Spatial Degrees of Freedom of Large Distributed MIMO Systems and Wireless Ad hoc Networks

Ayfer Özgür
Stanford University
aozgur@stanford.edu

Olivier Lévêque
EPFL, Switzerland
olivier.leveque@epfl.ch

David Tse
University of California at Berkeley
dtse@eecs.berkeley.edu

***Abstract*— We consider a large distributed MIMO system where wireless users with single transmit and receive antenna cooperate in clusters to form distributed transmit and receive antenna arrays. We characterize how the capacity of the distributed MIMO transmission scales with the number of cooperating users, the area of the clusters and the separation between them, in a line-of-sight propagation environment. We use this result to answer the following question: can distributed MIMO provide significant capacity gain over traditional multi-hop in large adhoc networks with n source-destination pairs randomly distributed over an area A? Two diametrically opposite answers [24] and [26] have emerged in the current literature. We show that neither of these two results are universal and their validity depends on the relation between the number of users n and $\sqrt{A}/\lambda$, which we identify as the spatial degrees of freedom in the network. λ is the carrier wavelength. When $\sqrt{A}/\lambda \geq n$, there are n degrees of freedom in the network and distributed MIMO with hierarchical cooperation can achieve a capacity scaling linearly in n as in [24], while capacity of multihop scales only as $\sqrt{n}$. On the other hand, when $\sqrt{A}/\lambda \leq \sqrt{n}$ as in [26], there are only $\sqrt{n}$ degrees of freedom in the network and they can be readily achieved by multihop. Our results also reveal a third regime where $\sqrt{n} \leq \sqrt{A}/\lambda \leq n$. Here, the number of degrees of freedom are smaller than n but larger than what can be achieved by multi-hop. We construct scaling optimal architectures for this intermediate regime.**

I. Introduction

Multiple-input multiple-output (MIMO) is one of the key technologies to achieve high-data rates in current wireless systems. Both the transmitter and the receiver are equipped with multiple antennas, which allows to spatially multiplex several streams of data and transmit them simultaneously. When the scattering environment is rich enough to allow receive antennas to separate out signals from different transmit antennas, MIMO channels offer large capacity gains: the capacity of such a MIMO channel with M transmit and receive antennas is proportional to M. This was established in the ground-breaking papers of Fochini-Gans [1] and Telatar [2] which assumed an i.i.d. fading model for channels between different antenna pairs. However, when there is insufficient scattering in the environment or when antennas are densely packed together in small hand-held devices, the channels between different antenna pairs become correlated. Such physical constraints prevent MIMO capacity from increasing indefinitely with M. The impact of correlated fading [3], [4], [5], [6], [7], the scattering environment [8], [9] and antenna coupling [10], [11] on MIMO capacity was studied by a large body of follow-up research, which also provided a characterization of the inherent spatial degrees of degrees of freedom in a MIMO system as a function of the area and the geometry of the antenna arrays and the angular spread of the physical environment [12], [13], [14], [15].

Recently, distributed MIMO communication arised as a promising tool to achieve large performance gains in wireless networks, similar to those provided by conventional MIMO in the point-to-point wireless channel. Here, wireless users with a single transmit and receive antenna cooperate in clusters to form distributed transmit and receive antenna arrays. A common assumption in the performance analysis of distributed MIMO systems is that channels between different pairs of nodes are subject to independent fading [16], [24]. Analogous to the point-to-point case, this leads to capacity gains linear in the number of nodes M contained in the transmit and receive clusters. In a distributed setting, nodes (or antennas) are typically much farther apart from each other as compared to classical MIMO, so an i.i.d. fading model may seem appropriate. However, the number of nodes M participating to the transmission can be also much larger in this case since there are no physical packing constraints like in MIMO. Capacity can not scale indefinitely with M and correlations between pairwise channels are expected to limit performance when M is large. In this paper, we provide a rigorous lower bound on the scaling of the capacity of a large distributed MIMO system with the area A_c of the transmit and receive clusters, the separation between the clusters d and the number of nodes M contained in each cluster assuming a line-of-sight propagation model and a random distribution of nodes over the cluster areas. We show that the capacity of distributed MIMO systems scales at least as

$$\begin{cases} \min\left\{M, \frac{A_c}{\lambda d}\right\} & \text{when} \quad \sqrt{A_c} \leq d \leq A_c/\lambda \\ \min\left\{M, \sqrt{A_c}/\lambda\right\} & \text{when} \quad 1 \leq d \leq \sqrt{A_c} \end{cases} \quad (1)$$

where λ is the carrier wavelength. This result identifies $\frac{A_c}{\lambda d}$ and $\sqrt{A_c}/\lambda$ as the spatial degrees of freedom in the distributed MIMO channel in the two corresponding regimes. The capacity of the channel scales linearly in M when the physical channel has more than M spatial degrees of freedom. Fortunately, this can be often the case for actual networks. Consider for example two clusters of area 100 m^2 separated by a distance of 100 m. When communication takes place around a carrier frequency of 3 GHz, $\sqrt{A_c}/\lambda = 1'000$. As long as there are less than $1'000$ users in each cluster, the line-of-sight channel has sufficient spatial degrees of freedom for all users. When the distance between the two clusters is 1 km, $\frac{A_c}{\lambda d}$ is still 100. The spatial degrees of freedom are expected to be even larger in scattering environments.

The distributed MIMO channel, and therefore its analysis, differs from the classical MIMO channel in a couple of ways. First, the distances between different pairs of users can be significantly different in the distributed case, which results in heterogeneous channel gains dictated by the geometry of the network. Second, while in classical MIMO, the separation between the antenna arrays d is typically much larger than the length (or the diameter) of the arrays $\sqrt{A_c}$, for distributed systems these two dimensions can be comparable. Our analysis takes into account these new aspects brought by the distributed nature of the problem. Mathematically, our approach differs significantly from existing results studying the spatial degrees of freedom of classical MIMO channels. While such results (for example, see [12], [14], [15]) are based on studying the singular values of the continuous propagation operator under *approximations* for the regime where $d \gg \sqrt{A_c}$, our analysis is *mathematically rigorous* and is based on random matrix analysis.

The characterization of the scaling of the capacity of distributed MIMO systems allows us to identify the number of spatial degrees of freedom in large wireless ad hoc networks, and to reconcile some seemingly contradicting results in the current literature on scaling laws for wireless networks. The study of the asymptotic regime where the number of users in a wireless network is large was initiated by the seminal paper [17] of Gupta and Kumar and received significant attention in the literature [18], [19], [20], [21], [22], [23]. Gupta and Kumar showed that the capacity of multihop cooperation scales as $\sqrt{n}$ with increasing number of user n in the network. In this traditional communication architecture, packets are routed from each source to its destination along a path where intermediate users act as relays. Each relay decodes the packets sent from the previous relay and forwards them to the next. A $\sqrt{n}$ scaling for the total capacity implies that the rate per user decreases as $1/\sqrt{n}$ with increasing system size n. Can more sophisticated cooperation between users significantly increase the capacity of large wireless networks? Two diametrically opposite answers have emerged in the recent literature:

1) Capacity can be significantly improved when users form distributed MIMO arrays via a hierarchical cooperation architecture [24]. In regimes where power is not a limiting factor [25], the capacity can scale almost linearly with n implying a constant rate per user.
2) The scaling of the capacity is upper bounded by $\sqrt{n}$ due to the spatial constraints imposed by the physical channel [26]. Nearest-neighbor multi-hop already achieves this scaling and more sophisticated cooperation is useless.

The key difference between these two results is their assumptions for the channel model between pairwise nodes. [24] assumes that the phases of the channel gains can be modeled as uniformly distributed random variables, independent across different pairs of nodes in the network. [26], on the other hand, starts from physical principles and regards the phases as functions of the locations of the nodes. While the physical channel model used in [26] is more fundamental, the i.i.d. phase model is also widely accepted in wireless communication engineering, particularly for nodes in far field from each other. What is the way to reconcile these two sets of results?

We answer this question in the second part of the paper building on the result in (1). We show that under the physical channel model of [26], the distributed MIMO based hierarchical cooperation architecture in [24] achieves a capacity scaling as

$$\max\left\{\sqrt{n}, \min\left\{n, \frac{\sqrt{A}}{\lambda}\right\}\right\}. \tag{2}$$

in a network of n source-destination pairs uniformly distributed over an area A and communicating around a carrier wavelength λ. The scaling of the capacity depends on how n compares to $\sqrt{A}/\lambda$, which can be interpreted as the spatial degrees of freedom available in the network. The two earlier results can be recovered as two special cases of this new result:

1) When $\sqrt{A}/\lambda \geq n$, the capacity scales linearly in n. In this regime, there are sufficient spatial degrees of freedom for all the n users in the network and they can be exploited by distributed MIMO communication. The i.i.d. fading assumption across different node pairs in [24] leads to n degrees of freedom and therefore [24] inherently assumes that the network operates in this regime.
2) When $\sqrt{A}/\lambda \leq \sqrt{n}$, the capacity scales as $\sqrt{n}$. In this regime, the spatial degrees of freedom available in the network are as few as $\sqrt{n}$, and therefore they can be readily achieved by multihop. By assuming that the density of nodes is fixed as the number of nodes n grows, [26] assumes that the number of spatial degrees of freedom $\sqrt{A}/\lambda$ is proportional to $\sqrt{n}$. Therefore, [26] inherently assumes that the network operates in this regime.

Therefore, neither of the two conclusions in [24] and [26] that more sophisticated cooperation can provide significant capacity gains or is useless are universal. They correspond to two different operating regimes of large wireless networks. (2) clarifies the conditions for a network to be in either of these two regimes. Indeed, (2) also uncovers a third regime where the network is partially limited in spatial degrees of freedom. When $\sqrt{n} \leq \sqrt{A}/\lambda \leq n$, the number of spatial degrees of freedom is smaller than n, so the spatial limitation is felt, but larger than what can be achieved by simple multi-hopping. (Multi-hop achieves $\sqrt{n}$ scaling independent of $\sqrt{A}/\lambda$.) We show that either a modification of the hierarchical cooperation scheme in [24] or a version of the MIMO-multihop scheme in [25] can achieve the $\sqrt{A}/\lambda$ available degrees of freedom and therefore the optimal scaling of the capacity in this regime. The main idea behind the first modification is to allow only a subset N of the source-destination pairs to communicate at a time using the hierarchical cooperation scheme, and then take turns among different subsets. $N = \sqrt{A}/\lambda$, the number of spatial degrees of freedom in the network, so each subset corresponds to a diluted network which is not limited in spatial degrees of freedom. The idea behind the second scheme is to form clusters of an intermediate size and hop across several clusters to reach destinations where each hop is performed via distributed MIMO transmissions. The cluster size is chosen critically to ensure linear scaling for the distributed MIMO transmissions.

Traditionally, the literature on scaling laws for wireless networks seeks the scaling of the capacity with the number of nodes n when all systems parameters are coupled with n in a specific way. One common approach is to assume that the area scales linearly in n, while all other parameters remain fixed, as in [26]. As we have already seen, this immediately implies that the spatial degrees of freedom in the network $\sqrt{A}/\lambda$ are proportional to $\sqrt{n}$. But the number of nodes and the area are two independent parameters of a network, each of which can take on a wide range of values. For actual networks, there can be a huge difference between $\sqrt{A}/\lambda$ and $\sqrt{n}$. Take an example of a network serving $n = 10'000$ users on a campus of 1 km^2, operating at 3 GHz: $\sqrt{A}/\lambda = 10'000$, while $\sqrt{n}$ is only 100, two orders of magnitude smaller. So while multi-hop can achieve a total throughput of the order of 100 bits/s/Hz, there is still a lot of potential for cooperation gain, since the spatial degrees of freedom are indeed $10'000$ and not 100 as given by (2).

For the classical MIMO channel, it is now well understood that there are a number of qualitatively different regimes. When the antennas are separated sufficiently apart, the capacity increases linearly in n. However for dense antenna arrays the capacity is limited by physical constraints and cannot increase linearly in n. To obtain an analogous understanding of the operating regimes of large wireless networks, we advocate in this paper a shift of the "large networks" research agenda from seeking a single "universal" scaling law, where parameters of the network are coupled to n in a specific way, to seeking a *multi-parameter family* of scaling laws, where the key parameters are decoupled and many different limits with respect to these parameters are taken. A single scaling law with a particular coupling between parameters is often arbitrary and too restrictive to cover the wide ranges that the multiple parameters of the network can take on. We have introduced this approach in [25] where we decoupled the number of nodes and the amount of power available. The current paper follows the spirit of [25], both mathematically and philosophically, but focuses on the number of nodes and the area of the network, while assuming there is a sufficient amount of power available that it is not limiting performance. A future goal of this research program is to investigate the dependence of the capacity on the number of nodes, the area of the network and the amount of power all together.

A version of this problem has been studied in an independent concurrent work [30], however both the formulation of the problem and the proposed architectures differ from the current paper: in [30], the network area A is taken either fixed or proportional to n, while the carrier wavelength λ scales down to zero with increasing n. The architecture in [30] is obatained by diluting the distributed MIMO transmissions at each level of the hierarchy in [24], as opposed to diluting the whole network as we do here. In the current paper, we also show that the same performances can be alternatively achieved with a MIMO-multihop strategy.

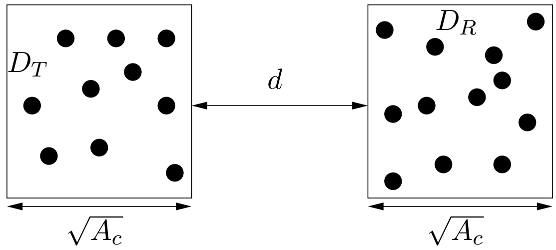

Fig. 1. Two square clusters of area A_c separated by distance d.

II. Spatial Degrees of Freedom of Distributed MIMO Systems

A. Model

We consider a distributed MIMO transmission between two square clusters of area A_c separated by distance d (see Figure 1), with each cluster containing M nodes distributed uniformly at random over the area A_c. Each node is equipped with one antenna, oriented in the direction perpendicular to the plane, with a given power budget P. We assume that communication takes place over a flat channel with bandwidth W and that the signal received by node i at the RX cluster at time-slot m is given by

$$y_i[m] = \sum_{k=1}^{M} h_{ik}\, x_k[m] + z_i[m]$$

where $x_k[m]$ is the signal sent by node k at the TX cluster at time m, $z_i[m]$ is additive white circularly symmetric Gaussian noise (AWGN) of power spectral density $N_0/2$ Watts/Hz. In a line-of-sight environment, the complex baseband-equivalent channel gain h_{ik} between transmit node k and receive node i at time m is given by

$$h_{ik} = \sqrt{G}\, \frac{\exp(2\pi j r_{ik}/\lambda)}{r_{ik}} \tag{3}$$

where λ is the carrier wavelength, r_{ik} is the distance between node i and node k and G is Friis' constant given by

$$G = G_t\, G_r\, \left(\frac{\lambda}{4\pi}\right)^2$$

with G_t and G_r being the transmit and receive antenna gains, respectively. Finally, we assume full channel state information at all the nodes, which is a reasonable assumption here, as the channel coefficients only depend on the node positions and these do not vary over time.

B. Main Result

Our goal in this section is to provide a lower bound on the spatial degrees of freedom of the system described above. It can be inferred from [26] that the spatial degrees of freedom of such a distributed MIMO system are upperbounded (up to logarithmic factors) by $\sqrt{A_c}/\lambda$ as A_c gets large, irrespective of the distance d. Actually, the spatial degrees of freedom decrease as d increases, because of the reduction of the aperture of the MIMO system. We prove below that the spatial degrees of freedom of the system are at least

$$\begin{cases} \min\{M, A_c/\lambda d\} & \text{if } \sqrt{A_c} \leq d \leq A_c/\lambda \\ \min\{M, \sqrt{A_c}/\lambda\} & \text{if } 1 \leq d \leq \sqrt{A_c} \end{cases}$$

(again, up to logarithmic factors). Notice that in the second regime, the obtained lower bound matches the upper bound found in [26]. Finally, notice that if $d \geq A_c/\lambda$, then the system has clearly at least one degree of freedom.

Let us now state our main result.

Theorem 2.1: Let $1 \leq d \leq A_c/\lambda$, and let the nodes in the transmit cluster D_T perform independent signaling with power P each, such that the long-distance SNR between these two clusters defined as

$$\text{SNR}(d) = M\,\frac{GP}{N_0 W d^2} \tag{4}$$

is greater than or equal to $0\,\text{dB}$. Then there exists a constant $K > 0$ independent of M, A_c, λ and d such that the capacity C_{MIMO} of the distributed MIMO channel from the transmit cluster D_T to the receive cluster D_R is lowerbounded by

$$C_M \geq \begin{cases} K\,\min\left\{M, \dfrac{A_c/\lambda d}{\log(A_c/\lambda d)}\right\} & \text{if } \sqrt{A_c} \leq d \leq A/\lambda \\ \\ K\,\min\left\{M, \dfrac{\sqrt{A_c}/\lambda}{\log(\sqrt{A_c}/\lambda)}\right\} & \text{if } 1 \leq d \leq \sqrt{A_c} \end{cases}$$

with high probability as M gets large.

The rest of this section is devoted to the proof of this result, which is made of three ingredients.

The first key ingredient provides a lower bound on the MIMO channel capacity averaged over the random node positions in the first regime where $\sqrt{A_c} \leq d \leq A_c/\lambda$.

Lemma 2.1: If $\sqrt{A_c} \leq d \leq A_c/\lambda$, then the expected capacity $\mathbb{E}(C_{\text{MIMO}})$, averaged over the random node positions, satisfies

$$\mathbb{E}(C_{\text{MIMO}}) \geq K\,\min\left\{M, \frac{A_c/\lambda d}{\log(A_c/\lambda d)}\right\} \tag{5}$$

Next, we show that the capacity of the distributed MIMO channel with given random node positions is close to its expected value with high probability. The proof relies on classical concentration arguments and is relegated to the end of the present section.

Lemma 2.2: In general, if M nodes participate to the MIMO transmission, then for all $\varepsilon > 0$, there exists $K > 0$ such that

$$|C_{\text{MIMO}} - \mathbb{E}(C_{\text{MIMO}})| \leq K\,M^{1/2+\varepsilon}$$

with high probability as M gets large.

Finally, we show that the study of the second regime ($1 \leq d \leq \sqrt{A_c}$) can be brought back to the case $d = \sqrt{A_c}$ by simply reducing the set of transmitting and receiving nodes by a factor 2, so that the intercluster distance becomes of the same order as their radius. Again, the proof is relegated to the end of the present section.

Lemma 2.3: If $1 \leq d \leq \sqrt{A_c}$, then the spatial degrees of freedom of the system are of the same order as when $d = \sqrt{A_c}$.

Combining these three lemmas yields the result given in Theorem 2.1: Assume first $\sqrt{A_c} \leq d \leq A_c/\lambda$. Then by Lemma 2.1,

$$\mathbb{E}(C_{\text{MIMO}}) \geq K\,\min\left\{M, \frac{A_c/\lambda d}{\log(A_c/\lambda d)}\right\}$$

If $M \leq \frac{A_c/\lambda d}{\log(A_c/\lambda d)}$, then Lemma 2.2 allows to conclude that C_{MIMO} itself is with high probability at least of order M. If on the other hand $M > \frac{A_c/\lambda d}{\log(A_c/\lambda d)}$, then it should be noticed that in this case, it is useless to have all the M nodes participating to the MIMO transmission. Only $M' = \frac{A_c/\lambda d}{\log(A_c/\lambda d)}$ suffice. Applying then the concentration result replacing M by M' allows to conclude. Finally, Lemma 2.3 shows that for all $1 \leq d \leq \sqrt{A_c}$, the degrees of freedom of the system are lowerbounded by

$$C_{\text{MIMO}} \geq K'\,\min\left\{M, \frac{\sqrt{A_c}/\lambda}{\log(\sqrt{A_c}/\lambda)}\right\}$$

with high probability for some other constant $K' > 0$. This concludes the proof. □

In the sequel, we provide the proof of Lemma 2.1. The proofs of Lemmas 2.2 and 2.3 are relegated to the end of the present section.

Proof of Lemma 2.1. For notational convenience, we start by defining

$$\begin{aligned} f_{ik} &= \frac{d}{\sqrt{G}}\,h_{ik} = \frac{d}{r_{ik}}\,\exp(2\pi j r_{ik}/\lambda) \\ &= \frac{d}{\|\boldsymbol{x}_k - \boldsymbol{w}_i\|}\,\exp(2\pi j\,\|\boldsymbol{x}_k - \boldsymbol{w}_i\|/\lambda) \end{aligned} \tag{6}$$

where r_{ik} denotes the distance between the nodes $k \in \mathcal{D}_T$ and $i \in \mathcal{D}_R$ located at positions $\boldsymbol{x}_k$ and $\boldsymbol{w}_i$, respectively. Notice that $d \leq r_{ik} \leq d\,(1 + 2\sqrt{2A_c}/d)$, so

$$c_0 \leq (1 + 2\sqrt{2A_c}/d)^{-1} \leq |f_{ik}| \leq 1, \tag{7}$$

where $c_0 = (1+2\sqrt{2})^{-1}$ and the first inequality follows from the fact that $\sqrt{A_c} \leq d$.

Remembering the definition of $\text{SNR}(d)$ given (4), we obtain the following expression for the average capacity of the distributed MIMO channel (where we recall here that H, F are the matrices with entries h_{ik}, f_{ik}, respectively):

$$\begin{aligned} \mathbb{E}(C_{\text{MIMO}}) &= \mathbb{E}\left(\log\det\left(I + \frac{P}{N_0 W}\,HH^\dagger\right)\right) \\ &= \mathbb{E}\left(\log\det\left(I + \text{SNR}(d)\,FF^\dagger/M\right)\right) \\ &= M\,\mathbb{E}\left(\log\left(1 + \text{SNR}(d)\,\lambda\right)\right) \end{aligned}$$

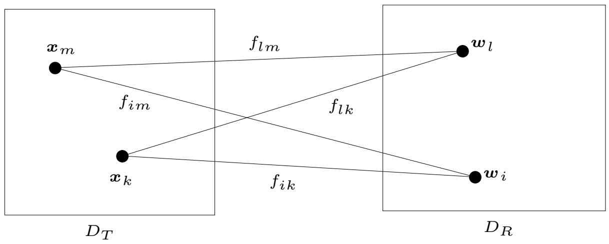

Fig. 2. $S = |\mathbb{E}(f_{ik}\, f_{lk}^*\, f_{lm}\, f_{im}^*)|$, $i \neq l$, $k \neq m$

where λ is an eigenvalue of $FF^\dagger/M$ picked uniformly at random. By applying now the Paley-Zygmund inequality, stating that for a non-negative random variable X,

$$\mathbb{P}(X \geq t) \geq \frac{(\mathbb{E}(X) - t)^2}{\mathbb{E}(X^2)}, \quad \forall 0 < t < \mathbb{E}(X)$$

we obtain that for $0 < t < \mathbb{E}(\lambda)$:

$$\begin{aligned}\mathbb{E}(C_{\text{MIMO}}) &\geq M\, \log\left(1 + \text{SNR}(d)\, t\right)\, \mathbb{P}(\lambda > t) \\ &\geq M\, \log\left(1 + \text{SNR}(d)\, t\right)\, \frac{(\mathbb{E}(\lambda) - t)^2}{\mathbb{E}(\lambda^2)}\end{aligned}$$

From (7), we further obtain

$$\mathbb{E}(\lambda) = \frac{1}{M^2}\, \mathbb{E}\left(\text{tr}(FF^\dagger)\right) = \frac{1}{M^2} \sum_{i,k=1}^{M} \mathbb{E}(|f_{ik}|^2) \geq c_0^2.$$

$$\begin{aligned}\mathbb{E}(\lambda^2) &= \frac{1}{M^3}\, \mathbb{E}(\text{tr}(FF^\dagger FF^\dagger)) \\ &= \frac{1}{M^3} \sum_{i,k,l,m=1}^{M} \mathbb{E}(f_{ik} f_{lk}^* f_{lm} f_{im}^*) \\ &\leq 2 + \frac{1}{M^3} \sum_{\substack{i,k,l,m=1 \\ i\neq l, k\neq m}}^{M} \mathbb{E}(f_{ik} f_{lk}^* f_{lm} f_{im}^*) \leq 2 + M\, S\end{aligned}$$

where $S = |\mathbb{E}(f_{ik}\, f_{lk}^*\, f_{lm}\, f_{im}^*)|$, with $i \neq l$ and $k \neq m$ (notice that S does not depend on the specific choice of $i \neq l$ and $k \neq m$). See Figure 2.

Choosing then $t = c_0^2/2$, we obtain

$$\begin{aligned}\mathbb{E}(C_{\text{MIMO}}) &\geq (M\, c_0^4/4)\, \log\left(1 + \text{SNR}(d)\, c_0^2/2\right)\, \frac{1}{2 + M\, S} \\ &\geq K\, \min\left\{M, \frac{1}{S}\right\}\end{aligned}$$

for a constant $K > 0$ independent of M and S, provided that $\text{SNR}(d) \geq 0$ dB, which was our initial assumption (4).

The quantity S, which takes values between 0 and 1, dictates therefore the capacity scaling. In the case where the channel matrix entries f_{ik} are i.i.d. and circularly-symmetric, $S = 0$, so the capacity is of order M. On the other hand, if we consider the line-of-sight channel model (6) in the scenario where nodes are placed on a single straight line, then a simple computation shows that S is of order 1, so that the capacity is also of order 1 in this case[1].

The problem we are looking at lies between these two extreme cases. Our aim in the following is to show the following lemma, which allows to conclude the proof of Lemma 2.1.

Lemma 2.4: If $\sqrt{A_c} \leq d \leq A_c/\lambda$, then there exists a constant $K > 0$ independent of A_c, λ and d, such that

$$S \leq K \left(\frac{\lambda d}{A_c}\right) \log\left(\frac{A_c}{\lambda d}\right). \tag{8}$$

We now give a proof idea for Lemma 2.4. Let us first explicitly write the expression for S. We have

$$\begin{aligned}S &= |\mathbb{E}(f_{ik}\, f_{lk}^*\, f_{lm}\, f_{im}^*)| \\ &= \left| \frac{1}{A_c^4} \int_{D_T} d\boldsymbol{x}_k \int_{D_T} d\boldsymbol{x}_m \int_{D_R} d\boldsymbol{w}_i \int_{D_R} d\boldsymbol{w}_l\, \rho\, e^{2\pi j \Delta/\lambda} \right|\end{aligned} \tag{9}$$

where

$$\Delta = \|\boldsymbol{x}_k - \boldsymbol{w}_i\| - \|\boldsymbol{x}_k - \boldsymbol{w}_l\| + \|\boldsymbol{x}_m - \boldsymbol{w}_l\| - \|\boldsymbol{x}_m - \boldsymbol{w}_i\| \tag{10}$$

and

$$\rho = \frac{d^4}{\|\boldsymbol{x}_k - \boldsymbol{w}_i\| \|\boldsymbol{x}_k - \boldsymbol{w}_l\| \|\boldsymbol{x}_m - \boldsymbol{w}_l\| \|\boldsymbol{x}_m - \boldsymbol{w}_i\|} \tag{11}$$

We first derive the result (8) by approximating the inter-node distances in the regime $\sqrt{A_c} \ll d \ll A_c$. This approximation, made already by various authors [12], [9], [14], [15] in different contexts, allows us to derive an upper bound on S, and correspondingly, a lower bound on the spatial degrees of freedom. In addition, we provide in the Appendix a rigorous derivation of the lower bound, which does not make use of the approximation. As far as we know, this derivation is new.

Consider two nodes at positions $\boldsymbol{x} = (-\sqrt{A_c}\, x, \sqrt{A_c} y) \in D_T$ and $\boldsymbol{w} = (d + \sqrt{A_c} w, \sqrt{A_c} z) \in D_R$, where $x, y, w, z \in [0,1]$ (see Figure 3). Using the assumption that $d \gg \sqrt{A_c}$, we obtain

$$\begin{aligned}\|\boldsymbol{x} - \boldsymbol{w}\| &= \sqrt{(d + \sqrt{A_c}\,(x+w))^2 + A_c\,(y-z)^2} \\ &\approx d + \sqrt{A_c}\,(x+w) + \frac{A_c}{2d}\,(y-z)^2\end{aligned}$$

which in turn implies

$$\begin{aligned}&\Delta = \|\boldsymbol{x}_k - \boldsymbol{w}_i\| - \|\boldsymbol{x}_k - \boldsymbol{w}_l\| + \|\boldsymbol{x}_m - \boldsymbol{w}_l\| - \|\boldsymbol{x}_m - \boldsymbol{w}_i\| \\ &\approx \frac{A_c}{2d}((y_k - z_i)^2 - (y_k - z_l)^2 + (y_m - z_l)^2 - (y_m - z_i)^2) \\ &= -\frac{A_c}{d}(y_m - y_k)(z_l - z_i)\end{aligned}$$

Next, let us also make the approximation that $\rho \approx 1$ in (11): this is actually assuming that the spatial degrees of freedom between the two clusters are mainly determined by the phases of the channel coefficients and not so much by the amplitudes. In the Appendix, we show that this intuition is correct.

[1]Strictly speaking, this only shows that the lower bound on the capacity is of order 1, but in this case, the matrix F can be shown to be essentially rank one, so the actual capacity is indeed of order 1.

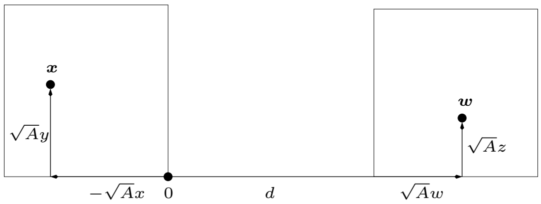

Fig. 3. Coordinate system.

These two successive approximations lead to the following approximation for S:

$$S \approx S_0$$
$$= \left| \int_0^1 dy_k \int_0^1 dy_m \int_0^1 dz_i \int_0^1 dz_l \, e^{-2\pi j \frac{A_c}{\lambda d}(y_m - y_k)(z_l - z_i)} \right|$$
$$= 2 \left| \int_0^1 dy_k \int_{y_k}^1 dy_m \int_0^1 dz_i \int_0^1 dz_l \, e^{-2\pi j \frac{A_c}{\lambda d}(y_m - y_k)(z_l - z_i)} \right|$$

where the second equation follows from the symmetry of the integrand. Note that this expression does not depend on the horizontal positions of the nodes. This can be interpreted as follows. Provided the above approximation is valid, the spatial degrees of freedom between two clusters of M nodes separated by a distance $d \gg \sqrt{A_c}$ are the same, be the nodes uniformly distributed on two squares of area A_c or on two parallel (vertical) lines[2] of length $\sqrt{A_c}$.

We show below that the above integral is of order $\lambda d / A_c$. Let us compute the first integral, which yields

$$\int_0^1 dz_l \, e^{-2\pi j \frac{A_c}{\lambda d}(y_m - y_k)(z_l - z_i)}$$
$$= -\frac{\lambda d}{2\pi j A_c (y_m - y_k)} \, e^{-2\pi j \frac{A_c}{\lambda d}(y_m - y_k)(z_l - z_i)} \Big|_{z_l = 0}^{z_l = 1}$$

This implies that

$$\left| \int_0^1 dz_l \, e^{-2\pi j \frac{A_c}{\lambda d}\,(y_m - y_k)\,(z_l - z_i)} \right| \leq K \, \frac{\lambda d}{A_c} \, \frac{1}{|y_m - y_k|}$$

for a constant $K > 0$ independent of A_c, λ and d. We can divide the integration over y_k and y_m into two parts, so

$$\int_0^1 dy_k \int_{y_k}^1 dy_m \int_0^1 dz_i \int_0^1 dz_l \, e^{-2\pi j \frac{A_c}{\lambda d}(y_m - y_k)(z_l - z_i)}$$
$$= \left(\int_0^1 dy_k \int_{y_k}^{(y_k + \varepsilon) \vee 1} dy_m + \int_0^{1 - \varepsilon} dy_k \int_{y_k + \varepsilon}^1 dy_m \right)$$
$$\times \int_0^1 dz_i \int_0^1 dz_l \, e^{-2\pi j \frac{A_c}{\lambda d}(y_m - y_k)(z_l - z_i)}$$

for any $0 < \varepsilon < 1$. The first term can be simply bounded by ε, which yields the following upper bound on S_0

$$S_0 \leq 2\varepsilon + 2K \, \frac{\lambda d}{A_c} \int_0^{1 - \varepsilon} dy_k \int_{y_k + \varepsilon}^1 dy_m \, \frac{1}{|y_m - y_k|}$$
$$\leq 2\varepsilon + 2K \, \frac{\lambda d}{A_c} \, \log \left(\frac{1}{\varepsilon} \right)$$

So choosing $\varepsilon = \lambda d / A_c$, we finally obtain

$$S \approx S_0 \leq K \left(\frac{\lambda d}{A_c} \right) \log \left(\frac{A_c}{\lambda d} \right)$$

for a constant $K > 0$ independent of A_c, λ and d. □

Proof of Lemma 2.2. First observe that the capacity of the distributed MIMO channel

$$C_{\text{MIMO}} = \log \det(I + \text{SNR}(d) \, F F^\dagger / M)$$

(where we reuse here the notation adopted in the proof of Lemma 2.1) can be seen as a function of the node positions $\boldsymbol{x}_k$ and $\boldsymbol{w}_i$:

$$C_{\text{MIMO}} = g_M(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_M; \boldsymbol{w}_1, \ldots, \boldsymbol{w}_M)$$

As we will see in the following, the capacity does not vary significantly with the node positions, which allows us to apply the following (simplified version of the) theorem by McDiarmid [28].

Theorem 2.2: Let $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_M$ be a family of i.i.d. random variables distributed in a bounded region $A \subset \mathbb{R}^2$, and let $f_M : \mathbb{R}^{2M} \to \mathbb{R}$ be a measurable function such that there is a constant c_M with

$$|f_M(\boldsymbol{x}_1, \ldots, \boldsymbol{x}'_k, \ldots, \boldsymbol{x}_M) - f_M(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k, \ldots \boldsymbol{x}_M)| \leq c_M$$

for all $1 \leq k \leq M$ and $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k, \boldsymbol{x}'_k, \ldots, \boldsymbol{x}_M \in A$. Then for all $t > 0$,

$$\mathbb{P}\left(|f_M(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_M) - \mathbb{E}(f_M(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_M))| > t\right)$$
$$\leq 2 \exp \left(-\frac{2t^2}{M \, c_M^2} \right)$$

In order to apply the above theorem (replacing f_M by g_M), we need to upperbound the differences

$$\Big| g_M(\boldsymbol{x}_1, \ldots, \boldsymbol{x}'_k, \ldots, \boldsymbol{x}_M; \boldsymbol{w}_1, \ldots, \boldsymbol{w}_M)$$
$$- g_M(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k, \ldots \boldsymbol{x}_M; \boldsymbol{w}_1, \ldots, \boldsymbol{w}_M) \Big|$$

and

$$\Big| g_M(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_M; \boldsymbol{w}_1, \ldots, \boldsymbol{w}'_k, \ldots, \boldsymbol{w}_M)$$
$$- g_M(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_M; \boldsymbol{w}_1, \ldots, \boldsymbol{w}_k, \ldots, \boldsymbol{w}_M) \Big|$$

As the problem is symmetric, we only consider the first case here. Fix $1 \leq k \leq M$. Notice first that modifying the vector $\boldsymbol{x}_k$ only modifies a single column of the matrix F. Let us define $\widetilde{F}$ as being the matrix F with column k removed (so $\widetilde{F}$ is an $M \times (M-1)$ matrix). Because of what was just observed, $F - \widetilde{F}$ is a rank one matrix, so using the interlacing property

[2]This result can be proved rigorously; actually, the rigorous argument following for the two squares applies equally likely to the case of two parallel lines

of the singular values of F and $\tilde{F}$ [29, Theorem 7.3.9], we obtain that for all $1 \le j \le M$,

$$\widetilde{\lambda}_j \le \lambda_j \quad \text{and} \quad \widetilde{\lambda}_j \ge \lambda_{j+1}$$

where $\lambda_1 \ge \ldots \ge \lambda_M$ are the eigenvalues of $FF^\dagger/M$ and $\widetilde{\lambda}_1 \ge \ldots \ge \widetilde{\lambda}_M$ are the eigenvalues of $\widetilde{F}\widetilde{F}^\dagger/M$. Remember now that

$$g_M(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_M; \boldsymbol{w}_1, \ldots, \boldsymbol{w}_M)$$
$$= \log\det(I + \mathrm{SNR}(d)\, FF^\dagger/M) = \sum_{j=1}^{M} \log(1 + \mathrm{SNR}(d)\lambda_j)$$

Defining

$$\begin{aligned} \widetilde{g}_M &= \log\det(I + \mathrm{SNR}(d)\, \widetilde{F}\widetilde{F}^\dagger/M) \\ &= \sum_{j=1}^{M} \log(1 + \mathrm{SNR}(d)\widetilde{\lambda}_j) \end{aligned}$$

and applying the above inequalities on the eigenvalues, we see that

$$|g_M(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_M; \boldsymbol{w}_1, \ldots, \boldsymbol{w}_M) - \widetilde{g}_M| \le \log(1 + \mathrm{SNR}(d)\, \lambda_1)$$

It can be easily seen $\lambda_1 \le M$. Besides, we are interested in the regime where the growth of $\mathrm{SNR}(d)$ is no more than polynomial in M (it is actually constant in the case of interest), so for all $\varepsilon > 0$, there exists a constant $K > 0$ such that

$$|g_M(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_M; \boldsymbol{w}_1, \ldots, \boldsymbol{w}_M) - \widetilde{g}_M| \le K \log M$$

so by the triangle inequality,

$$\Big| g_M(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k', \ldots, \boldsymbol{x}_M; \boldsymbol{w}_1, \ldots, \boldsymbol{w}_M) - g_M(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k, \ldots \boldsymbol{x}_M; \boldsymbol{w}_1, \ldots, \boldsymbol{w}_M) \Big| \le 2K \log M$$

This finally shows, via McDiarmid's theorem, that

$$\mathbb{P}\Big(\Big| g_M(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_M; \boldsymbol{w}_1, \ldots, \boldsymbol{w}_M) - \mathbb{E}(g_M(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_M; \boldsymbol{w}_1, \ldots, \boldsymbol{w}_M)) \Big| > t \Big) \le 2 \exp\left(-\frac{t^2}{2MK^2(\log M)^2} \right)$$

which gives the result, by setting $t = M^{1/2+\varepsilon}$. □

Proof of Lemma 2.3. The proof of this lemma is based on the following simple observation. When $d \le \sqrt{A_c}$, the two clusters are close to each other, as illustrated on Figure 4.

Consider now the two square subclusters of size $\frac{\sqrt{A_c}}{2} \times \frac{\sqrt{A_c}}{2}$ that are the most separated horizontally. These two subclusters are now separated by a distance $d + \sqrt{A_c}$, which is of the same order as $\sqrt{A_c}$. On the other hand, both the area and the number of nodes in these subclusters remain of the same order as in the original clusters. More precisely, the new area is $A_c/4$ and correspondingly, the number of nodes in each subcluster is around $M/4$ with high probability. By letting only the nodes in these two subclusters participate to the MIMO transmission, we therefore see that the same order of spatial degrees of freedom can be achieved as when $d = \sqrt{A_c}$. □

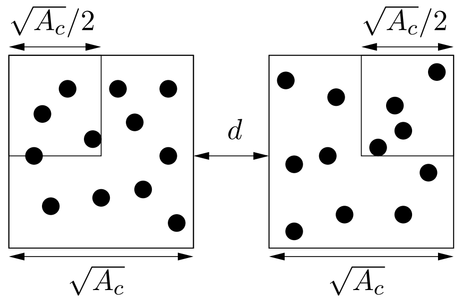

Fig. 4. The distance between the two highlighted subclusters is equal to $d + \sqrt{A_c}$.

III. AD HOC WIRELESS NETWORKS

An ad hoc network is a collection of wireless users which can self-organize and communicate among themselves without the help of any fixed infrastructure. Optimal cooperation architectures for such networks have received much focus in the asymptotic regime where the number of users is large. Consider an ad hoc network where n source-destination pairs are independently and uniformly distributed over an area A. Each source node wants to communicate to its corresponding destination at the same rate R bits/s/Hz. For this network, we have developed a hierarchical cooperation architecture in [24] that maintains a constant per-pair communication rate R even when the network serves a growing number of users, *provided that the channels between pairwise users are subject to i.i.d. fading.* More precisely, when the phases of the channel coefficients can be modeled as i.i.d. random variables independent of the node locations, hierarchical cooperation achieves an aggregate throughput $T = nR$ that scales linearly in n. Current communication architectures for ad hoc networks can not provide scalable performance. The traditional approach is to forward information from the source node to the destination by following a path, with intermediate nodes on the path acting as relays. The aggregate throughput of this multi-hop architecture scales as $\sqrt{n}$ with increasing network size n [17].

Distributed MIMO is key to the linear scaling of the architecture in [24]. Hierarchical cooperation allows nodes to efficiently organize in clusters and establish the n communications in the network via distributed MIMO transmissions between large clusters. Under i.i.d. fading, the capacity of these distributed MIMO transmissions are linear in the the number of nodes M contained in the transmit and receive clusters. In the earlier section, however, we have seen that the area of the clusters and the separation between them poses a limit on the linear scaling of distributed MIMO. In this section, we evaluate the performance of hierarchical cooperation under the physical model in (3) and show that it can still achieve linear scaling but provided that $\frac{\sqrt{A}}{\lambda} \ge n$. When $\frac{\sqrt{A}}{\lambda} < n$, we present two modifications of the hierarchical cooperation architecture that achieve the optimal scaling. Our result is formally summarized in the following theorem.

Theorem 3.1: Consider a wireless network of n nodes distributed uniformly at random over a square area A such that $\sqrt{A}/\lambda \geq \sqrt{n}$. Assume that each node is subject to an average power constraint P and the network is allocated a total bandwidth of W. The channels between pairwise users are governed by the LOS model in (3). Let us define the SNR of the network as

$$\text{SNR} = n\,\frac{GP}{N_0\,W\,A} \tag{12}$$

and assume that SNR ≥ 0 dB. For any $\varepsilon > 0$, there exists a constant $K > 0$ independent of n, λ and A such that an aggregate throughput

$$T \geq K\,\min\left\{n, \sqrt{A}/\lambda\right\}^{1-\varepsilon}$$

is achievable with high probability as n gets large.

We present the architectures that achieve this performance in the next section.

An upper bound on the best achievable capacity scaling under the model in (3) is developed in [26] for the case where the density of the nodes remains fixed as the number of nodes n grows. In this case, [26] shows that the capacity of the wireless network is upperbounded by

$$T \leq K_1\,\sqrt{n}\,(\log n)^2,$$

with high probability, where $K_1 > 0$ is a constant independent of n. The above result can at first lead to the conclusion that the best scaling achievable in wireless networks is $\sqrt{n}$ and therefore multi-hop is scaling optimal. However, a deeper look reveals that the conclusion that the capacity scales like $\sqrt{n}$ comes from the assumption that the density of nodes is fixed, so that $\sqrt{A}/\lambda$ is proportional to $\sqrt{n}$. A relatively straightforward generalization of the analysis in [26] gives the following refined upper bound on capacity:

$$T \leq \begin{cases} K_1\,n\,(\log n)^2 & \text{if } \frac{\sqrt{A}}{\lambda} \geq n \\ K_1\,\frac{\sqrt{A}}{\lambda}\,\left(\log \frac{\sqrt{A}}{\lambda}\right)^2 & \text{if } n > \frac{\sqrt{A}}{\lambda} > \sqrt{n} \\ K_1\,\sqrt{n}\,(\log n)^2 & \text{if } \frac{\sqrt{A}}{\lambda} \leq \sqrt{n} \end{cases} \tag{13}$$

with high probability, where $K_1 > 0$ is a constant independent of n, λ and A. For $\sqrt{A}/\lambda \leq \sqrt{n}$, this result says that the maximum achievable capacity is of order $\sqrt{n}$, which is achievable by a simple multihop. For $\sqrt{A}/\lambda > \sqrt{n}$, the achievability of the upper bound was an open problem and is now established in Theorem 3.1 when SNR ≥ 0 dB.

This leads to the conclusion that in the regime when n and $\sqrt{A}/\lambda$ are both large and SNR ≥ 0 dB, the capacity of the network is approximately

$$\max\left\{\sqrt{n}, \min\left\{n, \frac{\sqrt{A}}{\lambda}\right\}\right\}$$

Accordingly, the optimal operation of the network falls into three different operating regimes:

1) $\sqrt{A}/\lambda \leq \sqrt{n}$: The number of spatial degrees of freedom is too small, more sophisticated cooperation is useless and multihop is optimal.
2) $\sqrt{A}/\lambda \geq n$: The number of spatial degrees of freedom is n, the optimal performance can be achieved by the same hierarchical cooperation scheme introduced in [24]. Spatial degree of freedom limitation does not come into play and the performance is *as though* phases were i.i.d. uniform across node pairs.
3) $\sqrt{n} < \sqrt{A}/\lambda < n$: The number of degrees of freedom is smaller than n, so the spatial limitation is felt, but larger than what can be achieved by simple multi-hopping. A modification of the hierarchical cooperation scheme achieves the optimal scaling in this regime.

The SNR in (12) can be identified as the typical SNR between nearest neighbor nodes in the network under the channel model (3). Note that in a random network of n nodes distributed over an area A, the typical separation between nearest neighbor pairs is given by $\sqrt{A/n}$. The condition (12) ensures that these channels are in the high-SNR regime. Note that channels between pairs further away can be in low-SNR. Identifying optimal cooperation architectures for networks with SNR ≤ 0 dB under the physical channel model remains an open problem. Optimal architectures for such networks have been identified in [25] under the i.i.d. fading model.

A. Optimal cooperation in networks with limited spatial degrees of freedom

Capitalizing on the result of Theorem 2.1, in this section we prove Theorem 3.1 in three steps:

(A) When $\sqrt{A}/\lambda \geq n$, we verify that the performance of the hierarchical cooperation architecture in [24] scales linearly in n under the LOS model of (3).

(B) When $\sqrt{n} < \sqrt{A}/\lambda < n$, we show that a diluted hierarchical cooperation architecture achieves the scaling in Theorem 3.1. Here, only a randomly chosen subset of the source-destination pairs operate at a time while remaining nodes stay silent. This creates a diluted network for which case (1) holds so the network does not experience any limitation in spatial degrees of freedom. Different subsets take turns to operate.

(C) When $\sqrt{n} < \sqrt{A}/\lambda < n$, an alternative way to achieve the scaling in Theorem 3.1 is to use a hybrid architecture combining distributed MIMO with multi-hop, introduced in [25]. Here, nodes form MIMO clusters of an intermediate size and information is routed from one cluster to the next via successive distributed MIMO transmissions between adjacent clusters. The cluster size is critically chosen at the largest possible scale that allows for linear scaling of the distributed MIMO transmissions. When $\sqrt{n} = \sqrt{A}/\lambda$, this cluster size is a single node and the hybrid architecture reduces to pure multi-hop. When $\sqrt{A}/\lambda = n$, the cluster size is as large as n, and the architecture reduces to pure hierarchical cooperation.

The difference between the two strategies in (B) and (C) arises when we modify the channel model to

$$h_{ik} = \sqrt{G}\, \frac{e^{j2\pi r_{ik}/\lambda}}{r_{ik}^{\alpha/2}} \tag{14}$$

where α is the power path loss exponent determining how fast signal power decays with distance in the environment. Although not physical, this channel model provides a simple way to capture the impact of larger path loss attenuation due to multiple propagation paths, and at the same time it preserves the spatial correlation between channels by keeping the dependence of the phases to the geometric structure of the network. It has been shown in [26] that multiple paths do not change the scaling of the number of spatial degrees of freedom in a large network. Therefore it suffices to concentrate on a LOS model for the phases. However, multiple paths can have significant impact on the power. For example, with an additional reflected path from the ground plane the power path loss over distance r increases from r^2 to r^4. Under this new model, the power condition for achieving linear capacity for distributed MIMO becomes

$$\text{SNR}(d) = M\, \frac{GP}{N_0 W d^{\alpha}} \geq 0\,\text{dB}$$

in Theorem 2.1. Accordingly, the diluted hierarchical cooperation architecture in (B) achieves the performance in Theorem 3.1 when

$$\text{SNR}_l = n\, \frac{GP}{N_0\, W\, (\sqrt{A})^{\alpha}} \geq 0\,\text{dB}$$

SNR_l is defined as the long-range SNR of a network in [25]. This quantity can be identified as n times the received SNR in a point-to-point transmission over the largest scale in the network, the diameter A. The extra n comes from the network effect, it reflects the potential power gain due to cooperation over the global scale. For the multi-hop MIMO architecture in (3), the power requirement is given by

$$\text{SNR}(A_c) = M\, \frac{GP}{N_0\, W\, (A_c)^{\alpha}} \geq 0\,\text{dB}$$

where $\text{SNR}(A_c)$ is the analog of SNR_l but for a cluster of area A_c containing $M = A_c\, n/A$ nodes. In particular, we will choose M in the sequel such that $M = \frac{A}{\lambda^2\, n}$. When $\sqrt{n} < \sqrt{A}/\lambda < n$, we have $1 \leq M \leq n$. It is easy to verify that when $\alpha > 2$, $\text{SNR}(A_c) \geq \text{SNR}_l$ (see (3.12) in [25]), therefore the second condition is less stringent than the first. When $\alpha = 2$, $\text{SNR}(A_c) = \text{SNR}_l = \text{SNR}$ in (12). Therefore, for the LOS model, the two architectures are equivalent.

A detailed discussion on the relevance of the SNR parameters above in networks with i.i.d. fading is provided in [27]. The below discussion assumes that the reader is familiar with the hierarchical cooperation and the MIMO multihop architectures and their performance analysis. A detailed description of these strategies can be also found in [27].

1. Hierarchical cooperation when $\sqrt{A}/\lambda \geq n$ In this regime, the upper bound in (13) allows for throughput scaling linear in n. Potentially hierarchical cooperation can achieve arbitrarily close to linear scaling. One needs to check however that the MIMO transmissions taking place at all levels of the scheme are fully efficient, i.e. have capacity scaling linearly in the number of nodes in the clusters. This is easy to verify: consider a MIMO transmission between two clusters of area A_c and size $M = A_c\, n/A$. The separation d between these two clusters is upperbounded by the diameter of the network $\sqrt{A}$. Therefore

$$\frac{A_c}{\lambda d} \geq \frac{A_c}{\lambda\sqrt{A}} = \frac{\sqrt{A}}{\lambda n} M$$

Therefore when $\sqrt{A}/\lambda \geq n$, $\dfrac{A_c}{\lambda d} \geq M$, so by Eq. (5), distributed MIMO transmissions operate with full degrees of freedom (up to a logarithmic factor), just like in the case of i.i.d. phases. To compensate for the logarithmic factor, we argue that

$$\frac{A_c/\lambda d}{\log(A_c/\lambda d)} \geq M^{1-\varepsilon}$$

for any $\varepsilon > 0$ and sufficiently large M. This in turn implies that the capacity of the distributed MIMO transmissions scale as $M^{1-\varepsilon}$. The decrease by $M^{-\varepsilon}$ is captured in the $n^{-\varepsilon}$ degradation in the overall throughput in Theorem 2.1.

We also need to verify that the distributed MIMO transmissions have sufficient power as required in condition (4). In the hierarchical cooperation architecture, the MIMO transmission between clusters of area A_c and size M take place inside a larger cluster of area A_c' and size $M' = A_c'\, n/A$ in the next level of the hierarchy. Therefore the separation between the TX and RX clusters is upper bounded by $\sqrt{A_c'}$. During the MIMO transmissions each node transmits with elevated power $P_m = \frac{M'P}{M}$. This is because of the time-division between MIMO transmissions from different clusters. Each node transmits only a fraction M/M' of the time, therefore it can transmit with elevated power $\frac{M'P}{M}$ and still satisfy the average transmit power constraint P. See [27] for details. Therefore, the SNR for the MIMO transmissions is given by

$$\text{SNR}(d) = M\, \frac{GP_m}{N_0 W A_c'} = M' \frac{GP}{N_0 W A_c'} = n \frac{GP}{N_0 W A} \geq 0 \text{ dB}$$

where the last inequality is the power condition in Theorem 3.1. Therefore, MIMO transmissions at each level of the hierarchy have full degrees of freedom and sufficient power. Hierarchical cooperation achieves an aggregate throughput scaling arbitrarily close to linear in n in this case. □

2. Hierarchical cooperation when $\sqrt{n} \leq \sqrt{A}/\lambda < n$

In this regime, equation (13) shows that a linear throughput scaling is not achievable by any means. Nevertheless, the question remains whether one could outperform multi-hopping strategies, whose asymptotic performance $\sqrt{n}$ is strictly suboptimal compared to the upper bound $\sqrt{A}/\lambda$. A direct application of the hierarchical cooperation scheme fails to improve on multi-hop in this case, but it turns out that a simple adaptation of the scheme to this spatially limited situation achieves the optimal scaling.

The idea is the following: organize the communication of the n source-destination pairs into n/N sessions, each involving N source-destination pairs, where $N = \sqrt{A}/\lambda$. It is possible to choose here the nodes in a way such that each group of N nodes statistically occupies the total area of the network. This way, no group of N nodes considered alone feels the spatial limitation, as for this diluted network $N = \sqrt{A}/\lambda$ and we are in the case $\sqrt{A}/\lambda \geq N$ above. The sessions operate successively and the traffic in each session is handled using hierarchical cooperation where only the N chosen nodes are involved. The rest of the nodes remain silent. Since nodes are active only a fraction of N/n of the total time, when active they can transmit with elevated power $P_m = nP/N$ and still satisfy their individual power constraint P. Therefore, for the diluted network of N nodes in each session, the SNR is[3]

$$N \frac{GP_m}{N_0 WA} = n \frac{GP}{N_0 WA} \geq 0\,\text{dB}$$

Therefore, the diluted network is neither power nor space-limited and hierarchical cooperation achieves aggregate throughput of order $N^{1-\varepsilon} = (\sqrt{A}/\lambda)^{1-\varepsilon}$ for any fixed $\varepsilon > 0$. With time-division across different groups of nodes, the same throughput is achievable in the whole network. □

3. MIMO multi-hop when $\sqrt{n} \leq \sqrt{A}/\lambda < n$ Consider the MIMO multi-hop strategy described in Section 3.3 of [27]. On the global scale, this hybrid architecture is similar to multi-hop. The packets of each source-destination pair are transferred by hopping from one cluster to the next. At each hop, the packets are decoded and then re-encoded for the next hop. The architecture differs from multi-hop by the fact that each hop is performed via distributed MIMO transmissions assisted by hierarchical cooperation. Let us choose the cluster size M such that $\sqrt{A_c}/\lambda = M$ where $A_c = A\,M/n$ is the area of the cluster. This leads to $M = \frac{A}{\lambda^2 n}$. This choice of the cluster size ensures that the clusters are not limited in spatial degrees of freedom. Therefore, the capacity of distributed MIMO transmissions at each hop scales linearly in M provided that there is sufficient power. Since the distributed MIMO transmissions at each hop take place over a distance $\sqrt{A_c}$, the power condition in (4) yields

$$\text{SNR}(d) = M \frac{GP}{N_0 W A_c} \geq 0\,\text{dB}$$

This is equivalent to SNR ≥ 0 dB in Theorem 3.1[4], as $M/A_c = n/A$. When the capacity of the distributed MIMO transmissions at each hop scale linearly in M, the aggregate throughput of the MIMO multi-hop architecture is given by

$$\sqrt{n} M^{1/2-\varepsilon}$$

in Eq. (3.13) of [27]. Plugging our choice $M = \frac{A}{\lambda^2 n}$, we obtain an aggregate throughput scaling as $(\sqrt{A}/\lambda)^{1-\varepsilon}$. □

[3]With the model in (14), this condition becomes $\text{SNR}_l \geq 0$ dB.
[4]With the model in (14), this condition becomes $\text{SNR}(A_c) \geq 0$ dB.

IV. Acknowledgment

We would like to thank Marc Desgroseilliers for helping with the preparation of the present paper.

Appendix

Rigorous proof of Lemma 2.4. We now prove equation (8) without making use of approximations.

We start again with expression (9) for S. Notice that due to the symmetry of Δ and ρ in $\boldsymbol{w}_i$ and $\boldsymbol{w}_l$, we can upper bound (9) as

$$S \leq \frac{d^4}{A_c^4} \int_{D_T} d\boldsymbol{x}_k \int_{D_T} d\boldsymbol{x}_m \times \left| \int_{D_R} d\boldsymbol{w} \, \frac{e^{2\pi j\,(\|\boldsymbol{x}_k - \boldsymbol{w}\| - \|\boldsymbol{x}_m - \boldsymbol{w}\|)}}{\|\boldsymbol{x}_k - \boldsymbol{w}\|\,\|\boldsymbol{x}_m - \boldsymbol{w}\|} \right|^2$$

Expressing this upper bound more explicitly in the coordinate system shown on Figure 3, we obtain:

$$S \leq \int_0^1 dx_k \int_0^1 dy_k \int_0^1 dx_m \int_0^1 dy_m \times \left| \int_0^1 dw \int_0^1 dz \, \frac{e^{2\pi j g_{k,m}(w,z)}}{G_{k,m}(w,z)} \right|^2 \tag{15}$$

where

$$g_{k,m}(w,z) = \Big(\sqrt{(d + \sqrt{A_c}\,(x_k + w))^2 + A_c\,(y_k - z)^2} - \sqrt{(d + \sqrt{A_c}\,(x_m + w))^2 + A_c\,(y_m - z)^2} \Big)/\lambda$$

and

$$G_{k,m}(w,z) = d^{-2} \sqrt{(d + \sqrt{A_c}\,(x_k + w))^2 + A_c\,(y_k - z)^2} \times \sqrt{(d + \sqrt{A_c}\,(x_m + w))^2 + A_c\,(y_m - z)^2}$$

Let us first focus on the integral inside the square in (15). The key idea behind the next steps of the proof is contained in the following two lemmas.

Lemma 1.1: Let $g : [0,1] \to \mathbb{R}$ be a C^2 function such that $|g'(z)| \geq c_1 > 0$ for all $z \in [0,1]$ and g'' changes sign at most twice on $[0,1]$ (say e.g. $g''(z) \geq 0$ in $[z_-, z_+]$ and $g''(z) \leq 0$ outside). Let also $G : [0,1] \to \mathbb{R}$ be a C^1 function such that $|G(z)| \geq c_2 > 0$ and $G'(z)$ changes sign at most twice on $[0,1]$. Then

$$\left| \int_0^1 dz \, \frac{e^{2\pi j g(z)}}{G(z)} \right| \leq \frac{14}{\pi\, c_1\, c_2}.$$

Proof: By the integration by parts formula, we obtain

$$\int_0^1 dz \, \frac{e^{2\pi j g(z)}}{G(z)} = \int_0^1 dz \, \frac{2\pi j g'(z)}{2\pi j g'(z) G(z)} \, e^{2\pi j g(z)}$$
$$= \frac{e^{2\pi j g(z))}}{2\pi j g'(z) G(z)} \Big|_0^1 - \int_0^1 dz \, \frac{g''(z) G(z) + g'(z) G'(z)}{2\pi j (g'(z) G(z))^2} \, e^{2\pi j g(z)}$$

which in turn yields the upper bound

$$\left|\int_0^1 dz\, \frac{e^{2\pi j g(z)}}{G(z)}\right| \le \frac{1}{2\pi}\left(\frac{1}{|g'(1)||G(1)|} + \frac{1}{|g'(0)||G(0)|} + \int_0^1 dz\, \frac{|g''(z)|}{(g'(z))^2|G(z)|} + \int_0^1 dz\, \frac{|G'(z)|}{g'(z)(G(z))^2}\right)$$

By the assumptions made in the lemma, we have

$$\begin{aligned}\int_0^1 dz\, \frac{|g''(z)|}{(g'(z))^2|G(z)|} &\le \frac{1}{c_2}\int_0^1 dz\, \frac{|g''(z)|}{(g'(z))^2} \\ &= \frac{1}{c_2}\left(-\int_0^{z_-} dz\, \frac{g''(z)}{(g'(z))^2} + \int_{z_-}^{z_+} dz\, \frac{g''(z)}{(g'(z))^2} - \int_{z_+}^1 dz\, \frac{g''(z)}{(g'(z))^2}\right) \\ &= \frac{1}{c_2}\left(\frac{1}{g'(1)} - \frac{1}{g'(0)} + \frac{2}{g'(z_-)} - \frac{2}{g'(z_+)}\right)\end{aligned}$$

So

$$\int_0^1 dz\, \frac{|g''(z)|}{(g'(z))^2|G(z)|} \le \frac{6}{c_1\, c_2}.$$

We obtain in a similar manner that

$$\int_0^1 dz\, \frac{|G'(z)|}{g'(z)(G(z))^2} \le \frac{6}{c_1\, c_2}$$

Combining all the bounds, we finally get

$$\left|\int_0^1 dz\, \frac{e^{2\pi j g(z)}}{G(z)}\right| \le \frac{14}{\pi\, c_1\, c_2}$$

■

Lemma 1.2: Let $g : [0,1] \to \mathbb{R}$ be a C^2 function such that there exists $z_0 \in [0,1]$ and $c_1 > 0$ with $|g'(z)| \ge c_1\,|z - z_0|$ for all $z \in [0,1]$ and g'' changes sign at most twice on $[0,1]$. Let also $G : [0,1] \to \mathbb{R}$ be a C^1 function such that $|G(z)| \ge c_2 > 0$ and $G'(z)$ changes sign at most twice on $[0,1]$. Then

$$\left|\int_0^1 dz\, \frac{e^{2\pi j g(z)}}{G(z)}\right| \le \sqrt{\frac{14}{\pi\, c_1\, c_2}}.$$

Proof: The proof follows the steps of the previous lemma. In order to highlight the differences and for the sake of readability, we focus here on the simple case where $G(w,z) \equiv 1$. For any $\varepsilon > 0$, we have

$$\left|\int_0^1 dz\, e^{j\,2\pi g(z)}\right| = \left|\int_{z_0-\varepsilon}^{z_0+\varepsilon} dz\, e^{j\,2\pi g(z)}\right| + \left|\int_{[0,z_0-\varepsilon]\cup[z_0+\varepsilon,1]} dz\, \frac{j\,2\pi g'(z)}{j\,2\pi g'(z)}\, e^{j\,2\pi g(z)}\right|$$

Note that the first term can be simply upperbounded by 2ε. The second term can be bounded by the integration by parts formula, which yields.

$$\begin{aligned}&\int_{[0,z_0-\varepsilon]\cup[z_0+\varepsilon,1]} dz\, \frac{j\,2\pi g'(z)}{j\,2\pi g'(z)}\, e^{j\,2\pi g(z)} \\ &= \frac{e^{j\,2\pi g(z)}}{2\pi j g'(z)}\bigg|_0^{z_0-\varepsilon} + \int_0^{z_0-\varepsilon} dz\, \frac{g''(z)}{2\pi j (g'(z))^2}\, e^{j\,2\pi g(z)} \\ &\quad + \frac{e^{j\,2\pi g(z)}}{2\pi j g'(z)}\bigg|_{z_0+\varepsilon}^{1} + \int_{z_0+\varepsilon}^1 dz\, \frac{g''(z)}{2\pi j (g'(z))^2}\, e^{j\,2\pi g(z)}\end{aligned}$$

Using the assumptions on g' and g'' in the lemma and following similar steps to the proof of Lemma 1.1, we obtain

$$\begin{aligned}&\left|\int_0^1 dz\, \exp(2\pi j g(z))\right| \\ &\le 2\varepsilon + \frac{1}{2\pi}\left(\frac{2}{|g'(0)|} + \frac{2}{|g'(z_0-\varepsilon)|} + \frac{2}{|g'(z_0+\varepsilon)|} + \frac{2}{|g'(1)|}\right) \\ &\le 2\varepsilon + \frac{1}{2\pi}\,\frac{8}{c_1\varepsilon}\end{aligned}$$

Choosing $\varepsilon = \sqrt{\frac{2}{\pi c_1}}$ yields the desired result

$$\left|\int_0^1 dz\, \exp(2\pi j g(z))\right| \le \frac{8}{\sqrt{\pi\, c_1}}$$

which completes the proof. ■

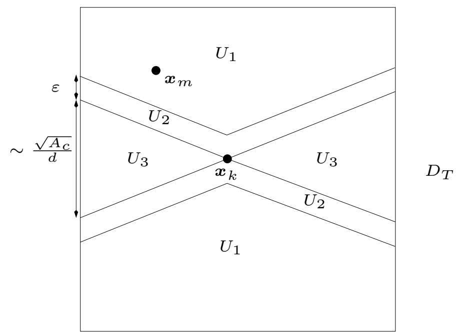

Fig. 5. Domains of integration: the relative positions of the points $\boldsymbol{x}_k$ and $\boldsymbol{x}_m$ determine in which domain one is (U_1 on the figure).

Let now $\varepsilon > 0$ and let us divide the integration domain $(x_k, x_m, y_k, y_m) \in [0,1]^4$ in (15) into three subdomains (see Figure 5):

$$\begin{aligned}U_1 &= \left\{|y_m - y_k| - (\sqrt{A_c}/d)\,|x_m - x_k| \ge \varepsilon\right\} \\ U_2 &= \left\{0 < |y_m - y_k| - (\sqrt{A_c}/d)\,|x_m - x_k| < \varepsilon\right\} \\ U_3 &= \left\{|y_m - y_k| \le (\sqrt{A_c}/d)\,|x_m - x_k|\right\}\end{aligned}$$

Consider first the integral over U_1. One can check that

$$\begin{aligned}&g_{k,m}(w,z) \\ &= -\frac{\sqrt{A_c}}{\lambda}\int_{x_k}^{x_m} \frac{(d/\sqrt{A_c} + x + w)\,dx}{\sqrt{(d/\sqrt{A_c} + x + w)^2 + (y_k - z)^2}} \\ &\quad + \frac{\sqrt{A_c}}{\lambda}\int_{y_k}^{y_m} \frac{(y - z)\,dy}{\sqrt{(d/\sqrt{A_c} + x_m + w)^2 + (y - z)^2}}\end{aligned}$$

So the first order partial derivative of $g_{k,m}(w,z)$ with respect to z is given by

$$\begin{aligned}
&\frac{\partial g_{k,m}}{\partial z}(w,z) \\
&= \frac{\sqrt{A_c}}{\lambda} \int_{x_k}^{x_m} \frac{(z-y_k)\,(d/\sqrt{A_c}+x+w)\;dx}{\left((d/\sqrt{A_c}+x+w)^2+(z-y_k)^2\right)^{3/2}} \\
&\quad + \frac{\sqrt{A_c}}{\lambda} \int_{y_k}^{y_m} \frac{(d/\sqrt{A_c}+x_m+w)^2\,dy}{\left((d/\sqrt{A_c}+x_m+w)^2+(z-y)^2\right)^{3/2}}
\end{aligned} \tag{16}$$

From this expression, we deduce that if $(x_k, x_m, y_k, y_m) \in U_1$, then

$$\left|\frac{\partial g_{k,m}}{\partial z}(w,z)\right| \geq K\,\frac{A_c}{\lambda d}\left(|y_m-y_k| - \frac{\sqrt{A_c}}{d}\,|x_m-x_k|\right)$$

for a constant $K > 0$ independent of A_c, λ and d. Notice next that $|G_{k,m}(y,z)| \geq 1$. It can further be checked that both $\frac{\partial^2 g_{k,m}}{\partial z^2}(w,z)$ and $\frac{\partial G_{k,m}}{\partial z}(w,z)$ change sign at most twice on the interval $z \in [0,1]$ (for w fixed). Therefore, applying Lemma 1.1, we conclude that

$$\begin{aligned}
&\left|\int_0^1 dw \int_0^1 dz \frac{e^{2\pi j g_{k,m}(w,z)}}{G_{k,m}(z)}\right| \\
&\leq \int_0^1 dw \left|\int_0^1 dz\, \frac{e^{2\pi j g_{k,m}(w,z)}}{G_{k,m}(y,z)}\right| \\
&\leq K\,\frac{\lambda d}{A_c}\,\frac{1}{|y_m-y_k| - (\sqrt{A_c}/d)\,|x_m-x_k|}
\end{aligned}$$

Since we know that this integral is also less than 1, this in turn implies

$$\begin{aligned}
&\int_{U_1} dx_k dx_m dy_k dy_m \left|\int_0^1 dw \int_0^1 dz\, \frac{e^{2\pi j g_{k,m}(w,z)}}{G_{k,m}(w,z)}\right|^2 \\
&\leq K\,\frac{\lambda d}{A_c} \int_{U_1} dx_k dx_m dy_k dy_m\, \frac{1}{|y_m-y_k| - (\sqrt{A_c}/d)\,|x_m-x_k|} \\
&= K\,\frac{\lambda d}{A_c}\,\log\left(\frac{1}{\varepsilon}\right)
\end{aligned}$$

Second, it is easy to check that

$$\int_{U_2} dx_k dx_m dy_k dy_m \left|\int_0^1 dw \int_0^1 dz\, \frac{e^{2\pi j g_{k,m}(w,z)}}{G_{k,m}(w,z)}\right|^2 \leq 2\varepsilon$$

The integral over the third domain of integration U_3 is more delicate. Notice first that the obvious bound

$$\int_{U_3} dx_k dx_m dy_k dy_m \left|\int_0^1 dw \int_0^1 dz\, \frac{e^{2\pi j g_{k,m}(w,z)}}{G_{k,m}(w,z)}\right|^2 \leq 2\frac{\sqrt{A_c}}{d}$$

allows to obtain

$$S \leq K\,\frac{\lambda d}{A_c}\,\log\left(\frac{1}{\varepsilon}\right) + 2\varepsilon + 2\,\frac{\sqrt{A_c}}{d}$$

which can be made smaller than $K\left(\frac{\lambda d}{A_c}\right)\log\left(\frac{A_c}{\lambda}\right)$ by choosing $\varepsilon = \frac{\lambda d}{A_c}$ when $\frac{A_c^{3/4}}{\sqrt{\lambda}} \leq d \leq \frac{A_c}{\lambda d}$ (as $\frac{\sqrt{A_c}}{d} \leq \frac{\lambda d}{A_c}$ in this case).

For the remainder of the proof, let us therefore assume that $\sqrt{A_c} \leq d \leq A_c^{3/4}/\sqrt{\lambda}$. As before, we focus on the integral inside the square in the following expression

$$\int_{U_3} dx_k dx_m dy_k dy_m \left|\int_0^1 dw \int_0^1 dz\, \frac{e^{2\pi j g_{k,m}(w,z)}}{G_{k,m}(w,z)}\right|^2 \tag{17}$$

Let us start by considering the simplest case where the points $\boldsymbol{x}_k$ and $\boldsymbol{x}_m$ are located on the same horizontal line, i.e. $y_k = y_m$. In this case, the second term in the expression (16) for $\frac{\partial g_{k,m}}{\partial z}(w,z)$ becomes zero, so we deduce the following lower bound:

$$\left|\frac{\partial g_{k,m}}{\partial z}(w,z)\right| \geq K\,\frac{A_c^{3/2}}{\lambda d^2}\,|x_m-x_k|\,|z-y_k|$$

This, together with the above mentioned properties of the functions $g_{k,m}$ and $G_{k,m}$, allows us to apply Lemma 1.2 so as to obtain

$$\left|\int_0^1 dw \int_0^1 dz\, \frac{e^{2\pi j g_{k,m}(w,z)}}{G_{k,m}(w,z)}\right| \leq K\,\frac{\sqrt{\lambda} d}{A_c^{3/4}}\,\frac{1}{\sqrt{|x_m-x_k|}}$$

for a constant $K > 0$ independent of A_c, λ and d. A slight generalization of this argument (see below for details) shows that not only when $y_k = y_m$ but for any $(x_k, x_m, y_k, y_m) \in U_3$, we have

$$\begin{aligned}
&\left|\int_0^1 dw \int_0^1 dz\, \frac{e^{2\pi j g_{k,m}(w,z)}}{G_{k,m}(w,z)}\right| \\
&\leq K\,\frac{\sqrt{\lambda} d}{A_c^{3/4}}\,\frac{1}{((x_m-x_k)^2+(y_m-y_k)^2)^{1/4}} \\
&\leq K\,\frac{\sqrt{\lambda} d}{A_c^{3/4}}\,\frac{1}{\sqrt{|x_m-x_k|}}
\end{aligned} \tag{18}$$

Since we also know that the above integral is less than 1, we further obtain

$$\begin{aligned}
&\left|\int_0^1 dw \int_0^1 dz\, \frac{e^{2\pi j g_{k,m}(w,z)}}{G_{k,m}(w,z)}\right|^2 \\
&\leq \min\left\{K\,\frac{\lambda d^2}{A_c^{3/2}}\,\frac{1}{|x_m-x_k|}\;,\;1\right\}
\end{aligned}$$

For any $0 < \eta < 1$, we can now upper bound (17) as follows:

$$\begin{aligned}
&\int_{U_3} dx_k dx_m dy_k dy_m \left|\int_0^1 dw \int_0^1 dz\, \frac{e^{2\pi j g_{k,m}(w,z)}}{G_{k,m}(w,z)}\right|^2 \\
&\leq |U_3 \cap \{|x_m-x_k| < \eta\}| \\
&+ K \int_{U_3 \cap \{|x_m-x_k| \geq \eta\}} dx_k dx_m dy_k dy_m\, \frac{\lambda d^2}{A_c^{3/2}}\,\frac{1}{|x_m-x_k|} \\
&\leq 2\eta + K\,\frac{\sqrt{A_c}}{d}\,\frac{\lambda d^2}{A_c^{3/2}}\,\log\left(\frac{1}{\eta}\right) = 2\eta + K\,\frac{\lambda d}{A_c}\,\log\left(\frac{1}{\eta}\right)
\end{aligned}$$

implying that

$$S \leq K\,\frac{\lambda d}{A_c}\,\log\left(\frac{1}{\varepsilon}\right) + 2\varepsilon + 2\eta + K\,\frac{\lambda d}{A_c}\,\log\left(\frac{1}{\eta}\right)$$

Choosing finally $\varepsilon = \eta = \lambda d/A_c$ allows to conclude that $S \leq K\left(\frac{\lambda d}{A_c}\right)\log\left(\frac{A_c}{\lambda d}\right)$ also in the case where $\sqrt{A_c} \leq d \leq A_c^{3/4}/\sqrt{\lambda}$. □

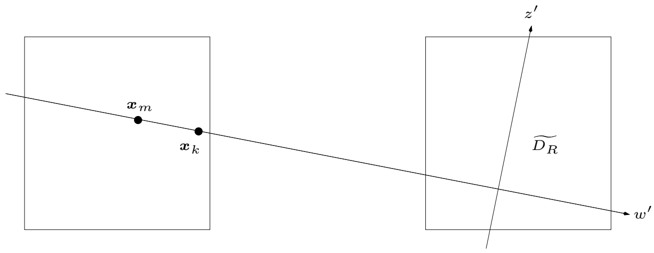

Fig. 6. Tilted reference frame.

Proof of equation (18). In order to prove (18), we need to make a change of coordinate system, replacing (w, z) by (w', z'), where w' is now in the direction of the vector $\boldsymbol{x}_k - \boldsymbol{x}_m$ and z' is perpendicular to it (see Figure 6). In this new coordinate system, the integral reads

$$\left| \int_{\widetilde{D_R}} dw'dz' \, \frac{e^{2\pi j g_{k,m}(w',z')}}{G_{k,m}(w',z')} \right|$$

where $g_{k,m}(w', z')$, $G_{k,m}(w', z')$ have the same form as $g_{k,m}(w, z)$, $G_{k,m}(w, z)$, but now, the domain of integration $\widetilde{D_R}$ is a tilted square, as indicated on the Figure 6. Using then the same argument as in the case $y_k = y_m$, we conclude that

$$\left| \int_{\widetilde{D_R}} dw'dz' \, \frac{e^{2\pi j g_{k,m}(w',z')}}{G_{k,m}(w',z')} \right| \le K \, \frac{\sqrt{\lambda}d}{A_c^{3/4}} \, \frac{1}{\sqrt{|x'_m - x'_k|}}$$

Noticing finally that $|x'_m - x'_k| = \sqrt{(x_m - x_k)^2 + (y_m - y_k)^2}$ allows to conclude that (18) holds. □

REFERENCES

[1] G. J. Foschini, *Layered Space-Time Architecture For Wireless Communication in a Fading Environment when Using Multi-Element Antennas*, AT&T Bell Labs Tech. Journal 1 (2), 41–59, 1996.

[2] E. Telatar, *Capacity of Multi-Antenna Gaussian Channels*, European Trans. on Telecommunications 10 (6), 585–596, November 1999.

[3] D. Shiu, G. J. Foschini, M. J. Gans, and J. M. Kahn, *Fading correlation and its effect on the capacity of multi-element antenna systems*, IEEE Trans. Communication, 48(3), pp.502-513, 2000.

[4] S. L. Loyka, *Channel capacity of MIMO architecture using the exponential correlation matrix*, IEEE Communication Letters, vol. 5, pp. 369-371, 2001

[5] , H. Bölcskei, D. Gesbert and A. J. Paulraj, *On the capacity of OFDM-based spatial multiplexing systems*, IEEE Trans. on Communications, 50 (2), pp. 225-234, 2002.

[6] , D. Gesbert, H. Bölcskei, D. A. Gore and A. J. Paulraj, *Outdoor MIMO wireless channels: models and performance predicition*, IEEE Trans. on Communications 50 (12), pp.1926-1934, 2002.

[7] K. Liu, V. Raghavan and A. Sayeed, *Capacity Scaling and Spectral Efficiency in Wideband Correlated MIMO Channels*, IEEE Transactions on Information Theory, pp. 2504-2526, Oct. 2003.

[8] G. G. Raleigh and J. M. Cioffi, *Spatio-temporal coding for wireless communication,* IEEE Trans. on Communications 46(3), pp. 357-366, 1998.

[9] A. M. Sayeed, *Deconstructing multi-antenna fading channels,* IEEE Trans. Signal Processing, vol. 50, pp. 2563-2579, Oct. 2002.

[10] R. Janaswamy, *Effect of element mutual coupling on the capacity of fixed length linear arrays,* IEEE Antennas Wireless Propagat. Lett., vol. 1, pp. 157160, 2002.

[11] Y. Fei, Y. Fan, B. K. Lau, and J. S. Thompson, *Optimal single-port matching impedance for capacity maximization in compact MIMO arrays,* IEEE Trans. Antennas Propagat., vol. 56, no. 11, pp. 35663575, Nov. 2008.

[12] D. A. B. Miller, *Communicating with waves between volumes: evaluating orthogonal spatial channels and limits on coupling strengths*, Applied Optics, 39 (11), pp. 1681 1699, 2000.

[13] T. S. Pollock, T. D. Abhayapala, and R. A. Kennedy, *Antenna saturation effects on MIMO capacity*, Proc. ICC, vol. 3, pp. 2301-2305, May 2003.

[14] A. S. Y. Poon, R. W. Brodersen and D. N. C. Tse, *Degrees of freedom in multiple-antenna channels: a signal space approach*, IEEE Trans. on Information Theory 51 (2), pp.523-536, 2005.

[15] L. Hanlen and M. Fu, *Wireless communication systems with spatial diversity: a volumetric model*, IEEE Trans. on Wireless Communication 5 (1), pp.133-142, 2006.

[16] S. Cui, A. Goldsmith and A. Bahai, *Energy-efficiency of MIMO and cooperative MIMO techniques in sensor networks,* IEEE Journal on Selected Areas in Communications, vol. 22(6), 2004.

[17] P. Gupta and P. R. Kumar, *The Capacity of Wireless Networks*, IEEE Trans. on Information Theory 42 (2), pp.388-404, 2000.

[18] L. -L. Xie and P. R. Kumar, *A Network Information Theory for Wireless Communications: Scaling Laws and Optimal Operation*, IEEE Trans. on Information Theory 50 (5), 2004, 748-767.

[19] A. Jovicic, P. Viswanath and S. R. Kulkarni, *Upper Bounds to Transport Capacity of Wireless Networks*, IEEE Trans. on Information Theory 50 (11), 2004, 2555-2565.

[20] O. Lévêque and E. Telatar, *Information Theoretic Upper Bounds on the Capacity of Large, Extended Ad-Hoc Wireless Networks*, IEEE Trans. on Information Theory 51 (3), 2005, 858- 865.

[21] L. -L. Xie, P. R. Kumar, *On the Path-Loss Attenuation Regime for Positive Cost and Linear Scaling of Transport Capacity in Wireless Networks*, IEEE Trans. on Information Theory, 52 (6), 2006, 2313-2328.

[22] A. Özgür, O. Lévêque, E. Preissmann, *Scaling laws for one and two-dimensional random wireless networks in the low attenuation regime*, IEEE Trans. on Information Theory 53 (10), 2007, 3573-3585.

[23] S. Aeron, V. Saligrama, *Wireless Ad hoc Networks: Strategies and Scaling Laws for the Fixed SNR Regime*, IEEE Trans. on Information Theory 53 (6), 2007, 2044 - 2059.

[24] A. Özgür, O. Lévêque, D. Tse, *Hierarchical Cooperation Achieves Optimal Capacity Scaling in Ad-Hoc Networks*, IEEE Trans. on Information Theory 53 (10), pp.3549-3572, 2007.

[25] A. Özgür, R. Johari, O. Lévêque, D. Tse, *Information Theoretic Operating Regimes of Large Wireless Networks*, IEEE Trans. on Information Theory 56 (1), pp.427-437, 2010.

[26] M. Franceschetti, M. D. Migliore, P. Minero, *The Capacity of Wireless Networks: Information-theoretic and Physical Limits*, IEEE Trans. on Information Theory 55 (8), 3413–3424, August 2009.

[27] A. Özgür, O. Lévêque and D. Tse, *Operating Regimes of Large Wireless Networks*, Foundations and Trends in Networking: Vol. 5: No 1, pp 1-107, http://dx.doi.org/10.1561/1300000016 .

[28] C. McDiarmid, *On the Method of Bounded Differences*, Surveys in Combinatorics, 1989.

[29] R. A. Horn and C. R. Johnson, *Matrix Analysis*, Cambridge University Press, 1985.

[30] S.-H. Lee and S.-Y. Chung, *Capacity scaling of wireless ad hoc networks: Shannon meets Maxwell*, IEEE Trans. on Information Theory 58 (3), 1702–1715, March 2012.