ source-destination pairs in the network, the total number of hops needed in the network is

$$\frac{n}{2} N_h = \Omega\left(\frac{n^{\frac{3}{2}-\epsilon}}{d_c}\right). \quad (10)$$

Therefore, there is at least one of the total n nodes, which has to relay $\Omega\left(\frac{n^{\frac{1}{2}-\epsilon}}{d_c}\right)$ connections. For the constant rate links, the rate per source-destination pair is upper bound by $R(n) = O\left(d_c n^{-\frac{1}{2}+\epsilon}\right)$, which leads to the following main result.

Theorem 1 (Long hop). *For the long hop strategy discussed in this section, the per source-destination rate is*

$$R(n) = O\left(\Psi(n)^{\frac{2}{\alpha}} n^{-\frac{1}{2} + \epsilon}\right). \quad (11)$$

Remark 2 (Scalability under long-hop strategy). *From Theorem 1, it is clear that in order to achieve $R(n) = O(1)$, we need $\Psi(n) = \Omega(n^{\frac{\alpha}{4} - \epsilon})$, which for a vanishingly small ϵ and $\alpha > 2$ is always higher than \sqrt{n} . In fact, for $\alpha = 4$, we need to scale the number of antennas linearly with n .*

IV. SHORT HOP: SPATIAL MULTIPLEXING

In this section, we consider the other extreme where the data for each source-destination pair is communicated through a series of *short hops*. By short hop, we specifically refer to the communication link between two “neighboring” nodes, each lying in the two consecutive small squares in Fig. 1. In the context of classical single-antenna wireless networks, it is well known that this strategy achieves $R(n) = \Omega\left(n^{-\frac{1}{2}}\right)$. Interested readers should refer to [3], which first studied such scaling results, or to [12] for a more pedagogical treatment of this topic. A natural question to ask now is what happens when these short hops are MIMO links capable of transmitting multiple independent streams by spatial multiplexing. If the number of streams remains constant independent of n , this does not affect the scaling results. However, if they scale up as $\Omega(\sqrt{n})$, it is clearly possible to achieve $R(n) = \Omega(1)$. Therefore, the goal is to find $\Psi(n)$ which enables each short hop MIMO link to achieve rate, say $R_{\text{sh}}(n)$, of $\Omega(\sqrt{n})$.

The main challenge in analyzing $R_{\text{sh}}(n)$ is the presence of LoS interference originating from other simultaneous MIMO transmissions. A closely related problem was recently studied in the literature as a part of the *hierarchical cooperation strategy* for ad hoc networks, where one of the intermediate steps is to derive the rate achievable by a distributed MIMO transmission in the presence of interference from other simultaneous MIMO transmissions [5]. While the procedure to handle this interference under i.i.d. MIMO channels is well understood, it is not the case when the interfering MIMO links are LoS [7], [8]. This problem is rigorously treated in [7], where the scaling of distributed MIMO link rate is derived by explicitly considering LoS MIMO interfering links, see [7, Lemma 2]. In this section, we take an alternate route and study the scaling of $\mathbb{E}[R_{\text{sh}}(n)]$, where $\mathbb{E}[\cdot]$ is over the antenna locations. We show that $\Psi(n) = \sqrt{n}$ is sufficient to achieve $\mathbb{E}[R_{\text{sh}}(n)] = \Omega(\sqrt{n})$. Our analysis is considerably simpler and involves a direct bound on the interference power, reducing the problem to finding the spatial DoFs of a single LoS MIMO link in isolation, which is given by (1). We first prove the following intermediate result. The proof is given in Appendix B.

Lemma 2. *For any continuous random variable $D \geq 0$*

$$\lim_{\Psi(n) \rightarrow \infty} \mathbb{E}[\exp(-j2\pi D\Psi(n))] = 0. \quad (12)$$

We now enrich our notation to study $R_{\text{sh}}(n)$. Assume $c\sqrt{5} \leq d_{\text{max}}$ so that each short-hop achieves full spatial DoF in the absence of interference. Denote the LoS MIMO channel of the desired link by \mathbf{H} and from i^{th} interferer to

the desired receiver by $\mathbf{H}^{(i)}$. Similarly, denote the transmit symbol of the desired and i^{th} interfering transmitters by \mathbf{x} and $\mathbf{x}^{(i)}$, respectively. We further assume that each transmitter distributes equal power across antennas, i.e., $\mathbb{E}[\mathbf{x}\mathbf{x}^\dagger] = \mathbb{E}[\mathbf{x}^{(i)}\mathbf{x}^{(i)\dagger}] = \frac{P}{\Psi(n)}\mathbf{I}$. For worst case analysis, we assume all the nodes are transmitting. We could have used typical time-division-multiple-access (TDMA) arguments to control interference temperature, e.g., see [5], but this does not affect our scaling results because of the bounded pathloss model. Denoting the set of interferers by \mathcal{I} , and the noise vector by \mathbf{z} , the received signal at the desired receiver is

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \sum_{i \in \mathcal{I}} \mathbf{H}^{(i)}\mathbf{x}^{(i)} + \mathbf{z}, \quad (13)$$

and the achievable rate can be expressed as

$$R_{\text{sh}}(n) \geq \log \det \left(I + \frac{P}{\Psi(n)} \mathbf{R}^{-1} \mathbf{H} \mathbf{H}^\dagger \right), \quad (14)$$

where \mathbf{R} is the covariance matrix of noise-plus-interference observed at the desired receiver and is given by

$$\mathbf{R} = \mathbf{I} + \frac{P}{\Psi(n)} \sum_{i \in \mathcal{I}} \mathbf{H}^{(i)} \mathbf{H}^{(i)\dagger}. \quad (15)$$

Since the antenna locations of the interferers only affect (14) through \mathbf{R} , by Jensen’s inequality we have

$$\mathbb{E}[R_{\text{sh}}(n)] \geq \log \det \left(I + \frac{P}{\Psi(n)} \mathbb{E}[\mathbf{R}]^{-1} \mathbf{H} \mathbf{H}^\dagger \right), \quad (16)$$

where the expectation is with respect to the antenna locations. Now the goal is to upper bound $\mathbb{E}[\mathbf{R}]$, which is

$$\mathbb{E}[\mathbf{R}] = \mathbf{I} + \frac{P}{\Psi(n)} \sum_{i \in \mathcal{I}} \mathbb{E} \left[\mathbf{H}^{(i)} \mathbf{H}^{(i)\dagger} \right], \quad (17)$$

where the $(k, m)^{\text{th}}$ element of $\mathbf{H}^{(i)} \mathbf{H}^{(i)\dagger}$ is

$$\sum_{l=1}^{\Psi(n)} \left[\min \left\{ 1, \left(d^{(i)} \right)^{-\frac{\alpha}{2}} \right\} \right]^2 e^{-j2\pi d_{kl}^{(i)} \Psi(n)} e^{j2\pi d_{ml}^{(i)} \Psi(n)}. \quad (18)$$

Taking expectation in (18), using Lemma 2 for large $\Psi(n)$, and substituting the result in (17), we get

$$\mathbb{E}[\mathbf{R}] = \left(1 + \sum_{i \in \mathcal{I}} P \left[\min \left\{ 1, \left(d^{(i)} \right)^{-\frac{\alpha}{2}} \right\} \right]^2 \right) \mathbf{I}. \quad (19)$$

A simple counting argument gives the following upper bound

$$\mathbb{E}[\mathbf{R}] \leq \left(1 + 8P + 8c^{-\alpha} P \sum_{i=2}^{\infty} i(i-1)^{-\alpha} \right) \mathbf{I}, \quad (20)$$

where the summation $\sum_{i=2}^{\infty} i(i-1)^{-\alpha} \leq \sum_{i=2}^{\infty} i^{1-\alpha}$ is clearly convergent for $\alpha > 2$. Therefore $\mathbb{E}[\mathbf{R}] \leq q\mathbf{I}$, where q is a constant independent of n . Substituting it back in (16), we get

$$\mathbb{E}[R_{\text{sh}}(n)] \geq \log \det \left(I + \frac{P}{q\Psi(n)} \mathbf{H} \mathbf{H}^\dagger \right) \stackrel{(a)}{=} \Omega(\Psi(n)), \quad (21)$$

where (a) follows by (1). This leads to our second main result.

Theorem 2 (Short hop). *For the short hop strategy discussed in this section, $\Psi(n) = \sqrt{n}$ achieves $\mathbb{E}[R(n)] = \Omega(1)$, where the expectation is over antenna locations.*

Remark 3 (Scalability and achievability). *First, comparing Theorems 1 and 2, we note that short hop strategy is significantly better for network scalability. While for any $\alpha > 2$ it requires only $\Psi(n) = \sqrt{n}$, the antenna requirement in long hop is always higher and keeps increasing further with α . Second, strictly speaking, Theorem 2 is not enough to claim achievability for any given realization of antenna locations. In addition to the achievability proof in [7, Lemma 2], another worthwhile direction for future investigation consistent with Theorem 2 is to use recent results of [10] on the spectral radius of a LoS MIMO channel to directly bound \mathbf{R} from above instead of invoking Jensen's inequality to bound $\mathbb{E}[\mathbf{R}]$.*

V. CONCLUSION

In this paper we studied the scalability of a wireless backhaul network by modeling it as an extended network with multi-antenna nodes in the LoS environment. Accounting for the fundamental limits on the DoF of a LoS MIMO channel, we compared two competing strategies in terms of the antennas per node required to make this network scalable. While the first minimizes the number of hops by forming thin beams and ideally achieving full power gain over each hop, the other communicates data from source to its destination through many short hops, each achieving full multiplexing gain. Although it may seem intuitive at first to minimize the number of hops by forming narrow beams and hence minimizing interference to neighboring nodes, we show that the short hop strategy is significantly better.

This work has numerous extensions. For instance, with these new found insights it is now important to establish information-theoretic performance bounds for this new regime. In the context of a practical deployment, it is important to study hybrid networks, where some nodes, representing legacy BSs, additionally have fixed capacity wired backhaul.

APPENDIX

A. Proof of Lemma 1

Recall the system setup depicted in Fig. 1, where the node locations are modeled by a perturbed lattice. Let $(X_i, Y_i) \in \mathbb{R}^2$ denote the location of the node in the i^{th} small square, where the ordering of squares is arbitrary. Let $X_i = X_{L_i} + X_{P_i}$, where X_{L_i} is the x coordinate of the lattice point and $-\frac{c}{2} \leq X_{P_i} \leq \frac{c}{2}$ denotes the perturbation along x -axis. Let the source-destination separation of a randomly chosen pair be D_{B_n} . Assuming the transmitter and receiver are respectively located in the i^{th} and k^{th} small squares, the cumulative distribution function (CDF) of D_{B_n} is $\mathbb{P}(D_{B_n} \leq d) =$

$$\begin{aligned} & \mathbb{P}((X_i - X_k)^2 + (Y_i - Y_k)^2 \leq d^2) \\ & \leq \mathbb{P}((X_i - X_k)^2 \leq d^2) = \mathbb{P}(-d + X_k \leq X_i \leq d + X_k) \\ & = \mathbb{P}(-d + X_k \leq X_{L_i} + X_{P_i} \leq d + X_k) \\ & \stackrel{(a)}{\leq} \mathbb{P}\left(-d + X_k - \frac{c}{2} \leq X_{L_i} \leq d + X_k + \frac{c}{2}\right) \stackrel{(b)}{\leq} \frac{2d + c}{c\sqrt{n}}, \end{aligned}$$

where (a) follows from $-\frac{c}{2} \leq X_{P_i} \leq \frac{c}{2}$, and the inequality in (b) is because we ignored the restrictions on the range of X_{L_i} . Clearly, for $\epsilon > 0$, $\lim_{n \rightarrow \infty} \mathbb{P}(D_{B_n} \leq n^{\frac{1}{2}-\epsilon}) = 0$, from which the result follows. \square

B. Proof of Lemma 2

First note that the exponential term can be equivalently expressed as $\exp(-j2\pi D\Psi(n)) = \exp(-j2\pi X)$, where $X = D\Psi(n) \bmod 1$, which clearly lies in $[0, 1]$. It is enough to show that as $\Psi(n)$ grows large, X tends to a uniform distribution between $[0, 1]$. Therefore for $x \in [0, 1]$

$$\mathbb{P}(X \leq x) = \sum_{i=0}^{\infty} \mathbb{P}(i \leq D\Psi(n) \leq i + x) \quad (22)$$

$$= \sum_{i=0}^{\infty} \mathbb{P}\left(\frac{i}{\Psi(n)} \leq D \leq \frac{i+x}{\Psi(n)}\right) \quad (23)$$

$$\stackrel{(a)}{=} x \sum_{i=0}^{\infty} \frac{1}{\Psi(n)} f_D\left(\frac{i}{\Psi(n)}\right) = x, \quad (24)$$

where (a) and its subsequent step hold under $\Psi(n) \rightarrow \infty$ and $f_D(\cdot)$ denotes the probability density function of D . Clearly X is uniformly distributed in $[0, 1]$. \square

REFERENCES

- [1] Small Cell Forum, "Backhaul technologies for small cells", white paper, document 049.02.01, Dec. 2013.
- [2] Z. Pi and F. Khan, "An introduction to millimeter-wave mobile broadband systems," *IEEE Commun. Magazine*, vol. 49, no. 6, pp. 101 – 107, Jun. 2011.
- [3] 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.
- [4] 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.
- [5] 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.
- [6] 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.
- [7] 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.
- [8] 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.
- [9] 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.
- [10] —, "Partially random matrices in line-of-sight wireless networks," in *Proc., IEEE Asilomar*, Pacific Grove, CA, Nov. 2013.
- [11] 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.
- [12] M. Franceschetti and R. Meester, *Random Networks for Communication: From Statistical Physics to Information Systems*. Cambridge University Press, 2007.
- [13] 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.
- [14] 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.
- [15] 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.