A Network Information Theory for Wireless Communication: Scaling Laws and Optimal Operation

Liang-Liang Xie, *Member, IEEE,* and P. R. Kumar, *Fellow, IEEE*

***Abstract*—How much information can be carried over a wireless network with a multiplicity of nodes, and how should the nodes cooperate to transfer information? To study these questions, we formulate a model of wireless networks that particularly takes into account the distances between nodes, and the resulting attenuation of radio signals, and study a performance measure that weights information by the distance over which it is transported.**

Consider a network with the following features.

i) n nodes located on a plane, with minimum separation distance $\rho_{\min} > 0$.

ii) A simplistic model of signal attenuation $\frac{e^{-\gamma\rho}}{\rho^{\delta}}$ over a distance ρ, where $\gamma \geq 0$ is the absorption constant (usually positive, unless over a vacuum), and $\delta > 0$ is the path loss exponent.

iii) All receptions subject to additive Gaussian noise of variance σ^2.

The performance measure we mainly, but not exclusively, study is the transport capacity $C_T := \sup \sum_{\ell=1}^{m} R_\ell \cdot \rho_\ell$, where the supremum is taken over m, and vectors $(R_1, R_2, \ldots, R_m)$ of feasible rates for m source–destination pairs, and ρ_ℓ is the distance between the ℓth source and its destination. It is the supremum distance-weighted sum of rates that the wireless network can deliver.

We show that there is a dichotomy between the cases of relatively high and relatively low attenuation. When $\gamma > 0$ or $\delta > 3$, the relatively high attenuation case, the transport capacity is bounded by a constant multiple of the sum of the transmit powers of the nodes in the network. This shows that there is a positive lower bound on the energy price in joules per bit-meter of information transport. If the nodes are individually power limited, the transport capacity consequently scales as $O(n)$. This order is, in fact, sharp, i.e., the transport capacity is $\Theta(n)$, for regular planar networks where the nodes are situated at integer lattice sites in a square. Consider now the "multihop" strategy where packets are routed over possibly multiple paths, and, along each path, packets are relayed from node to node with full decoding of each packet at each hop, treating all interference as noise, i.e., employing only point-to-point coding. This strategy is an order optimal strategy when the relaying burden can be balanced across the nodes, with no hop being too long. Or, in a randomly picked scenario, if nodes in a regular planar network randomly choose destination nodes, then the maximum common throughput that can be furnished to all nodes by multihop transport is nearly order optimal with respect to the transport capacity, differing only by a factor $1/\sqrt{\log n}$. Hence, up to order, there is no need for network coding or multiuser estimation. Thus, information theory can shed some light on what is an order-optimal architecture for wireless networks in situations where the load can be nearly balanced across nodes. In particular, the order optimality or near order optimality of multihop transport in such scenarios is of interest because much protocol development activity currently is actually aimed at realizing this strategy.

However, when $\gamma = 0$ and $\delta < 3/2$, the low-attenuation case, we show that there exist networks that can provide unbounded transport capacity for fixed total power, yielding zero energy priced communication. When nodes lie on a straight line and $\delta < 1$ (a physical impossibility in the three-dimensional world, but perhaps the examples can be generalized to a plane with larger values of δ), there are networks which can even attain superlinear scaling $\Theta(n^{\theta})$ for $\theta < 2$. Both these results are achieved by a strategy of coherent multistage relaying with interference subtraction. These examples show that nodes can profitably cooperate over large distances using coherence and multiuser estimation when the attenuation is low. These results are established by developing a coding scheme and an achievable rate for Gaussian multiple-relay channels, a result that may be of interest in its own right.

***Index Terms*—Ad hoc networks, capacity of wireless networks, multihop transport, multiuser information theory, network information theory, scaling laws, transport capacity, wireless networks.**

Manuscript received April 13, 2002; revised November 11, 2003. This work was supported in part by USARO under Contracts DAAD19-00-1-0466 and DAAD19-01010-465, DARPA under Contracts N00014-01-1-0576 and F33615-01-C-1905, AFOSR under Contract F49620-02-1-0217, DARPA/AFOSR under Contract F49620-02-1-0325, and NSF under Contracts NSF ANI 02-21357 and CCR-0325716. The material in this paper was presented in part at the IEEE HSN 2002: High-Speed Networking Workshop, New York, June 2002; the 2002 IEEE Information Theory Workshop, Bangalore, India, October 2002; the 41st IEEE Conference on Decision and Control, Las Vegas, NV, December 2002; the DIMACS Workshop on Network Information Theory, Rutgers University, Piscataway, NJ, March 2003, and the 2003 IEEE International Symposium on Information Theory, Yokohama, Japan, June/July 2003. This work was conducted while L.-L. Xie was visiting the Coordinated Science Laboratory, University of Illinois at Urbana-Champaign.

L.-L. Xie is with the Institute of Systems Science, Chinese Academy of Sciences, Beijing, 100080, China (e-mail: xie@mail.iss.ac.cn).

P. R. Kumar is with the University of Illinois at Urbana-Champaign, Coordinated Science Laboratory, Urbana, IL 61801-2307, USA (e-mail: prkumar@uiuc.edu).

Communicated by P. Narayan, Associate Editor for Shannon Theory.

Digital Object Identifier 10.1109/TIT.2004.826631

I. INTRODUCTION

THE focus of this paper is on wireless networks formed by nodes with radios. This includes *ad hoc* networks, currently the subject of great interest, the protocols for which are under intense development [1]–[7]. Since wireless networks may possibly play an important role in future communication networks, sensor networks, and sensor-actuator networks, it is important to understand what such networks are capable of doing, and how to operate them to maximize their capabilities. The goal of this paper is to develop an information theory for wireless networks to guide us in this process.

The two fundamental questions of interest are as follows.

i) How much information can wireless networks transport?

ii) How should one operate wireless networks?

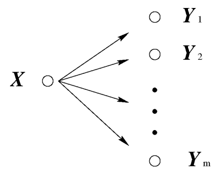

Fig. 1. The broadcast channel.

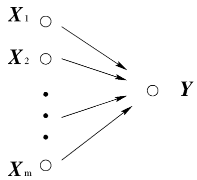

Fig. 2. The multiple-access channel.

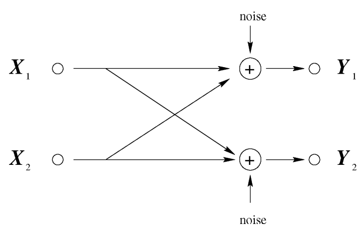

Fig. 3. A system with two transmitters and two receivers.

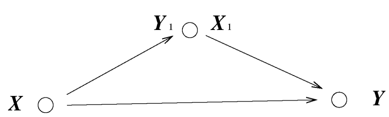

Fig. 4. The simplest relay channel.

A. Motivation and Background

It is a triumph of information theory that the capacity regions for even some systems have been characterized. The two prominent ones are the scalar Gaussian broadcast channel [10]–[13]; see Fig. 1, and the multiple-access channel [14], [15] shown in Fig. 2. More recently, for a network with a single source–destination pair, the asymptotic rate has been characterized as the number of nodes in a bounded domain is increased, while excluding them from open neighborhoods of the source and destination; see [16].

However, little else is known. The capacity region of even the simple four-node system with two sources and two receivers shown in Fig. 3, the so-called interference channel originally studied by Shannon (see [18], [19]), is unknown when the interfering powers are moderate rather than large or small. Also, unknown is the capacity of the simplest relay channel [20]–[22] shown in Fig. 4, consisting of just three nodes, a source, a relay, and a destination. Even in a simple four-node network with just two parallel relays, shown in Fig. 5, strategies which are very different in nature have to be considered for different parameter values [23].

In fact, when one turns to networks with even a few nodes, the possibilities for cooperation are very complex, and essentially nothing is known. Elementary modes of cooperation such as relaying, though already confounding efforts at a solution to date, only scratch the surface, and do not come close to exhausting the possibilities for interaction between a large number of nodes in a network. For example, a node could attempt to actively cancel the interference created by a second node at a third node. This is a form of cooperation where the first node attempts to decrease the denominator in the signal to interference plus noise ratio (SINR) of the third receiver rather than boosting the numerator. Nodes can simultaneously serve several functions such as relaying, broadcast, interference canceling, and other undreamt of possibilities. As observed in [17], the union between information theory and networks is not consummated, and in our view the two have actually been somewhat estranged.

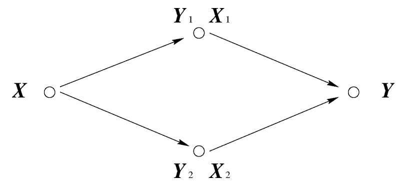

Fig. 5. A four-node network with two parallel relays.

Given this ocean of ignorance, what can one then say about much more complicated networks of the type shown later in Figs. 6 or 8, where there are several source–destination pairs among an arbitrarily large finite number of nodes, all cooperating in whatever ways are imaginable to maximize information transfer?

An attempt to address some of the issues raised above was made in [8] under an assumption on how the technology operates. However, to an information theorist, the answers are not conclusive as to what are the ultimate limits to feasibility. The reason is that, in [8], all interference is essentially regarded as noise, and models considered there presuppose that signals or packets are correctly received only if either there are no "collisions" with other packets being simultaneously transmitted by other nodes in the vicinity of the receiver, or the received signal-to-noise-plus-interference ratio is large enough, or the received rate is related to the SINR (see [9] for details on the latter), i.e., only point-to-point coding is considered. However, assumptions and constructs such as "collision," or "signal-to-noise-plus-interference ratio," are arbitrary. While they may well model how current technology operates, e.g., current wireless cards, and thus tell us what is feasible with such technology, they do not tell us what are the ultimate limits to information transfer in future wireless networks. The reason is simply that interference need not be interference—it can carry information. For example, it is well known from even the simple model of two transmitters and two receivers, see Fig. 3, that if there is excessive interference from an interfering transmitter, then that is in fact good, because the interfering signal can first be decoded perfectly, and then subtracted from the received signal, thus eliminating the interference.

Therefore, one wishes to study wireless networks without making arbitrary and preconceived assumptions about how they are to operate. Thus, it is that one turns to information theory for answers to the questions: How much information can wireless networks transport, and how should they be operated? This is the subject of the present paper.

B. Summary and Implications of the Results

To study these questions, we formulate an information-theoretic model of wireless networks which is somewhat richer than is usual in network information theory. We explicitly model distances between nodes, and attenuations of radio signals. We consider a model where nodes are located on a plane with a minimum separation distance between them, with each node having an individual power constraint (or a total power constraint is placed on all nodes). Attenuation of radio signals over a distance ρ is simplistically modeled as $\frac{e^{-\gamma\rho}}{\rho^{\delta}}$. Distance also plays an important role in the performance metric that we chiefly (but not exclusively) study, the transport capacity, which is the supremum distance weighted sum of rates that the network can support.

Briefly, our results are the following (see Theorems 3.1–3.12 for precise details).

i) With n denoting the number of nodes, the transport capacity grows like $O(n)$ when $\gamma > 0$ or $\delta > 3$. In fact, generally $\gamma > 0$, i.e., there is absorption, unless transmission is over a vacuum; see [24]. So this situation may be the commonly prevailing case, though γ may be small.

ii) The above is established by showing that the transport capacity is upper-bounded by a multiple of the total transmission powers of all the nodes, when $\gamma > 0$ or $\delta > 3$. Thus, there is a lower bound on the energy price in joules per bit-meter of information transport. This result may be of interest in its own right, since, given any traffic demand to be carried by the network and the distances between sources and destinations, it provides a lower bound on the energy consumption necessary to fulfill this communication demand. Such bounds may be useful in the future for energy-limited communication applications such as sensor networks, and the preconstant involved is worth sharpening.

iii) Currently, much protocol development activity is centered around realizing the following mode of operation, which we will refer to as "multihop" for brevity: Traffic is routed over possibly multiple paths, and, along each path, packets are relayed from node to node after being fully decoded and thus regenerated at each hop, treating all interference as noise, i.e., employing only point-to-point coding. The multihop mode is of interest because it is relatively simple to implement in that it does not require multiuser estimation or network coding or coherent cooperation. We show that this multihop strategy is an order-optimal or nearly order-optimal strategy for information transport, under some scenarios when $\gamma > 0$ or $\delta > 3$. As one instance of such a scenario, if nodes are located at integer lattice sites in a square, which we refer to as a regular planar network, and nodes randomly choose destinations, then multihop transport is nearly order optimal, differing at most by a factor $\frac{1}{\sqrt{\log n}}$ from the optimal order. Alternatively, if traffic can be load balanced across the network by multipath routing with bounded distance traversed at each hop, then the scaling law in i) is sharp, i.e., the transport capacity order is $\Theta(n)$, and it is achieved by multihop transport. Overall, the theme of these results is to provide justification for the use of multihop in situations where the load is balanced, or nearly balanceable across nodes by using multipath routing if necessary. These information-theoretic results thus attempt to shed some theoretical light on what is the architecture for information transport, in an order-optimal sense, for random or load-balanced scenarios. It should be noted that the situations network information theory has been successfully able to resolve in the past are the two extreme bottlenecked cases of the multiple-access channel and the scalar Gaussian broadcast channel. In the former, the destination is the bottleneck, while in the latter, it is the source. In contrast, in networks, one generally wishes to use all network-wide resources, and support traffic demands all over the network. Our results are aimed at such situations, and attempt to connect information theory to the world of networking by providing some strategic guidance for such contexts, albeit only in order-optimal sense with bounds on preconstants.

iv) There is a dichotomy between the relatively high and low attenuation cases. When $\gamma = 0$ and $\delta < 3/2$, there are networks where unbounded transport capacity can be obtained for a fixed total power. In fact, for nodes on a line with $\delta < 1$ (an impossibility in the three-dimensional world), there are networks where the transport capacity scales superlinearly like $\Theta(n^{\theta})$ for $1 < \theta < 2$. These results suggest that the low-attenuation regime may be quite different from the high-attenuation regime.

v) The strategy achieving the results in iv) is coherent multistage relaying with interference subtraction (CRIS). This result shows that nodes may be able to profitably cooperate over long distances by using coherence and multiuser estimation, when the attenuation is low. Thus, other architectures for information transport, different from multihop, arise as interesting candidates when the attenuation is low. Thus, information theory provides some strategic guidance.

vi) To exhibit the results in iv) and v) we provide a new coding scheme and an achievable rate for Gaussian multiple-relay channels. This scheme and result may be of interest in their own right.

The remainder of this paper is organized as follows. In Section II, we detail the models considered, and in Section III the main results, with nothing but proofs in Section IV. Some concluding remarks are made in Section V.

II. Model of Wireless Networks

We consider the following model of a wireless network, called a *planar network* (see Fig. 6).

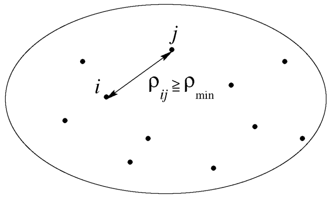

Fig. 6. A planar network: n nodes located on a two-dimensional plane, with minimum separation distance $\rho_{\min}$.

1) There is a finite set $\mathcal{N}$ of n nodes located on a plane.
2) There is a minimum positive separation distance $\rho_{\min}$ between nodes, i.e., $\rho_{\min} := \min_{i \neq j} \rho_{ij} > 0$, where ρ_{ij} is the distance between nodes $i, j \in \mathcal{N}$.
3) Every node has a receiver and a transmitter. At time instants $t = 1, 2, \ldots$, node $i \in \mathcal{N}$ sends $X_i(t)$, and receives $Y_i(t)$ with

$$Y_i(t) = \sum_{j \neq i} \frac{e^{-\gamma \rho_{ij}} X_j(t)}{\rho_{ij}^{\delta}} + Z_i(t)$$

where $Z_i(t)$, $i \in \mathcal{N}$, $t = 1, 2, \ldots$ are Gaussian independent and identically distributed (i.i.d.) random variables with mean zero and variance σ^2. The constant $\delta > 0$ will be called the *path loss exponent*, while $\gamma \geq 0$ will be called the *absorption constant*. A positive γ generally prevails except for transmission in a vacuum, and corresponds to a loss of $20\gamma \log_{10} e$ decibel per meter; see [24].
4) Denote by $P_i \geq 0$ the power used by node i. We will study two separate constraints on $\{P_1, P_2, \ldots, P_n\}$:

$$\text{Total Power Constraint } P_{\text{total}}: \sum_{i=1}^{n} P_i \leq P_{\text{total}}$$

or

$$\text{Individual Power Constraint } P_{\text{ind}}: P_i \leq P_{\text{ind}}, \quad \text{for } i = 1, 2, \ldots, n.$$

5) The network can have several source–destination pairs (s_ℓ, d_ℓ), $\ell = 1, \ldots, m$, where s_ℓ, d_ℓ are nodes in $\mathcal{N}$ with $s_\ell \neq d_\ell$, and $(s_\ell, d_\ell) \neq (s_j, d_j)$ for $\ell \neq j$. If $m = 1$, then there is only a single source–destination pair, which we will simply denote by (s, d).

Essentially, this is the network version of the classical additive white Gaussian noise (AWGN) channel, with signals attenuated by distance, and possibly multiple source–destination pairs. The model explicitly incorporates the distance between nodes, and signal attenuation as a function of distance. This feature is important for our results.

A special case is a regular planar network where the n nodes are located at the points (i, j) for $1 \leq i, j \leq \sqrt{n}$; see Fig. 7. This setting will be used mainly to exhibit achievability of some capacities, i.e., inner bounds.

Another special case is a *linear network* where the n nodes are located on a straight line, again with minimum separation distance $\rho_{\min}$; see Fig. 8. The chief reason for considering linear networks is that the proofs are easier to state and comprehend than in the planar case, and can be generalized to the planar case. Also, the linear case may have some utility for, say, networks of cars on a highway, since its scaling laws are different.

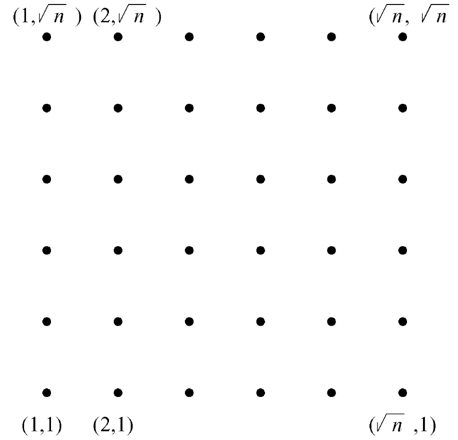

Fig. 7. A regular planar network: n nodes located on a plane at (i, j) with $1 \leq i, j \leq \sqrt{n}$.

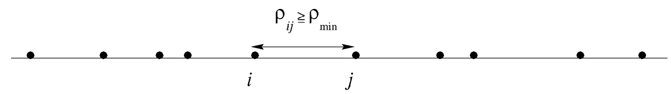

Fig. 8. A linear network: n nodes located on a line, with minimum separation distance $\rho_{\min}$.

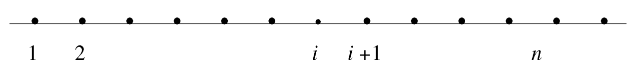

Fig. 9. A regular linear network: n nodes located on a line at $1, 2, \ldots, n$.

A special case of a linear network is a regular linear network where the n nodes are located at the positions $1, 2, \ldots, n$; see Fig. 9. This setting will also be used mainly to exhibit achievability results.

A. Definition of Feasible Rate Vectors

The following definition of feasible rates is standard. It captures the complicated interplays possible in a large number of nodes with multiple source–destination pairs, and intrinsically allows for all causal feedbacks, thus including all strategies for information transport. The distances between nodes, the attenuations, and the signal and noise powers, are all known *a priori* to all the nodes.

Definition 2.1: Consider a wireless network with multiple source–destination pairs (s_ℓ, d_ℓ), $\ell = 1, \ldots, m$, with $s_\ell \neq d_\ell$, and $(s_\ell, d_\ell) \neq (s_j, d_j)$ for $\ell \neq j$. Let $\mathcal{S} := \{s_\ell, \ell = 1, \ldots, m\}$ denote the set of source nodes. The number of nodes in $\mathcal{S}$ may be less than m, since we allow a node to have originating traffic for several destinations. Then a $((2^{TR_1}, \ldots, 2^{TR_m}), T, P_e^{(T)})$ code with total power constraint P_{total} consists of the following.

1) m independent random variables W_ℓ with $P(W_\ell = k_\ell) = \frac{1}{2^{TR_\ell}}$, for any $k_\ell \in \{1, 2, \ldots, 2^{TR_\ell}\}$, $\ell = 1, \ldots, m$. For any $i \in \mathcal{S}$, let

$$\overline{W}_i := \{W_\ell : s_\ell = i\} \quad \text{and} \quad \overline{R}_i := \sum_{\{\ell : s_\ell = i\}} R_\ell.$$

2) Functions

$$f_{i,t} : \mathbb{R}^{t-1} \times \{1, 2, \ldots, 2^{T\overline{R}_i}\} \to \mathbb{R}, \quad t = 1, 2, \ldots, T$$

for the source nodes $i \in \mathcal{S}$ and $f_{j,t} : \mathbb{R}^{t-1} \to \mathbb{R}, t = 2, \ldots, T$ for all the other nodes $j \notin \mathcal{S}$, such that

$$X_i(t) = f_{i,t}\big(Y_i(1), \ldots, Y_i(t-1), \overline{W}_i\big), \quad t = 1, 2, \ldots, T$$

$$X_j(1) = 0, \;\; X_j(t) = f_{j,t}\big(Y_j(1), \ldots, Y_j(t-1)\big), \quad t = 2, 3, \ldots, T$$

such that the following total power constraint holds:

$$\frac{1}{T} \sum_{t=1}^{T} \sum_{i \in \mathcal{N}} X_i^2(t) \le P_{\text{total}}, \quad \text{a.s.} \tag{1}$$

3) m decoding functions

$$g_\ell : \mathbb{R}^T \times \{1, 2, \ldots, |\overline{\mathcal{W}}_{d_\ell}|\} \to \{1, 2, \ldots, 2^{TR_\ell}\}$$

for the destination nodes of the m source–destination pairs $\{(s_\ell, d_\ell), \ell = 1, \ldots, m\}$, where $|\overline{\mathcal{W}}_{d_\ell}|$ is the number of different values $\overline{W}_{d_\ell}$ can take. Note that $\mathcal{W}_{d_\ell}$ may be empty.

4) The average probability of error:

$$P_e^{(T)} := \text{Prob}\big((\hat{W}_1, \hat{W}_2, \ldots, \hat{W}_m) \neq (W_1, W_2, \ldots, W_m)\big) \tag{2}$$

where $\hat{W}_\ell := g_\ell(Y_{d_\ell}^T, \overline{W}_{d_\ell})$, with

$$Y_{d_\ell}^T := \big(Y_{d_\ell}(1), Y_{d_\ell}(2), \ldots, Y_{d_\ell}(T)\big).$$

Definition 2.2: A rate vector $(R_1, \ldots, R_m)$ is said to be *feasible* for the m source–destination pairs $(s_\ell, d_\ell), \ell = 1, \ldots, m$, with total power constraint P_{total}, if there exists a sequence of $\big((2^{TR_1}, \ldots, 2^{TR_m}), T, P_e^{(T)}\big)$ codes satisfying the total power constraint P_{total}, such that $P_e^{(T)} \to 0$ as $T \to \infty$.

The preceding definitions are presented in the context of a total power constraint P_{total}. However, if an individual power constraint P_{ind} is placed on each node, then one simply needs to replace the constraint (1) by

$$\frac{1}{T} \sum_{t=1}^{T} X_i^2(t) \le P_{\text{ind}}, \quad \text{a.s., for } i \in \mathcal{N} \tag{3}$$

and correspondingly modify the rest of the definitions to define the set of feasible rate vectors under an individual power constraint.

B. The Transport Capacity

It is traditional in information theory to study the capacity *region*, which is the closure of the set of all such feasible vector rates. We will, however, focus mainly on the distance-weighted sum of rates introduced in [8].

Definition 2.3: The network's *transport capacity* C_T is

$$C_T := \sup_{(R_1, \ldots, R_m) \text{ feasible}} \sum_{\ell=1}^{m} R_\ell \cdot \rho_\ell$$

where for brevity $\rho_\ell := \rho_{s_\ell d_\ell}$ denotes the distance between s_ℓ and d_ℓ, and $R_\ell := R_{s_\ell d_\ell}$.

This is the supremum distance-weighted sum of rates that the network can deliver. The units in which it is measured is bit-meters per second, or bit-meters per slot. When one bit has been successfully received by a destination at a distance of one meter from the source of that bit, we say that the network has pumped one bit-meter.

This transport capacity is of interest for three different reasons. The first reason is that the transport capacity is indeed of interest in its own right as a natural measure of the distance hauling capacity of wireless networks. It is analogous to the man-miles/year metric used, for example, by airlines. It provides a single number which summarizes what a network can deliver, and is thus a useful quantity for designers to keep in mind.

Second, we will show under conditions detailed in the sequel that the transport capacity follows a scaling law. This is akin to a conservation law and is thus a hard constraint on what the wireless network can deliver, regardless of whether the transport capacity is of *prima facie* interest in its own right. Thus, for example, some rate vectors can be ruled out as impossible to support if their corresponding distance-weighted sum of rates exceeds the upper bound on the transport capacity.

Third, whenever a rate vector is feasible, and is such that its distance-weighted sum of rates is close to the transport capacity, then one can rest assured that the network is being operated close to capacity. We will see that this situation actually holds, up to order, for certain scenarios, and in such cases one is consequently assured that the strategy which supports that rate vector is order optimal. Thus we can provide some answers to what are order optimal strategies for wireless networks, and thereby attempt to bridge the gap between information theory and networking.

On the other hand, being a single number, the transport capacity does not provide full information on the complete set of rate vectors that can be supported. However such a complete characterization is difficult in general, and has certainly defied attempts to date.

III. The Theorems

Our main results can be grouped into four categories.

A. Upper bounds under high attenuation.
B. Multihop and feasible lower bounds under high attenuation.
C. The low-attenuation regime.
D. The Gaussian multiple-relay channel.

A. Upper Bounds Under High Attenuation

i) The transport capacity is bounded by the network's total transmission power in media with $\gamma > 0$ or $\delta > 3$.

It is well known from Shannon's work that for a single link (s,d), the *rate* R is bounded by the *received* power at d. What is interesting in wireless networks is that the transport capacity is upper-bounded by the total transmitted power P_{total} used by the entire network.

Theorem 3.1: In any planar network, with either positive absorption, i.e., $\gamma > 0$, or with path loss exponent $\delta > 3$

$$C_T \leq \frac{c_1(\gamma,\delta,\rho_{\min})}{\sigma^2} \cdot P_{\text{total}} \quad \text{where} \tag{4}$$

$$c_1(\gamma,\delta,\rho_{\min}) := \begin{cases} \frac{2^{2\delta+7}\log e}{\gamma^2 {\rho_{\min}}^{2\delta+1}} \frac{e^{-\gamma\rho_{\min}/2}(2-e^{-\gamma\rho_{\min}/2})}{(1-e^{-\gamma\rho_{\min}/2})}, & \text{if } \gamma > 0 \\ \frac{2^{2\delta+5}(3\delta-8)\log e}{(\delta-2)^2(\delta-3){\rho_{\min}}^{2\delta-1}}, & \text{if } \gamma=0 \text{ and } \delta>3. \end{cases} \tag{5}$$

ii) The transport capacity follows an $O(n)$ scaling law under the individual power constraint, in media with $\gamma > 0$ or $\delta > 3$.[1]

A corollary of the above theorem is as follows.

Theorem 3.2: Consider any planar network under the individual power constraint P_{ind}. Suppose that either there is some absorption in the medium, i.e., $\gamma > 0$, or there is no absorption at all but the path loss exponent $\delta > 3$. Then its transport capacity is upper-bounded as follows:

$$C_T \leq \frac{c_1(\gamma,\delta,\rho_{\min})P_{\text{ind}}}{\sigma^2} \cdot n \tag{6}$$

where $c_1(\gamma,\delta,\rho_{\min})$ is as in (5).

For n nodes located in an area of A square meters, it is shown in [8] that the transport capacity is of order $O(\sqrt{An})$ under a noninformation theoretic protocol model. If A itself grows like n, i.e., $A = \Theta(n)$, then the scaling law is $O(\sqrt{An}) = O(n)$, which coincides with the information-theoretic scaling law here. In fact, A must grow at least this rate since nodes are separated by a minimum distance $\rho_{\min} > 0$, i.e., $A = \Omega(n)$, and so the $O(n)$ result here is slightly stronger than the $O(\sqrt{An})$ result in [8].

iii) The following are the corresponding results for linear networks.

Theorem 3.3: If either $\gamma > 0$ or $\delta > 2$ in any linear network, then

$$C_T \leq \frac{c_2(\gamma,\delta,\rho_{\min})}{\sigma^2} \cdot P_{\text{total}}, \quad \text{where} \tag{7}$$

$$c_2(\gamma,\delta,\rho_{\min}) := \begin{cases} \frac{2e^{-2\gamma\rho_{\min}}\log e}{(1-e^{-\gamma\rho_{\min}})^2(1-e^{-2\gamma\rho_{\min}}){\rho_{\min}}^{2\delta-1}}, & \text{if } \gamma>0, \\ \frac{2\delta(\delta^2-\delta-1)\log e}{(\delta-1)^2(\delta-2){\rho_{\min}}^{2\delta-1}}, & \text{if } \gamma=0 \text{ and } \delta > 2. \end{cases} \tag{8}$$

[1]We use Knuth's notation: $f = O(g)$ if $\limsup_{n\to+\infty} \frac{f(n)}{g(n)} < +\infty$; $f = \Omega(g)$ if $g = O(f)$; $f = \Theta(g)$ if $f = O(g)$ as well as $g = O(f)$. Thus, all $O(\cdot)$ results are upper bounds, all $\Omega(\cdot)$ results are lower bounds, and all $\Theta(\cdot)$ results are sharp order estimates for the transport capacity.

Theorem 3.4: For any linear network, if either $\gamma > 0$ or $\delta > 2$, then the transport capacity is upper-bounded as follows:

$$C_T \leq \frac{c_2(\gamma,\delta,\rho_{\min})P_{\text{ind}}}{\sigma^2} \cdot n$$

where $c_2(\gamma,\delta,\rho_{\min})$ is as in (8).

B. Multihop and Feasible Lower Bounds Under High Attenuation

iv) The $O(n)$ upper bound on transport capacity is tight for regular planar networks in media with $\gamma > 0$ or $\delta > 3$, and it is achieved by multihop.

By a "multihop strategy" we mean the following. Let Π_ℓ denote the set of all paths from source s_ℓ to destination d_ℓ, where by such a path π we mean a sequence $(s_\ell = j_0, j_1, \ldots, \ldots, j_z = d_\ell)$ with $j_q \neq j_r$ for $q \neq r$. The total traffic rate R_ℓ to be provided to the source destination pair (s_ℓ, d_ℓ) is split over the paths in Π_ℓ in such a way that if traffic rate $\lambda_\pi \geq 0$ is to be carried over path π, then $\sum_{\pi\in\mathcal{P}_\ell} \lambda_\pi = R_\ell$. On each path π, packets are relayed from node to next node. On each such hop, each packet is fully decoded, treating all interference as noise. Thus, only point-to-point coding is used, and no network coding or multiuser estimation is employed. Such a strategy is of great interest and it is currently the object of much protocol development activity.

The following result implies that when $\gamma > 0$ or $\delta > 3$, the sharp order of the transport capacity for a regular planar network is $\Theta(n)$, and that it can be attained by multihop.

Theorem 3.5: In a regular planar network with either $\gamma > 0$ or $\delta > 1$, and individual power constraint P_{ind}

$$C_T \geq S\left(\frac{e^{-2\gamma}P_{\text{ind}}}{c_3(\gamma,\delta)P_{\text{ind}} + \sigma^2}\right) \cdot n, \quad \text{where}$$

$$c_3(\gamma,\delta) := \begin{cases} \frac{4(1+4\gamma)e^{-2\gamma} - 4e^{-4\gamma}}{2\gamma(1-e^{-2\gamma})}, & \text{if } \gamma > 0, \\ \frac{16\delta^2 + (2\pi-16)\delta - \pi}{(\delta-1)(2\delta-1)}, & \text{if } \gamma = 0 \text{ and } \delta > 1 \end{cases}$$

and $S(x)$ denotes the Shannon function

$$S(x) := \frac{1}{2}\log(1+x).$$

This order of distance weighted sum of rates is achievable by multihop.

v) Multihop is order optimal in a random scenario over a regular planar network in media with $\gamma > 0$ or $\delta > 3$, providing some theoretical justification for its use in situations where traffic is diffused over the network.

Theorem 3.6: Consider a regular planar network with either $\gamma > 0$ or $\delta > 1$, and individual power constraint P_{ind}. The n source–destination pairs are randomly chosen as follows: Every source s_ℓ is chosen as the node nearest to a randomly (uniformly i.i.d.) chosen point in the domain, and similarly for every destination d_ℓ. Then

$$\lim_{n\to\infty} \text{Prob}\Big(R_\ell = \frac{c}{\sqrt{n\log n}} \text{ is feasible for every } \ell \in \{1,2,\ldots,n\}\Big) = 1$$

for some $c > 0$. Consequently, a distance weighted sum of rates

$$C_T = \Omega(\frac{n}{\sqrt{\log n}})$$

is supported with probability approaching one as $n \to \infty$. This is within a factor $\frac{1}{\sqrt{\log n}}$ of the transport capacity $\Theta(n)$ possible when $\delta > 3$.

vi) A vector of rates $(R_1, R_2, \ldots, R_m)$ can be supported by multihop in a planar network in media with $\gamma > 0$ or $\delta > 1$, if the traffic can be load balanced such that no node is overloaded and no hop is too long.

This is a fairly straightforward result saying nothing about order optimality, and is provided only in support of the above theme that multihop is an appropriate architecture for balanceable scenarios.

Theorem 3.7: A set of rates $(R_1, R_2, \ldots, R_m)$ for a planar network can be supported by multihop if no hop is longer than a distance $\bar{\rho}$, and for every $1 \le i \le n$, the traffic to be relayed by node i

$$\sum_{\ell=1}^{m} \sum_{\{\pi \in \Pi_\ell : \text{ Node } i \text{ belongs to } \pi\}} \lambda_\pi$$

is less than

$$S\left(\frac{e^{-2\gamma\bar{\rho}} P_{\text{ind}}}{\bar{\rho}^{2\delta}(c_4(\gamma,\delta,\rho_{\min})P_{\text{ind}} + \sigma^2)}\right)$$

where

$$c_4(\gamma,\delta,\rho_{\min}) := \begin{cases} \frac{2^{3+2\delta} e^{-\gamma\rho_{\min}}}{\gamma\rho_{\min}^{1+2\delta}}, & \text{if } \gamma > 0 \\ \frac{2^{2+2\delta}}{\rho_{\min}^{2\delta}(\delta-1)}, & \text{if } \gamma = 0 \text{ and } \delta > 1. \end{cases}$$

C. The Low-Attenuation Regime

We now turn to situations where there is absolutely no absorption, i.e., $\gamma = 0$, and the path loss exponent is small. (The precise low values for which the results below hold vary, and are specified in the theorem statements.)

The following strategy for information transport, which we call coherent relaying with interference subtraction (CRIS), emerges as interesting in the scenarios that follow. For a source–destination pair (s, d), the nodes are divided into groups, with the first group containing only the source s, and the last group containing only the destination d. Call the higher numbered groups as "downstream" groups, though they need not actually be closer to the destination. Nodes in group i, for $1 \le i \le k-1$, dedicate a portion P_{ik} of their power to coherently transmit for the benefit of node k and its downstream nodes. Each node k employs interference subtraction during decoding to subtract out the known portion of its received signal being transmitted by its downstream nodes.

vii) In media with low attenuation, $\gamma = 0$ and $\delta < 3/2$, unbounded transport capacity can sometimes be obtained for bounded total power, by using CRIS.

The following theorem shows that in some scenarios, zero energy priced communication can be provided by CRIS in regular planar networks. This is in contrast with the high attenuation regime.

Theorem 3.8:

i) If there is no absorption, i.e., $\gamma = 0$, and the path loss exponent $\delta < 3/2$, then even with a fixed total power P_{total}, any arbitrarily large transport capacity can be supported by CRIS in a regular planar network with a large enough number of nodes n.

ii) If $\gamma = 0$ and $\delta < 1$, then even with a fixed total power P_{total}, CRIS can support a fixed rate $R_{\min} > 0$ for any single source–destination pair in any regular planar network, irrespective of the distance between them.

The following is the corresponding result for regular linear networks.

Theorem 3.9:

i) If $\gamma = 0$ and $\delta < 1$, then even with a fixed total power P_{total}, any arbitrarily large transport capacity can be supported by CRIS in a regular linear network with a large enough number of nodes n.

ii) If $\gamma = 0$ and $\delta < 1/2$, then even with a fixed total power P_{total}, CRIS can support a fixed rate $R_{\min} > 0$ for any single source–destination pair in any regular linear network, irrespective of the distance between them.

viii) A superlinear $\Theta(n^\theta)$ scaling law with $1 < \theta < 2$ is feasible for some linear networks when $\gamma = 0$ and $\delta < 1$.

Theorem 3.10: Consider $\gamma = 0$ and individual power constraint P_{ind}. For every $\frac{1}{2} < \delta < 1$,[2] and $1 < \theta < \frac{1}{\delta}$, there is a family of linear networks for which the transport capacity is

$$C_T = \Theta(n^\theta). \quad (9)$$

This order optimal transport capacity is attained in these networks by CRIS.

D. The Gaussian Multiple-Relay Channel

The results for the low-attenuation regime rely on the following results for the Gaussian multiple-relay channel. Consider a network of n nodes with α_{ij} the attenuation from node i to node j (the nodes need not lie on a plane, and in fact there need not even be a notion of distance), and i.i.d. additive $N(0, \sigma^2)$ noise at each receiver. Each node has an upper bound on the power available to it, which may differ from node to node. Suppose there is a single source–destination pair (s, d). We call this the Gaussian multiple relay channel.

ix) A new coding scheme and an explicit achievable rate for the Gaussian multiple relay channel.

[2]Any value of $\delta < 1$ is impossible in the three-dimensional world, since $\delta = 1$ corresponds to the ideal inverse square law. In this case, however, the nodes lie along a one-dimensional line, and perhaps the example here can be generalized to a planar network with a larger value of δ. One related issue to consider is Olber's paradox [25] on why the night sky is not bright. For a linear network with individual power constraints, even if there is no absorption ($\gamma = 0$), and even if there are an infinite number of nodes, the total received power is finite at every node if $\delta > 1$, or even if only $\delta > \frac{1}{2}$ provided the sources are incoherent. That is, the night sky is not bright for these path loss parameters.

These results may be of independent interest.

The first theorem addresses the case where each relaying group consists of only one node. The strategy used is CRIS.

Theorem 3.11: Consider the Gaussian multiple-relay channel with coherent multistage relaying and interference subtraction. Consider $M+1$ nodes, sequentially denoted by $0, 1, \dots, M$, with 0 as the source, M as the destination, and the other $M-1$ nodes serving as $M-1$ stages of relaying. Then any rate R satisfying the following inequality is achievable from 0 to M:

$$R < \min_{1\le j\le M} S\left(\frac{1}{\sigma^2}\sum_{k=1}^{j}\left(\sum_{i=0}^{k-1}\alpha_{ij}\sqrt{P_{ik}}\right)^2\right) \tag{10}$$

where $P_{ik} \ge 0$ satisfies $\sum_{k=i+1}^{M} P_{ik} \le P_i$.

Remark 3.1: For the network setting in Theorem 3.11, Theorem 3.1 in [26] shows that a rate R_0 is achievable if there exist some $\{R_1, R_2, \dots, R_{M-1}\}$ such that

$$R_{M-1} < S\left(\frac{P^R_{M,M-1}}{\sigma^2 + \sum_{\ell=0}^{M-2} P^R_{M,\ell}}\right) \quad \text{and}$$

$$R_m < \min\left\{ S\left(\frac{P^R_{m+1,m}}{\sigma^2 + \sum_{\ell=0}^{m-1} P^R_{m+1,\ell}}\right), \right.$$
$$\left. R_{m+1} + \min_{m+2\le k\le M} S\left(\frac{P^R_{k,m}}{\sigma^2 + \sum_{\ell=0}^{m-1} P^R_{k,\ell}}\right)\right\}$$

for each $m = 0, 1, \dots, M-2$, where

$$P^R_{k,\ell} := \left(\sum_{i=0}^{\ell}\alpha_{ik}\sqrt{P_{i,\ell+1}}\right)^2, \qquad 0 \le \ell < k \le M.$$

From the above, recursively for $m = M-2, M-1, \dots, 0$, it is easy to prove that

$$R_m < \min_{m+1\le j\le M} S\left(\frac{\sum_{k=m}^{j-1} P^R_{j,k}}{\sigma^2 + \sum_{\ell=0}^{m-1} P^R_{j,\ell}}\right).$$

For $m = 0$, this inequality is exactly (10), showing that we get a higher achievable rate in Theorem 3.11.

Remark 3.2: The right-hand side (RHS) in (10) can be maximized over the choice of order of the $M-1$ intermediate nodes.

The relaying can also be done by groups, and the next result addresses this. As above, maximization can be done over the assignment of nodes to the groups.

Theorem 3.12: Consider again the Gaussian multiple-relay channel using coherent multistage relaying with interference subtraction. Consider any $M+1$ groups of nodes sequentially denoted by $\mathcal{N}_0, \mathcal{N}_1, \dots, \mathcal{N}_M$ with $\mathcal{N}_0 = \{s\}$ as the source, $\mathcal{N}_M = \{d\}$ as the destination, and the other $M-1$ groups as $M-1$ stages of relay. Let n_i be the number of nodes in Group $\mathcal{N}_i$, $i \in \{0, 1, \dots, M\}$. Let the power constraint for each node in Group $\mathcal{N}_i$ be $\frac{P_i}{n_i} \ge 0$. Then any rate R satisfying the following inequality is achievable from s to d:

$$R < \min_{1\le j\le M} S\left(\frac{1}{\sigma^2}\sum_{k=1}^{j}\left(\sum_{i=0}^{k-1}\alpha_{\mathcal{N}_i\mathcal{N}_j}\sqrt{P_{ik}/n_i}\cdot n_i\right)^2\right) \tag{11}$$

where $P_{ik} \ge 0$ satisfies $\sum_{k=i+1}^{M} P_{ik} \le P_i$, and

$$\alpha_{\mathcal{N}_i\mathcal{N}_j} := \min\{\alpha_{k\ell} : k \in \mathcal{N}_i, \ell \in \mathcal{N}_j\}, \quad i, j \in \{0, 1, \dots, M\}.$$

IV. Nothing but Proofs

We begin with a max-flow min-cut bound.

A. A Max-Flow Min-Cut Lemma

This lemma has a similar spirit to that of Theorem 14.10.1 in [22]. But there are differences. First, here the bound (RHS of (12)) is in terms of received power instead of mutual information. Second, more importantly, there is an average over time in the bound here, which better takes care of the dynamic nature of the model we consider in this paper. When this lemma is applied later in proving the upper bounds in Theorems 3.1–3.3, we will see that this type of max-flow min-cut lemma is better suited for Gaussian wireless networks without resorting to mutual information terms.

Definition 4.1: Let $\mathcal{N}_1 \subset \mathcal{N}$. A source–destination pair (s_ℓ, d_ℓ) is said to *cut* $\mathcal{N}_1$ if $d_\ell \in \mathcal{N}_1$ but $s_\ell \notin \mathcal{N}_1$.

Lemma 4.1: Let $\mathcal{N}_1$ be any subset of $\mathcal{N}$. If $(R_1, \dots, R_m)$ is a feasible rate vector with a sequence of

$$\left((2^{TR_1}, \dots, 2^{TR_m}), T, P_e^{(T)}\right)$$

codes with $P_e^{(T)} \to 0$ as $T \to \infty$, then

$$\sum_{\{\ell: d_\ell\in\mathcal{N}_1, s_\ell\notin\mathcal{N}_1\}} R_\ell \le \frac{\log e}{2\sigma^2}\liminf_{T\to\infty} P^{\text{rec}}_{\mathcal{N}_1}(T) \tag{12}$$

where $P^{\text{rec}}_{\mathcal{N}_1}(T)$ is the average power received by $\mathcal{N}_1$, from outside $\mathcal{N}_1$, for the code $((2^{TR_1}, \dots, 2^{TR_m}), T, P_e^{(T)})$, i.e.,

$$P^{\text{rec}}_{\mathcal{N}_1}(T) := \frac{1}{T}\sum_{t=1}^{T}\sum_{i\in\mathcal{N}_1} E\left(\sum_{j\notin\mathcal{N}_1}\frac{e^{-\gamma\rho_{ij}}X_j(t)}{\rho_{ij}^{\delta}}\right)^2. \tag{13}$$

Proof: First we introduce some notation

$$U_i(t) := \sum_{j\notin\mathcal{N}_1}\frac{e^{-\gamma\rho_{ij}}X_j(t)}{\rho_{ij}^{\delta}}, \qquad i \in \mathcal{N}_1 \tag{14}$$

$$V_i(t) := U_i(t) + Z_i(t), \qquad i \in \mathcal{N}_1. \tag{15}$$

Denote

$$\overline{W}_{\mathcal{N}_1^{\text{dest-cut}}} := \{W_\ell : (s_\ell, d_\ell) \text{ cuts } \mathcal{N}_1\}$$
$$\mathcal{N}_1^{\text{source}} := \{s_\ell : s_\ell \in \mathcal{N}_1, \ell = 1, \dots, m\}$$
$$\overline{W}_{\mathcal{N}_1^{\text{source}}} := \{\overline{W}_i, i \in \mathcal{N}_1^{\text{source}}\}.$$

We adopt the notation

$$V_{\mathcal{N}_1}(t) := \{V_i(t), i \in \mathcal{N}_1\}$$
$$V^t_{\mathcal{N}_1} := \{V_{\mathcal{N}_1}(\tau), \tau = 1, \dots, t\}$$

and, similarly, for Y, U, X, and Z.

Now we prove that the following forms a Markov chain:

$$\overline{W}_{\mathcal{N}_1^{\text{dest-cut}}} \to \{V_{\mathcal{N}_1}^T, \overline{W}_{\mathcal{N}_1^{\text{source}}}\} \to \{Y_{\mathcal{N}_1}^T, \overline{W}_{\mathcal{N}_1^{\text{source}}}\} \quad (16)$$

by showing that any element in $Y_{\mathcal{N}_1}^T$ is a deterministic function of $\{V_{\mathcal{N}_1}^T, \overline{W}_{\mathcal{N}_1^{\text{source}}}\}$. This can be easily seen since for any $i \in \mathcal{N}_1$, $2 \le t \le T$

$$\begin{aligned} Y_i(t) &= V_i(t) + \sum_{\substack{j\in\mathcal{N}_1\\ j\ne i}} \frac{e^{-\gamma\rho_{ij}} X_j(t)}{\rho_{ij}^\delta} \\ &= V_i(t) + \sum_{\substack{j\in\mathcal{N}_1\setminus\mathcal{N}_1^{\text{source}}\\ j\ne i}} \frac{e^{-\gamma\rho_{ij}} f_{j,t}(Y_j^{t-1})}{\rho_{ij}^\delta} \\ &\quad + \sum_{\substack{j\in\mathcal{N}_1^{\text{source}}\\ j\ne i}} \frac{e^{-\gamma\rho_{ij}} f_{j,t}(Y_j^{t-1}, \overline{W}_j)}{\rho_{ij}^\delta} \end{aligned}$$

and for $t = 1$

$$Y_i(1) = V_i(1) + \sum_{\substack{j\in\mathcal{N}_1^{\text{source}}\\ j\ne i}} \frac{e^{-\gamma\rho_{ij}} f_{j,1}(\overline{W}_j)}{\rho_{ij}^\delta}.$$

Hence, by Fano's lemma and (16), we have

$$H(\overline{W}_{\mathcal{N}_1^{\text{dest-cut}}} | V_{\mathcal{N}_1}^T, \overline{W}_{\mathcal{N}_1^{\text{source}}}) \le 1 + TP_e^{(T)} \sum_{\{\ell: d_\ell\in\mathcal{N}_1, s_\ell\notin\mathcal{N}_1\}} R_\ell =: T\epsilon_T$$

where $\epsilon_T \to 0$ as $T \to \infty$.

Thus, we have the following chain of inequalities:

$$\begin{aligned} T &\sum_{\{\ell: d_\ell\in\mathcal{N}_1, s_\ell\notin\mathcal{N}_1\}} R_\ell \\ &= H(\overline{W}_{\mathcal{N}_1^{\text{dest-cut}}}) \\ &= I(\overline{W}_{\mathcal{N}_1^{\text{dest-cut}}}; V_{\mathcal{N}_1}^T, \overline{W}_{\mathcal{N}_1^{\text{source}}}) \\ &\quad + H(\overline{W}_{\mathcal{N}_1^{\text{dest-cut}}} | V_{\mathcal{N}_1}^T, \overline{W}_{\mathcal{N}_1^{\text{source}}}) \\ &\le I(\overline{W}_{\mathcal{N}_1^{\text{dest-cut}}}; V_{\mathcal{N}_1}^T, \overline{W}_{\mathcal{N}_1^{\text{source}}}) + T\epsilon_T \\ &= I(\overline{W}_{\mathcal{N}_1^{\text{dest-cut}}}; \overline{W}_{\mathcal{N}_1^{\text{source}}}) \\ &\quad + I(\overline{W}_{\mathcal{N}_1^{\text{dest-cut}}}; V_{\mathcal{N}_1}^T | \overline{W}_{\mathcal{N}_1^{\text{source}}}) + T\epsilon_T \\ &= 0 + h(V_{\mathcal{N}_1}^T | \overline{W}_{\mathcal{N}_1^{\text{source}}}) \\ &\quad - h(V_{\mathcal{N}_1}^T | \overline{W}_{\mathcal{N}_1^{\text{dest-cut}}}, \overline{W}_{\mathcal{N}_1^{\text{source}}}) + T\epsilon_T \\ &\le h(V_{\mathcal{N}_1}^T) - h(V_{\mathcal{N}_1}^T | \overline{W}_{\mathcal{N}_1^{\text{dest-cut}}}, \overline{W}_{\mathcal{N}_1^{\text{source}}}) + T\epsilon_T \end{aligned}$$

with

$$\begin{aligned} h(V_{\mathcal{N}_1}^T &| \overline{W}_{\mathcal{N}_1^{\text{dest-cut}}}, \overline{W}_{\mathcal{N}_1^{\text{source}}}) \\ &= \sum_{t=1}^T h\big(V_{\mathcal{N}_1}(t) | V_{\mathcal{N}_1}(1), \dots, V_{\mathcal{N}_1}(t-1), \\ &\qquad\qquad \overline{W}_{\mathcal{N}_1^{\text{dest-cut}}}, \overline{W}_{\mathcal{N}_1^{\text{source}}}\big) \\ &\ge \sum_{t=1}^T h\big(V_{\mathcal{N}_1}(t) | V_{\mathcal{N}_1}(1), \dots, V_{\mathcal{N}_1}(t-1), \\ &\qquad\qquad X_{\mathcal{N}_1^c}(t), \overline{W}_{\mathcal{N}_1^{\text{dest-cut}}}, \overline{W}_{\mathcal{N}_1^{\text{source}}}\big) \\ &= \sum_{t=1}^T h(V_{\mathcal{N}_1}(t) | X_{\mathcal{N}_1^c}(t)) \\ &\ge \sum_{t=1}^T h(V_{\mathcal{N}_1}(t) | U_{\mathcal{N}_1}(t)) \end{aligned}$$

where the last two (in)equalities follow from the following two Markov chains:

$$\begin{aligned} \{V_{\mathcal{N}_1}^{t-1}, \overline{W}_{\mathcal{N}_1^{\text{dest-cut}}}, \overline{W}_{\mathcal{N}_1^{\text{source}}}\} &\to X_{\mathcal{N}_1^c}(t) \to V_{\mathcal{N}_1}(t) \\ X_{\mathcal{N}_1^c}(t) &\to U_{\mathcal{N}_1}(t) \to V_{\mathcal{N}_1}(t). \end{aligned}$$

Hence, we have

$$\begin{aligned} T &\sum_{\{\ell: d_\ell\in\mathcal{N}_1, s_\ell\notin\mathcal{N}_1\}} R_\ell \\ &\le h(V_{\mathcal{N}_1}^T) - \sum_{t=1}^T h(V_{\mathcal{N}_1}(t)|U_{\mathcal{N}_1}(t)) + T\epsilon_T \\ &= h(V_{\mathcal{N}_1}^T) - \sum_{t=1}^T \sum_{i\in\mathcal{N}_1} h(Z_i(t)) + T\epsilon_T \\ &\le \sum_{t=1}^T \sum_{i\in\mathcal{N}_1} h(V_i(t)) - \sum_{t=1}^T \sum_{i\in\mathcal{N}_1} h(Z_i(t)) + T\epsilon_T \\ &= \sum_{t=1}^T \sum_{i\in\mathcal{N}_1} [h(V_i(t)) - h(V_i(t)|U_i(t))] + T\epsilon_T \\ &= \sum_{t=1}^T \sum_{i\in\mathcal{N}_1} I(U_i(t); V_i(t)) + T\epsilon_T \\ &\le \sum_{t=1}^T \sum_{i\in\mathcal{N}_1} \frac{1}{2}\log\left(1 + \frac{EU_i^2(t)}{\sigma^2}\right) + T\epsilon_T \\ &\le \frac{\log e}{2\sigma^2} \sum_{t=1}^T \sum_{i\in\mathcal{N}_1} EU_i^2(t) + T\epsilon_T. \end{aligned}$$

Finally, letting $T \to \infty$ in the above, and noticing $\epsilon_T \to 0$, we have (12). □

B. The Total Power Bound and Linear Scaling Law Under High Attenuation

We begin with the case of linear networks since the proof is easiest in that case.

Proof of Theorem 3.3: First we consider the case $\gamma = 0$ and $\delta > 2$.

Let $a_i \rho_{\min}$ denote the coordinate of the node i. Apply Lemma 4.1 to the following subsets:

$$\begin{aligned} \mathcal{N}_q^- &= \{i \in \mathcal{N} : -\infty < a_i \le q\} \\ \mathcal{N}_q^+ &= \{i \in \mathcal{N} : q \le a_i < \infty\}, \qquad q \in \mathbb{Z} \end{aligned} \quad (17)$$

and we have for any $q \in \mathbb{Z}$

$$\frac{2\sigma^2}{\log e} \cdot R_{\mathcal{N}_q^-} \le \liminf_{T\to\infty} \frac{1}{T} \sum_{t=1}^T \sum_{i\in\mathcal{N}_q^-} E\left(\sum_{j\notin\mathcal{N}_q^-} \frac{X_j(t)}{(a_j - a_i)^\delta \rho_{\min}^\delta}\right)^2 \quad (18)$$

$$\frac{2\sigma^2}{\log e} \cdot R_{\mathcal{N}_q^+} \leq \liminf_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T}\sum_{i\in\mathcal{N}_q^+} E\left(\sum_{j\notin\mathcal{N}_q^+} \frac{X_j(t)}{(a_i-a_j)^\delta \rho_{\min}^\delta}\right)^2. \quad (19)$$

Above, $R_{\mathcal{N}_q^-}$ is the sum of the rates of all the pairs which cut $\mathcal{N}_q^-$. $R_{\mathcal{N}_q^+}$ is similarly defined.

Now, any source–destination pair (s_ℓ, d_ℓ) with distance ρ_ℓ between s_ℓ and d_ℓ cuts at least $\lfloor \rho_\ell/\rho_{\min} \rfloor$ subsets among $\mathcal{N}_q^-, \mathcal{N}_q^+, q \in \mathbb{Z}$. For example, if $a_{d_\ell} = a$ and $a_{s_\ell} = a + \rho_\ell/\rho_{\min}$ (the case where $a_{s_\ell} < a_{d_\ell}$ being analyzed similarly), then (s_ℓ, d_ℓ) cuts the subsets

$$\mathcal{N}_q^-, q = \lceil a \rceil, \ldots, \lceil a + \rho_\ell/\rho_{\min} - 1 \rceil.$$

By definition, R_ℓ is a summand in every

$$R_{\mathcal{N}_q^-}, q = \lceil a \rceil, \ldots, \lceil a + \rho_\ell/\rho_{\min} - 1 \rceil.$$

Hence, we have (noting $\rho_\ell \geq \rho_{\min}$)

$$\begin{aligned}\sum_{\ell=1}^{m} R_\ell \cdot \rho_\ell &\leq 2\rho_{\min}\sum_{\ell=1}^{m} R_\ell \cdot \lfloor \rho_\ell/\rho_{\min} \rfloor \\ &\leq 2\rho_{\min}\sum_{q=-\infty}^{+\infty} R_{\mathcal{N}_q^-} + 2\rho_{\min}\sum_{q=-\infty}^{+\infty} R_{\mathcal{N}_q^+}. \quad (20)\end{aligned}$$

Now we prove that

$$\sum_{q=-\infty}^{+\infty} R_{\mathcal{N}_q^-} \leq \frac{c_2(\gamma,\delta,\rho_{\min})}{4\rho_{\min}\sigma^2} P_{\text{total}}. \quad (21)$$

By (18), we only need to show that

$$\frac{1}{T}\sum_{t=1}^{T}\sum_{q=-\infty}^{+\infty}\sum_{i\in\mathcal{N}_q^-} E\left(\sum_{j\notin\mathcal{N}_q^-}\frac{X_j(t)}{(a_j-a_i)^\delta}\right)^2 \leq \frac{c_2(\gamma,\delta,\rho_{\min}){\rho_{\min}}^{2\delta-1}}{2\log e} P_{\text{total}} \quad (22)$$

with $X_j(t)$ satisfying the total power constraint

$$\frac{1}{T}\sum_{t=1}^{T}\sum_{j\in\mathcal{N}} X_j^2(t) \leq P_{\text{total}}, \quad \text{a.s.} \quad (23)$$

The intuition behind the inequality (22) is that the summation of the received powers is upper-bounded by the total transmitted power.

We now establish (22) for the case where $\delta > 2$, as follows. By (23), for $\delta > 2$, we only need to prove that for any t

$$\sum_{q=-\infty}^{+\infty}\sum_{i\in\mathcal{N}_q^-}\left(\sum_{j\notin\mathcal{N}_q^-}\frac{X_j(t)}{(a_j-a_i)^\delta}\right)^2 \leq \frac{c_2(\gamma,\delta,\rho_{\min}){\rho_{\min}}^{2\delta-1}}{2\log e} P(t) \quad (24)$$

where

$$P(t) := \sum_{i\in\mathcal{N}} X_i^2(t). \quad (25)$$

First, we observe that the left-hand side (LHS) of (24) is a summation of infinite terms of the basic form $\beta_{jk}X_j(t)X_k(t)$, where β_{jk} is the appropriate coefficient. If every $X_j(t)X_k(t)$ is replaced with the larger value $\frac{1}{2}(X_j^2(t)+X_k^2(t))$, it is easy to see that

$$\text{LHS of (24)} \leq \sum_{k\in\mathcal{N}}\left(\sum_{q=-\infty}^{\lceil a_k\rceil-1}\sum_{i\in\mathcal{N}_q^-}\sum_{j\notin\mathcal{N}_q^-}\frac{1}{(a_j-a_i)^\delta(a_k-a_i)^\delta}\right)X_k^2(t).$$

This, together with (25), would imply (24), as long as for any $k \in \mathcal{N}$

$$\sum_{q=-\infty}^{\lceil a_k\rceil-1}\sum_{i\in\mathcal{N}_q^-}\sum_{j\notin\mathcal{N}_q^-}\frac{1}{(a_j-a_i)^\delta(a_k-a_i)^\delta} \leq \frac{c_2(\gamma,\delta,\rho_{\min}){\rho_{\min}}^{2\delta-1}}{2\log e}. \quad (26)$$

For $\delta > 2$, (26) is established by the following chain of inequalities: Letting $\underline{a}_q := \min_{j\notin\mathcal{N}_q^-} a_j$, and noting that $\min_{i\neq j}|a_i - a_j| \geq 1$, we have

$$\begin{aligned}&\text{LHS of (26)} \\ &\leq \sum_{q=-\infty}^{\lceil a_k\rceil-1}\sum_{i\in\mathcal{N}_q^-}\sum_{l=0}^{\infty}\frac{1}{(l+\underline{a}_q-a_i)^\delta}\frac{1}{(a_k-a_i)^\delta} \\ &\leq \sum_{q=-\infty}^{\lceil a_k\rceil-1}\sum_{i\in\mathcal{N}_q^-}\frac{\delta-1+\underline{a}_q-a_i}{(\delta-1)(\underline{a}_q-a_i)^\delta}\frac{1}{(a_k-a_i)^\delta} \\ &= \sum_{\{i:\, a_i<a_k\}}\sum_{q=\lceil a_i\rceil}^{\lceil a_k\rceil-1}\left[\frac{1}{(\underline{a}_q-a_i)^\delta}+\frac{1}{(\delta-1)(\underline{a}_q-a_i)^{\delta-1}}\right] \\ &\quad \times \frac{1}{(a_k-a_i)^\delta} \\ &\leq \sum_{\{i:\, a_i<a_k\}}\left[\sum_{l=1}^{\infty}\frac{1}{l^\delta}+\frac{1}{\delta-1}\sum_{l=1}^{\infty}\frac{1}{l^{\delta-1}}\right]\frac{1}{(a_k-a_i)^\delta} \\ &\leq \frac{\delta^3-\delta^2-\delta}{(\delta-1)^2(\delta-2)} \quad (27) \\ &= \frac{c_2(\gamma,\delta,\rho_{\min}){\rho_{\min}}^{2\delta-1}}{2\log e} \quad (28)\end{aligned}$$

where we have used the fact that for any real $a > 0$ and $\beta > 1$

$$\sum_{l=0}^{+\infty}\frac{1}{(l+a)^\beta} \leq \frac{1}{a^\beta} + \int_0^\infty \frac{1}{(a+x)^\beta}dx \leq \frac{\beta-1+a}{(\beta-1)a^\beta}. \quad (29)$$

Hence, (24) is established.

Thus, (21) follows. Similarly, we can prove

$$\sum_{q=-\infty}^{+\infty} R_{\mathcal{N}_q^+} \leq \frac{c_2(\gamma,\delta,\rho_{\min})}{4\rho_{\min}\sigma^2} P_{\text{total}}. \quad (30)$$

Finally, (7) follows from (20), (21) and (30).

Next we consider the case $\gamma > 0$. It is easy to see from the above that we only need to prove

$$\frac{1}{T}\sum_{t=1}^{T}\sum_{q=-\infty}^{+\infty}\sum_{i\in\mathcal{N}_q^-} E\left(\sum_{j\notin\mathcal{N}_q^-}\frac{e^{-\gamma(a_j-a_i)\rho_{\min}}X_j(t)}{(a_j-a_i)^\delta}\right)^2 \leq \frac{c_2(\gamma,\delta,\rho_{\min}){\rho_{\min}}^{2\delta-1}}{2\log e} P_{\text{total}}$$

which can be easily established since for any $k \in \mathcal{N}$

$$\sum_{q=-\infty}^{\lceil a_k \rceil -1} \sum_{i \in \mathcal{N}_q^-} \sum_{j \notin \mathcal{N}_q^-} \frac{e^{-\gamma(a_j-a_i)\rho_{\min}} e^{-\gamma(a_k-a_i)\rho_{\min}}}{(a_j-a_i)^\delta (a_k-a_i)^\delta} \le \frac{e^{-2\gamma\rho_{\min}}}{(1-e^{-\gamma\rho_{\min}})^2(1-e^{-2\gamma\rho_{\min}})} = \frac{c_2(\gamma,\delta,\rho_{\min})\rho_{\min}{}^{2\delta-1}}{2\log e}.$$

This completes the proof of Theorem 3.3. □

Now we turn to the planar case.

Proof of Theorem 3.1: The proof is similar to that of Theorem 3.3. Hence, we only mention the differences here.

Consider first the case $\gamma = 0$ and $\delta > 3$.

Let $(\frac{a_i \rho_{\min}}{2}, \frac{b_i \rho_{\min}}{2})$ denote the coordinates of node i. First, Lemma 4.1 is applied to the following four classes of subsets: for any $q \in \mathbb{Z}$

$$\mathcal{N}_{q,\infty}^- = \{i \in \mathcal{N} : -\infty < a_i \le q, -\infty < b_i < +\infty\} \quad (31)$$
$$\mathcal{N}_{q,\infty}^+ = \{i \in \mathcal{N} : q \le a_i < +\infty, -\infty < b_i < +\infty\}$$
$$\mathcal{N}_{\infty,q}^- = \{i \in \mathcal{N} : -\infty < a_i < +\infty, -\infty < b_i \le q\}$$
$$\mathcal{N}_{\infty,q}^+ = \{i \in \mathcal{N} : -\infty < a_i < +\infty, q \le b_i < +\infty\}.$$

For example, for the class (31), we have

$$\frac{2\sigma^2}{\log e} \cdot R_{\mathcal{N}_{q,\infty}^-} \le \liminf_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} \sum_{i \in \mathcal{N}_{q,\infty}^-} E\left(\sum_{j \notin \mathcal{N}_{q,\infty}^-} \frac{X_j(t)}{\rho_{ij}^\delta}\right)^2 \quad (32)$$

where $R_{\mathcal{N}_{q,\infty}^-}$ is defined similarly to $R_{\mathcal{N}_q^-}$ in the proof of Theorem 3.3. $R_{\mathcal{N}_{q,\infty}^+}$, $R_{\mathcal{N}_{\infty,q}^-}$, and $R_{\mathcal{N}_{\infty,q}^+}$ are also similarly defined.

In the planar case, for any source–destination pair (s_ℓ, d_ℓ) with distance ρ_ℓ between s_ℓ and d_ℓ, it is easy to see that it cuts at least $\lceil 2\rho_\ell/\rho_{\min} \rceil$ subsets among $\mathcal{N}_{q,\infty}^-, \mathcal{N}_{q,\infty}^+, \mathcal{N}_{\infty,q}^-, \mathcal{N}_{\infty,q}^+$, $q \in \mathbb{Z}$. Hence, we have the following inequality:

$$\sum_{\ell=1}^{m} R_\ell \cdot \rho_\ell \le \rho_{\min} \sum_{\ell=1}^{m} R_\ell \cdot \lceil 2\rho_\ell/\rho_{\min} \rceil \le \rho_{\min} \sum_{q=-\infty}^{+\infty} R_{\mathcal{N}_{q,\infty}^-} + \rho_{\min} \sum_{q=-\infty}^{+\infty} R_{\mathcal{N}_{q,\infty}^+} + \rho_{\min} \sum_{q=-\infty}^{+\infty} R_{\mathcal{N}_{\infty,q}^-} + \rho_{\min} \sum_{q=-\infty}^{+\infty} R_{\mathcal{N}_{\infty,q}^+}. \quad (33)$$

Now, we prove that

$$\sum_{q=-\infty}^{+\infty} R_{\mathcal{N}_{q,\infty}^-} \le \frac{c_1(\gamma,\delta,\rho_{\min})}{4\rho_{\min}\sigma^2} P_{\text{total}}. \quad (34)$$

By (32), we only need to show that

$$\frac{1}{T}\sum_{t=1}^{T} \sum_{q=-\infty}^{+\infty} \sum_{i \in \mathcal{N}_{q,\infty}^-} E\left(\sum_{j \notin \mathcal{N}_{q,\infty}^-} \frac{X_j(t)}{\rho_{ij}^\delta}\right)^2 \le \frac{c_1(\gamma,\delta,\rho_{\min})}{2\rho_{\min}\log e} P_{\text{total}} \quad (35)$$

or equivalently

$$\frac{1}{T}\sum_{t=1}^{T} \sum_{q=-\infty}^{+\infty} \sum_{i \in \mathcal{N}_{q,\infty}^-} E\left(\sum_{j \notin \mathcal{N}_{q,\infty}^-} \frac{X_j(t)}{[(a_j-a_i)^2+(b_j-b_i)^2]^{\delta/2}(\rho_{\min}/2)^\delta}\right)^2 \le \frac{c_1(\gamma,\delta,\rho_{\min})}{2\rho_{\min}\log e} P_{\text{total}} \quad (36)$$

with $X_j(t)$ satisfying the total power constraint

$$\frac{1}{T}\sum_{t=1}^{T}\sum_{j \in \mathcal{N}} X_j^2(t) \le P_{\text{total}} \quad \text{a.s.} \quad (37)$$

The intuition behind the inequality (36) is that the summation of the received powers is upper-bounded by the transmitted power.

We now establish (36) for the case where $\delta > 3$. By (37), for $\delta > 3$, we only need to prove that for any t

$$\sum_{q=-\infty}^{+\infty} \sum_{i \in \mathcal{N}_{q,\infty}^-} \left(\sum_{j \notin \mathcal{N}_{q,\infty}^-} \frac{X_j(t)}{[(a_j-a_i)^2+(b_j-b_i)^2]^{\delta/2}}\right)^2 \le \frac{c_1(\gamma,\delta,\rho_{\min})\rho_{\min}{}^{2\delta-1}}{2^{2\delta+1}\log e} P(t) \quad (38)$$

where

$$P(t) := \sum_{i \in \mathcal{N}} X_i^2(t). \quad (39)$$

After replacing each $X_j(t)X_k(t)$ by $\frac{1}{2}(X_j^2(t) + X_k^2(t))$ in the LHS of (38), we only need to prove that the coefficient of any $X_k^2(t)$ is bounded by $\frac{c_1(\gamma,\delta,\rho_{\min})\rho_{\min}{}^{2\delta-1}}{2^{2\delta+1}\log e}$, i.e., for any $k \in \mathcal{N}$

$$\sum_{q=-\infty}^{\lceil a_k \rceil -1} \sum_{i \in \mathcal{N}_{q,\infty}^-} \left(\sum_{j \notin \mathcal{N}_{q,\infty}^-} \frac{1}{[(a_j-a_i)^2+(b_j-b_i)^2]^{\delta/2}} \times \frac{1}{[(a_k-a_i)^2+(b_k-b_i)^2]^{\delta/2}}\right) \le \frac{c_1(\gamma,\delta,\rho_{\min})\rho_{\min}{}^{2\delta-1}}{2^{2\delta+1}\log e}. \quad (40)$$

Using the fact[3] that for any $d_0 \ge 2$

$$\frac{1}{d_0^\delta} \le \frac{4}{\pi}\int_{-\frac{\pi}{4}}^{\frac{\pi}{4}} \int_0^1 \frac{1}{(d_0^2+r^2-2rd_0\cos\theta)^{\frac{\delta}{2}}} r\,dr\,d\theta \quad (41)$$

since

$$\min_{i \ne j} [(a_j-a_i)^2+(b_j-b_i)^2]^{1/2} \ge 2$$

we have for any $i \in \mathcal{N}_{q,\infty}^-$, $\delta > 2$

$$\sum_{j \notin \mathcal{N}_{q,\infty}^-} \frac{1}{[(a_j-a_i)^2+(b_j-b_i)^2]^{\delta/2}}$$

[3] Consider a triangle with d_0 and r as the lengths of two sides with an angle θ between them. Then the third side has a length of $(d_0^2+r^2-2rd_0\cos\theta)^{1/2}$ by the triangle formula. When $d_0 \ge 2$, $\theta \in [-\frac{\pi}{4}, \frac{\pi}{4}]$ and $0 \le r \le 1$, the length of the third side is no more than d_0.

$$\leq \frac{4}{\pi}\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}}\int_{(\underline{a}_q-a_i+1)\vee 1}^{\infty}\frac{1}{x^\delta}xdxd\theta$$
$$\leq \frac{4}{(\delta-2)[(\underline{a}_q-a_i+1)\vee 1]^{\delta-2}}$$

where

$$\underline{a}_q := \min_{j\notin\mathcal{N}_{q,\infty}^-} a_j \quad \text{and} \quad x\vee y := \max\{x,y\}.$$

Then we have for $\delta > 3$

LHS of (40)

$$\leq \sum_{q=-\infty}^{\lceil a_k\rceil-1}\sum_{i\in\mathcal{N}_{q,\infty}^-}\frac{4}{(\delta-2)[(\underline{a}_q-a_i+1)\vee 1]^{\delta-2}} \times \frac{1}{[(a_k-a_i)^2+(b_k-b_i)^2]^{\delta/2}}$$
$$\leq \sum_{\{i:\,a_i<a_k\}}\sum_{q=\lceil a_i\rceil}^{\lceil a_k\rceil-1}\frac{4}{(\delta-2)[(\underline{a}_q-a_i+1)\vee 1]^{\delta-2}} \times \frac{1}{[(a_k-a_i)^2+(b_k-b_i)^2]^{\delta/2}}$$
$$\leq \sum_{\{i:\,a_i<a_k\}}\left(\frac{8}{\delta-2}+\frac{4}{\delta-3}\right)\frac{1}{[(a_k-a_i)^2+(b_k-b_i)^2]^{\delta/2}}$$
$$\leq \left(\frac{8}{\delta-2}+\frac{4}{\delta-3}\right)\frac{4}{\pi}\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}}\int_1^{\infty}\frac{1}{x^\delta}xdxd\theta$$
$$\leq \left(\frac{8}{\delta-2}+\frac{4}{\delta-3}\right)\frac{4}{\delta-2}$$
$$= \frac{c_1(\gamma,\delta,\rho_{\min}){\rho_{\min}}^{2\delta-1}}{2^{2\delta+1}\log e}.$$

Hence, (38) is proved. Thus, (34) follows. The remaining arguments are similar to the proof of Theorem 3.3.

Next we consider the case $\gamma > 0$. Similar to the linear case, we only need to show that for any $k \in \mathcal{N}$

$$\sum_{q=-\infty}^{\lceil a_k\rceil-1}\sum_{i\in\mathcal{N}_{q,\infty}^-}\left(\sum_{j\notin\mathcal{N}_{q,\infty}^-}\frac{e^{-\gamma[(a_j-a_i)^2+(b_j-b_i)^2]^{1/2}\rho_{\min}/2}}{[(a_j-a_i)^2+(b_j-b_i)^2]^{\delta/2}} \times \frac{e^{-\gamma[(a_k-a_i)^2+(b_k-b_i)^2]^{1/2}\rho_{\min}/2}}{[(a_k-a_i)^2+(b_k-b_i)^2]^{\delta/2}}\right)$$
$$\leq \frac{c_1(\gamma,\delta,\rho_{\min}){\rho_{\min}}^{2\delta-1}}{2^{2\delta+1}\log e}.$$

This holds for

$$c_1(\gamma,\delta,\rho_{\min}) = \frac{2^{2\delta+7}\log e}{\gamma^2{\rho_{\min}}^{2\delta+1}}\frac{e^{-\gamma\rho_{\min}/2}(2-e^{-\gamma\rho_{\min}/2})}{(1-e^{-\gamma\rho_{\min}/2})}.$$

This completes the proof of Theorem 3.1. □

The scaling law then follows.

Proofs of Theorems 3.2 and 3.4: The results for the case of individual power P_{ind} follow directly from the case of P_{total} in Theorems 3.3 and 3.1 by noting that $P_{\text{total}} = nP_{\text{ind}}$ is also a constraint. □

C. Multihop and Feasible Rates Under High Attenuation

First we show that $\Omega(n)$ is a feasible transport capacity in regular planar networks.

Proof of Theorem 3.5: We consider a regular planar network where every node ℓ is a source, with its destination d_ℓ chosen as one of its four nearest neighbors.

Each node independently generates its codebook with Gaussian distribution with variance $P = P_{\text{ind}} - \epsilon$, where $\epsilon > 0$. Every destination looks for the signals transmitted by its source, treating all the other transmissions as Gaussian noise. Hence, any rate R_ℓ satisfying the following is achievable for every source–destination pair (ℓ, d_ℓ)

$$R_\ell < S\left(\frac{e^{-2\gamma}P}{c_3(\gamma,\delta)P+\sigma^2}\right)$$

provided $c_3(\gamma,\delta)P$ is an upper bound on the interference, i.e.,

$$c_3(\gamma,\delta)P \geq \sum_{\substack{i\in\mathcal{N}\\ i\neq \ell,d_\ell}}\frac{e^{-2\gamma\rho_{id_\ell}}P}{\rho_{id_\ell}^{2\delta}}. \tag{42}$$

We now show this bound to be true irrespective of the number of nodes n in $\mathcal{N}$.

For the case $\gamma = 0$ and $\delta > 1$, this follows from the summability of the RHS of (42) for $\delta > 1$, since, irrespective of the number n of nodes in $\mathcal{N}$

$$\sum_{\substack{i\in\mathcal{N}\\ i\neq \ell,d_\ell}}\frac{P}{\rho_{id_\ell}^{2\delta}}$$
$$\leq 4\times\left(2\sum_{i=1}^{\infty}\frac{1}{i^{2\delta}}+\int_1^{\infty}\int_0^{\infty}\frac{1}{(x^2+y^2)^\delta}dxdy\right)P$$
$$\leq 4\times\left(2\cdot\frac{2\delta}{2\delta-1}+\frac{\pi}{4\delta-4}\right)P$$
$$\leq \frac{16\delta^2+(2\pi-16)\delta-\pi}{(\delta-1)(2\delta-1)}P$$
$$\leq c_3(\gamma,\delta)P.$$

Next consider the case $\gamma > 0$. Then

$$\sum_{\substack{i\in\mathcal{N}\\ i\neq \ell,d_\ell}}\frac{e^{-2\gamma\rho_{id_\ell}}P}{\rho_{id_\ell}^{2\delta}}$$
$$\leq 4\times\left(2\sum_{i=1}^{\infty}e^{-2\gamma i}+\int_1^{\infty}\int_0^{\infty}e^{-2\gamma(x^2+y^2)^{1/2}}dxdy\right)P$$
$$\leq 4\times\left(\frac{2e^{-2\gamma}}{1-e^{-2\gamma}}+\frac{e^{-2\gamma}}{2\gamma}\right)P$$
$$\leq \frac{4(1+4\gamma)e^{-2\gamma}-4e^{-4\gamma}}{2\gamma(1-e^{-2\gamma})}P$$
$$\leq c_3(\gamma,\delta)P.$$

Hence, the total achievable transport capacity is

$$n\cdot S\left(\frac{e^{-2\gamma}P}{c_3(\gamma,\delta)P+\sigma^2}\right)$$

for every $P < P_{\text{ind}}$, establishing the result of Theorem 3.5. □

The feasibility of a rate vector under multipath routing which can load balance, without requiring too long hops, is fairly straightforward.

Proof of Theorem 3.7: Note that the maximum distance that a signal has to travel on any hop is $\bar{\rho}$. This can used to lower bound the received signal strength. Moreover, we can prove that the total interference at any node j is bounded as follows:

Using the fact (similar to (41)) that for any $d_0 \geq \rho_{\min}$

$$\frac{1}{d_0^{2\delta}} \leq \frac{16}{\pi {\rho_{\min}}^2} \int_{-\frac{\pi}{4}}^{\frac{\pi}{4}} \int_0^{\frac{\rho_{\min}}{2}} \frac{1}{(d_0^2 + r^2 - 2rd_0 \cos\theta)^\delta} r dr d\theta$$

we have for $\gamma = 0$ and $\delta > 1$

$$\sum_{\substack{i \in \mathcal{N} \\ i \neq j}} \frac{P}{\rho_{ij}^{2\delta}} \leq \frac{16}{\pi {\rho_{\min}}^2} \int_{-\pi}^{\pi} \int_{\frac{\rho_{\min}}{2}}^{\infty} \frac{P}{x^{2\delta}} x dx d\theta = \frac{2^{2+2\delta}}{{\rho_{\min}}^{2\delta}(\delta - 1)} P$$

and for $\gamma > 0$

$$\sum_{\substack{i \in \mathcal{N} \\ i \neq j}} \frac{e^{-2\gamma\rho_{ij}} P}{\rho_{ij}^{2\delta}} \leq \frac{16P}{\pi {\rho_{\min}}^2} \int_{-\pi}^{\pi} \int_{\frac{\rho_{\min}}{2}}^{\infty} \frac{e^{-2\gamma x}}{x^{2\delta}} x dx d\theta \leq \frac{2^{4+2\delta} P}{{\rho_{\min}}^{1+2\delta}} \int_{\frac{\rho_{\min}}{2}}^{\infty} e^{-2\gamma x} dx = \frac{2^{3+2\delta} e^{-\gamma \rho_{\min}}}{\gamma {\rho_{\min}}^{1+2\delta}} P.$$

The rest of the proof follows as in the proof of Theorem 3.5. □

Now we turn to the random case for regular planar networks.

Proof of Theorem 3.6: Suppose that n source–destination pairs are randomly chosen as follows: Choose $2n$ points, $(a_1, a_2, \ldots, a_n)$ and $(b_1, b_2, \ldots, b_n)$, randomly (uniformly i.i.d.) in the domain of the regular planar network, which is a square of side $\sqrt{n} - 1$. Now let s_ℓ and d_ℓ be the nodes (which are located only at integral coordinates (i, j) with $1 \leq i, j \leq \sqrt{n}$) nearest to a_ℓ and b_ℓ, respectively. Then the n source–destination pairs are (s_ℓ, d_ℓ).

To route the traffic, we follow the scheme of [8]. Construct an axis parallel mini-square of side length 1 centered around each node. These mini-squares will play the role of the "cells" considered in [8]. Packets for a source–destination pair (s_ℓ, d_ℓ) will be relayed from node to node in the order that the straight line joining a_ℓ and b_ℓ intersects the mini-squares. (Diagonal hops occur with probability zero). Thus, each straight line (a_i, b_i) passing through a mini-square means that the node in the mini-square (which by construction has exactly one node in it) has to relay that route's traffic to one of its four nearest neighbors.

Note that the straight lines $\{(a_\ell, b_\ell) : 1 \leq \ell \leq n\}$ are i.i.d. (indeed, this is the reason for resorting to this construction of source–destination pairs). Also, the probability that a straight line passes through a given mini-square is less than $c\sqrt{\frac{\log n}{n}}$, for some constant c. Using the dimension bounds in [8] (since each square can be enclosed in a circumscribing circle) in the uniform weak law of large numbers of Vapnik–Chervonenkis [27], it follows that

$$\text{Prob}(\text{Every mini-square has no more than } c'\sqrt{n \log n} \text{ straight lines passing through it}) \to 1, \quad \text{as } n \to \infty.$$

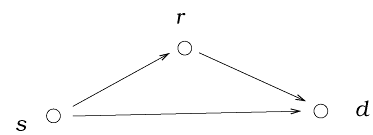

Fig. 10. The single-relay channel.

Now suppose that every source–destination pair carries a traffic of rate less than $\frac{R_{\min}}{c'\sqrt{n \log n}}$. Then

$$\text{Prob}(\text{Every node needs to send no more than rate } R_{\min} \text{ to one of its four nearest neighbors}) \to 1, \quad \text{as } n \to \infty.$$

However, as already shown in the proof of Theorem 3.5, in a regular planar network, every node can indeed send at a fixed positive rate $R_{\min} > 0$ to any one of its four nearest neighbors.

Thus, a rate of $\frac{R_{\min}}{c'\sqrt{n \log n}}$ can indeed be supported for all the source–destination pairs simultaneously, with probability approaching one as $n \to \infty$.

Finally, since there are n sources, and the mean distance between a source and its destination is $\Omega(\sqrt{n})$, it follows that a transport capacity of $\Omega(\frac{n}{\sqrt{\log n}})$ is supported, again with probability approaching 1 as $n \to \infty$. □

D. The Gaussian Multiple-Relay Channel and Coherent Relaying With Interference Subtraction

We now address the channel considered in Theorems 3.11 and 3.12, featuring a multitude of relays. Each stage of relay can be either one node or a group of nodes.

To see the basic idea of our coding scheme, it is enough to consider the simplest single-relay channel consisting of a source node s, relay node r, and destination node d, as in Fig. 10. Let α_{sr}, α_{sd}, and α_{rd} denote the corresponding signal attenuation factors.

The whole transmission time is divided equally into a sequence of blocks. In each block (except the first and the last), node s divides its power P_s into two parts: θP_s and $(1-\theta)P_s$, $0 \leq \theta \leq 1$, for different purposes. The part θP_s, used to inform node r of its intention for the next block, can achieve any rate R (by Shannon's formula) satisfying

$$R < S\left(\frac{\alpha_{\text{sr}}^2 \theta P_s}{\sigma^2}\right). \tag{43}$$

The other part $(1-\theta)P_s$ is used to cooperate with the total power P_r of the relay node r to coherently transmit signals to node d. This cooperation is possible since node r has gotten to know the intention of node s from the previous block. Then what node d receives is the addition of three components: i) the signal consisting of the coherent cooperation of s and r with power $(\alpha_{sd}\sqrt{(1-\theta)P_s} + \alpha_{rd}\sqrt{P_r})^2$; ii) the signal sent by s intended mainly for r, with power $\alpha_{sd}^2 \theta P_s$; iii) the noise with power σ^2. Now, for the decoding at node d, the following procedure is used (the rigorous decoding argument uses jointly typical sequences;

here, we only provide the intuition): Node d decodes at the end of each block *simultaneously* taking the first part in this block and the second part in the previous block as signal (note that they represent the same information), with the first part in the previous block deducted (this is done after the decoding at the end of the previous block). Then the following rate is achievable:

$$R < S\left(\frac{\left(\alpha_{sd}\sqrt{(1-\theta)P_s}+\alpha_{rd}\sqrt{P_r}\right)^2}{\alpha_{sd}^2\theta P_s+\sigma^2}\right)+S\left(\frac{\alpha_{sd}^2\theta P_s}{\sigma^2}\right) = S\left(\frac{\alpha_{sd}^2 P_s+\alpha_{rd}^2 P_r+2\alpha_{sd}\alpha_{rd}\sqrt{(1-\theta)P_sP_r}}{\sigma^2}\right).$$

Together with the constraint (43), this leads to the following achievable rate:

$$R < \max_{0\le\theta\le 1}\min\left\{S\left(\frac{\alpha_{\text{sr}}^2\theta P_s}{\sigma^2}\right), S\left(\frac{\alpha_{sd}^2 P_s+\alpha_{rd}^2 P_r+2\alpha_{sd}\alpha_{rd}\sqrt{(1-\theta)P_sP_r}}{\sigma^2}\right)\right\}. \quad (44)$$

We should remark that the above coding–decoding scheme is different from that of [21], though we still use a block-coding argument. Unlike earlier, see [22], we do not partition the 2^{TR} messages into 2^{TR_0} cells, where the destination first decides in which cell the message lies, and then determines which exact message it is based on the help from the relay in the next block. We combine these two steps into one: The destination waits and does the decoding only when it has received all the related signals, and determines the message directly once and for all. We will see that our scheme is simpler and avoids some inconvenient techniques (e.g., Slepian–Wolf partitioning), although it coincides with [21] in giving the same achievable rate formula for the one-relay case. More importantly, we will also see that this new coding scheme is easier to extend to the multilevel relay case and generally achieves higher rates than those proved in [26].

We use some standard results for jointly typical sequences which we gather together here; see [22, Secs. 8.6, 9.2, and 10.1].

Definition 4.2: The set $A_\epsilon^{(T)}$ of jointly typical sequences $\{(x^T, y^T)\}$ with respect to the joint density function $f(x,y)$ is the set of T sequences with empirical entropies ϵ close to the true entropies, i.e.,

$$A_\epsilon^{(T)} = \left\{(x^T,y^T)\in\mathbb{R}^T\times\mathbb{R}^T : \left|-\frac{1}{T}\log f(x^T)-h(X)\right|<\epsilon, \left|-\frac{1}{T}\log f(y^T)-h(Y)\right|<\epsilon, \left|-\frac{1}{T}\log f(x^T,y^T)-h(X,Y)\right|<\epsilon\right\}$$

where

$$f(x^T,y^T)=\prod_{i=1}^T f(x_i,y_i).$$

Definition 4.3: $A_\epsilon^{(T)}(P,N)$ denotes the set $A_\epsilon^{(T)}$ with respect to the joint density function

$$f(x,y) = g_P(x)g_N(y-x) = \frac{1}{\sqrt{2\pi P}}\exp\left(-\frac{x^2}{2P}\right)\cdot\frac{1}{\sqrt{2\pi N}}\exp\left(-\frac{(y-x)^2}{2N}\right).$$

Lemma 4.2: Let (X^T, Y^T) be sequences of length T drawn i.i.d. according to

$$f(x^T,y^T)=\prod_{i=1}^T g_P(x_i)g_N(y_i-x_i).$$

Then

1) $\text{Prob}((X^T,Y^T)\in A_\epsilon^{(T)}(P,N))\to 1$ as $T\to\infty$.
2)
$$\int_{(x^T,y^T)\in A_\epsilon^{(T)}(P,N)} dx^T dy^T \le 2^{T(h(X,Y)+\epsilon)}$$
where $h(X,Y)$ denotes the differential entropy.
3) If
$$(\tilde{X}^T,\tilde{Y}^T)\sim\prod_{i=1}^T g_P(x_i)g_{P+N}(y_i)$$
i.e., $\tilde{X}^T$ and $\tilde{Y}^T$ are independent with the same marginals as (X^T,Y^T), then
$$\text{Prob}((\tilde{X}^T,\tilde{Y}^T)\in A_\epsilon^{(T)}(P,N))\le 2^{-T(S(\frac{P}{N})-3\epsilon)}.$$
Also, for sufficiently large T
$$\text{Prob}((\tilde{X}^T,\tilde{Y}^T)\in A_\epsilon^{(T)}(P,N))\ge(1-\epsilon)2^{-T(S(\frac{P}{N})+3\epsilon)}.$$

Proof of Theorem 3.11: We consider B blocks of transmission, each of T transmission slots. A sequence of $B-M+1$ indexes, $w_b\in\{1,\dots,2^{TR}\}, b=1,2,\dots,B-M+1$, will be sent over in TB transmission slots. (Note that as $B\to\infty$, the rate $TR(B-M+1)/TB$ is arbitrarily close to R for any T.)

Generation of Codebooks: Randomly generate M^2 matrices $\mathcal{X}_k(b_0)$ for $k=1,\dots,M$, and $b_0=1,\dots,M$, each of size $2^{TR}\times T$, with every element independently chosen with Gaussian distribution $N(0,1-\varepsilon_1)$. These are the randomly generated codebooks. In the same block, different nodes need use independent codewords from different codebooks. Since it takes M blocks to transmit one complete message, every node is assigned M independent codebooks. The total number of codebooks needed is M^2 since there are M nodes. The M^2 matrices are revealed to all the $M+1$ nodes. Let

$$\mathcal{X}_k(b) := \mathcal{X}_k(b \bmod M), \qquad b=1,2,\dots,B.$$

Denote by $x_k(b,w)$ the wth row of the matrix $\mathcal{X}_k(b)$, for $w\in\{1,\dots,2^{TR}\}$. It denotes the wth codeword.

Encoding: At the beginning of each block $b\in\{1,\dots,B\}$, every node $i\in\{0,\dots,M-1\}$ has estimates (see the sequel)

$\widehat{w}_{b-k+1,i}$ of w_{b-k+1}, $k \geq i+1$ (with $\widehat{w}_{b-k+1,0} = w_{b-k+1}$) and sends the following vector of length T in the block:

$$\vec{X}_i(b) := \sum_{k=i+1}^{M} \sqrt{P_{ik}} x_k(b, \widehat{w}_{b-k+1,i}).$$

We set

$$\widehat{w}_{b_1,i} := w_{b_1} := 0 \text{ for any } b_1 \leq 0, \text{ and } x_k(b,0) := 0. \quad (45)$$

Every node $k \in \{1, \ldots, M\}$ thus receives the vector

$$\begin{aligned}\vec{Y}_k(b) &= \sum_{\substack{0 \leq i \leq M-1 \\ i \neq k}} \alpha_{ik} \vec{X}_i(b) + \vec{Z}_k(b) \\ &= \sum_{\substack{0 \leq i \leq M-1 \\ i \neq k}} \sum_{l=i+1}^{M} \alpha_{ik} \sqrt{P_{il}} x_l(b, \widehat{w}_{b-l+1,i}) + \vec{Z}_k(b) \\ &= \left(\sum_{l=1}^{k} \sum_{i=0}^{l-1} + \sum_{l=k+1}^{M} \sum_{\substack{0 \leq i \leq l-1 \\ i \neq k}} \right) \alpha_{ik} \sqrt{P_{il}} x_l(b, \widehat{w}_{b-l+1,i}) \\ &\quad + \vec{Z}_k(b). \end{aligned} \quad (46)$$

Let

$$\widehat{\vec{Y}}_k(b) := \vec{Y}_k(b) - \sum_{l=k+1}^{M} \sum_{\substack{0 \leq i \leq l-1 \\ i \neq k}} \alpha_{ik} \sqrt{P_{il}}\, x_l(b, \widehat{w}_{b-l+1,k}). \quad (47)$$

As we show in the sequel, this will serve as an estimate by node k of

$$\sum_{l=1}^{k} \sum_{i=0}^{l-1} \alpha_{ik} \sqrt{P_{il}} x_l(b, \widehat{w}_{b-l+1,i}).$$

Decoding: At the *end* of each block $b \in \{1, \ldots, B\}$, every node $k \in \{1, \ldots, M\}$ (for $b-k+1 \geq 1$) declares $\widehat{w}_{b-k+1,k} = w$ if w is the unique value in $\{1, \ldots, 2^{TR}\}$ such that in *all* the blocks $b-j, j = 0, 1, \ldots, k-1$

$$\begin{aligned}&\left\{ \sum_{i=0}^{k-j-1} \alpha_{ik} \sqrt{P_{i,k-j}}\, x_{k-j}(b-j, w), \right. \\ &\quad \left. \widehat{\vec{Y}}_k(b-j) - \sum_{l=k-j+1}^{k} \sum_{i=0}^{l-1} \alpha_{ik} \sqrt{P_{il}}\, x_l(b-j, \widehat{w}_{b-j-l+1,k}) \right\} \\ &\in A_\epsilon^{(T)}(\bar{P}_{k,j}, N_{k,j}) \end{aligned} \quad (48)$$

where

$$\bar{P}_{k,j} := \left(\sum_{i=0}^{k-j-1} \alpha_{ik} \sqrt{P_{i,k-j}} \right)^2 (1-\varepsilon_1)$$

$$N_{k,j} := \sum_{l=1}^{k-j-1} \left(\sum_{i=0}^{l-1} \alpha_{ik} \sqrt{P_{il}} \right)^2 (1-\varepsilon_1) + \sigma^2.$$

Otherwise, if a unique w as above does not exist, an error is declared and $\widehat{w}_{b-k+1,k}$ is set to 0.

Analysis of Probability of Error: Denote the event that no decoding error is made in the first b blocks by

$$A_c(b) := \big\{ \widehat{w}_{b_1-k+1,k} = w_{b_1-k+1}, \text{for all } b_1 \in \{1, \ldots, b\} \text{ and } k \in \{1, \ldots, M\} \big\}$$

and let its probability be $P_c(b) := \mathrm{Prob}(A_c(b))$, with $P_c(0) := 1$.

Then the probability that some decoding error is made at some node $k \in \{1, \ldots, M\}$ in some block $b \in \{1, \ldots, B\}$ is

$$\begin{aligned} P_e &:= \mathrm{Prob}\big(\widehat{w}_{b-k+1,k} \neq w_{b-k+1}, \\ &\qquad \text{for some } k \in \{1, \ldots, M\}, b \in \{1, \ldots, B\}\big) \\ &= \sum_{b=1}^{B} \mathrm{Prob}\big(\widehat{w}_{b-k+1,k} \neq w_{b-k+1} \\ &\qquad \text{for some } k \in \{1, \ldots, M\} | A_c(b-1)\big) \cdot P_c(b-1) \\ &\leq \sum_{b=1}^{B} \sum_{k=1}^{M} \mathrm{Prob}\big(\widehat{w}_{b-k+1,k} \neq w_{b-k+1} | A_c(b-1)\big) \\ &\qquad \cdot P_c(b-1) \\ &= \sum_{b=1}^{B} \sum_{k=1}^{M} P_e(b,k) \cdot P_c(b-1) \end{aligned} \quad (49)$$

where

$$P_e(b,k) := \mathrm{Prob}(\widehat{w}_{b-k+1,k} \neq w_{b-k+1} | A_c(b-1)).$$

Hence, $P_e(b,k)$ is the probability that a decoding error happens at node k in block b, conditioned on the event that no decoding error was made in the former $b-1$ blocks.

Next, we calculate $P_e(b,k)$. Since $A_c(b-1)$ is presumed to hold, for any node k we have

$$\widehat{w}_{b_1-k+1,k} = w_{b_1-k+1}, \qquad \text{for } k \leq b_1 \leq b-1.$$

Hence, noting (45), $\widehat{w}_{b_2,k} = w_{b_2}$ whenever $b_2 + k \leq b$. Then, by (46) and (47), for all $b-j$ with $j \geq 0$

$$\begin{aligned}&\vec{Y}_k(b-j) \\ &= \left(\sum_{l=1}^{k} \sum_{i=0}^{l-1} + \sum_{l=k+1}^{M} \sum_{\substack{0 \leq i \leq l-1 \\ i \neq k}} \right) \alpha_{ik} \sqrt{P_{il}}\, x_l(b-j, w_{b-j-l+1}) \\ &\quad + \vec{Z}_k(b-j)\end{aligned}$$

and

$$\begin{aligned}&\widehat{\vec{Y}}_k(b-j) \\ &\quad = \vec{Y}_k(b-j) - \sum_{l=k+1}^{M} \sum_{\substack{0 \leq i \leq l-1 \\ i \neq k}} \alpha_{ik} \sqrt{P_{il}}\, x_l(b-j, w_{b-j-l+1}) \\ &\quad = \sum_{l=1}^{k} \sum_{i=0}^{l-1} \alpha_{ik} \sqrt{P_{il}}\, x_l(b-j, w_{b-j-l+1}) + \vec{Z}_k(b-j).\end{aligned}$$

So

$$\begin{aligned}&\widehat{\vec{Y}}_k(b-j) - \sum_{l=k-j+1}^{k} \sum_{i=0}^{l-1} \alpha_{ik} \sqrt{P_{il}}\, x_l(b-j, \widehat{w}_{b-j-l+1,k}) \\ &\qquad = \sum_{l=1}^{k-j} \sum_{i=0}^{l-1} \alpha_{ik} \sqrt{P_{il}}\, x_l(b-j, w_{b-j-l+1}) + \vec{Z}_k(b-j).\end{aligned}$$

Hence, under the condition $A_c(b-1)$, the decoding rule (48) is equivalent to the following: Each node $k \in \{1,\ldots,M\}$ (when $b-k+1 \geq 1$) declares $\widehat{w}_{b-k+1,k} = w$ if w is the unique value in $\{1,\ldots,2^{TR}\}$ such that in all the blocks $b-j$, for $j = 0,1,\ldots,k-1$

$$\left\{ \sum_{i=0}^{k-j-1} \alpha_{ik}\sqrt{P_{i,k-j}}x_{k-j}(b-j,w), \sum_{l=1}^{k-j}\sum_{i=0}^{l-1} \alpha_{ik}\sqrt{P_{il}}x_l(b-j,w_{b-j-l+1}) + \vec{Z}_k(b-j) \right\} \in A_\epsilon^{(T)}(\bar{P}_{k,j}, N_{k,j}). \tag{50}$$

Let

$$\mathcal{W}_{b,k,j} := \{w \in \{1,\ldots,2^{TR}\} : w \text{ satisfies (50)}\}$$
$$\mathcal{W}_{b,k} := \bigcap_{j=0}^{k-1} \mathcal{W}_{b,k,j}.$$

Then, $P_e(b,k)$ is the probability that $w_{b-k+1} \notin \mathcal{W}_{b,k}$, or some $w(\neq w_{b-k+1}) \in \mathcal{W}_{b,k}$, conditioned on the event that no decoding error was made in the former $b-1$ blocks. Thus,

$$\begin{aligned} P_e(b,k) &= \text{Prob}\big(w_{b-k+1} \notin \mathcal{W}_{b,k}, \text{ or } w \in \mathcal{W}_{b,k} \text{ for some } w \neq w_{b-k+1} | A_c(b-1)\big) \\ &\leq \text{Prob}\big(w_{b-k+1} \notin \mathcal{W}_{b,k} | A_c(b-1)\big) + \text{Prob}\big(w \in \mathcal{W}_{b,k} \text{ for some } w \neq w_{b-k+1} | A_c(b-1)\big) \\ &\leq \frac{\text{Prob}(w_{b-k+1} \notin \mathcal{W}_{b,k})}{P_c(b-1)} + \frac{\text{Prob}(w \in \mathcal{W}_{b,k} \text{ for some } w \neq w_{b-k+1})}{P_c(b-1)}. \end{aligned}$$

Hence, by (49)

$$P_e \leq \sum_{b=1}^{B}\sum_{k=1}^{M} [\text{Prob}(w_{b-k+1} \notin \mathcal{W}_{b,k}) + \text{Prob}(w \in \mathcal{W}_{b,k} \text{ for some } w \neq w_{b-k+1})]. \tag{51}$$

Now, by Lemma 4.2, for T large enough, we have for $j = 0,1,\ldots,k-1$

$$\text{Prob}(w_{b-k+1} \notin \mathcal{W}_{b,k,j}) \leq \epsilon$$

and for any $w' \neq w_{b-k+1}$

$$\text{Prob}(w' \in \mathcal{W}_{b,k,j}) \leq 2^{-T(S(\frac{\bar{P}_{k,j}}{N_{k,j}})-3\epsilon)}.$$

Hence,

$$\begin{aligned} \text{Prob}(w_{b-k+1} \notin \mathcal{W}_{b,k}) &\leq \sum_{j=0}^{k-1} \text{Prob}(w_{b-k+1} \notin \mathcal{W}_{b,k,j}) \\ &\leq \sum_{j=0}^{k-1} \epsilon = k\epsilon \leq M\epsilon \end{aligned} \tag{52}$$

and

$$\begin{aligned} \text{Prob}(w \in \mathcal{W}_{b,k} \text{ for some } w \neq w_{b-k+1}) &\leq \sum_{\substack{w' \in \{1,\ldots,2^{TR}\} \\ w' \neq w_{b-k+1}}} \text{Prob}(w' \in \mathcal{W}_{b,k}) \\ &= \sum_{\substack{w' \in \{1,\ldots,2^{TR}\} \\ w' \neq w_{b-k+1}}} \prod_{j=0}^{k-1} \text{Prob}(w' \in \mathcal{W}_{b,k,j}) \\ &\leq (2^{TR}-1) \prod_{j=0}^{k-1} 2^{-T(S(\frac{\bar{P}_{k,j}}{N_{k,j}})-3\epsilon)} \\ &= (2^{TR}-1) 2^{-T(S(\frac{\bar{P}_k}{\sigma^2})-3k\epsilon)}. \end{aligned} \tag{53}$$

The equality (53) follows from the independence of the rows $x_k(b,w)$ and also the transmissions w_b, the fact that

$$\sum_{j=0}^{k-1} S\left(\frac{\bar{P}_{k,j}}{N_{k,j}}\right) = S\left(\frac{\bar{P}_k}{\sigma^2}\right)$$

as well as

$$\bar{P}_k = \sum_{l=1}^{k}\left(\sum_{i=0}^{l-1} \alpha_{ik}\sqrt{P_{il}}\right)^2 (1-\varepsilon_1).$$

For any R satisfying (10), by choosing T large enough, we can make ε_1 and ϵ small enough such that for any $\varepsilon_2 > 0$

$$\text{Prob}(w \in \mathcal{W}_{b,k} \text{ for some } w \neq w_{b-k+1}) \leq (2^{TR}-1) 2^{-T(S(\frac{\bar{P}_k}{\sigma^2})-3k\epsilon)} < \varepsilon_2. \tag{54}$$

Hence, by (51), (52), and (54)

$$\begin{aligned} P_e &\leq \sum_{b=1}^{B}\sum_{k=1}^{M} (M\epsilon + \varepsilon_2) \\ &\leq BM^2\epsilon + BM\varepsilon_2 \end{aligned}$$

which can be made arbitrarily small by letting $T \to \infty$. □

Next we address the case of relaying by groups.

Proof of Theorem 3.12: The proof follows similarly to that of Theorem 3.11. The only difference is that now all the n_i nodes in each group $\mathcal{N}_i$ equally share the same power P_{ik} and transmit coherently. We take the maximum attenuation $\alpha_{\mathcal{N}_i\mathcal{N}_j}$ to ensure that every node in each group can successfully do the decoding. □

E. The Low-Attenuation Regime

First we show how unbounded transport capacity can be obtained for fixed total power in linear networks.

Proof of Theorem 3.9: We consider one source–destination pair where the source node is located at 0, and the destination node is located at n. Let the $n-1$ nodes in between, located at $1,2,\ldots,n-1$, be the $n-1$ stages of relay. Then, by Theorem 3.11, the following rate is achievable:

$$R < \min_{1\leq j\leq n} S\left(\frac{1}{\sigma^2}\sum_{k=1}^{j}\left(\sum_{i=0}^{k-1}\frac{\sqrt{P_{ik}}}{(j-i)^\delta}\right)^2\right) \tag{55}$$

with the total power constraint

$$\sum_{k=1}^{n}\sum_{i=0}^{k-1} P_{ik} \le P_{\text{total}}.$$

The intuitive interpretation of P_{ik} is the part of the power used by node i intended directly for node k.

We specifically choose

$$P_{ik} := \frac{P}{(k-i)^{\alpha} k^{\beta}}, \qquad 0 \le i < k \le n \tag{56}$$

where $\alpha > 1, \beta > 1$ are two constants to be determined later, and

$$P := \frac{(\alpha-1)(\beta-1)}{\alpha\beta} P_{\text{total}}. \tag{57}$$

Using (29), it is easy to check that the total power constraint P_{total} holds.

For $3-\alpha-\beta > 0$, we now establish the following lower bound usable in the RHS of (55):

$$\sum_{k=1}^{j}\left(\sum_{i=0}^{k-1}\frac{\sqrt{P}}{(k-i)^{\alpha/2}k^{\beta/2}(j-i)^{\delta}}\right)^2 = \Omega(j^{3-\alpha-\beta-2\delta}). \tag{58}$$

For $3-\alpha-\beta > 0$, we have

$$\begin{aligned}
&\sum_{k=1}^{j}\left(\sum_{i=0}^{k-1}\frac{\sqrt{P}}{(k-i)^{\alpha/2}k^{\beta/2}(j-i)^{\delta}}\right)^2\\
&\ge \frac{P}{j^{2\delta}}\sum_{k=1}^{j}\left(\sum_{i=0}^{k-1}\frac{1}{(k-i)^{\alpha/2}}\right)^2\frac{1}{k^{\beta}}\\
&= \frac{P}{j^{2\delta}}\sum_{k=1}^{j}\left(\frac{1}{k^{\alpha/2}}\sum_{i=0}^{k-1}\frac{1}{(1-i/k)^{\alpha/2}}\right)^2\frac{1}{k^{\beta}}\\
&\ge \frac{P}{j^{2\delta}}\sum_{k=1}^{j}\left(\frac{1}{k^{\alpha/2}}\int_0^{k-1}\frac{1}{(1-x/k)^{\alpha/2}}dx\right)^2\frac{1}{k^{\beta}}\\
&= \frac{P}{j^{2\delta}}\sum_{k=1}^{j}\left(\frac{k}{k^{\alpha/2}}\int_0^{\frac{k-1}{k}}\frac{1}{(1-y)^{\alpha/2}}dy\right)^2\frac{1}{k^{\beta}}\\
&\ge \frac{c_0 P}{j^{2\delta}}\sum_{k=2}^{j}k^{2-\alpha-\beta} \ge \frac{c_0 P}{j^{2\delta}}\int_2^j x^{2-\alpha-\beta}dx \qquad (\text{for } j\ge 2)\\
&= \frac{c_0 P}{j^{2\delta}}\frac{j^{3-\alpha-\beta}-2^{3-\alpha-\beta}}{3-\alpha-\beta}\\
&= \Omega(j^{3-\alpha-\beta-2\delta})
\end{aligned}$$

where $c_0 := \left(\int_0^{1/2}\frac{1}{(1-y)^{\alpha/2}}dy\right)^2 > 0$. This establishes (58).

Now we proceed by analyzing two cases.

Case 1. $\delta < \frac{1}{2}$: In this case, we specifically choose $\alpha > 1$ and $\beta > 1$ such that

$$3-\alpha-\beta-2\delta > 0. \tag{59}$$

Then by (58) and (59), there exists some $\underline{P} > 0$ such that for any j

$$\sum_{k=1}^{j}\left(\sum_{i=0}^{k-1}\frac{\sqrt{P}}{(k-i)^{\alpha/2}k^{\beta/2}(j-i)^{\delta}}\right)^2 \ge \underline{P}.$$

Thus, by (55), for any n, any $R < S(\frac{P}{\sigma^2})$ is achievable. Without loss of generality, this means that any $R < S(\frac{P}{\sigma^2})$ is achievable with power constraint P_{total} for any single source–destination pair. Furthermore, since $\rho_{0,n} = n$, $R\cdot n$ is an achievable network transport with power constraint P_{total}, which tends to infinity as $n \to \infty$.

Case 2. $\frac{1}{2} \le \delta < 1$: In this case, we specifically choose $\alpha > 1$ and $\beta > 1$ such that

$$4-\alpha-\beta-2\delta > 0. \tag{60}$$

Note that $3-\alpha-\beta-2\delta < 0$. Hence, the minimum of (58) over $j = 1, 2, \ldots, n$ is attained at $j = n$. So by (60), we have

$$\begin{aligned}
n \min_{1\le j\le n} S\left(\frac{1}{\sigma^2}\sum_{k=1}^{j}\left(\sum_{i=0}^{k-1}\frac{\sqrt{P}}{(k-i)^{\alpha/2}k^{\beta/2}(j-i)^{\delta}}\right)^2\right)&\\
= \Omega(n^{4-\alpha-\beta-2\delta}) \to \infty, \qquad \text{as } n\to\infty.&
\end{aligned}$$

This means that an arbitrarily large transport capacity is achievable with a fixed total power constraint P_{total}. □

Now we show that unbounded transport capacity can be obtained for fixed total power in planar networks.

Proof of Theorem 3.8: The idea of the proof is similar to that of the linear case in Theorem 3.9. The only difference is that in the planar case there are more nodes to help.

We still consider one source–destination pair where the source node s is located at $(0,0)$ and the destination node d is located at $(r^q, 0)$, with q a positive integer to be determined.

We need the cooperation of $r-1$ groups of relay nodes: Group $\mathcal{N}_i$ consists of n_i nodes in a neighborhood of the node $(i^q, 0)$, for $i = 1, \ldots, r-1$, with $\mathcal{N}_0 = \{s\}$, $n_0 = 1$. Each Group $\mathcal{N}_i$ corresponds to the node i in the linear case: The n_i nodes equally share the power P_{ik} defined in (56) and coherently transmit.

Then by Theorem 3.12, the following rate is achievable:

$$R < \min_{1\le j\le r} S\left(\frac{1}{\sigma^2}\sum_{k=1}^{j}\left(\sum_{i=0}^{k-1}\frac{\sqrt{P_{ik}/n_i}\cdot n_i}{\rho^{\delta}_{\mathcal{N}_i\mathcal{N}_j}}\right)^2\right) \tag{61}$$

where $\rho_{\mathcal{N}_i\mathcal{N}_j}$ is the maximum distance between any node in Group $\mathcal{N}_i$ and any node in Group $\mathcal{N}_j$.

For any $i = 1, 2, \ldots, r-1$, we specifically choose Group $\mathcal{N}_i$ to be the set of nodes:

$$\{(u,v) : i^q \le u \le i^q + i^{q-1} - 1, -i^{q-1} \le v \le i^{q-1}\}.$$

It is easy to check that these groups are disjoint from each other and $n_i > i^{2(q-1)}$. Furthermore, for any $0 \le i < j < r$

$$\rho_{ij} < j^q - i^q + i^{q-1} + j^{q-1} + j^{q-1} < 3j^q.$$

Hence, by (61), the following rate is achievable:

$$R < \min_{1\le j\le r} S\left(\frac{1}{\sigma^2}\sum_{k=1}^{j}\left(\sum_{i=0}^{k-1}\frac{\sqrt{P_{ik}}\cdot i^{q-1}}{3^{\delta}j^{q\delta}}\right)^2\right). \tag{62}$$

Similarly to the linear case, for $1+2q-\alpha-\beta>0$, we can prove the following lower bound usable in the RHS of (62):

$$\sum_{k=1}^{j}\left(\sum_{i=0}^{k-1}\frac{\sqrt{P}\cdot i^{q-1}}{(k-i)^{\alpha/2}k^{\beta/2}3^{\delta}j^{q\delta}}\right)^2=\Omega(j^{1+2q-\alpha-\beta-2q\delta}). \tag{63}$$

For $1+2q-\alpha-\beta>0$, we have

$$\begin{aligned}&\sum_{k=1}^{j}\left(\sum_{i=0}^{k-1}\frac{\sqrt{P}\cdot i^{q-1}}{(k-i)^{\alpha/2}k^{\beta/2}3^{\delta}j^{q\delta}}\right)^2\\
&\geq\frac{P}{3^{2\delta}j^{2q\delta}}\sum_{k=1}^{j}\left(\sum_{i=0}^{k-1}\frac{i^{q-1}}{(k-i)^{\alpha/2}}\right)^2\frac{1}{k^{\beta}}\\
&=\frac{P}{3^{2\delta}j^{2q\delta}}\sum_{k=1}^{j}\left(\frac{k^{q-1}}{k^{\alpha/2}}\sum_{i=1}^{k-1}\frac{(i/k)^{q-1}}{(1-i/k)^{\alpha/2}}\right)^2\frac{1}{k^{\beta}}\\
&\geq\frac{P}{3^{2\delta}j^{2q\delta}}\sum_{k=1}^{j}\left(\frac{k^{q-1}}{k^{\alpha/2}}\int_0^{k-1}\frac{(x/k)^{q-1}}{(1-x/k)^{\alpha/2}}dx\right)^2\frac{1}{k^{\beta}}\\
&=\frac{P}{3^{2\delta}j^{2q\delta}}\sum_{k=1}^{j}\left(\frac{k^{q}}{k^{\alpha/2}}\int_0^{\frac{k-1}{k}}\frac{y^{q-1}}{(1-y)^{\alpha/2}}dy\right)^2\frac{1}{k^{\beta}}\\
&\geq\frac{c_0P}{3^{2\delta}j^{2q\delta}}\sum_{k=2}^{j}k^{2q-\alpha-\beta}\quad(\text{for } j\geq 2)\\
&\geq\frac{c_0P}{3^{2\delta}j^{2q\delta}}\int_2^{j}x^{2q-\alpha-\beta}dx \qquad (64)\\
&=\frac{c_0P}{3^{2\delta}j^{2q\delta}}\frac{j^{1+2q-\alpha-\beta}-2^{1+2q-\alpha-\beta}}{1+2q-\alpha-\beta}\\
&=\Omega(j^{1+2q-\alpha-\beta-2q\delta}) \qquad (65)\end{aligned}$$

where

$$c_0:=\left(\int_0^{1/2}\frac{y^{q-1}}{(1-y)^{\alpha/2}}dy\right)^2>0.$$

Note that the inequality in (64) holds for any value of $2q-\alpha-\beta$. This establishes (63).

Now we proceed with two cases.

Case 1. $\delta<1$: In this case, we choose q such that

$$1+2q-\alpha-\beta-2q\delta>0. \tag{66}$$

Then by (63) and (66), there exists some $\underline{P}>0$ such that for any j

$$\sum_{k=1}^{j}\left(\sum_{i=0}^{k-1}\frac{\sqrt{P}\cdot i^{q-1}}{(k-i)^{\alpha/2}k^{\beta/2}3^{\delta}j^{q\delta}}\right)^2\geq\underline{P}.$$

Then by (62), for any r, any $R<S(\frac{P}{\sigma^2})$ is achievable. Without loss of generality, this means that any $R<S(\frac{P}{\sigma^2})$ is achievable with power constraint P_{total} for any single source–destination pair. Furthermore, $R\cdot r^q$ is an achievable network transport with power constraint P_{total}, which tends to infinity as $r\to\infty$.

Case 2. $1\leq\delta<\frac{3}{2}$: In this case, we choose q such that

$$1+3q-\alpha-\beta-2q\delta>0. \tag{67}$$

Then by (63) and (67), we have

$$\begin{aligned}&r^q\min_{1\leq j\leq r}S\left(\frac{1}{\sigma^2}\sum_{k=1}^{j}\left(\sum_{i=0}^{k-1}\frac{\sqrt{P}\cdot i^{q-1}}{(k-i)^{\alpha/2}k^{\beta/2}3^{\delta}j^{q\delta}}\right)^2\right)\\
&\qquad=\Omega(r^{1+3q-\alpha-\beta-2q\delta})\to\infty,\quad\text{as } r\to\infty.\end{aligned}$$

This means that an arbitrarily large transport capacity is achievable with a fixed total power constraint P_{total}. □

We now exhibit networks that allow a $\Theta(n^{\theta})$ scaling law under low attenuation.

Proof of Theorem 3.10: We consider the case of one source–destination pair, where the source node is located at 0 and the destination node is located at n^{θ}. Let the $n-1$ relay nodes be located at i^{θ}, $i=1,2,\ldots,n-1$. Then by Theorem 3.11, the following rate is achievable:

$$R<\min_{1\leq j\leq n}S\left(\frac{1}{\sigma^2}\sum_{k=1}^{j}\left(\sum_{i=0}^{k-1}\frac{\sqrt{P'_{ik}}}{(j^{\theta}-i^{\theta})^{\delta}}\right)^2\right) \tag{68}$$

with

$$P'_{ik}:=\frac{P'}{(k-i)^{\alpha}},\qquad 0\leq i<k\leq n$$

where $1<\alpha<3-2\theta\delta$ is some constant and $P':=\frac{\alpha-1}{\alpha}P_{\text{ind}}$ is such that the power constraint for every node is satisfied.

Similarly to (58), we can prove the following lower bound on the RHS of (68):

$$\sum_{k=1}^{j}\left(\sum_{i=0}^{k-1}\frac{\sqrt{P'}}{(k-i)^{\alpha/2}(j^{\theta}-i^{\theta})^{\delta}}\right)^2=\Omega(j^{3-\alpha-2\theta\delta}).$$

If $3-\alpha-2\theta\delta>0$, then the minimum over $1\leq j\leq n$ occurs at $j=1$, and is positive. Thus, a positive rate is achievable provided one can satisfy $3-\alpha-2\theta\delta>0$, as well as $\alpha>1$.

To satisfy the above inequalities, we simply choose any small $\epsilon>0$, and consider a network with $\theta:=\frac{1}{\delta}-\epsilon$. Then we choose $\alpha=1+\epsilon\delta$. Such a network can provide a fixed positive rate from source 0 to destination n, irrespective of n. Since the distance between source and destination is n^{θ}, it yields a transport capacity of $\Omega(n^{\theta})$.

To show the optimality of this order, we now prove that $O(n^{\theta})$ is also an upper bound. First, we note that the total power received by all the other nodes, from any candidate source node j, is bounded

$$\begin{aligned}&\sum_{i=0}^{j-1}\frac{P_{\text{ind}}}{(j^{\theta}-i^{\theta})^{2\delta}}+\sum_{i=j+1}^{n}\frac{P_{\text{ind}}}{(i^{\theta}-j^{\theta})^{2\delta}}\\
&\quad\leq\sum_{i=0}^{j-1}\frac{P_{\text{ind}}}{(j-i)^{2\delta}}+\sum_{i=j+1}^{n}\frac{P_{\text{ind}}}{(i-j)^{2\delta}}\\
&\quad\leq\frac{4\delta}{2\delta-1}P_{\text{ind}}<\infty.\end{aligned}$$

Hence, if we take the cut-set around the candidate source node j and apply Lemma 4.1, it follows that the achievable rate is bounded above. Noting that the source–destination distance is at most n^{θ}, we have $O(n^{\theta})$ as an upper bound on the optimal scaling law for this one source case.

Hence, $\Theta(n^{\theta})$ is the optimal scaling law. It is achieved by coherent multistage relaying with interference subtraction, which is therefore the order optimal strategy for information transmission in the networks. □

V. Concluding Remarks

We have examined the problem of how much information can be transported over wireless networks, and what are the order optimal strategies for doing so. In the tradition of information theory, one wishes to determine the ultimate limit to what is achievable without presupposing that packets destructively "collide" if they are from nearby transmitters, or that they can be received only if signal-to-interference ratio is large, etc. The difficulty is that a multitude of nodes can cooperate in very complicated and sophisticated ways, and standard modes of cooperation such as broadcast, multiple access, or relaying do not begin to exhaust the realm of the possible. Worse still, even simple networks, such as the three-node relay channel, or the two-by-two interference channel, are unsolved to date.

Given this state of affairs, we make progress by first enriching the model to explicitly take into account distances between nodes, and by formulating simple models of attenuation as a function of distance. Second, we make progress by asking for less. Instead of studying the capacity region, which is the closure of the set of all vectors of feasible rates, we study a scalar, the supremal distance-weighted sum of rates $\sum R_{\ell} \cdot \rho_{\ell}$, which we have called the transport capacity. There is another sense in which we ask for less. Instead of studying the exact transport capacity, we study the scaling laws for it as the number n of nodes in the network grows. Our contention is that any time there is a scaling law, one needs to first characterize the rate of growth. The preconstant in the scaling law is also important, though it is secondary to the rate of growth, and we provide bounds for it. Through this process we shed light on what wireless networks can achieve in terms of their information transport capacity for every n, at least in some load balanced scenarios.

There is a dichotomy between the cases of relatively high and relatively low attenuation. When either there is absorption $(\gamma > 0)$, or the path loss exponent $\delta > 3$, the transport capacity is bounded by a multiple of the total transmission powers of all the nodes. This allows us to obtain energy upper bounds for information transport across distances in networks. From this it follows that $O(n)$ is an upper bound on the transport capacity of all planar networks when each node has an individual power constraint. In some random scenarios over regular planar networks, this order of the upper bound can be nearly realized by multihop operation, which is consequently the order optimal strategy for the nodes to cooperate. In some other scenarios, the multihop transport strategy attains this order, proving not only that it is order optimal, but also that $\Theta(n)$ is a sharp estimate of the transport capacity. The thrust here is that while network information theory has successfully resolved the two extremes of the bottlenecked scenarios, the multiple-access and scalar Gaussian broadcast channels, in the past, the operating regime where one uses all network resources and supports traffic demands diffused across the network has been open. Our results provide scaling laws addressed to such situations, and provide some justification for the multihop mode of operation in situations where the load can be nearly balanced across the nodes. These results are of interest because the multihop mode of operation is currently the subject of much attention in the protocol development community, using only point-to-point coding, and making no use of network coding or multiuser estimation. Thus, they attempt to bridge the gap between information theory and the area of networking.

In contrast, when there is absolutely no absorption $(\gamma = 0)$ and the attenuation is low with $\delta < 3/2$, we show that there are networks which can attain unbounded transport capacity for fixed total power. We also show that when nodes are on a line and $\delta < 1$, a physical impossibility in the three-dimensional world, there are networks where the transport capacity scales superlinearly as $\Theta(n^{\theta})$ for $\theta < 2$. The strategy which realizes this, and which is consequently an order optimal strategy, is coherent multistage relaying with interference subtraction: At each stage of relaying, all upstream nodes coherently transmit, and all receivers use interference subtraction at each stage. What these results indicate is that when attenuation is low, nodes can profit from long-distance cooperation in manners different from multihop transport, and thus one needs to consider different architectures for information transport. These latter results are obtained by developing a new coding scheme and establishing an achievable rate, superior to earlier results, for the Gaussian multiple-relay channel, a result that may be of interest in its own right.

Open questions abound. An important one is to study the transition regime and characterize what happens for intermediate values of the path loss exponent, when there is absolutely no absorption. Our channel model is simplistic. The constants need to be sharpened. Much remains to be done.

References

[1] C. E. Perkins and P. Bhagwat, "Highly dynamic destination-sequenced distance-vector routing (DSDV) for mobile computers," *Computer Commun. Rev.*, vol. 24, no. 4, pp. 234–244, 1994.

[2] C. Perkins, E. Royer, and S. Das. (1999) Ad Hoc on Demand Distance Vector (AODV) Routing. [Online]. Available: http://www.ietf.org/internet-drafts/draft-ietf-manet-aodv-03.txt

[3] J. Broch, D. Johnson, and D. Maltz. (1998) The Dynamic Source Routing Protocol for Mobile Ad Hoc Networks. [Online]. Available: http://www.ietf.org/internetdrafts/draft-ietf-manet-dsr-01.txt

[4] *Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications*, IEEE Std. 802.11–1997, 1997.

[5] S. Narayanaswamy, V. Kawadia, R. S. Sreenivas, and P. R. Kumar, "Power control in ad-hoc networks: Theory, architecture, algorithm and implementation of the COMPOW protocol," in *Proc. European Wireless Conf.—Next Generation Wireless Networks: Technologies, Protocols, Services and Applications*, Florence, Italy, Feb. 25–28, 2002, pp. 156–162.

[6] V. Kawadia and P. R. Kumar, "Power control and clustering in ad hoc networks," in *Proc. INFOCOM2003*, San Francisco, CA, Mar./Apr. 2003.

[7] R. Rozovsky and P. R. Kumar, "SEEDEX: A MAC protocol for ad hoc networks," in *Proc. 2001 ACM Int. Symp. Mobile Ad Hoc Networking and Computing*, Long Beach, CA, Oct. 4–5, 2001, pp. 67–75.

[8] P. Gupta and P. R. Kumar, "The capacity of wireless networks," *IEEE Trans. Inform. Theory*, vol. 46, pp. 388–404, Mar. 2000.

[9] P. Gupta, "Design and performance analysis of wireless networks," Ph.D. dissertation, Univ. Illinois at Urbana-Champaign, Dept. Elec. Computer Eng., Urbana, IL, 2000.

[10] T. Cover, "Broadcast channels," *IEEE Trans. Inform. Theory*, vol. IT-18, pp. 2–14, Jan. 1972.

[11] P. Bergmans, "Random coding theorem for broadcast channels with degraded components," *IEEE Trans. Inform. Theory*, vol. IT-19, pp. 197–207, Mar. 1973.

[12] ——, "A simple converse for broadcast channels with additive white Gaussian noise," *IEEE Trans. Inform. Theory*, vol. IT-20, pp. 279–280, Mar. 1974.

[13] T. Cover, "Comments on broadcast channels," *IEEE Trans. Inform. Theory*, vol. 44, pp. 2524–2530, Oct. 1998.

[14] R. Ahlswede, "Multi-way communication channels," in *Proc. 2nd Int. Symp. Information Theory (Tsahkadsor, Armenian S.S.R.)*. Budapest, Hungary: Akademiai Kiado, 1971, pp. 23–52.

[15] H. Liao, "Multiple access channels," Ph.D. dissertation, Univ. Hawaii, Dept. Elec. Eng., Honolulu, HI, 1972.

[16] M. Gastpar and M. Vetterli, "On the capacity of wireless networks: The relay case," in *Proc. IEEE INFOCOM*, New York, June 2002.

[17] A. Ephremides and B. Hajek, "Information theory and communication networks: An unconsummated union," *IEEE Trans. Inform. Theory*, vol. 44, pp. 2416–2434, Oct. 1998.

[18] C. E. Shannon, "Two-way communication channels," in *Proc. 4th Berkeley Symp. Mathematical Statistics Probability*, Berkeley, CA, 1961, pp. 611–644.

[19] A. El Gamal and M. H. M. Costa, "The capacity region of a class of deterministic interference channels," *IEEE Trans. Inform. Theory*, vol. IT-28, pp. 343–346, Mar. 1982.

[20] E. C. Van der Meulen, "Three-terminal communication channels," *Adv. Appl. Probab.*, vol. 3, pp. 120–154, 1971.

[21] T. M. Cover and A. A. El Gamal, "Capacity theorems for the relay channel," *IEEE Trans. Inform. Theory*, vol. IT-25, pp. 572–584, Sept. 1979.

[22] T. Cover and J. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.

[23] B. Schein and R. Gallager, "The Gaussian parallel relay network," in *Proc. 2000 IEEE Int. Symp. Information Theory*, Sorrento, Italy, 2000, p. 22.

[24] M. Franceschetti, J. Bruck, and L. Schulman, "Microcellular systems, random walks and wave propagation," in *Proc. IEEE Symp. Antennas and Propagation Soc.*, San Antonio, TX, June 2002.

[25] E. Harrison, *Darkness at Night: A Riddle of the Universe*. Cambridge, MA: Harvard Univ. Press, 1987.

[26] P. Gupta and P. R. Kumar, "Toward an information theory of large networks: An achievable rate region," *IEEE Trans. Inform. Theory*, vol. 49, pp. 1877–1894, Aug. 2003.

[27] V. N. Vapnik and A. Chervonenkis, "On the uniform convergence of relative frequencies of events to their probabilities," *Theory of Probab. its Applic.*, vol. 16, no. 2, pp. 264–280, 1971.