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

Liang-Liang Xie‡ and P. R. Kumar§
Department of Electrical and Computer Engineering, and
Coordinated Science Laboratory
University of Illinois, Urbana-Champaign

Submitted to *IEEE Transactions on Information Theory*

April 12, 2002

Abstract

How much information can be carried over a wireless network with a multiplicity of nodes? What are the optimal strategies for information transmission and cooperation among the nodes? We obtain sharp information theoretic scaling laws under some conditions. We also establish the optimality of multi-hop operation in some situations, and a strategy of coherent multi-stage relaying with interference cancellation in some others.

Consider a network with:

(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.

---

*Please address all correspondence to the second author.

†This material is based upon work partially supported by the USARO under Contracts DAAD19-00-1-0466 and DAAD 19-01010-465, DARPA under Contracts F33615-01-C-1905 and N00014-01-1-0576, ONR under Contract N00014-99-1-0696, and AFOSR under Contract AF-DC-5-36128.

‡Institute of Systems Science, Chinese Academy of Sciences, Beijing 100080, China. This work was conducted while he was visiting the Coordinated Science Laboratory, University of Illinois.

§University of Illinois, Coordinated Science Laboratory, 1308 West Main St., Urbana, IL 61801-2307, USA. Email: prkumar@uiuc.edu. Web: http://black.csl.uiuc.edu/~prkumar.

Let $C_T := \sup \sum_{\ell=1}^{m} R_\ell \cdot \rho_\ell$, where the supremum is taken over vectors $(R_1, R_2, \ldots, R_m)$ of feasible rates for m source-destination pairs, and ρ_ℓ is the distance between the ℓ-th source and its destination. Call this distance-weighted sum-capacity the *transport capacity* of the network. We show that:

(i) $C_T = O(n)$ is an upper bound on the capacity of any planar network if either there is any absorption, i.e., $\gamma > 0$, or the path loss exponent $\delta > 3$, and each node is subject to an individual power constraint P_{ind}. For a regular planar network where the nodes are located on a square integer lattice, the optimal transport capacity is $C_T = \Theta(n)$. It can be achieved by multi-hop operation: Packets can simply be relayed from node to node, with all interference simply treated as noise at each hop.

(ii) For every $\frac{1}{2} < \delta < 1$, and $1 < \theta < \frac{1}{\delta}$, there is a family of networks with nodes located on a line, such that $C_T = \Theta(n^\theta)$ is the optimal transport capacity when there is no absorption, i.e., $\gamma = 0$. The optimal strategy is coherent multi-stage relaying with interference cancellation – all upstream nodes coherently transmit to help each stage of relaying, and all receivers employ interference cancellation at each stage.

(iii) A given rate vector for a set of source-destination pairs can be supported in a planar network if the traffic can be routed in a multi-hop way such that the total traffic to be relayed by any node is less than a certain $c(\bar{\rho})$, where $\bar{\rho}$ is an upper bound on the distance of all hops, under the individual power constraint P_{ind}, when $\gamma > 0$ or $\delta > 1$. If n source-destination pairs are randomly chosen, then a regular planar network with n nodes can simultaneously support a rate $R_\ell = \Omega(\frac{1}{\sqrt{n \log n}})$ for every source-destination pair ℓ, with probability approaching one as $n \to \infty$.

(iv) The total power used by the entire network bounds the transport capacity: $C_T \leq cP_{total}$ if $\gamma > 0$ or $\delta > 3$.

(v) However, the transport capacity can be unbounded even at fixed total power if $\gamma = 0$ and $\delta < \frac{3}{2}$.

(vi) We provide an explicit rate for the general Gaussian multiple relay channel with a single source-destination pair, which is achievable by coherent multi-stage relaying with interference cancellation.

Similar results are provided for linear networks.

1 Introduction

The focus of this paper is on wireless networks, that is, on 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, 2, 3, 4, 5, 6]. As their name implies, ad hoc networks can be set up without any pre-existing wired infrastructure that may be either capital intensive or simply not feasible in a mobile environment, as for example in a network for automobiles. Wireless networks may also be used to interconnect embedded devices whose proliferation rate is faster than PCs. With each embedded device functioning as a sensor or an actuator, in addition to having computational capability, the future may see large orchestras of control systems played over the ether and controlling our physical environment [7].

Since so much of this depends on wireless networking, it is important to understand what such networks are capable of doing, and how to operate them to maximize their capabilities. Thus we seek an information theory for wireless networks to guide us in this process – the goal of this paper.

In contrast to wireline networks, wireless networks do not come with ready made links. Instead, they only consist of nodes radiating energy. Links, if such a notion even exists, have to be fashioned out of the ether by nodes choosing signals and power levels for radiation. Two fundamental questions that arise are:

i) How much information can wireless networks transport?

ii) How should one operate wireless networks?

1.1 The ocean of ignorance

An attempt to address these issues was made in [8] under an assumption on how the technology operates. However, to an information theorist, the answers there 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 (SINR) is large enough, or the received rate is related to the SINR (see Gupta [9] for more on the latter). 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., WaveLan 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 Figure 3, that if there is excessive interference from an interfering transmitter, then that is good, because the interfering signal can first be decoded perfectly, and then subtracted from the received signal, thus eliminating the interference.

Thus, one wishes to study wireless networks without making preconceived assumptions about how they are to operate. There is however a universe of possibilities. Nodes may broadcast simultaneously to several receivers, or several receivers may simultaneously transmit to a certain receiver (multiple access), or a node can serve as a relay, etc. However, these modes of cooperation only scratch the surface, and do not come close to exhausting the possibilities for interaction between a large number of nodes in a network. A group of nodes could cooperate somehow in cancelling the interference of another group of transmitters at a third group of receivers, and so on. Nodes can simultaneously serve several functions of relaying, broadcast, interference cancelling, etc. There are just too many ways in which a plethora of nodes can cooperate with each other. More possibilities exist than can be dreamt of.

Thus it is that one turns to information theory for an answer to the question: How much information can wireless networks transport?

It is a triumph of information theory that the capacity regions for some systems have been characterized, as for example the Gaussian broadcast channel [10, 11, 12, 13] shown in Figure 1, and the Gaussian multiple access channel [14, 15] shown in Figure 2. 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].

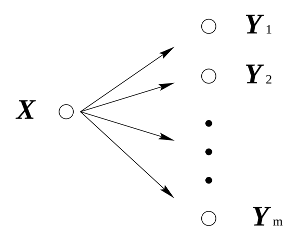

Figure 1: The Gaussian broadcast channel.

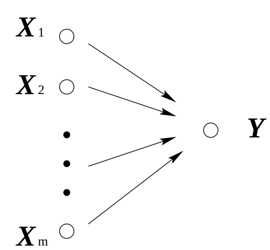

Figure 2: The Gaussian multiple access channel.

However, as observed in [17], the union between information theory and networks is not wholly consummated. The capacity region of even the simple four node system with two sources and two receivers shown in Figure 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, 21, 22] shown in Figure 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 Figure 5, strategies which are quite different in nature have to be considered for different parameter values [23].

Given this ocean of ignorance, what can one then say about much more complicated networks of the type shown in Figures 6 or 7, where there are several source-destination pairs among an arbitrarily large finite number of nodes on the plane or line, all cooperating in whatever ways are imaginable to maximize information transfer? One really needs a more large-scale

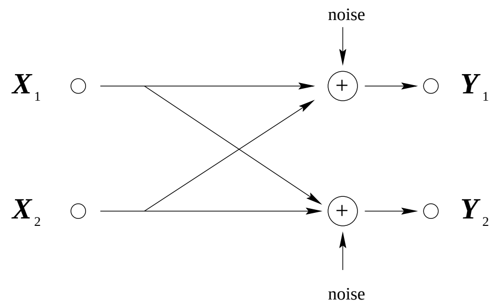

Figure 3: A system with two transmitters and two receivers.

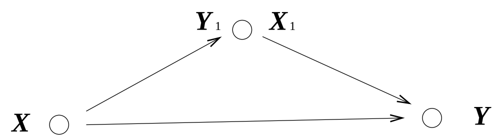

Figure 4: The simplest relay channel.

information theory that can inform us as to what are the limits to information transfer in networks and also, importantly, how one is to operate them. This motivates the subject of the present paper, where our goal is to precisely address complex wireless networks of the type shown in Figures 6 or 7.

The remainder of this paper is organized as follows. In Section 2, we detail the models considered, and in Section 3 the main results, with nothing but proofs in Section 4. Some concluding remarks are made in Section 5, and some open issues, which bear examination and which may lead to a more complete theory, are mentioned.

2 Models considered

The wireless network models considered in this paper have the following ingredients:

1. A finite set $\mathcal{N}$ of n nodes located on a plane.

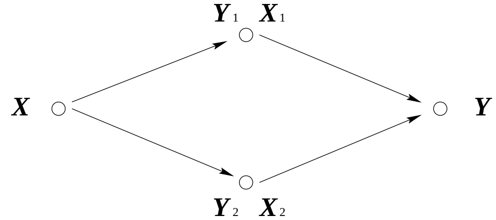

Figure 5: A four node network with two parallel relays.

2. Let ρ_{ij} be the distance between any two nodes $i, j \in \mathcal{N}$ with minimum separation distance $\rho_{\min} := \min_{i \neq j} \rho_{ij} > 0$.

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 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 $20\gamma \log_{10} e$ db/meter; see [24].

4. Denote by $P_i \geq 0$ the power used by node i. We will study two separate types of constraints on $\{P_1, P_2, \ldots, P_n\}$:

   Total Power Constraint P_{total}: $\sum_{i=1}^{n} P_i \leq P_{total}$,
   or
   Individual Power Constraint P_{ind}: $P_i \leq P_{ind}$ 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 $\mathcal{N}$. 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 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.

2.1 The planar and linear settings considered

We will consider four settings for wireless networks.

Planar networks

We consider n nodes located on a two-dimensional plane, with the only restriction on the locations being that the minimum separation between any two nodes is $\rho_{\min} > 0$; see Figure 6. We call such a network a *planar network*.

Linear networks

In this case we suppose that the n nodes are located on a straight line, again with minimum separation distance $\rho_{\min}$; see Figure 7. We call such a network a *linear network*. The chief

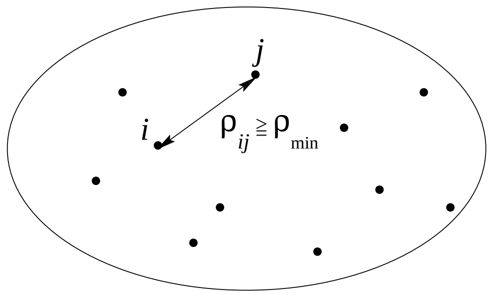

Figure 6: A planar network: n nodes located on a two-dimensional plane, with minimum separation distance $\rho_{\min}$.

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.

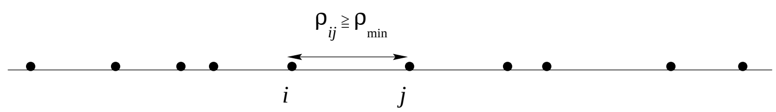

Figure 7: A linear network: n nodes located on a line, with minimum separation distance $\rho_{\min}$.

Regular planar networks

By this we mean a square containing n nodes located at the points (i, j) for $1 \leq i, j \leq \sqrt{n}$; see Figure 8. This setting will be used mainly to exhibit achievability of some capacities, i.e., inner bounds.

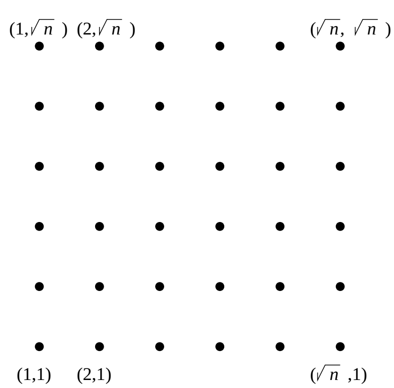

Figure 8: A regular planar network: n nodes located on a plane at (i, j) with $1 \leq i, j \leq \sqrt{n}$.

Regular linear networks

Here we consider n nodes located on a straight line, at positions $1, 2, \ldots, n$; see Figure 9. This setting will also be used mainly to exhibit achievability results.

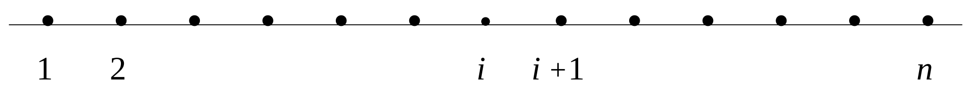

Figure 9: A regular linear network: n nodes located on a line at $1, 2, \ldots, n$.

2.2 The transport capacity

Let $(R_{s_1d_1}, R_{s_2d_2}, \ldots, R_{s_md_m})$ be a vector of feasible rates for the m source-destination pairs. (The precise definition of a feasible rate vector is given in Section 3.1). For brevity, we will denote $R_\ell := R_{s_\ell d_\ell}$, and $\rho_\ell := \rho_{s_\ell d_\ell}$. It is traditional in information theory to study the capacity *region*, which is the set of all such feasible vector rates.

We will however also consider the distance-weighted sum-capacity introduced in [8],

$$C_T := \sup_{(R_1, R_2, \ldots, R_m)} \sum_{\ell=1}^{m} R_\ell \cdot \rho_\ell,$$

and call it the *transport capacity* of the network. The units in which it is measured is bit-meters/second, or bit-meters/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. It is analogous to the man-miles/year metric considered, for example, by airlines.

This transport capacity is of interest for two different reasons. First, we will show, under conditions detailed in the sequel, that regardless of how many and which source-destination pairs are active, and the throughput requirements of each pair, the transport capacity follows a scaling law. That is, it satisfies a conservation law and is thus a constraint on what the wireless network can deliver, regardless of whether it is of prima facie interest in its own right.

The second reason is that it is indeed of interest in its own right. It is a natural quantity that allows us to compare apples with apples in multi-hop networks, and avoids double counting the rate supplied to a longer path as two separate rates for two of its sub-paths.

3 The main results

Our main results are the following:

(i) The best case transport capacity for planar networks with individual power constraints follows a $\Theta(n)$ scaling law when either there is absorption, i.e.,

$\gamma > 0$, **or the path loss exponent** $\delta > 3$.[1]

The following result shows that $O(n)$ is an upper bound for all planar networks:

Theorem 3.1 *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:*

$$C_T \leq \frac{c_1(\gamma, \delta, \rho_{\min}) P_{ind}}{\sigma^2} \cdot n, \text{ where} \tag{1}$$

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

This proves that the "square-root scaling law" of [8] continues to hold without making any assumptions on how the network is to operate.[2] It thus captures the ultimate limits of what is achievable without making any pre-conceived assumptions on the nature of technology.

That this $O(n)$ upper bound is tight is captured by the following theorem, which shows that it can be achieved by a regular planar network. Let $S(x)$ denote the Shannon function:

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

Theorem 3.2 *In a regular planar network with either $\gamma > 0$ or $\delta > 1$, and individual power constraint P_{ind}, the following network transport capacity is achievable:*

$$C_T \geq S\left(\frac{e^{-2\gamma} P_{ind}}{c_2(\gamma, \delta) P_{ind} + \sigma^2}\right) \cdot n,$$

[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 estimates for the optimal capacity.

[2] The reason why the *square-root* itself is missing in our statement is due to the fact that here we have not normalized the area of the domain to be 1 square meter. In [8], the precise order is $O(\sqrt{An})$, where A is the area of the domain. Due to the minimum distance between nodes being positive, the area for a planar network has to grow at least linearly in the number of nodes n.

where

$$c_2(\gamma,\delta) \;:=\; \frac{4(1+4\gamma)e^{-2\gamma}-4e^{-4\gamma}}{2\gamma(1-e^{-2\gamma})} \text{ if } \gamma>0,$$
$$:=\; \frac{16\delta^2+(2\pi-16)\delta-\pi}{(\delta-1)(2\delta-1)} \text{ if } \gamma=0 \text{ and } \delta>3.$$

From the above two results we see that, in the very best case, a planar network can attain a transport capacity scaling law of $\Theta(n)$.

When can a set of rates $(R_1, R_2, \ldots, R_m)$ be supported by a planar network? It is enough that the traffic can be routed in a multi-hop manner over a variety of routes such that no node is overloaded.

Theorem 3.3 *A set of rates $(R_1, R_2, \ldots, R_m)$ can be supported by a planar network if, for some $\bar{\rho}$, the traffic can be routed in a multi-hop manner (with a single source-destination pair's traffic possibly carried over many paths) such that no hop is longer than $\bar{\rho}$, and every node has to relay less than $S\left(\frac{e^{-2\gamma\bar{\rho}}P_{ind}}{\bar{\rho}^{2\delta}(c_3(\gamma,\delta,\rho_{\min})P_{ind}+\sigma^2)}\right)$, where*

$$c_3(\gamma,\delta,\rho_{\min}) \;:=\; \frac{2^{3+2\delta}e^{-\gamma\rho_{\min}}}{\gamma\rho_{\min}^{1+2\delta}} \text{ if } \gamma>0,$$
$$:=\; \frac{2^{2+2\delta}}{\rho_{\min}^{2\delta}(\delta-1)} \text{ if } \gamma=0 \text{ and } \delta>1.$$

What is the situation when n source-destination pairs are randomly chosen? The following result shows that every one of the n randomly chosen source-destination pairs in a regular planar network can be provided a rate $R_\ell = \Omega(\frac{1}{\sqrt{n\log n}})$ for $\ell = 1, 2, \ldots, n$, with probability approaching one as $n \to \infty$, yielding a transport capacity $\Omega\left(\frac{n}{\sqrt{\log n}}\right)$ as a consequence.

Theorem 3.4 *Consider a regular planar network with either $\gamma > 0$ or $\delta > 3$, 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*

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

for some $c > 0$. Consequently, a transport capacity of

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

is supported with probability approaching one as $n \to \infty$.

The following is the corresponding result for linear networks.

Theorem 3.5 *For any linear network, if either $\gamma > 0$ or $\delta > 2$, then the transport capacity is upper bounded as follows:*

$$C_T \leq \frac{c_4(\gamma, \delta, \rho_{\min}) P_{ind}}{\sigma^2} \cdot n, \ \textit{where}$$

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

(ii) Multi-hop operation is optimal when attenuation is large.

We can use the above results to answer how wireless networks should be operated. For example, should one operate wireless networks in a multi-hop mode where packets are simply relayed from node to node, with all interference simply regarded as noise at each hop? Or should one use more sophisticated strategies where nodes coherently cooperate and use interference cancellation?

This is a fundamental question because multi-hop operation brings with it several problems, e.g., the routing problem [1, 2, 3], the media access control problem [4, 6], the power control problem [5], etc., protocols for which are under consideration in their appropriate research and development communities. On the other hand, other strategies may require multi-user detection or interference cancellation or network-wide coherent cooperation. The strategies in the two cases are thus violently different. Answering this question is fundamental to determining how to operate wireless networks, directly affects the basic design of the overall system, and determines all the subsequent communication protocols of interest. It is precisely the kind of question that one hopes to resolve by network information theory.

We show in Theorem 3.2 that when either there is any absorption in the medium, i.e., $\gamma > 0$, or the path loss exponent $\delta > 3$, then multi-hop operation is scaling law optimal. At each hop the packet is decoded with all interference simply regarded as noise, and then relayed to the next node.

Such operation can be achieved with current off-the-shelf technology, which is thus provably optimal in an information theoretic sense in at least certain circumstances.

(iii) A $\Theta(n^\theta)$ scaling law with $1 < \theta < 2$ is feasible under low attenuation with individual power constraints.

When the attenuation is low, one can attain a transport capacity of $\Theta(n^\theta)$ for $1 < \theta < 2$ even for linear networks. Obviously, this result holds for planar networks too, since they include linear networks as a special case.

Theorem 3.6 *Consider $\gamma = 0$ and individual power constraint P_{ind}. For every $\frac{1}{2} < \delta < 1$,*[3] *and $1 < \theta < \frac{1}{\delta}$, there is a family of linear networks for which the transport capacity is*

$$C_T \quad = \quad \Theta(n^\theta). \tag{4}$$

This optimal transport capacity is attained by coherent multi-stage relaying with interference cancellation – *all upstream nodes coherently transmit to help at each stage of relaying, and all receivers employ interference cancellation at each stage. This is therefore an optimal strategy for information transmission.*

(iv) Coherent multi-stage relaying with interference cancellation is optimal for a single source-destination pair, when attenuation is low.

For the networks in Theorem 3.6, since the optimal transport capacity order of $\Theta(n^\theta)$ is attained by coherent multi-stage relaying with interference cancellation, such a scheme is the optimal strategy for information transmission.

Thus we see an interesting bifurcation: At high attenuation, multi-hop operation is optimal, where all interference can simply be regarded as noise. At low attenuation, coherent multi-stage relaying with interference cancellation is optimal.

(v) The transport capacity is bounded by the total power in networks with high attenuation.

It is well known from Shannon's work that for a source-destination pair (s_ℓ, d_ℓ), the *rate* R_ℓ is bounded by the *received* power at d_ℓ.

What is interesting is that, in networks, there is a fundamental relationship between the *total transmitted power* P_{total} used by the entire network, and the transport capacity of the network:

> The *transport capacity* is bounded by the *total transmitted power*, when the attenuation is large.

This is true irrespective of how the nodes are located, subject only to a minimum separation distance $\rho_{\min} > 0$, how many source-destination pairs exist, and how they are chosen. Thus, the total power P_{total} available to the entire network plays a key role.

Theorem 3.7 *In any planar network, with either positive absorption, i.e., $\gamma > 0$, or with path loss exponent $\delta > 3$,*

$$C_T \quad \leq \quad \frac{c_1(\gamma, \delta, \rho_{\min})}{\sigma^2} \cdot P_{total}, \tag{5}$$

where $c_1(\gamma, \delta, \rho_{\min})$ is as in (2).

[3] 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 dark.

The following is the corresponding result for linear networks.

Theorem 3.8 *If either $\gamma > 0$ or $\delta > 2$ in any linear network, then*

$$C_T \leq \frac{c_4(\gamma, \delta, \rho_{\min})}{\sigma^2} \cdot P_{total}, \tag{6}$$

where $c_4(\gamma, \delta, \rho_{\min})$ is as in (3).

(vi) At low attenuation unbounded transport capacity can be obtained for bounded total power.

In contrast to the high attenuation case, when the attenuation is low, the transport capacity can be unbounded even with finite total power.

Theorem 3.9 *(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 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}, any regular planar network can support a fixed rate $R_{\min} > 0$ for any single source-destination pair, irrespective of the distance between them.

The following is the corresponding result for linear networks.

Theorem 3.10 *(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 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}, any regular linear network can support a fixed rate $R_{\min} > 0$ for any single source-destination pair, irrespective of the distance between them.

(vii) The Gaussian multiple relay channel with a single source-destination pair: Coherent relaying with interference cancellation, and an explicit achievable rate.

Consider a network of n nodes with α_{ij} the attenuation from node i to node j (the nodes need not be on a plane, and in fact there need not 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*.

Consider the following strategy for cooperation: 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 of their power P_{ik} to coherently transmit for the benefit of the nodes in downstream groups. Each node j employs interference cancellation, and uses jointly typical decoding which conforms with all the coherent transmissions of its upstream nodes. We call this strategy *coherent multi-stage relaying with interference cancellation.*

We provide the following explicit expressions for the achievable rate R. The first theorem addresses the case where each relaying group consists of only one node.

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

$$R \;<\; \min_{1 \le j \le M} S\left(\frac{1}{\sigma^2}\sum_{k=1}^{j}\left(\sum_{i=0}^{k-1}\alpha_{ij}\sqrt{P_{ik}}\right)^2\right) \tag{7}$$

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 [25] shows that a rate R_0 is achievable if there exist some $\{R_1, R_2, \cdots, R_{M-1}\}$ such that*

$$\begin{aligned} R_{M-1} &< S\left(\frac{P^R_{M,M-1}}{\sigma^2 + \sum_{\ell=0}^{M-2} P^R_{M,\ell}}\right), \qquad \textit{and} \\ R_m &< \min\left\{ S\left(\frac{P^R_{m+1,m}}{\sigma^2 + \sum_{\ell=0}^{m-1} P^R_{m+1,\ell}}\right), R_{m+1} + \min_{m+2 \le k \le M} S\left(\frac{P^R_{k,m}}{\sigma^2 + \sum_{\ell=0}^{m-1} P^R_{k,\ell}}\right)\right\} \end{aligned}$$

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

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

From the above, recursively for $m = M-2, M-1, \cdots, 0$, 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 (7), showing that we get a higher achievable rate in Theorem 3.11.

Theorem 3.12 *Consider again the Gaussian multiple relay channel using coherent multi-stage relaying with interference cancellation. 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 $\frac{P_i}{n_i} \geq 0$. Then any rate R satisfying the following inequality is achievable from s to d:*

$$R \quad < \quad \min_{1 \leq j \leq M} S\left(\frac{1}{\sigma^2} \sum_{k=1}^{j} \left(\sum_{i=0}^{k-1} \alpha_{\mathcal{N}_i \mathcal{N}_j} \sqrt{P_{ik}/n_i} \cdot n_i\right)^2\right) \tag{8}$$

where $P_{ik} \geq 0$ satisfies $\sum_{k=i+1}^{M} P_{ik} \leq P_i$, and $\alpha_{\mathcal{N}_i \mathcal{N}_j} := \min\{\alpha_{k\ell} : k \in \mathcal{N}_i, \ell \in \mathcal{N}_j\}$, $i, j \in \{0, 1, \dots, M\}$.

3.1 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.

Definition 3.1 *In a wireless network with multiple source-destination pairs (s_ℓ, d_ℓ), $\ell = 1, \dots, m$, a $((2^{TR_1}, \dots, 2^{TR_m}), T, \lambda_T)$ code with total power constraint P_{total} consists of the following:*

1. *m random variables W_ℓ with $P(W_\ell = k_\ell) = \frac{1}{2^{TR_\ell}}$, for any $k_\ell \in \{1, 2, \dots, 2^{TR_\ell}\}$, $\ell = 1, \dots, m$.*
2. *Functions $f_{s_\ell, t} : \mathbb{R}^{t-1} \times \{1, 2, \dots, 2^{TR_\ell}\} \to \mathbb{R}^1, t = 1, 2, \dots, T$ for the source nodes $s_\ell, \ell = 1, \dots, m$ and $f_{i,t} : \mathbb{R}^{t-1} \to \mathbb{R}^1, t = 2, \dots, T$ for all the other nodes $i \notin \{s_\ell, \ell = 1, \dots, m\}$, such that*

$$X_{s_\ell}(t) \quad = \quad f_{s_\ell, t}(Y_{s_\ell}(1), \dots Y_{s_\ell}(t-1), W_\ell), \quad t = 1, 2, \dots, T, \quad \ell = 1, \dots, m;$$

$$X_i(1) = 0, \quad X_i(t) = f_{i,t}(Y_i(1), \dots, Y_i(t-1)), \quad t = 2, 3, \dots, T, \text{ for } i \notin \{s_\ell, \ell = 1, \dots, m\},$$

such that the following total power constraint holds:

$$\frac{1}{T} \sum_{t=1}^{T} \sum_{i \in \mathcal{N}} X_i^2(t) \quad \leq \quad P_{total}, \quad a.s. \tag{9}$$

3. *m decoding functions $g_{d_\ell} : \mathbb{R}^T \to \{1, 2, \dots, 2^{TR_\ell}\}$ for the destination nodes $d_\ell, \ell = 1, \dots, m$.*

4. The maximal probability of error:

$$\lambda_T = \max_{\substack{k_\ell \in \{1,2,\dots,2^{TR_\ell}\} \\ \ell=1,2,\dots,m}} Prob\{g_{d_\ell}(Y_{d_\ell}^T) \neq k_\ell | W_\ell = k_\ell\}, \tag{10}$$

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

Definition 3.2 *A rate vector* $(R_1, \dots, R_m)$ *is said to be* feasible *for the* m *source-destination pairs* $(s_\ell, d_\ell), \ell = 1, \dots, m$, *with total power constraint* P_{total}, *if there exists a sequence of* $((2^{TR_1}, \dots, 2^{TR_m}), T, \lambda_T)$ *codes satisfying the total power constraint* P_{total}, *such that* $\lambda_T \to 0$ *as* $T \to \infty$.

Next is the definition of the transport capacity of a network.

Definition 3.3 *The network transport capacity* C_T *with constraint* P_{total} *is*

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

where ρ_ℓ *is the distance between* s_ℓ *and* d_ℓ.

The above definitions are presented in the context of a total power constraint P_{total}. With individual power constraint P_{ind}, one simply needs to replace the constraint (9) by

$$\frac{1}{T} \sum_{t=1}^{T} X_i^2(t) \leq P_{ind}, \quad \text{a.s., for } i \in \mathcal{N}, \tag{11}$$

and correspondingly modify the rest of the definitions.

4 Nothing but proofs

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

4.1 A max-flow min-cut lemma

The following max-flow min-cut bound will be used to establish certain upper bounds on the feasible rate vectors.

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 $(R_1, \ldots, R_m)$ is a feasible rate vector with a sequence of $((2^{TR_1}, \ldots, 2^{TR_m}), T, \lambda_T)$ codes with $\lambda_T \to 0$ as $T \to \infty$, then*

$$\sum_{\{\ell : d_\ell \in \mathcal{N}_1, s_\ell \notin \mathcal{N}_1\}} R_\ell \quad \leq \quad \frac{1}{2\sigma^2} \liminf_{T \to \infty} P^{rec}_{\mathcal{N}_1}(T), \tag{12}$$

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

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

Proof. First we introduce some notation:

$$U_i(t) \quad := \quad \sum_{j \notin \mathcal{N}_1} \frac{X_j(t)}{\rho^{\delta}_{ij}}, \quad i \in \mathcal{N}_1; \tag{14}$$

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

Denote $W_{\mathcal{N}_1^{dest-cut}} := \{W_\ell : (s_\ell, d_\ell) \text{ cuts } \mathcal{N}_1\}$, $\mathcal{N}_1^{source} := \{s_\ell : s_\ell \in \mathcal{N}_1, \ell = 1, \ldots, m\}$ and $W_{\mathcal{N}_1^{source}} := \{W_i, i \in \mathcal{N}_1^{source}\}$. We adopt the notation:

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

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

Now we prove that the following forms a Markov chain

$$W_{\mathcal{N}_1^{dest-cut}} \to \{V^T_{\mathcal{N}_1}, W_{\mathcal{N}_1^{source}}\} \quad \to \quad \{Y^T_{\mathcal{N}_1}, W_{\mathcal{N}_1^{source}}\}, \tag{16}$$

by showing that any element in $Y^T_{\mathcal{N}_1}$ is a deterministic function of $\{V^T_{\mathcal{N}_1}, W_{\mathcal{N}_1^{source}}\}$. This can be easily seen since for any $i \in \mathcal{N}_1$, $2 \leq t \leq T$,

$$\begin{aligned} Y_i(t) \quad &= \quad V_i(t) + \sum_{\substack{j \in \mathcal{N}_1 \\ j \neq i}} \frac{X_j(t)}{\rho^{\delta}_{ij}} \\ &= \quad V_i(t) + \sum_{\substack{j \in \mathcal{N}_1 \setminus \mathcal{N}_1^{source} \\ j \neq i}} \frac{f_{j,t}(Y_j^{t-1})}{\rho^{\delta}_{ij}} + \sum_{\substack{j \in \mathcal{N}_1^{source} \\ j \neq i}} \frac{f_{j,t}(Y_j^{t-1}, W_j)}{\rho^{\delta}_{ij}}, \end{aligned}$$

and for $t = 1$,

$$Y_i(1) \quad = \quad V_i(1) + \sum_{\substack{j \in \mathcal{N}_1^{source} \\ j \neq i}} \frac{f_{j,1}(W_j)}{\rho^{\delta}_{ij}}.$$

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

$$H(W_{\mathcal{N}_1^{dest-cut}}|V_{\mathcal{N}_1}^T, W_{\mathcal{N}_1^{source}}) \quad \leq \quad 1 + T\lambda_T \quad =: \quad T\epsilon_T,$$

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

Thus, we have the following chain of inequalities:

$$\begin{aligned}
T\sum_{k=1}^{m_1} R_{\ell_k} \quad = \quad H(W_{\mathcal{N}_1^{dest-cut}}) \quad &= \quad I(W_{\mathcal{N}_1^{dest-cut}}; V_{\mathcal{N}_1}^T, W_{\mathcal{N}_1^{source}}) + H(W_{\mathcal{N}_1^{dest-cut}}|V_{\mathcal{N}_1}^T, W_{\mathcal{N}_1^{source}}) \\
&\leq \quad I(W_{\mathcal{N}_1^{dest-cut}}; V_{\mathcal{N}_1}^T, W_{\mathcal{N}_1^{source}}) + T\epsilon_T \\
&= \quad I(W_{\mathcal{N}_1^{dest-cut}}; W_{\mathcal{N}_1^{source}}) + I(W_{\mathcal{N}_1^{dest-cut}}; V_{\mathcal{N}_1}^T|W_{\mathcal{N}_1^{source}}) + T\epsilon_T \\
&= \quad 0 + h(V_{\mathcal{N}_1}^T|W_{\mathcal{N}_1^{source}}) - h(V_{\mathcal{N}_1}^T|W_{\mathcal{N}_1^{dest-cut}}, W_{\mathcal{N}_1^{source}}) + T\epsilon_T \\
&\leq \quad h(V_{\mathcal{N}_1}^T) - h(V_{\mathcal{N}_1}^T|W_{\mathcal{N}_1^{dest-cut}}, W_{\mathcal{N}_1^{source}}) + T\epsilon_T,
\end{aligned}$$

with

$$\begin{aligned}
& h(V_{\mathcal{N}_1}^T|W_{\mathcal{N}_1^{dest-cut}}, W_{\mathcal{N}_1^{source}}) \\
= \quad & \sum_{t=1}^{T} h(V_{\mathcal{N}_1}(t)|V_{\mathcal{N}_1}(1), \ldots, V_{\mathcal{N}_1}(t-1), W_{\mathcal{N}_1^{dest-cut}}, W_{\mathcal{N}_1^{source}}) \\
\geq \quad & \sum_{t=1}^{T} h(V_{\mathcal{N}_1}(t)|V_{\mathcal{N}_1}(1), \ldots, V_{\mathcal{N}_1}(t-1), X_{\mathcal{N}_1^{dest-cut}}(t), W_{\mathcal{N}_1^{dest-cut}}, W_{\mathcal{N}_1^{source}}) \\
= \quad & \sum_{t=1}^{T} h(V_{\mathcal{N}_1}(t)|X_{\mathcal{N}_1^{dest-cut}}(t)) \\
\geq \quad & \sum_{t=1}^{T} h(V_{\mathcal{N}_1}(t)|U_{\mathcal{N}_1}(t)),
\end{aligned}$$

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

$$\{V_{\mathcal{N}_1}^{t-1}, W_{\mathcal{N}_1^{dest-cut}}, W_{\mathcal{N}_1^{source}}\} \to X_{\mathcal{N}_1^{dest-cut}}(t) \to V_{\mathcal{N}_1}(t); \qquad X_{\mathcal{N}_1^{dest-cut}}(t) \to U_{\mathcal{N}_1}(t) \to V_{\mathcal{N}_1}(t).$$

Hence, we have

$$\begin{aligned} T\sum_{k=1}^{m_1} R_{\ell_k} &\leq h(V_{\mathcal{N}_1}^T) - \sum_{t=1}^{T} h(V_{\mathcal{N}_1}(t)|U_{\mathcal{N}_1}(t)) + T\epsilon_T \\ &= h(V_{\mathcal{N}_1}^T) - \sum_{t=1}^{T}\sum_{i\in\mathcal{N}_1} h(Z_i(t)) + T\epsilon_T \\ &\leq \sum_{t=1}^{T}\sum_{i\in\mathcal{N}_1} h(V_i(t)) - \sum_{t=1}^{T}\sum_{i\in\mathcal{N}_1} h(Z_i(t)) + T\epsilon_T \\ &= \sum_{t=1}^{T}\sum_{i\in\mathcal{N}_1} [h(V_i(t)) - h(V_i(t)|U_i(t))] + T\epsilon_T \\ &= \sum_{t=1}^{T}\sum_{i\in\mathcal{N}_1} I(U_i(t);V_i(t)) + T\epsilon_T \\ &\leq \sum_{t=1}^{T}\sum_{i\in\mathcal{N}_1} \frac{1}{2}\log\left(1+\frac{EU_i^2(t)}{\sigma^2}\right) + T\epsilon_T \\ &\leq \frac{1}{2\sigma^2}\sum_{t=1}^{T}\sum_{i\in\mathcal{N}_1} EU_i^2(t) + T\epsilon_T. \end{aligned}$$

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

4.2 The total power bounds the transport capacity

We begin with the case of linear networks.

Proof of Theorem 3.8 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:

$$\mathcal{N}_q^- = \{i\in\mathcal{N} : -\infty < a_i \leq q\}, \quad \mathcal{N}_q^+ = \{i\in\mathcal{N} : q \leq a_i < \infty\}, \quad q\in\mathbb{Z}, \tag{17}$$

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

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

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

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 $\mathcal{N}_q^-, q = \lfloor a+1 \rfloor, \ldots, \lfloor a + \rho_\ell/\rho_{\min} \rfloor$. By definition, R_ℓ is a summand in every $R_{\mathcal{N}_q^-}, q = \lfloor a+1 \rfloor, \ldots, \lfloor a + \rho_\ell/\rho_{\min} \rfloor$. Hence we have (noting $\rho_\ell \geq \rho_{\min}$)

$$\sum_{\ell=1}^{m} R_\ell \cdot \rho_\ell \leq 2\rho_{\min} \sum_{\ell=1}^{m} R_\ell \cdot \lfloor \rho_\ell/\rho_{\min} \rfloor \leq 2\rho_{\min} \sum_{q=-\infty}^{+\infty} R_{\mathcal{N}_q^-} + 2\rho_{\min} \sum_{q=-\infty}^{+\infty} R_{\mathcal{N}_q^+}. \tag{20}$$

Now we prove that

$$\sum_{q=-\infty}^{+\infty} R_{\mathcal{N}_q^-} \leq \frac{c_4(\gamma, \delta, \rho_{\min})}{4\rho_{\min}\sigma^2} P_{total}. \tag{21}$$

By (18), we only need to show that

$$\frac{1}{T} \sum_{t=1}^{T} \sum_{q=-\infty}^{+\infty} \sum_{i \in \mathcal{N}_q^-} E \left(\sum_{j \notin \mathcal{N}_q^-} \frac{X_j(t)}{(a_j - a_i)^\delta} \right)^2 \leq \frac{c_4(\gamma, \delta, \rho_{\min}) \rho_{\min}^{2\delta-1}}{2} P_{total}, \tag{22}$$

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

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

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,

$$\sum_{q=-\infty}^{+\infty} \sum_{i \in \mathcal{N}_q^-} \left(\sum_{j \notin \mathcal{N}_q^-} \frac{X_j(t)}{(a_j - a_i)^\delta} \right)^2 \leq \frac{c_4(\gamma, \delta, \rho_{\min}) \rho_{\min}^{2\delta-1}}{2} P(t), \tag{24}$$

where

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

First, we observe that the L.H.S. 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

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

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

$$\sum_{q=-\infty}^{\lceil a_k \rceil -1} \sum_{i\in\mathcal{N}_q^-} \sum_{j\notin\mathcal{N}_q^-} \frac{1}{(a_j-a_i)^\delta (a_k-a_i)^\delta} \leq \frac{c_4(\gamma,\delta,\rho_{\min})\rho_{\min}^{2\delta-1}}{2}. \tag{26}$$

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

$$\begin{aligned}
&\text{L.H.S. of (26)} \\
\leq\ & \sum_{q=-\infty}^{\lceil a_k \rceil -1} \sum_{i\in\mathcal{N}_q^-} \sum_{l=0}^{\infty} \frac{1}{(l+\underline{a}_q-a_i)^\delta} \frac{1}{(a_k-a_i)^\delta} \\
\leq\ & \sum_{q=-\infty}^{\lceil a_k \rceil -1} \sum_{i\in\mathcal{N}_q^-} \frac{\delta-1+\underline{a}_q-a_i}{(\delta-1)(\underline{a}_q-a_i)^\delta} \frac{1}{(a_k-a_i)^\delta} \\
=\ & \sum_{\{i:\, a_i<a_k\}} \sum_{q=\lceil a_i \rceil}^{\lceil a_k \rceil -1} \left[\frac{1}{(\underline{a}_q-a_i)^\delta} + \frac{1}{(\delta-1)(\underline{a}_q-a_i)^{\delta-1}} \right] \frac{1}{(a_k-a_i)^\delta} \\
\leq\ & \sum_{\{i:\, a_i<a_k\}} \left[\sum_{l=1}^{\infty} \frac{1}{l^\delta} + \frac{1}{\delta-1} \sum_{l=1}^{\infty} \frac{1}{l^{\delta-1}} \right] \frac{1}{(a_k-a_i)^\delta} \\
\leq\ & \frac{\delta^3-\delta^2-\delta}{(\delta-1)^2(\delta-2)} \qquad (27) \\
=\ & \frac{c_4(\gamma,\delta,\rho_{\min})\rho_{\min}^{2\delta-1}}{2}, \qquad (28)
\end{aligned}$$

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

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

Hence (24) is established.

Thus (21) follows. Similarly, we can prove

$$\sum_{q=-\infty}^{+\infty} R_{\mathcal{N}_q^+} \leq \frac{c_4(\gamma,\delta,\rho_{\min})}{4\rho_{\min}\sigma^2} P_{total}. \qquad (30)$$

Finally, (6) 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

$$\frac{1}{T}\sum_{t=1}^{T}\sum_{q=-\infty}^{+\infty}\sum_{i\in\mathcal{N}_q^-} E\left(\sum_{j\notin\mathcal{N}_q^-}\frac{e^{-\gamma(a_j-a_i)\rho_{\min}}X_j(t)}{(a_j-a_i)^\delta}\right)^2 \leq \frac{c_4(\gamma,\delta,\rho_{\min})\rho_{\min}^{2\delta-1}}{2} P_{total},$$

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

$$\begin{aligned}\sum_{q=-\infty}^{\lceil a_k\rceil-1}\sum_{i\in\mathcal{N}_q^-}\sum_{j\notin\mathcal{N}_q^-}\frac{e^{-\gamma(a_j-a_i)\rho_{\min}}e^{-\gamma(a_k-a_i)\rho_{\min}}}{(a_j-a_i)^\delta(a_k-a_i)^\delta} &\leq \frac{e^{-2\gamma\rho_{\min}}}{(1-e^{-\gamma\rho_{\min}})^2(1-e^{-2\gamma\rho_{\min}})} \\ &= \frac{c_4(\gamma,\delta,\rho_{\min})\rho_{\min}^{2\delta-1}}{2}.\end{aligned}$$

□

Proof of Theorem 3.7. The proof is similar to that of Theorem 3.8. 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:

$$\begin{aligned}\mathcal{N}_{q,\infty}^- &= \{i\in\mathcal{N}: -\infty < a_i \leq q, -\infty < b_i < +\infty\}, \quad q\in\mathbb{Z}; \qquad (31)\\ \mathcal{N}_{q,\infty}^+ &= \{i\in\mathcal{N}: q \leq a_i < +\infty, -\infty < b_i < +\infty\}, \quad q\in\mathbb{Z};\\ \mathcal{N}_{\infty,q}^- &= \{i\in\mathcal{N}: -\infty < a_i < +\infty, -\infty < b_i \leq q\}, \quad q\in\mathbb{Z};\\ \mathcal{N}_{\infty,q}^+ &= \{i\in\mathcal{N}: -\infty < a_i < +\infty, q \leq b_i < +\infty\}, \quad q\in\mathbb{Z}.\end{aligned}$$

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

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

where $R_{\mathcal{N}_{q,\infty}^-}$ is defined similarly to $R_{\mathcal{N}_q^-}$ in the proof of Theorem 3.8. $R_{\mathcal{N}_{q,\infty}^+}$, $R_{\mathcal{N}_{\infty,q}^-}$ and $R_{\mathcal{N}_{\infty,q}^+}$ are also similarly defined.

Now 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:

$$\sum_{\ell=1}^{m} R_\ell \cdot \rho_\ell \leq \rho_{\min} \sum_{\ell=1}^{m} R_\ell \cdot \lceil 2\rho_\ell/\rho_{\min} \rceil$$

$$\leq \rho_{\min} \sum_{q=-\infty}^{+\infty} R_{\mathcal{N}_{q,\infty}^-} + \rho_{\min} \sum_{q=-\infty}^{+\infty} R_{\mathcal{N}_{q,\infty}^+} + \rho_{\min} \sum_{q=-\infty}^{+\infty} R_{\mathcal{N}_{\infty,q}^-} + \rho_{\min} \sum_{q=-\infty}^{+\infty} R_{\mathcal{N}_{\infty,q}^+}. \quad (33)$$

Now, we prove that

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

By (32), we only need to show that

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

or equivalently,

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

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

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

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,

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

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

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

Using the fact that for any $d_0\geq 2$

$$\frac{1}{d_0^\delta}\leq\frac{4}{\pi}\int_{-\frac{\pi}{4}}^{\frac{\pi}{4}}\int_0^1\frac{1}{(d_0^2+r^2-2rd_0\cos\theta)^{\frac{\delta}{2}}}rdrd\theta,$$

since $\min_{i\neq j}[(a_j-a_i)^2+(b_j-b_i)^2]^{1/2}\geq 2$, we have for any $i\in\mathcal{N}_{q,\infty}^-$, $\delta>2$

$$\begin{aligned}\sum_{j\notin\mathcal{N}_{q,\infty}^-}\frac{1}{[(a_j-a_i)^2+(b_j-b_i)^2]^{\delta/2}} &\leq \frac{4}{\pi}\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}}\int_{(\underline{a}_q-a_i+1)\vee 1}^{\infty}\frac{1}{x^\delta}xdxd\theta\\ &\leq \frac{4}{(\delta-2)[(\underline{a}_q-a_i+1)\vee 1]^{\delta-2}},\end{aligned}$$

where $\underline{a}_q:=\min_{j\notin\mathcal{N}_{q,\infty}^-}a_j$.

Then we have for $\delta>3$,

$$\begin{aligned}&\text{L.H.S. of (40)}\\ \leq\ & \sum_{q=-\infty}^{\lceil a_k\rceil-1}\sum_{i\in\mathcal{N}_{q,\infty}^-}\frac{4}{(\delta-2)[(\underline{a}_q-a_i+1)\vee 1]^{\delta-2}}\frac{1}{[(a_k-a_i)^2+(b_k-b_i)^2]^{\delta/2}}\\ \leq\ & \sum_{\{i:\,a_i<a_k\}}\sum_{q=\lceil a_i\rceil}^{\lceil a_k\rceil-1}\frac{4}{(\delta-2)[(\underline{a}_q-a_i+1)\vee 1]^{\delta-2}}\frac{1}{[(a_k-a_i)^2+(b_k-b_i)^2]^{\delta/2}}\\ \leq\ & \sum_{\{i:\,a_i<a_k\}}\left(\frac{8}{\delta-2}+\frac{4}{\delta-3}\right)\frac{1}{[(a_k-a_i)^2+(b_k-b_i)^2]^{\delta/2}}\\ \leq\ & \left(\frac{8}{\delta-2}+\frac{4}{\delta-3}\right)\frac{4}{\pi}\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}}\int_1^\infty\frac{1}{x^\delta}xdxd\theta\\ \leq\ & \left(\frac{8}{\delta-2}+\frac{4}{\delta-3}\right)\frac{4}{\delta-2}\\ =\ & \frac{c_1(\gamma,\delta,\rho_{\min})\rho_{\min}^{2\delta-1}}{2^{2\delta+1}}.\end{aligned}$$

Hence (38) is proved. Thus (34) follows. The remaining arguments are similar to the proof of Theorem 3.8.

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

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

This holds for

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

Proofs of Theorems 3.1 and 3.5. The results for the case of individual power P_{ind} follow directly from the case of P_{total} in Theorems 3.8 and 3.7 by noting that $P_{total} = nP_{ind}$ is also a constraint. □

4.3 The Gaussian multiple relay channel: The strategy of coherent multi-stage relaying with interference cancellation and an achievable rate

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.

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

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.,*

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

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. *$Prob((X^T, Y^T) \in A_\epsilon^{(T)}(P,N)) \to 1$ 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*

   $$Prob((\tilde{X}^T, \tilde{Y}^T) \in A_\epsilon^{(T)}(P,N)) \le 2^{-T(S(\frac{P}{N})-3\epsilon)}.$$

   Also, for sufficiently large T,

   $$Prob((\tilde{X}^T, \tilde{Y}^T) \in A_\epsilon^{(T)}(P,N)) \ge (1-\epsilon)2^{-T(S(\frac{P}{N})+3\epsilon)}.$$

Proof of Theorem 3.11. The coding-decoding scheme is different from that of [21], though we still use a block coding argument. We consider B blocks of transmission, each of T transmission slots. A sequence of $B-M+1$ indices, $w_b \in \{1,\ldots,2^{TR}\}, b = 1,2,\ldots,B-M+1$ 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 $k = 1,\ldots,M$, and $b_0 = 1,\ldots,M$, each of size $2^{TR} \times T$, with every element independently chosen with Gaussian distribution $N(0, 1-\varepsilon_1)$. The M^2 matrices are revealed to all the $M+1$ nodes. Let $\mathcal{X}_k(b) := \mathcal{X}_k(b \bmod M), b = 1,2,\ldots,B$. Denote by $x_k(b,w)$ the w-th row of the matrix $\mathcal{X}_k(b)$, for $w \in \{1,\ldots,2^{TR}\}$.

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

$$\widehat{w}_{b_1,i} \;:=\; w_{b_1} := 0 \text{ for any } b_1 \leq 0, \text{ and } x_k(b,0) := 0. \tag{41}$$

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

$$\begin{aligned}
\vec{Y}_k(b) &= \sum_{\substack{0 \leq i \leq M-1 \\ i \neq k}} \alpha_{ik} \vec{X}_i(b) + \vec{Z}_k(b) \\
&= \sum_{\substack{0 \leq i \leq M-1 \\ i \neq k}} \sum_{l=i+1}^{M} \alpha_{ik} \sqrt{P_{il}} x_l(b, \widehat{w}_{b-l+1,i}) + \vec{Z}_k(b) \\
&= \left(\sum_{l=1}^{k} \sum_{i=0}^{l-1} + \sum_{l=k+1}^{M} \sum_{\substack{0 \leq i \leq l-1 \\ i \neq k}} \right) \alpha_{ik} \sqrt{P_{il}} x_l(b, \widehat{w}_{b-l+1,i}) + \vec{Z}_k(b).
\end{aligned} \tag{42}$$

Let

$$\widehat{\vec{Y}}_k(b) \;:=\; \vec{Y}_k(b) - \sum_{l=k+1}^{M} \sum_{\substack{0 \leq i \leq l-1 \\ i \neq k}} \alpha_{ik} \sqrt{P_{il}}\, x_l(b, \widehat{w}_{b-l+1,k}). \tag{43}$$

This will serve as an estimate by node k of

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

as we show in the sequel.

Decoding

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

$b-j, j = 0,1,\ldots,k-1$:

$$\left\{ \sum_{i=0}^{k-j-1} \alpha_{ik} \sqrt{P_{i,k-j}}\, x_{k-j}(b-j,w), \hat{\vec{Y}}_k(b-j) - \sum_{l=k-j+1}^{k} \sum_{i=0}^{l-1} \alpha_{ik} \sqrt{P_{il}}\, x_l(b-j, \widehat{w}_{b-j-l+1,k}) \right\}$$

$$\in A_\epsilon^{(T)}(\bar{P}_{k,j}, N_{k,j}), \quad (44)$$

where

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

Otherwise, if an 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

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

and let its probability be $P_c(b) := \text{Prob}(A_c(b))$, 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

$$\begin{aligned} P_e &:= \text{Prob}(\widehat{w}_{b-k+1,k} \neq w_{b-k+1}, \text{ for some } k \in \{1,\ldots,M\}, b \in \{1,\ldots,B\}) \\ &= \sum_{b=1}^{B} \text{Prob}(\widehat{w}_{b-k+1,k} \neq w_{b-k+1} \text{ for some } k \in \{1,\ldots,M\} | A_c(b-1)) \cdot P_c(b-1) \\ &\leq \sum_{b=1}^{B} \sum_{k=1}^{M} \text{Prob}(\widehat{w}_{b-k+1,k} \neq w_{b-k+1} | A_c(b-1)) \cdot P_c(b-1) \\ &= \sum_{b=1}^{B} \sum_{k=1}^{M} P_e(b,k) \cdot P_c(b-1), \end{aligned} \quad (45)$$

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

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

Hence, noting (41), $\widehat{w}_{b_2,k} = w_{b_2}$ whenever $b_2 + k \leq b$. Then, by (42) and (43), for all $b-j$ with $j \geq 0$,

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

and

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

So,

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

Hence, under the condition $A_c(b-1)$ the decoding rule (44) is equivalent to: Each node $k \in \{1,\dots,M\}$ (when $b-k+1 \ge 1$) declares $\widehat{w}_{b-k+1,k} = w$ if w is the unique value in $\{1,\dots,2^{2R}\}$ such that in all the blocks $b-j$, for $j = 0, 1, \dots, k-1$:

$$\left\{\sum_{i=0}^{k-j-1} \alpha_{ik}\sqrt{P_{i,k-j}}x_{k-j}(b-j, w), \sum_{l=1}^{k-j}\sum_{i=0}^{l-1} \alpha_{ik}\sqrt{P_{il}}x_l(b-j, w_{b-j-l+1}) + \vec{Z}_k(b-j)\right\} \in A_\epsilon^{(T)}(\bar{P}_{k,j}, N_{k,j}). \qquad (46)$$

Let

$$\mathcal{W}_{b,k,j} := \{w \in \{1,\dots,2^{TR}\} : w \text{ satisfies } (46)\};$$
$$\mathcal{W}_{b,k} := \bigcap_{j=0}^{k-1} \mathcal{W}_{b,k,j}.$$

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,

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

Hence, by (45),

$$P_e = \sum_{b=1}^{B}\sum_{k=1}^{M} [\text{Prob}(w_{b-k+1} \notin \mathcal{W}_{b,k}) + \text{Prob}(w \in \mathcal{W}_{b,k} \text{ for some } w \neq w_{b-k+1})]. \qquad (47)$$

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,

$$\text{Prob}(w_{b-k+1} \notin \mathcal{W}_{b,k}) \;\leq\; \sum_{j=0}^{k-1} \text{Prob}(w_{b-k+1} \notin \mathcal{W}_{b,k,j}) \leq \sum_{j=0}^{k-1} \epsilon = k\epsilon \leq M\epsilon, \tag{48}$$

and

$$\begin{aligned}
\text{Prob}(w \in \mathcal{W}_{b,k} \text{ for some } w \neq w_{b-k+1}) \;&\leq\; \sum_{\substack{w' \in \{1,\ldots,2^{TR}\} \\ w' \neq w_{b-k+1}}} \text{Prob}(w' \in \mathcal{W}_{b,k}) \\
&= \sum_{\substack{w' \in \{1,\ldots,2^{TR}\} \\ w' \neq w_{b-k+1}}} \prod_{j=0}^{k-1} \text{Prob}(w' \in \mathcal{W}_{b,k,j}) \qquad (49) \\
&\leq (2^{TR}-1) \prod_{j=0}^{k-1} 2^{-T(S(\frac{\bar{P}_{k,j}}{N_{k,j}})-3\epsilon)} \\
&= (2^{TR}-1) 2^{-T(S(\frac{\bar{P}_k}{\sigma^2})-3k\epsilon)}.
\end{aligned}$$

The equality (49) 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 (7), by choosing T large enough, we can make ε_1 and ϵ small enough such that for any $\varepsilon_2 > 0$

$$\text{Prob}(w \in \mathcal{W}_{b,k} \text{ for some } w \neq w_{b-k+1}) \;\leq\; (2^{TR}-1) 2^{-T(S(\frac{\bar{P}_k}{\sigma^2})-3k\epsilon)} < \varepsilon_2. \tag{50}$$

Hence, by (47), (48) and (50),

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

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

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. □

4.4 An unbounded transport capacity can be obtained for bounded total power when attenuation is low

First we consider the linear case.

Proof of Theorem 3.10. 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:

$$R < \min_{1\leq j\leq n} S\left(\frac{1}{\sigma^2}\sum_{k=1}^{j}\left(\sum_{i=0}^{k-1}\frac{\sqrt{P_{ik}}}{(j-i)^{\delta}}\right)^2\right), \quad (51)$$

with the total power constraint $\sum_{k=1}^{n}\sum_{i=0}^{k-1} P_{ik} \leq P_{total}$. The intuitive interpretation of P_{ik} is the part of the power used by node i intended directly for node k.

We specifically choose

$$P_{ik} := \frac{P}{(k-i)^{\alpha}k^{\beta}}, \quad 0 \leq i < k \leq n, \quad (52)$$

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

$$P := \frac{(\alpha-1)(\beta-1)}{\alpha\beta}P_{total}. \quad (53)$$

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 R.H.S. of (51):

$$\sum_{k=1}^{j}\left(\sum_{i=0}^{k-1}\frac{\sqrt{P}}{(k-i)^{\alpha/2}k^{\beta/2}(j-i)^{\delta}}\right)^2 = \Omega(j^{3-\alpha-\beta-2\delta}). \quad (54)$$

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

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

where $c_0:=\left(\int_0^{1/2}\frac{1}{(1-y)^{\alpha/2}}dy\right)^2>0$. This establishes (54).

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

$$3-\alpha-\beta-2\delta>0. \tag{55}$$

Then by (54) and (55), there exists some $\underline{P}>0$ such that for any j,

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

Thus, by (51), for any n, any $R<S(\frac{\underline{P}}{\sigma^2})$ 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

$$4 - \alpha - \beta - 2\delta > 0. \tag{56}$$

Note that $3 - \alpha - \beta - 2\delta < 0$. Hence the minimum of (54) over $j = 1, 2, \ldots, n$ is attained at $j = n$. So by (56), we have

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

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

Now we turn to the planar case.

Proof of Theorem 3.9. The idea of the proof is similar to that of the linear case in Theorem 3.10. 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 (52) and coherently transmit.

Then by Theorem 3.12, the following rate is achievable

$$R < \min_{1 \le j \le r} S\left(\frac{1}{\sigma^2}\sum_{k=1}^{j}\left(\sum_{i=0}^{k-1}\frac{\sqrt{P_{ik}/n_i} \cdot n_i}{\rho^{\delta}_{\mathcal{N}_i\mathcal{N}_j}}\right)^2\right), \tag{57}$$

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 (57), the following rate is achievable:

$$R < \min_{1 \le j \le r} S\left(\frac{1}{\sigma^2}\sum_{k=1}^{j}\left(\sum_{i=0}^{k-1}\frac{\sqrt{P_{ik}} \cdot i^{q-1}}{3^{\delta} j^{q\delta}}\right)^2\right). \tag{58}$$

Similarly to the linear case, for $1+2q-\alpha-\beta>0$, we can prove the following lower bound usable in the R.H.S. of (58):

$$\sum_{k=1}^{j}\left(\sum_{i=0}^{k-1}\frac{\sqrt{P}\cdot i^{q-1}}{(k-i)^{\alpha/2}k^{\beta/2}3^{\delta}j^{q\delta}}\right)^2=\Omega(j^{1+2q-\alpha-\beta-2q\delta}). \tag{59}$$

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

$$\begin{aligned}
&\sum_{k=1}^{j}\left(\sum_{i=0}^{k-1}\frac{\sqrt{P}\cdot i^{q-1}}{(k-i)^{\alpha/2}k^{\beta/2}3^{\delta}j^{q\delta}}\right)^2\\
\geq\ & \frac{P}{3^{2\delta}j^{2q\delta}}\sum_{k=1}^{j}\left(\sum_{i=0}^{k-1}\frac{i^{q-1}}{(k-i)^{\alpha/2}}\right)^2\frac{1}{k^{\beta}}\\
=\ & \frac{P}{3^{2\delta}j^{2q\delta}}\sum_{k=1}^{j}\left(\frac{k^{q-1}}{k^{\alpha/2}}\sum_{i=1}^{k-1}\frac{(i/k)^{q-1}}{(1-i/k)^{\alpha/2}}\right)^2\frac{1}{k^{\beta}}\\
\geq\ & \frac{P}{3^{2\delta}j^{2q\delta}}\sum_{k=1}^{j}\left(\frac{k^{q-1}}{k^{\alpha/2}}\int_0^{k-1}\frac{(x/k)^{q-1}}{(1-x/k)^{\alpha/2}}dx\right)^2\frac{1}{k^{\beta}}\\
=\ & \frac{P}{3^{2\delta}j^{2q\delta}}\sum_{k=1}^{j}\left(\frac{k^{q}}{k^{\alpha/2}}\int_0^{\frac{k-1}{k}}\frac{y^{q-1}}{(1-y)^{\alpha/2}}dy\right)^2\frac{1}{k^{\beta}}\\
\geq\ & \frac{c_0P}{3^{2\delta}j^{2q\delta}}\sum_{k=2}^{j}k^{2q-\alpha-\beta}\geq\frac{c_0P}{3^{2\delta}j^{2q\delta}}\int_2^j x^{2q-\alpha-\beta}dx\quad(\text{for } j\geq 2) &(60)\\
=\ & \frac{c_0P}{3^{2\delta}j^{2q\delta}}\frac{j^{1+2q-\alpha-\beta}-2^{1+2q-\alpha-\beta}}{1+2q-\alpha-\beta}\\
=\ & \Omega(j^{1+2q-\alpha-\beta-2q\delta}), &(61)
\end{aligned}$$

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 (60) holds for any value of $2q-\alpha-\beta$. This establishes (59).

Now we proceed with two cases.

Case 1. $\delta<1$.

In this case, we choose q such that

$$1+2q-\alpha-\beta-2q\delta>0. \tag{62}$$

Then by (59) and (62), 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 (58), 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

$$1 + 3q - \alpha - \beta - 2q\delta > 0. \tag{63}$$

Then by (59) and (63), we have

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

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

4.5 A transport capacity of $\Omega(n)$ is achievable in planar networks, and feasible rates

First we show what is achievable in a regular planar network.

Proof of Theorem 3.2. 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 $P = P_{ind} - \epsilon$, 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):

$$R_\ell < S\left(\frac{e^{-2\gamma}P}{c_2(\gamma,\delta)P + \sigma^2}\right),$$

provided $c_2(\gamma, \delta)P$ is an upper bound on the interference, i.e.,

$$c_2(\gamma,\delta)P \geq \sum_{\substack{i \in \mathcal{N} \\ i \neq \ell, d_\ell}} \frac{e^{-2\gamma\rho_{id_\ell}}P}{\rho_{id_\ell}^{2\delta}}, \tag{64}$$

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 right hand side of (64) for $\delta > 1$, since, irrespective of the number n of nodes in $\mathcal{N}$,

$$\begin{aligned}
\sum_{\substack{i\in\mathcal{N}\\ i\neq \ell, d_\ell}} \frac{P}{\rho_{id_\ell}^{2\delta}} &\leq 4\times\left(2\sum_{i=1}^{\infty}\frac{1}{i^{2\delta}} + \int_1^\infty\int_0^\infty \frac{1}{(x^2+y^2)^\delta}dxdy\right)P \\
&\leq 4\times\left(2\cdot\frac{2\delta}{2\delta-1}+\frac{\pi}{4\delta-4}\right)P \\
&\leq \frac{16\delta^2+(2\pi-16)\delta-\pi}{(\delta-1)(2\delta-1)}P \\
&\leq c_2(\gamma,\delta)P.
\end{aligned}$$

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

$$\begin{aligned}
\sum_{\substack{i\in\mathcal{N}\\ i\neq \ell, d_\ell}} \frac{e^{-2\gamma\rho_{id_\ell}}P}{\rho_{id_\ell}^{2\delta}} &\leq 4\times\left(2\sum_{i=1}^{\infty}e^{-2\gamma i} + \int_1^\infty\int_0^\infty e^{-2\gamma(x^2+y^2)^{1/2}}dxdy\right)P \\
&\leq 4\times\left(\frac{2e^{-2\gamma}}{1-e^{-2\gamma}}+\frac{e^{-2\gamma}}{2\gamma}\right)P \\
&\leq \frac{4(1+4\gamma)e^{-2\gamma}-4e^{-4\gamma}}{2\gamma(1-e^{-2\gamma})}P \\
&\leq c_2(\gamma,\delta)P.
\end{aligned}$$

Hence the total achievable transport capacity is $n\cdot S\left(\frac{e^{-2\gamma}P}{c_2(\gamma,\delta)P+\sigma^2}\right)$, for every $P < P_{ind}$, establishing the result of Theorem 3.2. □

Proof of Theorem 3.3. 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 that for any $d_0 \geq \rho_{\min}$,

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

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

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

and for $\gamma > 0$,

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

The rest of the proof follows as above in Theorem 3.2. □

Proof of Theorem 3.4. 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). (Since a node may serve as a source for several destinations, or as a destination for several sources, the definition of feasible rate in Section 3.1 has to be modified appropriately).

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 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] in the uniform weak law of large numbers of Vapnik-Chervonenkis [26], it follows that Prob(Every mini-square has no more than $c'\sqrt{n\log n}$ straight lines passing through it) $\to 1$, as $n \to \infty$. Now suppose that every source-destination pair carries a traffic of rate less than $\frac{R_{\min}}{c'\sqrt{n\log n}}$. Then Prob(Every node needs to send no more than rate $R_{\min}$ to one of its four nearest neighbors) $\to 1$, as $n \to \infty$.

However, as already shown in the proof of Theorem 3.2, 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 $n \to \infty$. □

4.6 Networks with transport capacity $C_T = \Theta(n^\theta)$ for $1 < \theta < 2$ under low attenuation, and the optimality of coherent multi-stage relaying with interference cancellation

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

Proof of Theorem 3.6. 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

$$R < \min_{1 \le j \le n} S\left(\frac{1}{\sigma^2}\sum_{k=1}^{j}\left(\sum_{i=0}^{k-1}\frac{\sqrt{P'_{ik}}}{(j^\theta - i^\theta)^\delta}\right)^2\right), \tag{65}$$

with

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

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

Similarly to (54), we can prove the following lower bound on the R.H.S. of (65):

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

If $3 - \alpha - 2\theta\delta > 0$, then the minimum over $1 \le j \le n$ 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:

$$\sum_{i=0}^{j-1}\frac{P_{ind}}{(j^\theta - i^\theta)^{2\delta}} + \sum_{i=j+1}^{n}\frac{P_{ind}}{(i^\theta - j^\theta)^{2\delta}} \le \sum_{i=0}^{j-1}\frac{P_{ind}}{(j-i)^{2\delta}} + \sum_{i=j+1}^{n}\frac{P_{ind}}{(i-j)^{2\delta}} \le \frac{4\delta}{2\delta - 1}P_{ind} < \infty.$$

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 for this one source case.

Hence $\Theta(n^\theta)$ is the optimal scaling. It is achieved by coherent multi-stage relaying with interference cancellation, which is therefore the optimal strategy for information transmission in the networks. □

5 Concluding remarks

We have examined the problem of how much information can be transported over wireless networks, and what are the optimal strategies for doing so. In the best tradition of information theory, one wishes to determine the ultimate limits 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, only scratch the surface. They do not begin to exhaust the realm of the possible. Also, even simple networks, such as the three node relay channel, or the two-by-two interference cannel, are unsolved to date.

We make progress in this area by asking for less. Instead of studying just the capacity region, which is the set of all vectors of feasible rates, we study the distance-weighted sum of rates $\sum R_\ell \cdot \rho_\ell$, which we have called the transport capacity. There is a second sense in which we ask for less. We study scaling laws for the transport capacity as the number n of nodes in the network grows. The preconstant in the scaling law is of course important, but it is secondary to the rate of growth. In any case, we provide bounds for the preconstant for every n, thus characterizing the optimal achievable, at least in some scenarios. Finally, distance plays an explicit role in our theory in that we explicitly model signal attenuation.

Two broad results may be worthy of note. When either there is absorption ($\gamma > 0$), or the path loss exponent $\delta > 3$, $O(n)$ is an upper bound on the transport capacity of all planar networks. This upper bound can be realized in regular planar networks by multi-hop operation, which is consequently the optimal strategy for the nodes to cooperate, at least up to order. Packets need only be relayed from node to node, with all interference simply being regarded as noise at each hop. This mode of operation is currently the subject of much attention in the protocol development community.

In contrast, when there is absolutely no absorption ($\gamma = 0$) and the attenuation is very low with path loss exponent $\frac{1}{2} < \delta < 1$, there are networks where the transport capacity is $\Theta(n^{\frac{1}{\delta}-\epsilon})$. The strategy which realizes this, and which is consequently an optimal strategy, is coherent multi-stage relaying with interference cancellation: At each stage of relaying, all upstream nodes coherently transmit, and all receivers use interference cancellation at each stage. An achievable rate, superior to earlier results, is given for the Gaussian multiple relay channel with a single source-destination pair employing such a strategy.

Open questions abound. What happens for intermediate values of the path loss exponent, when there is absolutely no absorption, is still unresolved. Our channel model is simplistic. 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 Communications Review*, vol. 24, no. 4, pp. 234–244, 1994. (ACM SIGCOMM '94).

[2] C. Perkins, E. Royer, and S. Das, "Ad hoc on demand distance vector (AODV) routing." http://www.ietf.org/internet-drafts/draft-ietf-manet-aodv-03.txt. IETF Internet Draft (work in progress), June 1999.

[3] J. Broch, D. Johnson, and D. Maltz, "The dynamic source routing protocol for mobile ad hoc networks." http://www.ietf.org/internetdrafts/draft-ietf-manet-dsr-01.txt, December 1998. IETF Internet Draft (work in progress).

[4] IEEE Computer Society LAN MAN Standards Committee, "Wireless LAN medium access control (MAC) and physical layer (PHY) specifications." IEEE Standard 802.11-1997. The Institute of Electrical and Electronics Engineers, New York, NY, 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 *European Wireless Conference – Next Generation Wireless Networks: Technologies, Protocols, Services and Applications*, (Florence, Italy), pp. 156–162, February 25–28, 2002.

[6] R. Rozovsky and P. R. Kumar, "SEEDEX: A MAC protocol for ad hoc networks," in *Proceedings of the 2001 ACM International Symposium on Mobile Ad Hoc Networking and Computing*, (Long Beach, CA), pp. 67–75, October 4-5, 2001.

[7] P. R. Kumar, "New technological vistas for systems and control: The example of wireless networks," *IEEE Control Systems Magazine*, vol. 21, pp. 24–37, February 2001.

[8] P. Gupta and P. R. Kumar, "The capacity of wireless networks," *IEEE Transactions on Information Theory*, vol. IT-46, pp. 388–404, March 2000.

[9] P. Gupta, *Design and Performance Analysis of Wireless Networks.* PhD thesis, University of Illinois, Urbana-Champaign, August 2000. Department of Electrical and Computer Engineering.

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

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

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

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

[14] R. Ahlswede, "Multi-way communication channels," in *Proceedings of the 2nd Int. Symp. Inform. Theory (Tsahkadsor, Armenian S.S.R.)*, (Prague), pp. 23–52, Publishing House of the Hungarian Academy of Sciences, 1971.

[15] H. Liao, *Multiple access channels.* PhD thesis, University of Hawaii, Honolulu, HA, 1972. Department of Electrical Engineering.

[16] M. Gastpar and M. Vetterli, "On the capacity of wireless networks: The relay case." Proceedings of INFOCOM 2002. To appear.

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

[18] C. E. Shannon, "Two-way communication channels," in *Proceedings of the 4th Berkeley Symp. Math. Stat. Prob.*, (Berkeley, CA), pp. 611–644, University of California Press, 1961.

[19] A. E. Gamal, "The capacity region of a class of deterministic interference channels," *IEEE Transactions on Information Theory*, vol. 28, pp. 343–346, 1982.

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

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

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

[23] B. Schein and R. Gallager, "The Gaussian parallel relay network," in *Proceedings of the 2000 IEEE Int. Symp. Inform. Theory*, (Sorrento, Italy), p. 22, 2000.

[24] M. Franceschetti, J. Bruck, and L. Schulman, "Microcellular systems, random walks and wave propagation." To appear in *Proceedings of the IEEE Symposium on Antennas and Propagation Society* (IEEE AP-S 2002, San Antonio, TX), June 2002.

[25] P. Gupta and P. R. Kumar, "Towards an information theory of large networks: An achievable rate region." Preprint, September 28 2001. http://black.csl.uiuc.edu/∼prkumar/ps_files/net_info.ps.

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