

Interference Networks with Point-to-Point Codes

Francois Baccelli, Abbas El Gamal, and David Tse
INRIA-ENS, Stanford University, UC Berkeley

Abstract—The paper establishes the capacity region of the Gaussian interference channel with many transmitter-receiver pairs constrained to use point-to-point codes. The capacity region is shown to be strictly larger in general than the achievable rate regions when treating interference as noise, using successive interference cancellation decoding, and using joint decoding. The gains in coverage and achievable rate using the optimal decoder are analyzed in terms of ensemble averages using stochastic geometry. In a spatial network where the nodes are distributed according to a Poisson point process and the channel path loss exponent is $\beta > 2$, it is shown that the density of users that can be supported by treating interference as noise can scale no faster than $B^{2/\beta}$ as the bandwidth B grows, while the density of users can scale linearly with B under optimal decoding.

Index Terms—Network information theory, interference, successive interference cancellation, joint decoding, stochastic geometry, coverage, ad hoc network, stochastic network, performance evaluation.

I. INTRODUCTION

Most wireless communication systems employ point-to-point codes with receivers that treat interference as noise (IAN). This architecture is also assumed in most wireless networking studies. While using point-to-point codes has several advantages, including leveraging many years of development of good codes and receiver design for the point-to-point AWGN channel and requiring no significant coordination between the transmitters, treating interference as noise is not necessarily the optimal decoding rule. Motivated by results in network information theory, recent wireless networking studies have considered point-to-point codes with successive interference cancellation decoding (SIC) (e.g., see [8]), where each receiver decodes and cancels the interfering codewords from other transmitters one at a time before decoding the codeword from its tagged transmitter, and joint decoding [2] (JD), where the receiver treats the

network as a multiple access channel and decodes all the messages jointly.

In this paper, we ask a more fundamental question: given that transmitters use point-to-point codes, what is the performance achievable by the optimal decoding rule? The context we consider is a wireless network of multiple transmitter-receiver pairs, modeled as a Gaussian interference channel. The first result we establish in this direction is the capacity region of this channel when all the transmitters use Gaussian point-to-point codes. We show that none of the above decoding rules alone is optimal. Rather, a combination of treating interference as noise and joint decoding is shown to be capacity-achieving. Second, we show that this result can be extended to the case when the transmitters are only constrained to use codes that are capacity-achieving for the point-to-point and multiple access channels, but not necessarily Gaussian-like.

We then specialize the results to find a simple formula for computing the symmetric capacity for these codes. Assuming a wireless network model with users distributed according to a spatial Poisson process, we use simulations to study the gain in achievable symmetric rate and coverage when the receivers use the optimal decoding rule (OPT) for point-to-point Gaussian codes as compared to treating interference as noise, successive cancellation decoding, and joint decoding. We then use stochastic geometry techniques to study the performance in the wideband limit, where a high density of users share a very wide bandwidth. Under a channel model where the attenuation with distance is of the form $r^{-\beta}$ with $\beta > 2$, it is shown that the density of users that can be supported by treating interference as noise can scale no faster than $B^{2/\beta}$ as the bandwidth B grows, while the density of users can scale *linearly* with B under optimal decoding. For an attenuation of the form $(k+r)^{-\beta}$, the density of users scales linearly with B , but when the distance between the tagged transmitter and its receiver tends to infinity, the rate for OPT scales like the wideband capacity of a point-to-point Gaussian channel without interference.

F. Baccelli was a Miller Professor at UC Berkeley when this research was initiated. His work is supported in part by a grant of the INRIA@SiliconValley programme. The research of A. El Gamal is supported in part by DARPA ITMANET. The research of D. Tse is supported in part by the National Science Foundation under grant 0722032 and by the AFOSR under grant number FA9550-09-1-0317.

II. CAPACITY REGION WITH GAUSSIAN POINT-TO-POINT CODES

Consider a Gaussian interference channel with $K + 1$ transmitter-receiver pairs, where each transmitter $j \in [0 : K]$ wishes to send an independent message $M_j \in [1 : 2^{nR_j}]$ to its corresponding receiver j at rate R_j (in the unit of bits/s/Hz). The signal at receiver j when the complex signals $\mathbf{X} = (X_0, X_1, \dots, X_K)$ are transmitted is

$$Y_j = \sum_{l=0}^K g_{jl} X_l + Z_j \quad \text{for } j \in [0 : K],$$

where g_{jl} are the complex channel gains and $Z_j \sim \mathcal{CN}(0, 1)$ is a complex circularly symmetric Gaussian noise with an average power of 1. We assume each transmitter is subject to the same power constraint Q (in the unit of Watts/Hz). Define the received power from transmitter l at receiver j as $P_{jl} = |g_{jl}|^2 Q$. Without further constraints on the transmitters' codes, the capacity region of this channel is not known even for the two transmitter-receiver pair case (see [6] for known results on this problem). In this section we establish the capacity region using Gaussian generated point-to-point codes for an arbitrary number of transmitter-receiver pairs.

We define an $(n, 2^{nR_0}, \dots, 2^{nR_K})$ Gaussian point-to-point (G-ptp) code¹ to consist of a set of randomly and independently generated codewords $x_j^n(m_j) = (x_{j_1}, x_{j_2}, \dots, x_{j_n})(m_j)$, $m_j \in [1 : 2^{nR_j}]$, $j \in [0 : K]$, each according to an i.i.d. $\mathcal{CN}(0, \sigma^2)$ sequence, for some $0 < \sigma^2 \leq Q$. We assume each transmitter in the Gaussian interference channel uses such a code with each receiver $j \in [0 : K]$ assigning an estimate $\hat{m}_j(y_j^n) \in [1 : 2^{nR_j}]$ of message m_j to each received sequence y_j^n . We define the probability of error for a G-ptp code as

$$p_n = \frac{1}{K+1} \sum_{j=0}^K P\{\hat{M}_j \neq M_j\}.$$

We denote the average of this probability of error over G-ptp codes as \bar{p}_n . A rate tuple $\mathbf{R} = (R_0, R_1, \dots, R_K)$ is said to be achievable via a sequence of $(n, 2^{nR_0}, \dots, 2^{nR_K})$ G-ptp codes if $\bar{p}_n \rightarrow 0$ as $n \rightarrow \infty$. The capacity region with G-ptp is the closure of the set of achievable rate tuples (R_0, R_1, \dots, R_K) .

Remarks:

- 1) Our definition of codes precludes the use of time sharing and power control (although in general one can use time sharing with ptp codes). The

justification is that time sharing (or the special cases of time/frequency division) require additional coordination.

- 2) Note that if a rate tuple is achievable via a sequence of G-ptp codes then there exists a sequence of (deterministic) codes that achieves this rate tuple. We use the definition of achievability via the average probability of error over codes to simplify the proof of the converse. The results, however, can be shown to apply to sequences of G-ptp codes almost surely, and to an even more general class of (deterministic) codes in Section III.

Let \mathcal{S} be a nonempty subset of $[0 : K]$ and $\mathcal{S}^c = [0 : K] \setminus \mathcal{S}$ be its complement. Define $X_{\mathcal{S}}$ to be the vector of transmitted signals X_l such that $l \in \mathcal{S}$, and define the sum $X_j(\mathcal{S}) = \sum_{l \in \mathcal{S}} g_{jl} X_l$. Similarly define $P_j(\mathcal{S}) = \sum_{l \in \mathcal{S}} P_{jl}$, $R_{\mathcal{S}}$, and $R(\mathcal{S}) = \sum_{l \in \mathcal{S}} R_l$.

Consider a Gaussian multiple access channel (MAC) with transmitters $X_{\mathcal{S}}$, receiver Y_j , where $j \in \mathcal{S}$, and additive Gaussian noise power $P_j(\mathcal{S}^c) + 1$. Recall that the capacity region $\mathcal{A}_j(\mathcal{S})$ of this MAC is

$$\left\{ R_{\mathcal{S}} : R(\mathcal{T}) \leq C \left(\frac{P_j(\mathcal{T})}{1 + P_j(\mathcal{S}^c)} \right) \text{ for every } \mathcal{T} \subseteq \mathcal{S} \right\},$$

where $C(x) = \log(1 + x)$ for $x \geq 0$. All logarithms are base 2 in this paper.

Now, define the rate regions

$$\mathcal{C}_j = \{\mathbf{R} : R_{\mathcal{S}} \in \mathcal{A}_j(\mathcal{S}) \text{ for some } \mathcal{S} \text{ containing } j\}.$$

and

$$\mathcal{C} = \bigcap_{j=0}^K \mathcal{C}_j. \quad (1)$$

One of the main results in this paper is establishing the capacity region of the Gaussian interference channel with G-ptp codes.

Theorem 1: The capacity region of the Gaussian $K + 1$ transmitter-receiver pair interference channel with G-ptp codes is \mathcal{C} .

By symmetry of the capacity expression, we only need to establish achievability and the converse for the rate region \mathcal{C}_0 , which ensures reliable decoding of transmitter 0's message at receiver 0. Hence from this point onward, we focus on receiver 0. We will refer to this receiver and its corresponding transmitter 0 as *tagged*. We also refer to other transmitters as *interferers*. We relabel the signal from the tagged receiver, its gains, and additive noise as

$$Y = \sum_{l=0}^K g_l X_l + Z.$$

We also relabel the received power from the tagged transmitter 0 as P_0 and the received power from interferer

¹By a code here we just mean the message set and the codebook.

$j, j \geq 1$, as I_j (for interference). For any subset of interferers \mathcal{T} , we denote $I(\mathcal{T})$ as the sum of the received power from these interferers. We will also drop the index 0 from the notations $\mathcal{A}_0(\mathcal{S})$ and $X_0(\mathcal{S})$.

For clarity of presentation, first consider the case of $K = 1$. Here the signal of the tagged receiver is

$$Y = g_0 X_0 + g_1 X_1 + Z.$$

For this receiver, there are two subsets to consider, $\mathcal{S} = \{0\}$ and $\mathcal{S} = \{0, 1\}$. The region $\mathcal{A}(\{0\})$ is the set of rate pairs (R_0, R_1) such that

$$R_0 \leq C\left(\frac{P_0}{1 + I_1}\right),$$

and the region $\mathcal{A}(\{0, 1\})$ is the set of rate pairs (R_0, R_1) such that

$$\begin{aligned} R_0 &\leq C(P_0), \\ R_1 &\leq C(I_1), \end{aligned}$$

$$R_0 + R_1 \leq C(P_0 + I_1).$$

Hence, the region \mathcal{C}_0 for the tagged receiver is the union of these two regions.

It is interesting to compare \mathcal{C}_0 to the achievable rate regions for other schemes that use point-to-point codes. Define the rate regions:

$$\begin{aligned} \mathcal{R}_{\text{IAN}} &= \mathcal{A}(\{0\}), \\ \mathcal{R}_{\text{SIC}} &= \left\{ (R_0, R_1) : R_0 \leq C(P_0), R_1 \leq C\left(\frac{I_1}{1 + P_0}\right) \right\}, \\ \mathcal{R}_{\text{JD}} &= \mathcal{A}(\{0, 1\}). \end{aligned}$$

The region \mathcal{R}_{IAN} is achieved by a receiver that decodes the tagged transmitter's message while treating interference as Gaussian noise. The region \mathcal{R}_{SIC} is achieved by the successive interference cancellation receiver; the interferer's message is first decoded, treating the tagged transmitter's signal as Gaussian noise with power P , and then the message from the tagged receiver is decoded after canceling the interferer's signal. The region \mathcal{R}_{JD} is the two transmitter-receiver pair Gaussian MAC capacity region. It is the set of achievable rates when the receiver insists on correctly decoding both messages, which is the achievable region using joint decoding in Blomer and Jindal [2].

It is not difficult to see that the following relationships between the regions hold (see Figure 1):

$$\begin{aligned} \mathcal{R}_{\text{IAN}} &\subset \mathcal{C}_0, \\ \mathcal{R}_{\text{SIC}} &\subset \mathcal{R}_{\text{JD}} \subset \mathcal{C}_0, \\ \mathcal{C}_0 &= \mathcal{R}_{\text{IAN}} \cup \mathcal{R}_{\text{JD}}. \end{aligned}$$

Note that the last relationship above says that the

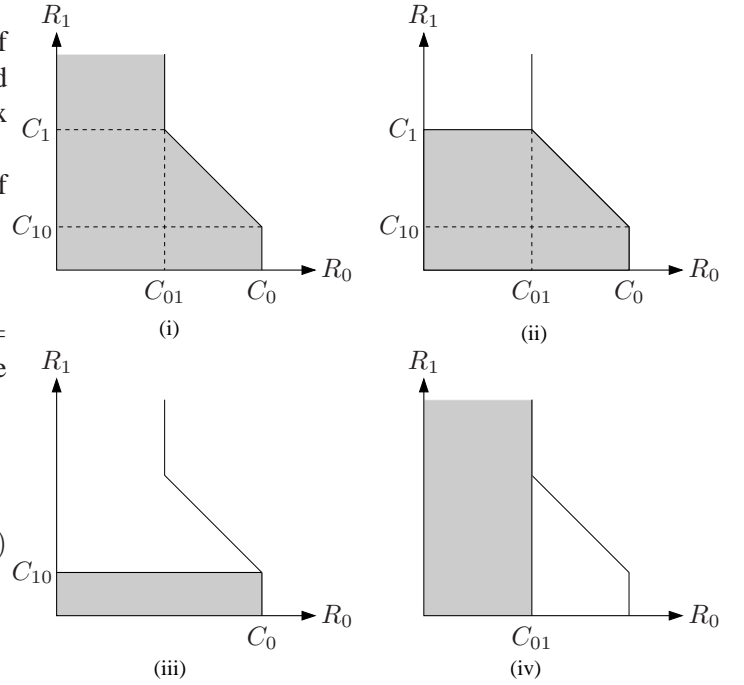


Fig. 1. \mathcal{C}_0 is the shaded region in Figure (i). The \mathcal{R}_{JD} region is depicted on Figure (ii). \mathcal{R}_{SIC} is on Figure (iii) and \mathcal{R}_{IAN} is on (iv). $C_0 = C(P_0)$, $C_1 = C(I_1)$, $C_{01} = C(P_0/(1 + I_1))$, $C_{10} = C(I_1/(1 + P_0))$.

receiver can do no better than treating interference as Gaussian noise or jointly decoding the messages from the tagged transmitter and the interferer.

In the following, we first establish the capacity region for the case $K = 1$, and then extend the result to arbitrary K . In Section III, we also show that our results extend to the class of *MAC capacity-achieving codes*.

A. Proof of Theorem 1 for $K = 1$

Proof of Achievability. To prove the achievability of any rate pair in the interior of \mathcal{C}_0 , we use Gaussian ptp codes with average power $Q(1 - \delta)$ and joint typicality decoding as in [4]. Further, we use simultaneous decoding [6] in which receiver 0 declares that the message \hat{m}_0 is sent if it is the unique message such that $(x_0^n(\hat{m}_0), y^n)$ is jointly typical or $(x_0^n(\hat{m}_0), x_1^n(\hat{m}_1), y^n)$ is jointly typical for some $m_1 \in [1 : 2^{nR_1}]$. A straightforward analysis of the average probability of error shows that $\bar{p} \rightarrow 0$ as $n \rightarrow \infty$ if either

$$R_0 < C\left(\frac{P_0}{1 + I_1}\right), \quad (2)$$

or

$$\begin{aligned} R_0 &< C(P_0), \\ R_0 + R_1 &< C(P_0 + I_1). \end{aligned}$$

The first constraint (2) is $\mathcal{A}(\{0\})$, the IAN region. Denote the region defined by the second set of constraints by $\mathcal{F}(\{0, 1\})$; it is the same as the MAC region $\mathcal{A}(\{0, 1\})$ but with the constraint on R_1 removed. Hence, the resulting achievable rate region $\bar{\mathcal{C}}_0 = \mathcal{A}(\{0\}) \cup \mathcal{F}(\{0, 1\})$ appears to be larger than $\mathcal{C}_0 = \mathcal{A}(\{0\}) \cup \mathcal{A}(\{0, 1\})$. It is easy to see from Figure 1, however, that it actually coincides with \mathcal{C}_0 . Hence, receiver 0 can correctly decode M_1 if treating interference as noise fails but simultaneous decoding succeeds even though it does not require it. We establish the converse for the original characterization of \mathcal{C}_0 , hence providing an alternative proof that the two regions coincide.

Remark: Although we presented the decoding rule as a two-step procedure, since the receiver knows the transmission rates, it already knows whether to apply IAN or simultaneous decoding.

Proof of the converse. To prove the converse, suppose we are given a sequence of random G-ptp codes and decoders with rate pair (R_0, R_1) and such that the average probability of error approaches 0 as $n \rightarrow \infty$. We want to show that $(R_0, R_1) \in \mathcal{C}_0$. Consider two cases:

- 1) $R_1 < C(I_1)$: Under this condition and by the assumption that the tagged receiver can reliably decode its message, the tagged receiver can cancel off the received signal from the tagged transmitter and then reliably decode the message from transmitter 1. Hence (R_0, R_1) is in the capacity region of the MAC with transmitters (X_0, X_1) and receiver Y , and hence in \mathcal{C}_0 .
- 2) $R_1 \geq C(I_1)$: Fix an $\epsilon > 0$, and let $Z = U + V$, where U and V are independent Gaussian noise components with variances N and $1 - N$, respectively, such that

$$C\left(\frac{I_1}{N}\right) = R_1 + \epsilon.$$

Consider the AWGN channel

$$W = g_1 X_1 + U. \quad (3)$$

Since we are assuming G-ptp codes and $R_1 < C(I_1/N)$, the average probability of decoding error over this channel approaches zero as $n \rightarrow \infty$. Hence, by Fano's inequality, the mutual information over a block of n symbols, averaged over G-ptp codes, is

$$\begin{aligned} \bar{I}(X_1^n; W^n) &= \bar{h}(X_1^n) - \bar{h}(X_1^n | W^n) \\ &\geq nR_1 - n\delta_n, \end{aligned}$$

where $\delta_n \rightarrow 0$ as $n \rightarrow \infty$. Denoting by $\bar{h}(W^n)$ the differential entropy of W^n averaged over the

G-ptp codes, this implies that

$$\begin{aligned} \bar{h}(W^n) &\geq nR_1 - n\delta_n + h(U^n) \\ &= nR_1 - n\delta_n + n \log(\pi e N) \\ &= nC\left(\frac{I_1}{N}\right) - n\epsilon - n\delta_n + n \log(\pi e N) \\ &= n \log(\pi e(I_1 + N)) - n\epsilon - n\delta_n. \end{aligned}$$

Now, let $\tilde{W}^n = W^n + V^n$. By the conditional entropy power inequality, we have

$$\begin{aligned} 2^{\frac{1}{n}\bar{h}(\tilde{W}^n)} &\geq 2^{\frac{1}{n}\bar{h}(W^n)} + 2^{\frac{1}{n}\bar{h}(V^n)} \\ &\geq 2^{\log(\pi e(I_1 + N)) - \delta_n - \epsilon} + \pi e(1 - N) \\ &= \pi e(I_1 + N)2^{-\delta_n - \epsilon} + \pi e(1 - N). \end{aligned}$$

Hence,

$$\bar{h}(\tilde{W}^n) \geq n \log\left(\pi e(I_1 + N)2^{-\delta_n - \epsilon} + \pi e(1 - N)\right).$$

The fact that $\bar{h}(Y^n) \leq n \log(\pi e(P_0 + I_1 + 1))$ and the last lower bound give an upper bound on the average mutual information for the tagged transmitter-receiver pair

$$\begin{aligned} \bar{I}(X_0^n; Y^n) &= \bar{h}(Y^n) - \bar{h}(\tilde{W}^n) \\ &\leq n \log(\pi e(P_0 + I_1 + 1)) \\ &\quad - n \log(\pi e(I_1 + N)2^{-\delta_n - \epsilon} + \pi e(1 - N)). \end{aligned}$$

Since this is true for all $\epsilon > 0$, we have

$$\begin{aligned} \bar{I}(X_0^n; Y^n) &\leq n \log(\pi e(P_0 + I_1 + 1)) \\ &\quad - n \log\left(\pi e(I_1 + N)2^{-\delta_n} + \pi e(1 - N)\right) \\ &\leq nC\left(\frac{P_0}{1 + I_1}\right) + n\tilde{\delta}_n. \end{aligned}$$

Since we assume the tagged receiver can decode its intended message, $R_0 < C(P_0/(1 + I_1))$, and hence $(R_0, R_1) \in \mathcal{C}_0$. This completes the proof of Theorem 1 for $K = 1$.

Remarks:

- 1) What the above proof showed is that if the message of transmitter 0 is reliably decoded, then either: (1) the interferer's message can be jointly decoded as well, in which case the rate vector \mathbf{R} is in the 2-transmitter MAC capacity region, or (2) the interference plus the background noise is close to i.i.d. Gaussian, in which case decoding transmitter 0's message treating transmitter 1's interference plus background noise as Gaussian is optimal.
- 2) One may think that since the interferer uses a Gaussian random code, the interference must be Gaussian and hence the interference plus background noise must also be Gaussian. This thinking

is misguided, however, since what is important to the communication problem are the statistics of the interference plus noise *conditional on a realization of the interferer's random code*. Given a realization of the code, the interference is discrete, coming from a code, and hence it is not in general true that the interference plus noise is close to i.i.d. Gaussian. What we showed in the above converse is that this holds when the message from the interferer cannot be jointly decoded with the message from transmitter 0.

B. Proof of Theorem 1 for arbitrary K

Now, consider the general case with $K+1$ transmitter-receiver pairs.

Proof of achievability. The proof is a straightforward generalization of the proof for $K = 1$, and the condition for the probability of error to approach 0 is that the rate vector \mathbf{R} lies in the region:

$$\bar{\mathcal{C}}_0 := \{\mathbf{R} : R_{\mathcal{S}} \in \mathcal{F}(\mathcal{S}) \text{ for some subset } \mathcal{S} \text{ with } 0 \in \mathcal{S}\}, \quad (4)$$

where

$$\mathcal{F}(\mathcal{S}) = \left\{ \mathbf{R} : R(\mathcal{T} \cup \{0\}) < C \left(\frac{P_0 + I(\mathcal{T})}{1 + I(\mathcal{S}^c)} \right) \right. \\ \left. \text{for every } \mathcal{T} \subseteq \mathcal{S} \setminus \{0\} \right\}$$

is the *augmented* MAC region for the subset of transmitters \mathcal{S} treating the transmitters in \mathcal{S}^c as Gaussian noise.

As in the $K = 1$ case, the region $\bar{\mathcal{C}}_0$ appears to be larger than \mathcal{C}_0 . We again establish the converse for the original characterization of \mathcal{C}_0 , hence showing that $\bar{\mathcal{C}}_0$ coincides with \mathcal{C}_0 .

Proof of the converse. The proof for the $K = 1$ case identifies, for a given a rate vector, a *maximal* set of interferers whose messages can be jointly decoded with the tagged transmitter's message. This set depends on the given rates of the interferer; if $R_1 < C(I_1)$, the set is $\{1\}$, otherwise it is \emptyset . The key to the proof is to show that whichever the case may be, the residual interference created by the transmitters whose messages are *not* decoded plus the background noise must be asymptotically i.i.d. Gaussian. We generalize this proof to an arbitrary number of interferers. In this general setting, however, *explicitly* identifying a maximal set of interferers whose messages can be jointly decoded with the tagged transmitter's message is a combinatorially difficult task. Instead, we identify it *existentially*.

Suppose the transmission rate vector is \mathbf{R} and the average probability of error for the tagged receiver

approaches zero as $n \rightarrow \infty$. Consider the set of subsets of interferers

$$\mathcal{D} = \{\mathcal{T} : 0 \notin \mathcal{T}, R_{\mathcal{T}} \in \mathcal{A}(\mathcal{T})\}.$$

Intuitively, these are all the subsets of interferers whose messages can be jointly decoded after decoding M_0 while treating the other transmitted signals as Gaussian noise. Let \mathcal{T}^* be a *maximal* set in \mathcal{D} , i.e., there is no larger subset $\mathcal{T} \in \mathcal{D}$ that contains \mathcal{T}^* . Since the message M_0 is decodable by the assumption of the converse, the tagged receiver can cancel off the tagged transmitter's signal. Next, the messages of the interferers in \mathcal{T}^* can be decoded, treating the interference from the remaining interferers plus the background noise as Gaussian. This is because by assumption $R_{\mathcal{T}^*} \in \mathcal{A}(\mathcal{T}^*)$ and all interferers are using G-tp codes. After canceling off the signals from the interferers in \mathcal{T}^* , the tagged receiver is left with interferers in $(\mathcal{T}^* \cup \{0\})^c$. Since no further messages can be decoded treating the rest as Gaussian noise (by the maximality of \mathcal{T}^*), it follows that for any subset $\mathcal{S} \subset (\mathcal{T}^* \cup \{0\})^c$, $R_{\mathcal{S}}$ is not in the capacity region of the MAC with transmitters in \mathcal{S} and Gaussian noise with power $I((\mathcal{T}^* \cup \{0\})^c \setminus \mathcal{S}) + 1$. Let

$$W = X((\mathcal{T}^* \cup \{0\})^c) + Z.$$

In the $K = 1$ scenario, \mathcal{T}^* is either $\{1\}$ or \emptyset . In the first case, both messages are decoded, hence the power of the residual interference plus that of the background noise is automatically Gaussian. In the second case, the interferer's message is not decoded, and our earlier argument shows the interferer must be communicating above the capacity of the point-to-point Gaussian channel to receiver 0. Hence the aggregate interference plus noise must be asymptotically i.i.d. Gaussian. In the general scenario with K interferers, there may be more than one residual interferer left after decoding a maximal set \mathcal{T}^* . The following lemma, which is proved in the following subsection, shows that this situation generalizes appropriately.

Lemma 1: Consider a k -transmitter MAC

$$Y = \sum_{j=1}^k g_j X_j + Z,$$

where the received power from transmitter j is P_j and $Z \sim \mathcal{CN}(0, 1)$. Let

$$\mathcal{B} = \{\mathbf{R} : R_{\mathcal{S}} \in \mathcal{A}(\mathcal{S}) \text{ for some nonempty } \mathcal{S}\}. \quad (5)$$

If the transmitters use G-tp codes at rate vector \mathbf{R} and $\mathbf{R} \notin \mathcal{B}$, then

$$\lim_{n \rightarrow \infty} \frac{1}{n} \bar{h}(Y^n) = \log \left(\pi e \left(\sum_{j=1}^k P_j + 1 \right) \right),$$

that is, the received sequence Y^n is asymptotically i.i.d. Gaussian.

Lemma 1 shows that the interference after decoding the interferers in $\mathcal{T}^* \cup \{0\}$ plus the background noise is asymptotically i.i.d. Gaussian. Hence, $R_{\mathcal{T}^* \cup \{0\}} \in \mathcal{A}(\mathcal{T}^* \cup \{0\})$, and we can conclude that $\mathbf{R} \in \mathcal{C}_0$. This completes the converse proof of Theorem 1 for arbitrary K .

C. Proof of Lemma 1

The proof needs the following fact about \mathcal{B} . Recall that the boundary of the MAC capacity region consists of multiple faces. We refer to the one corresponding to the constraint on the total sum rate as the *sum rate face*.

Fact 1: Let \mathbf{R} be a rate vector such that $R_{\mathcal{S}}$ is on the boundary of $\mathcal{A}(\mathcal{S})$ for some \mathcal{S} but not on its sum-rate face. Then \mathbf{R} cannot be on the boundary of \mathcal{B} . In other words, the non-sum-rate faces of the MAC regions $\mathcal{A}(\mathcal{S})$ are never exposed on the boundary of \mathcal{B} .

Figure 2 depicts \mathcal{B} for $K = 2$. Here, the boundary of \mathcal{B} consists of three segments, each of which is a sum-rate face of a MAC region. The two non-sum-rate faces of $\mathcal{A}(\{1, 2\})$ are not exposed.

Proof of Fact 1: Let \mathbf{R} be a rate vector such that $R_{\mathcal{S}}$ is on the boundary of $\mathcal{A}(\mathcal{S})$ for some \mathcal{S} but not on its sum rate face. Then there is a subset \mathcal{T} of \mathcal{S} such that

$$R(\mathcal{T}) = C \left(\frac{P(\mathcal{T})}{1 + P(\mathcal{S}^c)} \right) \quad (6)$$

and for all subsets \mathcal{V} strictly containing \mathcal{T} and inside \mathcal{S} ,

$$R(\mathcal{V}) < C \left(\frac{P(\mathcal{V})}{1 + P(\mathcal{S}^c)} \right). \quad (7)$$

Subtracting (6) from (7) implies that for all such sets \mathcal{V} ,

$$R(\mathcal{V} \setminus \mathcal{T}) < C \left(\frac{P(\mathcal{V} \setminus \mathcal{T})}{1 + P(\mathcal{T}) + P(\mathcal{S}^c)} \right).$$

This implies that $R_{\mathcal{S} \setminus \mathcal{T}}$ is in the strict interior of $\mathcal{A}(\mathcal{S} \setminus \mathcal{T})$. Hence, \mathbf{R} cannot be on the boundary of \mathcal{B} . This completes the proof of Fact 1. ■

Proof of Lemma 1: The proof is by induction on the number of transmitters k .

$k = 1$: this just says that for a point-to-point Gaussian channel, if we transmit at a rate above capacity using a G-ptp code, then the output is Gaussian. This is a well-known fact.

Assume the lemma holds for all $j < k$. Consider the case with k transmitters.

Express $Z = U + V$, where U and V are independent Gaussians with variances N and $1 - N$, respectively,

where N is chosen such that \mathbf{R} is on the boundary of $\tilde{\mathcal{B}}$ for the MAC

$$\tilde{W} = \sum_{j=1}^k g_j X_j + U.$$

Here, $\tilde{\mathcal{B}}$ is the same as \mathcal{B} except that the background noise power 1 is replaced by N . Let \mathcal{E} be the collection of all subsets $\mathcal{S} \subset [1 : k]$ for which $R_{\mathcal{S}} \in \tilde{\mathcal{A}}(\mathcal{S})$ ($\tilde{\mathcal{A}}(\mathcal{S})$ is the same as $\mathcal{A}(\mathcal{S})$ except that the background noise power 1 is replaced by N). Pick a maximal subset \mathcal{S} from that collection. By Fact 1, $R_{\mathcal{S}}$ must be on the sum-rate face of $\tilde{\mathcal{A}}(\mathcal{S})$. The MAC can be decomposed as

$$\tilde{W} = X(\mathcal{S}) + X(\mathcal{S}^c) + U.$$

By the maximality of \mathcal{S} , no further transmitted messages can be decoded beyond the ones for the transmitters in \mathcal{S} (otherwise, there would exist a bigger subset \mathcal{S}' containing \mathcal{S} and for which $R_{\mathcal{S}'} \in \tilde{\mathcal{A}}(\mathcal{S}')$). This implies in particular that for any subset $\mathcal{T} \subset \mathcal{S}^c$, the rate vector $R_{\mathcal{T}}$ cannot be in the region $\tilde{\mathcal{A}}(\mathcal{T})$; otherwise if such a \mathcal{T} exists, the receiver could have first decoded the messages of transmitters in \mathcal{S} , cancelled their signals, and then decoded the messages of the transmitters in \mathcal{T} , treating the residual interference plus noise as Gaussian. Hence if we consider the smaller MAC

$$W = X(\mathcal{S}^c) + U,$$

we can apply the induction hypothesis to show that W^n is asymptotically i.i.d. Gaussian. So now we have a Gaussian MAC for transmitters in \mathcal{S}

$$\tilde{W} = X(\mathcal{S}) + W$$

and since the rate vector $R_{\mathcal{S}}$ lies on the sum rate boundary of this MAC, we now have a situation of a super-transmitter, i.e., a combination of all transmitters in \mathcal{S} , sending at the capacity of this Gaussian channel. Using a very similar argument as in the $K = 1$ proof, one can show that W^n is asymptotically i.i.d. Gaussian. Adding back the removed noise V yields the desired conclusion. This completes the proof of Lemma 1. ■

III. CAPACITY REGION WITH MAC-CAPACITY-ACHIEVING CODES

The converse in Theorem 1 says that if the transmitters use Gaussian random codes, then one can do no better than treating interference as Gaussian noise or joint decoding. The present section shows that this converse result generalizes to a certain class of (deterministic) ‘‘MAC-capacity-achieving’’ codes, to be defined

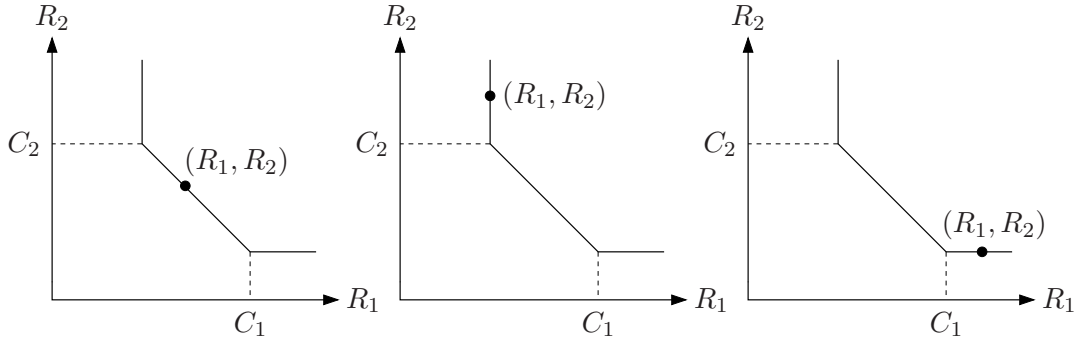


Fig. 2. The boundary of \mathcal{B} for $K = 2$ has three segments, all of which are sum-rate faces. A rate-tuple on the boundary of \mathcal{B} can lie on one of them.

precisely below. We first focus on the two-transmitter-receiver pair case and then generalize to the $K + 1$ -transmitter case.

An $(n, 2^{nR})$ (deterministic) single-user code satisfying the transmit power constraint Q is said to achieve a rate R over a point-to-point Gaussian channel $Y = gX + Z$ if the probability of error $p_n \rightarrow 0$ as the block length $n \rightarrow \infty$. An $(n, 2^{nR})$ code is said to be *point-to-point (ptp) capacity-achieving* if it achieves a rate of R over *every* point-to-point Gaussian channel with capacity greater than R .

Now consider the two transmitter-receiver pair Gaussian interference channel. A rate-pair (R_0, R_1) is said to be achievable over the interference channel via a sequence of ptp-capacity-achieving codes if there exists a sequence of such codes for each transmitter such that the probability of error

$$p_n = \frac{1}{2} \left(P\{\hat{M}_0 \neq M_0\} + P\{\hat{M}_1 \neq M_1\} \right)$$

approaches 0 as $n \rightarrow \infty$. The capacity region with ptp-capacity-achieving codes is the closure of the set of achievable rates. The theorem below is a counterpart to the converse in Theorem 1 for G-tp codes.

Theorem 2: The capacity region of the two transmitter-receiver pair interference channel with ptp-capacity achieving codes is no larger than \mathcal{C} , as defined in (1) for $K = 2$.

Proof: The result follows from the observation that in the proof of the converse for Theorem 1, the only property we used about the G-tp codes is that the average decoding error probability of the interferer's message after canceling the message of the intended transmitter goes to zero whenever $R_1 < C(I_1)$. This property remains true if the interferer uses a ptp-capacity-achieving code instead of a G-tp code. ■

Theorem 2 says that as long as the codes of the transmitters are designed to optimize point-to-point performance, the region \mathcal{C} is the fundamental limit on their

performance over the interference channel. This is true even if the codes do not “look like” randomly generated Gaussian codes.

Now let us consider the $K + 1$ -transmitter interference channel for general K . Is \mathcal{C} still an outer bound to the capacity region if all the transmitters use ptp-capacity-achieving codes? The answer is no. A counter-example can be found in [3] (Section IIB), which considers a 3-transmitter many-to-one interference channel with interference occurring only at receiver 0. There, it is shown that if each of the transmitters uses a lattice code, which is ptp-capacity-achieving, one can do better than *both* joint decoding all transmitters' messages and decoding just transmitter 0's message treating the rest of the signal as Gaussian noise at receiver 0. The key is to use lattice codes for transmitter 1 and 2, and have them *align* at receiver 0 so that the two interferers appear as one interferer. Hence, it is no longer necessary for receiver 0 to decode the messages of *both* interferers in order to decode the message from transmitter 0; decoding the *sum* of the two interferers is sufficient. At the same time, treating the interference from 1 and 2 as Gaussian noise is also strictly sub-optimal.

In this counter-example, the transmitters' codes are ptp-capacity-achieving but not “MAC capacity-achieving” in the sense that receiver 0 cannot jointly decode the individual messages of the interferers. A careful examination of the proof of the converse in Theorem 1 for general K reveals that the converse in fact holds whenever the codes of the transmitters satisfy such a MAC-capacity-achieving property.

Consider a k -transmitter Gaussian MAC

$$Y = \sum_{j=1}^k g_j X_j + Z$$

and a subset $\mathcal{S} \subset [1 : k]$. A $(n, 2^{nR_1}, \dots, 2^{nR_k})$ (deterministic) code for this MAC, where each transmitter satisfies the same transmit power constraint Q , is said to

achieve the rate-tuple R_S over the MAC if the probability of error

$$p_n(\mathcal{S}) = \frac{1}{|\mathcal{S}|} \sum_{j \in \mathcal{S}} P\{\hat{M}_j \neq M_j\}$$

approaches 0 as $n \rightarrow \infty$. An $(n, 2^{nR_1}, \dots, 2^{nR_k})$ code is said to be *MAC-capacity-achieving* if for every $\mathcal{S} \subset [1 : k]$, it achieves a rate R_S over every Gaussian MAC whose capacity region $\mathcal{A}(\mathcal{S})$ contains R_S . Recall that the region $\mathcal{A}(\mathcal{S})$ is the capacity region of the MAC with transmitters X_S and the signals from the rest of the transmitters treated as Gaussian noise. Thus this definition says that a MAC capacity-achieving code is good enough to achieve this performance for any subset \mathcal{S} of transmitters.

Now consider the $K+1$ transmitter-receiver pair Gaussian interference channel. A rate-tuple \mathbf{R} is said to be achievable on the interference channel via a sequence of MAC-capacity-achieving codes if there exists a sequence of MAC-capacity-achieving codes for every subset containing K transmitters such that the probability of error

$$p_n = \frac{1}{K+1} \sum_{j=0}^K P\{\hat{M}_j \neq M_j\}$$

approaches zero as $n \rightarrow \infty$. The capacity region with MAC-capacity-achieving codes is the closure of all such rates.

Theorem 3: The capacity region of the Gaussian $K+1$ -transmitter interference channel with MAC-capacity achieving codes is no larger than \mathcal{C} , as defined in (1).

Proof:

The result follows from the observation that in the proof of the converse in Theorem 1, the only property that was used about the G-ptp codes of the transmitters is precisely the MAC-capacity-achieving property defined above. ■

The counter-example above shows that one can indeed do better than the region \mathcal{C} , for example using interference alignment. Interference alignment, however, requires careful coordination and accurate channel knowledge at the transmitters. On the other hand, one can satisfy the MAC-capacity-achieving property without the need of such careful coordination. So, if one takes the MAC-capacity-achieving property as a definition of lack of coordination between the transmitters, then the above theorem delineates the fundamental limit to the performance on the interference channel if the transmitters are not coordinated.

IV. SYMMETRIC RATE

We specialize the results in the previous sections to the case when all messages have the same rate R .

This will help us compare the network performance of the optimal decoder to other decoders for Gaussian ptp codes. Throughout the section, we assume that $I_1 \geq I_2 \geq \dots \geq I_K$, and define $I[j : k] = \sum_{i=j}^k I_i$ and $I = \sum_{i=1}^K I_i$. When $K = \infty$, we will assume that I is finite, hence $I_i \rightarrow 0$ as $i \rightarrow \infty$.

A. Optimal Decoder

Focusing again on the tagged receiver 0, define the *symmetric rate* R_{sym} as the supremum over R such that $(R, R, \dots, R) \in \mathcal{C}_0$. We can express the symmetric rate R_{sym} as the solution of a simple optimization problem.

Lemma 2: The symmetric rate under G-ptp codes is

$$R_{\text{sym}} = \max_{k \in [0:K]} \min_{l \in [0:k]} \frac{1}{l+1} C \left(\frac{P_0 + I[k-l+1:k]}{1 + I[k+1:K]} \right). \quad (8)$$

Proof: From the reduced characterization of \mathcal{C}_0 in (4), we have

$$R_{\text{sym}} = \max_{\mathcal{S}:0 \in \mathcal{S}} R_{\text{sym}}(\mathcal{S}) = \max_{k \in [0:K]} R_{\text{sym}}([0:k]),$$

where $R_{\text{sym}}(\mathcal{S})$ is the symmetric rate of the region $\mathcal{F}(\mathcal{S})$. The second equality follows from the observation that the reduced MAC region $\mathcal{F}(\mathcal{S})$ is monotonically increasing in the received powers from the transmitters in \mathcal{S} and decreasing in the interference power from transmitters in \mathcal{S}^c . Hence, among all subsets \mathcal{S} of size $k+1$, the one with the largest symmetric rate is $[0:k]$ (the one with the highest powered transmitters and lowest powered interferers).

Taking into account all 2^k constraints of the region $\mathcal{F}([0:k])$, we have

$$\begin{aligned} R_{\text{sym}}([0:k]) &= \min_{\mathcal{T} \subset [1:k]} \frac{1}{|\mathcal{T}|+1} C \left(\frac{P_0 + I(\mathcal{T})}{1 + I[k+1:K]} \right) \\ &= \min_{l \in [0:k]} \min_{\mathcal{T} \subset [1:k], |\mathcal{T}|=l} \frac{1}{l+1} C \left(\frac{P_0 + I(\mathcal{T})}{1 + I[k+1:K]} \right). \end{aligned}$$

The desired result (8) now follows from the fact that among all the subsets \mathcal{T} of size l , the one with the smallest total power $I(\mathcal{T})$ is $[k-l+1:k]$. ■

B. Other Decoders

We will use the following nomenclature for the rest of the paper:

- IAN refers to treating interference as noise decoding. The condition for IAN is

$$R < C \left(\frac{P_0}{1+I} \right).$$

- SIC(k) refers to successive interference cancellation in which the tagged receiver sequentially decodes and cancels the signals from the k strongest transmitters treating other signals as noise and then decodes the message from the tagged transmitter while treating the remaining signals as Gaussian noise. The conditions for SIC are

$$R < C\left(\frac{I_l}{1 + P_0 + I[l + 1 : K]}\right) \text{ for } l \in [1 : k],$$

$$R < C\left(\frac{P_0}{1 + I[k + 1 : K]}\right).$$

- JD(k) refers to joint decoding of the messages of the first $k + 1$ transmitters and treating the rest as Gaussian noise. The conditions for JD(k) are

$$R < \frac{1}{l+1} C\left(\frac{P_0 + I[k - l + 1 : k]}{1 + I[k + 1 : K]}\right) \text{ for } l \in [0 : k]. \quad (9)$$

The JD(k) conditions are not monotonic in k , that is, the fact that JD(k) holds neither implies that JD(k') holds for $k' < k$ nor for $k' > k$ in general.

- OPT(k) refers to the optimal decoder used in the proof of Theorem 1 if there were only k interferers. OPT(K) or simply OPT refers to the optimal decoding rule. The conditions for OPT(K) are $R < R_{\text{sym}}$ with R_{sym} given by (8). Since the condition for OPT is the union of the JD(l) conditions for $0 \leq l \leq k$, if OPT(k) holds, so does OPT(k') for all $k' > k$.

C. Number of Interferer Messages Decoded

Lemma 2 shows that, for K finite, the optimal decoding strategy is to use JD(k_{opt}) with

$$k_{\text{opt}} = \operatorname{argmax}_{k \in [0:K]} \xi(k), \quad (10)$$

where

$$\xi(k) = \min_{l \in [0:k]} \frac{1}{l+1} C\left(\frac{P_0 + I[k - l + 1 : k]}{1 + I[k + 1 : K]}\right) \quad (11)$$

provided the argmax in question is uniquely defined.

The following lemma is focused on the case $K = \infty$, which will be considered in the next sections, and where one may fear that the maximum is not defined in (8), i.e., the argmax in (10) is not defined. Fortunately, this is not the case.

Lemma 3: If $K = \infty$, $P_0 > 0$ and $I < \infty$, then $k_{\text{opt}} < \infty$.

Proof: We have

$$0 \leq \xi(k) \leq \min_{l \in [0:k]} \frac{1}{l+1} C(P_0 + I) = \frac{1}{k+1} C(P_0 + I).$$

Hence $k \rightarrow \xi(k)$ is a positive function bounded from above by a function that tends to 0 when k tends to

infinity. The values where ξ is maximal are then all finite unless it is 0 everywhere. But this is not the case since our assumptions on P_0 and I imply that $\xi(0) > 0$. ■

The following lemma will be used later.

Lemma 4: Let

$$\underline{k} = \min\{k \geq 1 \text{ such that } I_k < P_0\}.$$

Then, a sufficient condition for achievability by OPT at rate R is that

$$\underline{k}R < C\left(\frac{\underline{k}P_0}{1 + I[\underline{k} : \infty]}\right). \quad (12)$$

Further, if this condition holds, then the conditions for JD($\underline{k} - 1$) are met.

Proof: We can derive a lower bound for the symmetric rate in (8) in terms of \underline{k} :

$$\begin{aligned} & \max_{k \in [0:K]} \min_{l \in [0:k]} \frac{1}{l+1} C\left(\frac{P_0 + I[k - l + 1 : k]}{1 + I[k + 1 : K]}\right) \\ & \geq \min_{l \in [0:\underline{k}-1]} \frac{1}{l+1} C\left(\frac{P_0 + I[\underline{k} - l : \underline{k} - 1]}{1 + I[\underline{k} : K]}\right) \\ & \geq \min_{l \in [0:\underline{k}-1]} \frac{1}{l+1} C\left(\frac{(l+1)P_0}{1 + I[\underline{k} : K]}\right) \\ & = \frac{1}{\underline{k}} C\left(\frac{\underline{k}P_0}{1 + I[\underline{k} : K]}\right). \end{aligned}$$

The first inequality is obtained by choosing $k = \underline{k} - 1$ in the outer maximization; the second inequality is obtained by lower bounding the received powers of all the interferers with index $\leq \underline{k}$ by P_0 ; the last equality follows from the fact that $C(x)/x$ is a monotonically decreasing function of x . The sufficient condition (12) for achievability is now obtained by requiring the target rate R to be less than this lower bound. ■

Lemma 4 gives a guideline on how to select the set of interferers to jointly decode: under the condition (12), the success of joint decoding at rate R is guaranteed when decoding all interferers with a received power larger than that of the tagged transmitter. This is only a bound, however, and as we will see in the simulation section, one can often succeed in decoding more than $\underline{k} - 1$ transmitters.

V. SPATIAL NETWORK MODELS AND SIMULATION RESULTS

The aim of the simulations we provide in this section is to illustrate the performance improvements of OPT versus IAN and JD. The framework chosen for these simulations is a spatial network with a denumerable collection of randomly located nodes. In the following section we also use this spatial network model for mathematical analysis.

A. Spatial Network Models

All the spatial network models considered below feature a set of transmitter nodes located in the Euclidean plane. The channel gains defined in Section II, or equivalently the received signal power P_0 and the interference powers I_j , $j \in [1 : K]$, at the tagged receiver are evaluated using a path loss function $l(r)$, where r is distance. Here are two examples used in the literature (and in some examples below):

- $l(r) = r^{-\beta}$, with $\beta > 2$ (case with pole),
- $l(r) = (k+r)^{-\beta}$, with $\beta > 2$ and k a constant (case without pole); it makes sense to take k equal to the wavelength.

More precisely, if we denote the locations of the transmitters by T_j , $j \in [0 : K]$ and that of the tagged receiver by y and if we assume that the tagged receiver selects the interferers with the strongest received powers to be jointly decoded, then $|g_{00}|^2 = l(|T_0 - y|)$ and $|g_{0j}|^2 = l(|T_j - y|)$, or equivalently $P_0 = l(|T_0 - y|)Q$ and $I_j = l(|T_j - y|)Q$ for $j \in [1 : K]$. Here Q denotes the transmit power. Since we assume that $I_1 \geq I_2 \dots \geq I_K$, the strongest interferer is the closest one to y (excluding the tagged transmitter). Let $I(y)$ be the total interference at the tagged receiver, namely $I(y) = \sum_{j \neq 0} I_j$.

The simulations also consider the following extensions of this basic model:

- The fading case, where the channel gain is further multiplied by $F_j(y)$, where $F_j(y)$ represents the effect of fading from transmitter j to y . In this case, the strongest interferer is not necessarily the closest to y .
- The case where the power constraint is not the same for all transmitters. Then $P_0 = l(|T_0 - y|)Q_0$ and $I_j = l(|T_j - y|)Q_j$ for $j \in [1 : K]$, with Q_j the power constraint of transmitter j .

B. IAN, SIC(1), JD(1), and OPT(1) Cells

1) *Definitions:* Fix some rate R . For each decoding rule A (i.e., IAN, ...) as defined above, let Ξ^A be the set of locations in the plane where the conditions for rule A are met with respect to the tagged transmitter and for R . We refer to this set as the A cell for rate R . The main objects of interest are hence the cells Ξ^{IAN} , $\Xi^{\text{SIC}(1)}$, $\Xi^{\text{JD}(1)}$ and $\Xi^{\text{OPT}(1)}$.

2) *Inclusions:* Rather than looking at the increase of rate obtained when moving from a decoding rule to another, we fix R and compare the cells of the two decoding rules. In view of the comparison results in

Section II, we have

$$\begin{aligned} \Xi^{\text{IAN}} &\subset \Xi^{\text{OPT}(1)}, \\ \Xi^{\text{SIC}(1)} &\subset \Xi^{\text{JD}(1)} \subset \Xi^{\text{OPT}(1)}, \\ \Xi^{\text{OPT}(1)} &= \Xi^{\text{IAN}} \cup \Xi^{\text{JD}(1)}. \end{aligned}$$

For all pairs of conditions A and B , we define $\Xi^{A \setminus B}$ to be the set of locations in the plane where the condition for A is met but the condition for B is not met. For instance,

$$\Xi^{\text{SIC}(1) \setminus \text{IAN}} = \Xi^{\text{SIC}(1)} \setminus \Xi^{\text{IAN}}.$$

3) *Simulations:* In the simulation plots below, the transmitters are randomly located according to a Poisson point process. The attenuation function is of the form $l(r) = (1+r)^{-\beta}$ or $l(r) = r^{-\beta}$.

Figure 3 compares $\Xi^{\text{SIC}(1)}$ and $\Xi^{\text{OPT}(1)}$. Notice that SIC does not increase the region that is *covered* compared to IAN, whereas OPT(1) does.

Figure 4 compares OPT(1) to JD(1) and IAN. Note that there is no gain moving from JD(1) to OPT(1) outside the IAN cell. Also, in such a spatial network, one of the practical weaknesses of JD(1) is its lack of coverage continuity (the JD(1) cell has holes and may even lack connectivity as shown in the plots). These holes are due to the unnecessary symmetry between the tagged transmitter and the strongest interferer, which penalizes the former.

C. SIC(1) versus OPT(1)

There are interesting differences between SIC(1) and OPT(1). Let

$$\Theta^A = \bigcup_j \Xi_j^A,$$

where Ξ_j^A is the cell of transmitter j using decoding rule A (IAN, OPT(1), SIC(1)). Also let

$$\nu^A(y) = \sum_j \mathbf{1}_{y \in \Xi_j^A}$$

denote the number of transmitters covering location y under condition A . Consider the following observations.

1) We have

$$\Theta^{\text{SIC}(1)} = \Theta^{\text{IAN}}, \quad (13)$$

that is, the region of in the plane covered when treating interference as noise is *identical* to that of successive interference cancellation. This follows from the condition for SIC(1), which implies that the location under consideration is included in the cell of another transmitter in the symmetrical rate case. The gain of SIC(1) is hence only in the

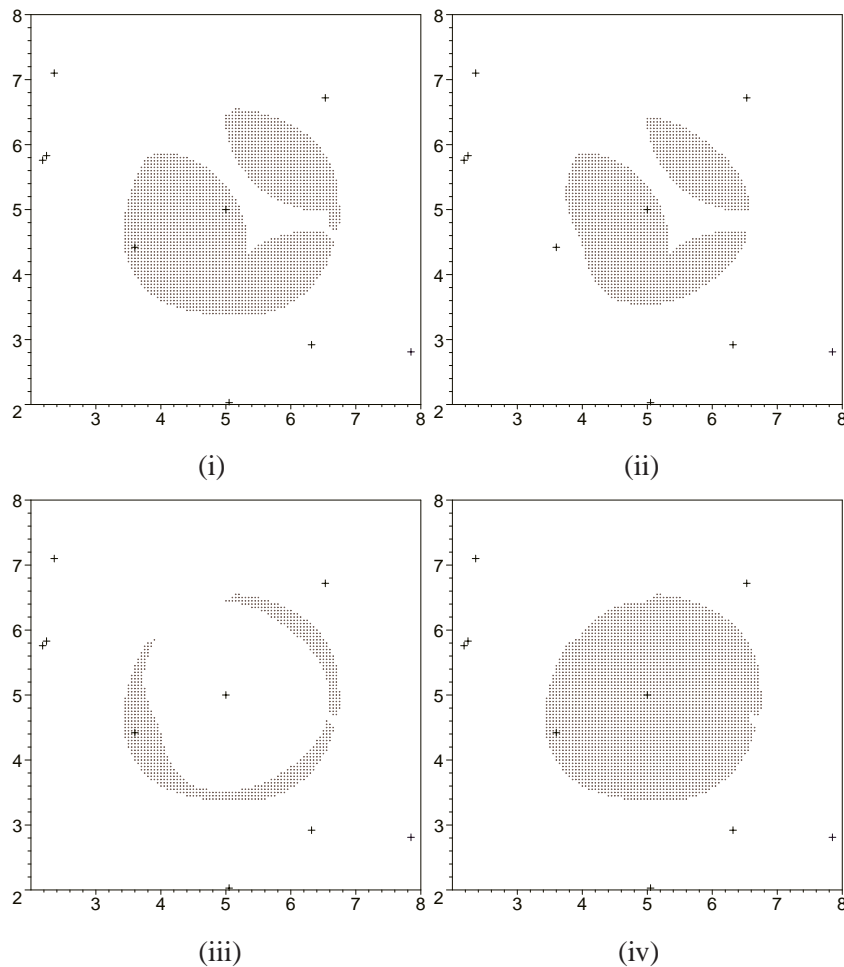


Fig. 4. Figure (i) depicts $\Xi^{\text{JD}(1)}$ for the tagged transmitter, located at $[5, 5]$, and Figure (ii) $\Xi^{\text{JD}(1)} \cap \Xi^{\text{IAN}}$. Figure (iii) shows $\Xi^{\text{JD}(1) \setminus \text{IAN}}$ and Figure (iv) $\Xi^{\text{OPT}(1)}$. The path loss exponent is $\beta = 2.5$, the power constraint is $Q = 100$ for all users; the threshold is $R = 0.2$ bits/s/Hz, and the user density is $\lambda = 0.3$. The attenuation is $l(r) = (1 + r)^{-\beta}$.

diversity of transmitters that can be received at any location y , i.e.,

$$\nu^{\text{SIC}(1)}(y) \geq \nu^{\text{IAN}}(y) \text{ for every } y. \quad (14)$$

2) We have

$$\Theta^{\text{OPT}(1)} \supset \Theta^{\text{IAN}}. \quad (15)$$

As we see in Figure 3, this inclusion is strict for some parameter values, that is, *optimal decoding increases global coverage, whereas successive interference cancellation does not*. We also have

$$\nu^{\text{OPT}(1)}(y) \geq \nu^{\text{IAN}}(y). \quad (16)$$

3) Finally, we have

$$\Theta^{\text{OPT}(1)} \supset \Theta^{\text{SIC}(1)}. \quad (17)$$

There is no general comparison between $\nu^{\text{SIC}(1)}(y)$ and $\nu^{\text{OPT}(1)}(y)$, however.

D. The $\text{OPT}(k)$ Cell

We now explore the performance of $\text{OPT}(k)$, that is, when the tagged receiver jointly decodes up to the strongest k interferers and treats the rest as noise. In Figure 5, we give samples of the region $\Xi^{\text{OPT}(2) \setminus \text{OPT}(1)}$, which is the additional area covered by moving from $\text{OPT}(1)$ to $\text{OPT}(2)$.

Since $\Xi^{\text{OPT}(k)} \subset \Xi^{\text{OPT}(k+1)}$ for all k , there exists a limit set $\Xi^{\text{OPT}(\infty)}$, which is the set of locations where the tagged receiver can decode the message of the tagged transmitter jointly with some set of other interferers messages at rate R (the existence follows from monotonicity and a boundedness argument using the assumption that the noise power is positive).

E. Optimal Number of Interferer Messages Decoded

We now illustrate the optimal number of jointly decoded interferer messages k_{opt} defined in (10). Consider

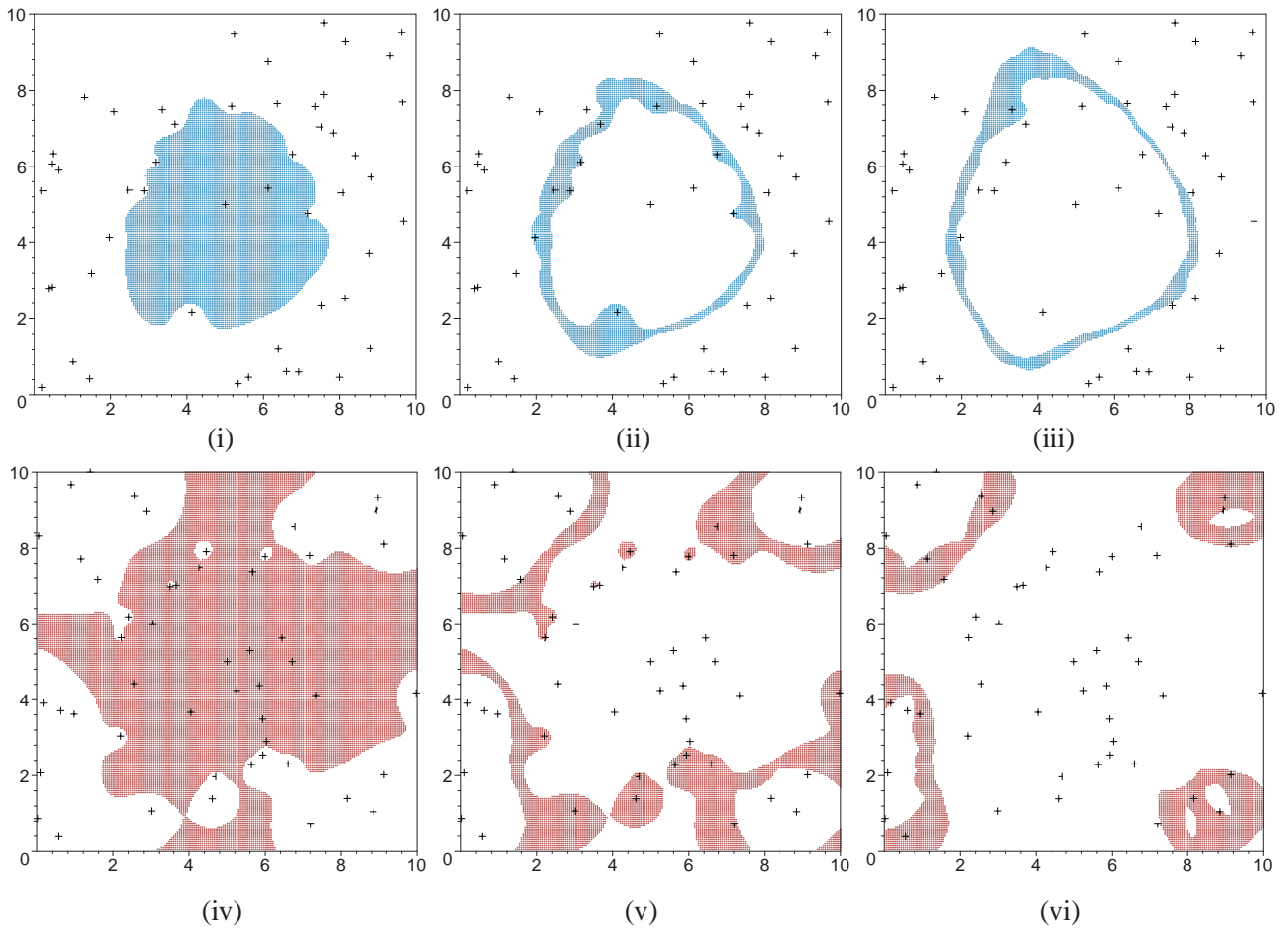


Fig. 5. On (i) and (iv), the dashed region is Ξ^{IAN} for the tagged transmitter. The dashed regions of (ii) and (v) give $\Xi^{\text{OPT(1)} \setminus \text{IAN}}$ for the tagged transmitter; those of (iii) and (vi) give $\Xi^{\text{OPT(2)} \setminus \text{OPT(1)}}$ for the tagged transmitter. The spatial user density is 0.5 and the power constraints are constant and all equal to $Q = 100$. Here $\beta = 3$. The tagged transmitter is at the center of the plot. The attenuation is $l(r) = (1 + r)^{-\beta}$. The top plots are for $R = 0.03$ bits/s/Hz and the bottom ones are for $R = 0.015$ bits/s/Hz.

transmitters distributed according a spatial Poisson process with intensity $\lambda = 10$. Assume there is no fading. In *Scenario I*, the attenuation function is that with a pole and we assume that the path loss exponent is $\beta = 3$. Each transmitter has a power constraint of $Q = 100$ and $P_0 = 5$ (this means that the distance between the tagged transmitter and its receiver is appr. 2.71.). Figure 6 plots the function $k \rightarrow \xi(k)$ defined in (11) for a sample of a Poisson point process of interferers. The maximum is reached for $k_{\text{opt}} = 230$ (we just plot the informative part of the curve here) and R_{sym} is approximately 0.00595. The number of interferers with power larger than P_0 is 229. *Scenario II* is the same but for the attenuation function without pole. In this case, the number of transmitters with a power larger than P_0 is 90 and $k_{\text{opt}} = 90$.

F. Single Hop in Ad Hoc Networks

In order to further illustrate the differences between IAN and OPT, we consider a simulation setting extending that considered above. We assume a tagged transmitter and its receiver and a collection of other transmitters located according to a Poisson point process of intensity λ that represent the nodes of an ad hoc network that interfere with the tagged transmission. We wish to compare the largest distance between the tagged transmitter and its receiver under IAN and OPT.

To do so, we first fix a distance between the tagged transmitter and its receiver, which determines P_0 (as in the last subsection). We use a Monte Carlo simulation to determine the optimal number of decoded transmitters k_{opt} for a sample of the Poisson point process of interferers. We then use (8) to determine the largest possible achievable rate R under OPT.

We then consider the largest distance r between the tagged transmitter and receiver such that the rate R is

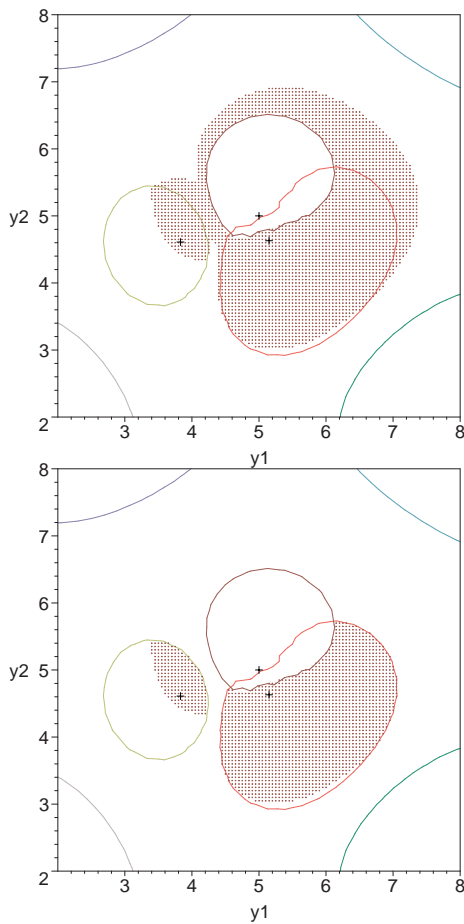


Fig. 3. The top plot is for $\Xi^{\text{OPT}(1)\backslash\text{IAN}}$ and the bottom plot is for $\Xi^{\text{SIC}(1)\backslash\text{IAN}}$ for the tagged transmitter. The transmitters are denoted by crosses. The contours denote the boundaries of the IAN cells of different transmitters. The spatial user density is 0.1. The power constraints Q_i are here randomly chosen according to a uniform distribution over $[0, 2000]$. Variable transmission powers show up when devices are heterogeneous or power controlled. $R = 0.73$ bits/s/Hz and $\beta = 3$. The tagged transmitter is at the center of the plot (at $[5, 5]$). The attenuation is $l(r) = (1 + r)^{-\beta}$.

achievable using IAN, i.e., such that $R < C(Qr^{-\beta}/(1 + I))$, for the interference I created by the same point process of interferers as above.

Figure 7 shows the locations of the interferers (obtained by sampling a Poisson point process), the tagged receiver (located at the center) and the tagged transmitter (at the other end of the long segment). The setting is that of Scenario I of Subsection V-E.

The long segment represents the distance between the tagged transmitter to its receiver (tagged link) under OPT. Its length is approximately 2.71. The short segment (displayed here for comparison) is the longest possible IAN link to the same receiver for the same rate $R = 0.00595$ of the OPT link. The length of the latter link is 0.543. Hence, for this setting, about five times longer single hops can be supported at rate R when

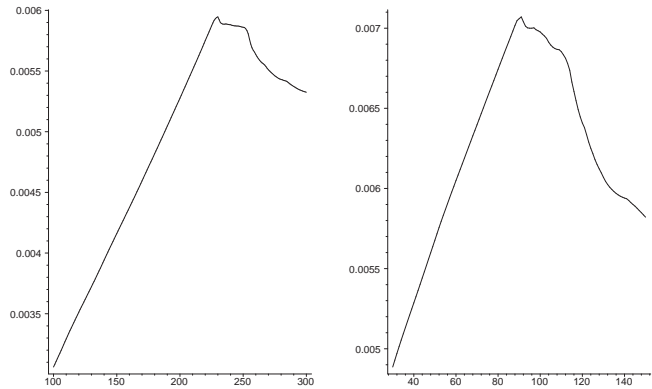


Fig. 6. A sample of the function $k \rightarrow \xi(k)$ for a Poisson collection of interferers. The x -axis is that of the k variable. The maximum of this function provides k_{opt} . Left: the attenuation is $l(r) = (1 + r)^{-\beta}$. Right: $l(r) = r^{-\beta}$.

moving from IAN to OPT.

VI. ASYMPTOTIC ANALYSIS IN THE WIDEBAND REGIME

This section is devoted to the analysis of the gain offered by using the optimal decoder $\text{OPT}(K)$ compared to IAN in large networks using the stochastic geometry approach [1]. The setting is that of Section IV, namely we consider a tagged transmitter-receiver pair and a denumerable collection of interferers. We assume here that these interferers are located according to some homogeneous Poisson process in the plane. We focus on the wideband regime, where all users share a large bandwidth and the density of users is large.

More precisely, the wideband limit is the regime where the bandwidth $B \rightarrow \infty$, the average transmit power is fixed at \bar{Q} Watts, and the target data rate for each transmitter is fixed at \bar{R} bits/s. This means that the transmit power per Hz $Q = \bar{Q}/B$ and the data rate per Hz $R = \bar{R}/B$ both tend to zero as $B \rightarrow \infty$. We also assume that the tagged transmitter and receiver are at a fixed *non-random* distance r_0 from each other, so that the *received* power from the tagged transmitter at the tagged receiver is fixed at $\bar{P}_0 = l(r_0)\bar{Q}$ Watts. On the other hand, the received power from interferer j , $\bar{I}_j = l(T_j - y)\bar{Q}$, is random. The noise power is assumed to be 1 Watt/Hz, as before.

As the bandwidth B increases, we would like the network to support an increasing user density λ_B so that the spectral efficiency of the system is kept non-vanishing. The following two theorems compare the performance of the IAN decoder and the OPT decoder

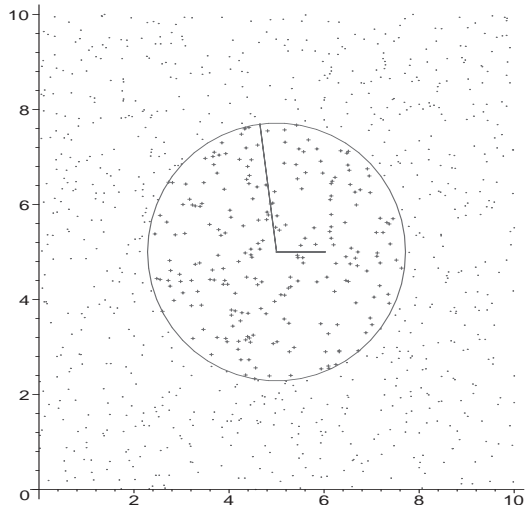


Fig. 7. Comparison of IAN and OPT in an ad hoc network. The receiver is at the center of the plot (at [5,5]). The long segment is a link of fixed length using OPT. The rate R is the largest rate that R_{sym} can be sustained on this link. The short segment gives the longest link that can be sustained at the same rate R under IAN at this receiver. The same thermal noise at the receiver, the transmission power at the transmitters and the set of interferers are the same in the two cases. The set of red points (inside the circle) is the optimal set of transmitters that are jointly decoded by the receiver under OPT, whereas the set of blue points (outside the circle) features the other transmitters that the tagged receiver considers as noise. The attenuation is $l(r) = r^{-\beta}$.

in terms of how fast the density λ_B can scale with the bandwidth B while still reliably decoding the tagged transmitter's message.

Theorem 4: Consider the path loss model $l(r) = r^{-\beta}$. If $\lambda_B = \kappa B^p$ with $p > \frac{2}{\beta}$ and $\kappa > 0$, then for every target rate $\bar{R} > 0$, the IAN decoding condition for the tagged receiver cannot be satisfied almost surely as B grows.

Theorem 5: Consider the path loss model $l(r) = r^{-\beta}$. If $\lambda_B = \kappa B$ for $\kappa > 0$ and if

$$\rho \bar{R} < C \left(\frac{\rho \bar{P}_0}{1 + 2\rho \bar{P}_0 / (\beta - 2)} \right) = C \left(\frac{\rho \bar{P}_0}{1 + \frac{2\pi\kappa Q}{r_0^\beta (\beta - 2)}} \right), \quad (18)$$

then almost surely the OPT decoding condition is satisfied as the bandwidth B grows. Here, $\rho = \kappa \pi r_0^2$ is the expected number of interferers per Hz within the communication radius r_0 from the tagged receiver.

Theorem 4 says that one needs a *sub-linear* scaling of the user density to guarantee a positive rate \bar{R} under the IAN decoder. In particular, the classical strong law of large numbers shows that for all $\kappa > 0$, the linear scaling $\lambda_B = \kappa B$ leads to a zero achievable rate under

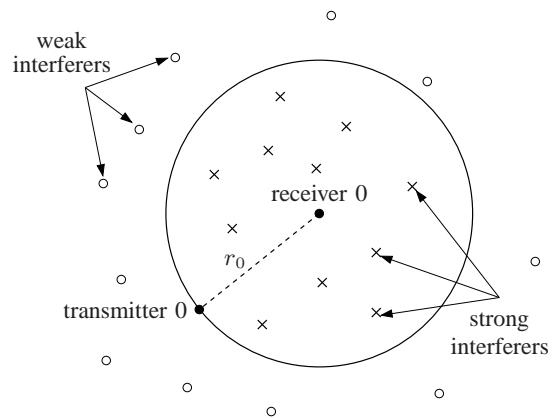


Fig. 8. The tagged transmitter and the tagged receiver are at a distance r_0 from each other. The strong interferers are within a distance of r_0 from the tagged receiver.

IAN. This implies that the network spectral efficiency in terms of total bits per second per Hz per unit area goes to zero under IAN. For the OPT decoder, on the other hand, Theorem 5 says that a linear scaling of user density can be supported and hence a positive spectral efficiency can be achieved in the wideband limit.

Before proving the above theorems, we provide some intuition as to why the user density scalings of these two decoders differ quite dramatically. The situation is depicted in Figure 8. The received interference power from each of the strong interferers inside the circle is larger than the signal power P_0 . In fact, as the user density increases, there will be more and more interferers very close to the tagged receiver with much larger received powers than P_0 . Their effect is fatal to the IAN decoder, which treats all interference as noise. The OPT decoder, on the other hand, can take advantage of these interferers' high received powers to jointly decode their messages together with that of the tagged transmitter. This effectively turns their interference power into useful signal energy. The proof of Theorem 5 shows that the total useful received power from these strong interferers is at least comparable to the total harmful received power of the interferers outside the disk; hence reliable communication at a positive rate \bar{R} bits/s for the tagged receiver (and for everyone else). In fact, the term $\rho \bar{P}_0$ in (18) is a lower bound on the total power per Hz received from the strong interferers, and the term $2\rho \bar{P}_0 / (\beta - 2)$ is the total power per Hz received from the weak interferers outside the disk.

We are now ready to prove the above theorems.

Proof of Theorem 4: The feasibility of the rate \bar{R} is equivalent to

$$C \left(\frac{\bar{P}_0}{B + \bar{I}} \right) \geq \frac{\bar{R}}{B}, \quad (19)$$

where \bar{I} denotes the total interference power at the tagged receiver (in Watts) in a Poisson network of intensity λ_B . The last condition is equivalent to

$$\frac{B\bar{P}_0}{B + \bar{I}} = \frac{\bar{P}_0}{1 + (\kappa/\lambda_B)^{1/p}\bar{I}} \geq B \left[2^{\bar{R}/B} - 1 \right].$$

Let $n = \lfloor \lambda_B \rfloor$. We have $\bar{I} > \sum_{j=1}^n \bar{I}_j$ with \bar{I}_j a collection of i.i.d. shot noise processes of intensity 1. The random variable \bar{I}_i has a stable distribution with parameter $2/\beta$ [7], and hence its moments of order p are infinite for $p > 2/\beta$. The Marcinkiewicz-Siegmund strong law of large number [5] then implies that for all $p > 2/\beta$,

$$\limsup_{n \rightarrow \infty} \frac{1}{n^{1/p}} \sum_{j=1}^n \bar{I}_j = \infty$$

in an almost sure sense. Since $B \left(2^{\bar{R}/B} - 1 \right) \rightarrow \bar{R} \ln 2 > 0$, this shows that the condition (19) cannot hold true for sufficiently large B . ■

Proof of Theorem 5: With the notation of Lemma 4, recall that \underline{k} is the index of the first interferer whose received power at receiver 0 is less than the received signal power P_0 . Equivalently, $\underline{k} - 1$ is the number of interferers in a disk of radius r_0 from receiver 0.

From Lemma 4, for the rate \bar{R} to be feasible, it is enough to show that

$$\frac{k\bar{R}}{B} < C \left(\frac{k\bar{P}_0}{B + \bar{I}[\underline{k} : \infty]} \right). \quad (20)$$

Now, almost surely, when B tends to infinity,

$$\frac{k}{B} \rightarrow \kappa\pi r_0^2 \quad (21)$$

and

$$\lim_{B \rightarrow \infty} \frac{1}{B} \bar{I}[\underline{k} : \infty] = \kappa m(\bar{P}_0) < \infty, \quad (22)$$

with

$$m(\bar{P}_0) = \frac{2\pi r_0^2 \bar{P}_0}{\beta - 2}.$$

In order to show (22), we represent $\bar{I}[\underline{k} : \infty]$ as the sum of n i.i.d. random variables $\tilde{I}_1, \dots, \tilde{I}_n$, where \tilde{I}_1 is the shot noise for the attenuation function r^β and for a spatial Poisson point process with intensity 1 outside a disk of radius r_0 and 0 inside. Since $\mathbb{E}(\tilde{I}_1) < \infty$, (22) follows from the strong law of large numbers. From Campbell's formula [1], we obtain

$$m(\bar{P}_0) = \mathbb{E}(\tilde{I}_1) = \bar{Q} 2\pi \int_{r_0}^{\infty} \frac{1}{r^\beta} r dr = \frac{2\pi r_0^2 \bar{P}_0}{\beta - 2},$$

using the fact that $\bar{P}_0 = \bar{Q} r_0^{-\beta}$. Substituting (22) and (21) into (20) and simplifying yields the condition (18)

for OPT to decode successfully almost surely for B large enough. ■

Note that the linear user density scaling achieved by OPT cannot be achieved by the decoder OPT(k) for any fixed k . One has to jointly decode the messages from an increasing number of interferers as the bandwidth and the user density increase.

When the distance r_0 between the tagged transmitter and its receiver tends to infinity, the received power $P_0 \rightarrow 0$, and (18) reads $\bar{R} < \bar{P}_0 \log e$, which is the wideband capacity of a point-to-point Gaussian channel without interference.

Remark: This result may seem surprising at first glance. In the ad hoc network setting of Section V-F, this result suggests that when OPT(k) of high order k is used in a wideband system, one can maintain a channel from a tagged transmitter to a tagged receiver, say at distance r_0 , with a positive rate (determined by Theorem 5) when the user density tends to infinity. For instance, in the ad hoc setting of the simulation section, this means that one can maintain simultaneous single hop channels that "jump" over a very large number of nodes of the ad hoc network. In contrast, in the IAN case, in order to maintain the same rate, one has to set a multihop route over a number of relay nodes that tends to infinity as the user density tends to infinity. However, this is perhaps not so surprising since in this setting, one could in principle organize some sort of TDMA or FDMA IAN scheme (which silences a large collection of nodes when the tagged transmission takes place) that has asymptotic performance of the same kind as that exhibited by OPT. So, OPT can in fact be seen as a way of obtaining good performance without a priori partitioning the users into different time or frequency slots.

Notice that the last comparison results rely on the assumption that the loss function is the one with a pole. In the case without a pole, we can obtain the following results using very similar arguments based on the classical strong law of large numbers (and are easily extended to general attenuation functions such that $\int_{\mathbb{R}^+} l(r) r dr < \infty$).

Theorem 6: Consider the path loss model $l(r) = (k + r)^{-\beta}$. When $\lambda_B = \kappa B$, and the bandwidth B tends to infinity, the IAN decoding condition is satisfied almost

surely iff

$$\begin{aligned} \bar{R} &< C \left(\frac{\bar{P}_0}{1 + 2\rho\bar{P}_0 \int_0^\infty \left(\frac{k+r_0}{k+r}\right)^\beta \frac{r}{r_0^2} dr} \right) \\ &= C \left(\frac{\bar{P}_0}{1 + 2\pi\kappa\bar{Q} \int_0^\infty \left(\frac{1}{k+r}\right)^\beta r dr} \right), \end{aligned} \quad (23)$$

and the OPT condition almost surely if

$$\begin{aligned} \rho\bar{R} &< C \left(\frac{\rho\bar{P}_0}{1 + 2\rho\bar{P}_0 \int_{r_0}^\infty \left(\frac{k+r_0}{k+r}\right)^\beta \frac{r}{r_0^2} dr} \right) \\ &= C \left(\frac{\rho\bar{P}_0}{1 + 2\pi\bar{Q} \int_{r_0}^\infty \left(\frac{1}{k+r}\right)^\beta r dr} \right), \end{aligned} \quad (24)$$

where again $\rho = \kappa\pi r_0^2$.

As a direct corollary of the last formulas, when the distance r_0 between the tagged transmitter and its receiver tends to infinity so that the received power $P_0 \rightarrow 0$, the right hand side of (24) tends to zero like $\bar{P}_0 \log e$, which is the wideband capacity of a point-to-point Gaussian channel without interference. On the other hand, the effect of interference never disappears for IAN. Hence, in this limiting regime, for the case without a pole and with a given linear user growth rate, a positive rate is feasible for both IAN and OPT in the limit, but with different values. When the tagged transmitter and receiver are far away from each other, the scaling of this feasible rate under OPT is as though there were no interferers.

Remark: Above, we focused on the case where the node density tends to infinity. For the finite density case, the performance of the decoding strategies considered can be evaluated from the joint distributions of the total interference and of the order statistics I_1, I_2, \dots using the tools described in [1].

VII. CONCLUSION

In this paper, we studied the optimal performance achievable in a Gaussian interference network when the transmitters are constrained to use uncoordinated point-to-point codes. While recent results have shown that to achieve the *ultimate* capacity of such networks, techniques such as superposition coding and interference alignment are needed, such techniques require significantly more complex codes and coordination between the transmitters. What our results suggest is that using simple point-to-point codes and no coordination between the transmitters, one can achieve quite significant gains.

Moreover, since many existing wireless networks already use near-capacity-achieving point-to-point coding, our results also point to the possibility of significant performance gain from just upgrading the receivers and not the transmitters. This provides an evolutionary path to improving the performance of existing wireless networks. It would also be interesting to extend our results to establish the capacity region for MAC-capacity-achieving codes with limited coordination, such as frequency/time partitioning and power control.

An interesting future direction is to explore how to design a distributed medium-access protocol when receivers employ optimal decoding. There would be two important components to such a protocol. The first component is *interferer sensing* by the receivers. Each receiver senses the powers and the identities of its interferers. This can be implemented through some beaconing scheme. The second component is a *backoff* procedure by the transmitter. Each user, when it has data to transmit, needs to sense when its receiver can accommodate its transmission. This in turn depends on the number and powers of the interferers who are transmitting. In conventional protocols such as Carrier Sense Multiple Access (CSMA), transmission occurs when the level of interference is below a certain threshold. This makes sense for an IAN receiver. However, under optimal decoding, sometimes having a strong interferer is advantageous as it enables joint decoding. Hence, the backoff procedure will have to be more elaborate.

Another interesting direction is to explore the implementation of optimal decoding. When the number of interferers whose messages are jointly decoded becomes large, one might fear an exponential growth of the combination of codewords to be tested by each decoder when decoding. However, it is not completely clear that this exponential growth is necessary to achieve capacity. For example, at the corner points, SIC, with a complexity that only grows linearly with the number of decoded interferers, is sufficient. It may be possible to reduce the complexity of decoding for the points in the interior of the sum rate face as well.

REFERENCES

- [1] F. Baccelli and B. Błaszczyszyn. *Stochastic Geometry and Wireless Networks, Volume 1: Theory*. Foundations and Trends in Networking. NOW Publishers, 2009. Web version available at: <http://hal.inria.fr/inria-00403039>.
- [2] J. Blomer and N. Jindal. Transmission capacity of wireless ad hoc networks: successive interference cancellation vs. joint detection. In *Proceedings of the IEEE International Conference on Communications*, 2009.

- [3] G. Bresler, A. Parekh, and D. Tse. Approximate capacity of many-to-one and one-to-many interference channels. *IEEE Transactions on Information Theory*, 56(9):4566–4592, Sept. 2010.
- [4] T. Cover and J. Thomas. *Elements of Information Theory*. John Wiley & Sons, Inc., 2006.
- [5] P. Embrechts, C. Klupperberg, and T. Mikosh. *Modelling Extremal Events*. Springer, 1991.
- [6] A. El Gamal and Y.H. Kim. *Lecture Notes on Network Information Theory*. January 2010. web version: <http://arxiv.org/abs/1001.3404>.
- [7] M. Haenggi and R.K. Ganti. *Interference in Large Wireless Networks*. Foundations and Trends in Networking. NOW Publishers, 2010.
- [8] S. Weber, J.G. Andrews, X. Yang, and G. de Veciana. Transmission capacity of wireless ad hoc networks with successive interference cancellation. *IEEE Transactions on Information Theory*, 53(8):2799–2814, August 2007.