On The Limitations of Cooperation in Wireless Networks

Angel Lozano
Dept. of Information and Communication Technologies
Universitat Pompeu Fabra (UPF)
08018 Barcelona, Spain
Email: angel.lozano@upf.edu

Jeffrey G. Andrews and Robert W. Heath, Jr.
Department of Electrical and Computer Eng.
The University of Texas at Austin
Austin, TX 78712
Email: {jandrews,rheath}@ece.utexas.edu

***Abstract*—Cooperation in a large wireless network (such as a cellular system) is shown to have certain fundamental limitations: namely, even perfect cooperation cannot in general change an interference-limited network to a noise-limited one. In contrast to existing literature that routinely assumes that the spectral efficiency scales with $\log P$ as the transmit power P grows large, we show the existence of a spectral efficiency upper bound that does not grow with P. The result uses well-accepted principles of information theory to reach the conclusion that it is not possible (or even helpful) to fully coordinate a large wireless network. Rather than simply low- and high-power regimes, there are three distinct network operating regimes: low-power; a *DoF* regime, where the $\log P$ scaling holds; and a *saturation* regime where the spectral efficiency hits a ceiling that is independent of P. Using a cellular system example, it is demonstrated that the transition to the saturation regime is operationally relevant and perhaps explains the lackluster gains from cooperation observed in practice.**

I. Introduction

It is well-known that wireless networks with many uncoordinated transmitters and receivers using the same spectrum become interference-limited, which means that increasing the transmit power of each node does not improve the spectral efficiency once the transmit power is sufficiently high. In cellular systems, for example, this kicks in quickly, and results in a large fraction of users having low SINR (signal-to-interference-and-noise ratio) regardless of the transmit powers, unless bandwidth-wasting methods such as frequency reuse are implemented. It is widely believed that this limitation is not fundamental, but rather an artifact of each BS (base station) transmitting/receiving autonomously. If the BSs cooperated, the logic goes, the "interference channel" could be converted to a massive broadcast channel (BC, for the downlink) or multiple access channel (MAC, for the uplink), with all the BSs jointly decoding (MAC) or encoding (BC). The bit rate and latency of the backhaul links are current technology limitations, but if and when those are overcome, then it seems that an arbitrary number of BSs could cooperate to achieve enormous spectral efficiency gains over the lone-BS model [1], with the only limitation being the amount of coordination that is practical [2].

This paper shows that this belief is false. On the contrary, it is fundamentally impossible to convert a large interference network to a BC or MAC regardless of the amount of cooperation. Above a certain point, additional cooperation gives at best rapidly diminishing improvements, and the spectral efficiency saturates to a value that is bounded with the SNR (signal-to-noise ratio). The result is general and includes any type of cooperation, reception or transmission, and is valid for any sufficiently large network with a finite channel coherence. For concreteness, we focus on the uplink cellular case with non-coherent detection, so no specific channel estimation procedure is needed.

A. Status Quo

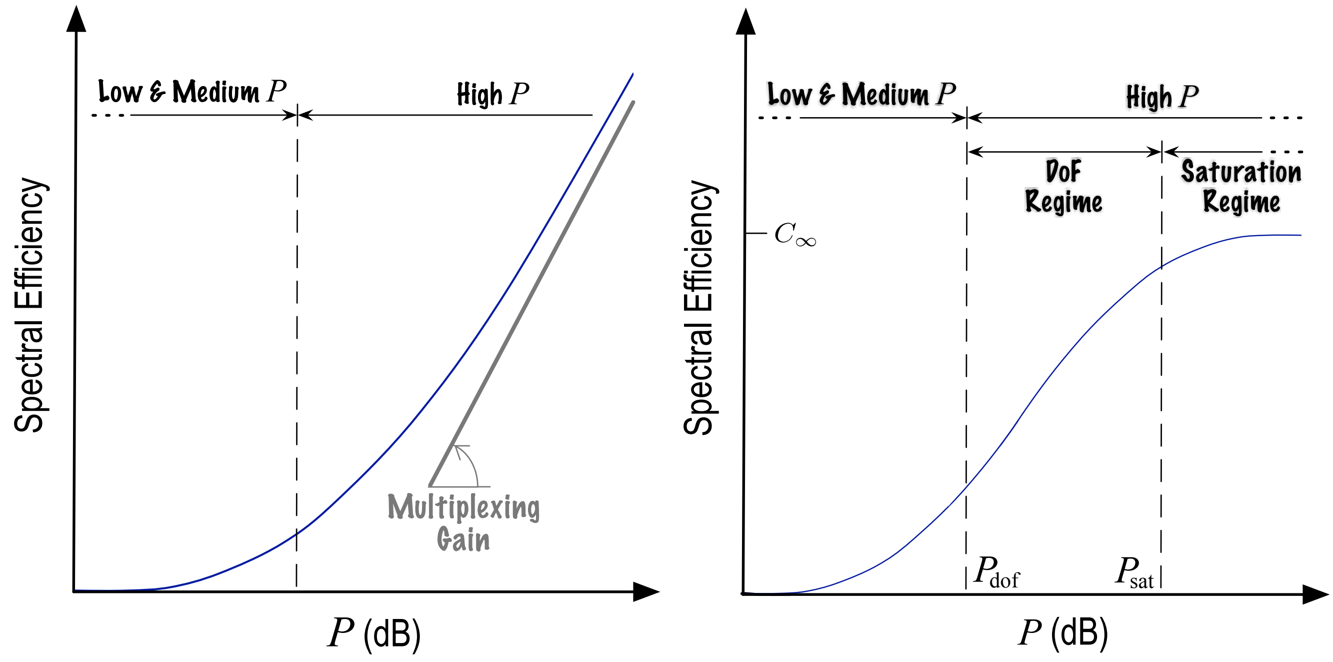
There are currently about 3 million macrocellular BSs worldwide, but that number will climb above 50 million by around 2015 once small cells (pico and femto cells) that share the same spectrum are counted [3]. A typical urban area will have thousands to tens of thousands of BSs occupying the same spectrum. Evaluating the performance of such large systems is a complex task, which one inevitably associates with massive computer simulations. To develop insights and understand the fundamentals, ideas are generally incubated and developed in much simpler settings that represent a fragment of a system. These simpler settings are typically described (at a given time epoch) by the following relationship.

Relationship 1 (Traditional System Model)
The observation at receiver n is

$$Y_n = \sum_{k=1}^{K} H_{nk}\sqrt{P}X_k + Z_n \qquad n = 1, \ldots, N \qquad (1)$$

where K and N are the number of transmitters and receivers, respectively, X_k is the signal generated by transmitter k, normalized such that P is its power, and H_{nk} is the channel from transmitter k to receiver n. The term Z_n is the noise at receiver n, typically white and Gaussian, with some normalized variance.

The signals and channel coefficients in the above relationship can be scalars, vectors and/or matrices, depending on the number of antennas or frequency subchannels. For the sake of exposition, we henceforth consider single-antenna transmitters



(a) (b)

Fig. 1: Spectral efficiency vs. power in dB: (a) Relationship 1, (b) Relationship 2.

and receivers in a single frequency band, so all these values are scalars.

Relationship 1 subsumes most basic information-theoretic channel settings, including the single-user channel ($K = N = 1$), MAC ($K > 1$ and $N = 1$), BC ($K = 1$ and $N > 1$), and K-user interference channel ($K = N$ with $K, N > 1$). Thus, Relationship 1 is quite general and can be used to describe any network with K transmitters and N receivers. It has led to heralded results on the capacity of the BC and MAC [4], definition of the number of DoF in a network, and the idea of IA (interference alignment) [5], [6]. It has been used to describe a large—possibly coordinated—cellular network [7].

The aforementioned results take place in the high-power regime, which is where the issue of interference comes to the fore. For any setting conforming to Relationship 1, the high-power behavior is as illustrated in Fig. 1(a). If the channel coefficients $\{H_{nk}\}$ are known by the corresponding receivers and possibly also the transmitters, the spectral efficiency achieved by a given user grows, for $P \to \infty$, linearly with $\log(P)$ with a slope given by the so-called multiplexing gain. The multiplexing gain cannot exceed the number of DoF, and in fact it must be smaller if diversity is to be achieved as well. Cooperative techniques such as network MIMO or IA then aim at maximizing the number of DoF available in the network.

The above interference-channel developments, spawned by Relationship 1, have spurred a vast amount of publications that promise very large gains in spectral efficiency through cooperation, see e.g. [8], [9] and references therein. In subsequent system-level simulations and field trials, however, these gains do not seem to materialize, and typical gains reported by industry are meager, on the order of 10-20% [10]–[13]. These major discrepancies point to a disconnect, to some fundamental way in which a fragment of a cellular system is not properly modeled by Relationship 1.

B. Claims

An obvious problem with Relationship 1 is that there is a cutoff of K (possibly cooperating) transmitters, and all other interference is ignored. The K cooperating transmitters are typically assumed to be geographic neighbors, and are referred to as a *cluster*. It is usually assumed that any interference from outside the cluster can be lumped into the Z_n noise terms. It cannot, because Z_n has a fixed variance that does not depend on P whereas the external interference power is proportional to P. This can be illustrated by generalizing Relationship 1 as follows, to more accurately describe a wireless network.

Relationship 2 (Proposed System Model)
The observation at receiver n is

$$Y_n = \sum_{k=1}^{K} H_{nk}\sqrt{P}X_k + \sum_{k=K+1}^{\tilde{K}} H_{nk}\sqrt{P}X_k + Z_n \quad n = 1, \ldots, N \tag{2}$$

where $\tilde{K}$ and $\tilde{N}$ the total numbers of transmitters and receivers in the system, and K and N are the ones cooperating. Defining $Z'_n = \sum_{k=K+1}^{\tilde{K}} H_{nk}X_k$ as the external interference *at receiver n, a cluster is described by*

$$Y_n = \sum_{k=1}^{K} H_{nk}\sqrt{P}X_k + \sqrt{P}Z'_n + Z_n \quad n = 1, \ldots, N \tag{3}$$

Thus, as $P \to \infty$, the external interference does as well, if the total network size is greater than K. This gives rise to a *saturation regime* as shown in Fig. 1(b), where further increasing the power above some value P_{sat} does not noticeably increase the spectral efficiency because the external interference Z' is larger than the noise Z.

A possible counter to this observation is that this saturation only occurs because the cluster size K, i.e., the number of cooperating transmitters, is too small. If the cluster K were large enough, the thinking goes, this saturation effect should only become noticeable at very high SNR, or if $K \to \tilde{K}$, there would in theory be no saturation as all users would be cooperating. In large systems this is not the case, however, as we show in the sequel. Using straightforward applications of information theory and communication theory, this paper establishes the following main points:

1) As a result of external interference, a cluster within a cellular system is described by Relationship 2 rather than Relationship 1. The performance of *any* cooperative scheme is described by Fig. 1(b), not the more familiar Fig. 1(a).
2) The existence of external interference (whose power scales with P) is inevitable in large systems. This holds regardless of whether channel coefficients are explicitly estimated (coherent) or not (noncoherent), and regardless of the size of the clusters within the system.
3) For noncoherent detection, the spectral efficiency improves as the cluster size K grows up to a certain point, and is at all times less than a finite bounded value as K and P grow large.[1]
4) Three distinct regimes can be identified: (i) a *low-power* regime where cooperation does not play a major role since noise is the main limitation; (ii) a *DoF* regime where the spectral efficiency scales with $\log P$, which occurs when the noise Z is small relative to P but still much larger than the external interference Z', i.e., $Z \gg Z'$, and (iii) a *saturation* regime where $Z' > Z$ and the spectral efficiency no longer increases (significantly) with P. The notion of DoF, and with it any standard high-SNR result, is only meaningful in the DoF regime, over a bounded range of P.
5) New quantities of interest emerge: the values of P at which the two transitions occur, i.e., P_{dof} and P_{sat}, and the limiting spectral efficiency, C_∞. The values of these new quantities depend on the system topology, the channel propagation laws, and the degree of user mobility. In cases of practical interest, the transition to the saturation regime takes place within the range of operational interest.

[1]For coherent detection the spectral efficiency is not even monotonically increasing with K, since the cost of estimating $K - 1$ channels for each receiver soon becomes larger than the gain from increased cooperation. This case is explored in [14].

II. System Modeling

This section is devoted to refining Relationship 2, defining necessary concepts, and to describing the models used to embody it in the remainder of the paper.

Definition 1 Geometry Profile
Consider a set of values $\{G_{nk}\}$, for $k = 1, 2, ..., K$ and $n = 1, 2, ..., N$, where a given G_{nk} is the average channel power gain resulting from distance-based decay, shadow fading, building penetration losses, antenna patterns, and possibly other factors, between transmitter k and receiver n. The value G_{nk} is unitless but is the magnitude square of the channel gain, and hence it is termed a power gain. The geometry profile for every receiver n is the set of normalized channel power gains

$$g_{nk} = \frac{G_{nk}}{\sum_{k=1}^{K} G_{nk}} \tag{4}$$

such that, for every n,

$$\sum_{k=1}^{K} g_{nk} = 1. \tag{5}$$

Therefore, the geometry profile $\{g_{nk}\}$ signifies the *share* of receiver n's signal power that corresponds to each transmitter k and encapsulates all the relative gains in the network. It can also account for differences in user transmit powers, e.g., uplink power control relative to a baseline value P. In short, the geometry profile provides a compact but very general representation of all the channels and channel-dependent effects in the network.

We can absorb the various normalizations for each receiver n into a signal-to-noise ratio (SNR_n) that scales with P and a signal-to-external-interference ratio (SIR_n) that does not scale with P.

Definition 2 SNR and SIR.
The SNR and SIR at receiver n are are

$$\mathsf{SNR}_n = \frac{\sum_{k=1}^{K} G_{nk} P}{N_0 B}, \tag{6}$$

$$\mathsf{SIR}_n = \frac{\sum_{k=1}^{K} G_{nk}}{\sum_{k=K+1}^{\tilde{K}} G_{nk}}, \tag{7}$$

where N_0 is the noise spectral density and B the bandwidth.

Other definitions of SNR and SIR are possible, but we will adopt these. Using these definitions, (3) can be rewritten as

$$y_n = \sqrt{\mathsf{SNR}_n} \sum_{k=1}^{K} \sqrt{g_{nk}} h_{nk} x_k + \sqrt{\frac{\mathsf{SNR}_n}{\mathsf{SIR}_n}} z'_n + z_n \quad n = 1, \ldots, N \tag{8}$$

where the noise terms $\{z_n\}$, the external interference terms $\{z'_n\}$, the signals $\{x_k\}$, and the fading channel coefficients

$\{h_{nk}\}$ are all mutually independent random variables normalized to be unit-variance.

If both the noise and the residual interference are Gaussian, then (8) becomes

$$y_n = \sqrt{\mathsf{SNR}_n} \sum_{k=1}^{K} \sqrt{g_{nk}} h_{nk} x_k + \sqrt{1 + \frac{\mathsf{SNR}_n}{\mathsf{SIR}_n}} z''_n \quad n = 1, \ldots, N \tag{9}$$

where z''_n is the aggregate noise-plus-interference, also Gaussian and unit variance, i.e., $z''_n \sim \mathcal{N}_{\mathbb{C}}(0, 1)$. The noise is typically of thermal origin and is therefore Gaussian. The external interference is made up of a large number of independent terms and thus its distribution tends to be approximately Gaussian too. We thus focus on (9), with $z''_n \sim \mathcal{N}_{\mathbb{C}}(0, 1)$, although all of the points made henceforth apply (qualitatively) in the wider generality of (8), with z_n and z'_n having different distributions.

Relationship 3 *The observation at receiver n is given by*

$$y_n = \sqrt{\frac{\mathsf{SNR}_n \cdot \mathsf{SIR}_n}{\mathsf{SNR}_n + \mathsf{SIR}_n}} \sum_{k=1}^{K} \sqrt{g_{nk}} h_{nk} x_k + z''_n \tag{10}$$

$$= \sqrt{\mathsf{SINR}_n} \sum_{k=1}^{K} \sqrt{g_{nk}} h_{nk} x_k + z''_n \tag{11}$$

where SINR_n is the signal-to-interference-plus-noise ratio at receiver n, which equals the harmonic mean of SNR_n and SIR_n,

$$\frac{1}{\mathsf{SINR}_n} = \frac{1}{\mathsf{SNR}_n} + \frac{1}{\mathsf{SIR}_n}. \tag{12}$$

For small P, then, $\mathsf{SINR}_n \approx \mathsf{SNR}_n$ whereas for $P \to \infty$, $\mathsf{SINR}_n \to \mathsf{SIR}_n$. The formulation on the basis of SNR_n and SIR_n is very general in that it captures not only the scaling with P, but also with other parameters such as cell size or noise variance.

The fact that the interference power should scale with P is acknowledged in [15], [16], but the representation therein differs from Relationships 2–3 in two important ways:

- In [15], [16], the channel gains are deterministic and thereby known perfectly—at no cost—by all transmitters and receivers. In Relationships 2–3, in contrast, the channel gains outside the cluster of interest cannot be known and those within the cluster have to be learned (explicitly or not).
- In [15], [16], the ratios SNR_n and SIR_n are constrained to relate as $\mathsf{SIR}_n = \mathsf{SNR}_n^{1-\beta}$ where $\beta \geq 0$ is a fixed parameter. In Relationships 2–3, in contrast, they can be arbitrary.

The assumptions in [15], [16] render the analysis therein applicable to small systems satisfying $K = \tilde{K} \ll L$ and $\mathsf{SIR}_n = \mathsf{SNR}_n^{1-\beta}$ and the main results are that the spectral efficiency still scales indefinitely with $\log(P)$ with a modified DoF notion dubbed *generalized DoF*. Our focus, alternatively, is on large systems where fading is of essence.

The fact that the interference power should scale with P is also acknowledged in [17] albeit in the context of a small (19-cell) system.

Returning to Relationship 3, the small-scale fading is modeled as Rayleigh block fading and thus the normalized channel coefficients satisfy $h_{nk} \sim \mathcal{N}_{\mathbb{C}}(0, 1)$. With block fading, the channels remain constant for L symbols and then change to a different value in an IID (independent identically distributed) fashion. If we denote by $T_{\rm c}$ and $B_{\rm c}$ the channel coherence in time and frequency, then—irrespective of how the signaling is arranged along the time and frequency dimensions—the number of symbols over which the channel remains coherent is roughly $L = \lfloor B_{\rm c} T_{\rm c} \rfloor$. Block fading is a coarse but effective approximation to continuous fading, and remarkable equivalences between the two have been uncovered [18]. According to these equivalences, $L = 0.5\, \lambda B / v$ where λ is the wavelength and v the velocity. To put numbers to this, at a typical cellular wavelength of 15 cm (about 2 GHz), $L \approx 20,000$ for pedestrian velocities and $L \approx 1000$ for vehicular velocities.

To close this section, let us introduce an exemplary system that shall be utilized throughout in order to be able to give some concrete numbers for C_∞ and $P_{\rm sat}$.

Example 1 *The exemplary system has tri-sector hexagonal cells of size R. Two tiers of cells are shown in Fig. 2; we consider cells extending infinitely in all directions. Each sector's antenna has a uniform gain over the $120°$ span of the sector and a gain that is uniformly $Q|_{\rm dB}$ lower outside that span. Orthogonal signaling resources (time slots and frequency bands) are allocated to the users within each sector. On any given resource, thus, there is a single user per sector and hence $K = N$. Such user is centered in azimuth within its sector and at a distance $2R/3$ from the BS. The signals experience distance-dependent decay with an exponent γ in addition to Rayleigh fading.*

The aforedescribed example is representative of a cellular system, while having the virtue of being isotropic and of having regular user locations. Eq. (9) can be embodied for any desired values of $K = N$, i.e., for clusters of arbitrary size, and the external interference can be easily summed for $\tilde{K}, \tilde{N} \to \infty$ thanks to the regular user locations.

III. Main Results

In this section, we prove the claims made in Section I-B using the models and concepts that have now been introduced.

Let us postulate the use of non-coherent detection without explicit channel estimation. It is important to realize that—from an information-theoretic perspective—this subsumes as a special case the procedure of transmitting pilot symbols, explicitly estimating the channel coefficients, and detecting the data coherently. That is, pilot symbols are a specific form of redundancy and explicit channel estimation followed by coherent data detection is a specific detection strategy. Presentation of the coherent case is deferred to [14].

In the absence of CSI (channel-state information) at the receivers, the capacity-achieving signals and the capacity itself are generally unknown—even for a single-user fading channel.

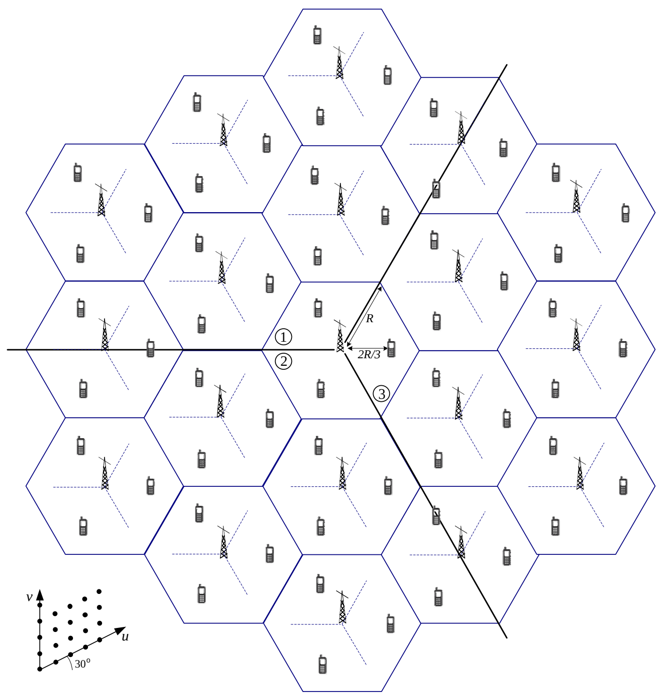


Fig. 2: Regular hexagonal universe with tri-sector hexagonal cells of size R. Each user is centered within its sector and at distance $2R/3$ from the BS.

Given the difficulty in computing the exact spectral efficiency achievable without CSI at the receivers, what we shall do is show that its limiting value, C_∞, is bounded by above by a quantity that does not depend on P; hence, C cannot grow without bound with P. With clusters of limited size, the presence of external interference is sure to bring about a finite C_∞, with its value depending on K, perhaps relative to $\tilde{K}$. The question that is unanswered is whether this changes for $K = \tilde{K}$ and $N = \tilde{N}$, i.e., when the entire system cooperates as one cluster.

Consider the uplink. With block fading, the transmit-receive relationship can be vectorized for the entire system and all the symbols in a fading block as

$$\mathbf{Y} = \mathrm{diag}\left\{\sqrt{\mathsf{SNR}_1}, \ldots, \sqrt{\mathsf{SNR}_N}\right\}\mathbf{H}\mathbf{X} + \mathbf{Z} \tag{13}$$

where $\mathbf{Y}$ and $\mathbf{Z}$ are $N \times L$, $\mathbf{H}$ is $N \times K$ and $\mathbf{X}$ is $K \times L$. The entries of $\mathbf{Z}$ are IID with $\mathbf{Z}_{n\ell} \sim \mathcal{N}_{\mathbb{C}}(0,1)$ whereas the entries of $\mathbf{H}$ are independent with $\mathbf{H}_{nk} \sim \mathcal{N}_{\mathbb{C}}(0, g_{nk})$; notice that the normalized power gains $\{g_{nk}\}$ are directly incorporated as the variances of the entries of $\mathbf{H}$. The entries of $\mathbf{X}$ are unit variance and, since the kth row of $\mathbf{X}$ contains the signal sequence transmitted by user k over the L symbols of a fading block, the rows of $\mathbf{X}$ are independent.

Proposition 1 *Consider the uplink of a cellular system subject to block-fading with $L < K$. Define, for $n = 1, \ldots, N$, respective diagonal matrices $\mathbf{G}_n = \mathrm{diag}\{g_{n1}, \ldots, g_{nK}\}$. Each $\mathbf{G}_n$ contains along its diagonal the normalized power gains between the K transmitters and the nth receiver and thus $\mathrm{Tr}\{\mathbf{G}_n\} = 1$. If $\mathbf{X}$ is full rank, the average spectral efficiency that can be achieved reliably for $P \to \infty$ satisfies $C_\infty \leq C_\infty^{\mathsf{UB}}$ with*

$$C_\infty^{\mathsf{UB}} = -\frac{1}{K}\sum_{n=1}^{N}\frac{1}{L}\,\mathbb{E}\left[\log_2 \det\left(\mathbf{X}^\dagger \mathbf{G}_n \mathbf{X}\right)\right]. \tag{14}$$

Proof : Using the chain rule, the mutual information between $\mathbf{X}$ and $\mathbf{Y}$ (in bits per coherence block) can be expressed as

$$I(\mathbf{X};\mathbf{Y}) = I(\mathbf{H}\mathbf{X};\mathbf{Y}) - I(\mathbf{Y};\mathbf{H}|\mathbf{X}). \tag{15}$$

Invoking (5) and the zero-mean unit-variance nature of the entries of $\mathbf{X}$, it can be verified that the entries of $\mathbf{H}\mathbf{X}$ are also zero-mean and unit-variance. The term $I(\mathbf{H}\mathbf{X};\mathbf{Y})$ is upper-bounded by the value it would take if those entries were IID $\mathcal{N}_{\mathbb{C}}(0,1)$. Thus,

$$I(\mathbf{X};\mathbf{Y}) \leq \sum_{n=1}^{N} L \log_2(1 + \mathsf{SNR}_n) - I(\mathbf{Y};\mathbf{H}|\mathbf{X}). \tag{16}$$

Turning now our attention to $I(\mathbf{Y};\mathbf{H}|\mathbf{X})$, and denoting differential entropy operator by $\mathfrak{h}(\cdot)$,

$$I(\mathbf{Y};\mathbf{H}|\mathbf{X}) = \mathfrak{h}(\mathbf{Y}|\mathbf{X}) - \mathfrak{h}(\mathbf{Y}|\mathbf{H},\mathbf{X}). \tag{17}$$

Conditioned on $\mathbf{X}$, the rows of $\mathbf{Y}$ are independent (both conditionally and unconditionally on $\mathbf{H}$) and thus both differential entropy terms in (17) can be computed row-wise and simply added. It follows that

$$I(\mathbf{Y};\mathbf{H}|\mathbf{X}) = \sum_{n=1}^{N} I(\mathbf{y}_n;\mathbf{h}_n|\mathbf{X}) \tag{18}$$

where $\mathbf{y}_n = \sqrt{\mathsf{SNR}_n}\,\mathbf{h}_n\mathbf{X} + \mathbf{z}_n$ with $\mathbf{h}_n$ and $\mathbf{z}_n$ the nth rows of $\mathbf{H}$ and $\mathbf{Z}$, respectively. Since $\mathbf{h}_n$ and $\mathbf{z}_n$ are complex Gaussian vectors,

$$I(\mathbf{Y};\mathbf{H}|\mathbf{X}) = \sum_{n=1}^{N}\mathbb{E}\left[\log_2 \det\left(\mathbf{I} + \mathsf{SNR}_n \mathbf{X}^\dagger \mathbf{G}_n \mathbf{X}\right)\right] \tag{19}$$

where $\mathbf{G}_n = \mathrm{diag}\{g_{n1}, \ldots, g_{nK}\}$.

Since $I(\mathbf{X};\mathbf{Y})$ increases monotonically with P, we concentrate on upper-bounding it for $P \to \infty$.

Suppose first that we had $K < L$. Then, for large SNR_n, $n = 1, \ldots, N$, we would have

$$I(\mathbf{Y};\mathbf{H}|\mathbf{X}) = \sum_{n=1}^{N}\left(K \log_2 \mathsf{SNR}_n + \mathbb{E}\left[\log_2 \det\left(\mathbf{X}\mathbf{X}^\dagger \mathbf{G}_n\right)\right]\right) + o(1) \tag{20}$$

and, from (16),

$$I(\mathbf{X};\mathbf{Y}) \leq \sum_{n=1}^{N}\left((L-K)\log_2 \mathsf{SNR}_n - \mathbb{E}\left[\log_2 \det\left(\mathbf{X}\mathbf{X}^\dagger \mathbf{G}_n\right)\right]\right) + o(1) \tag{21}$$

whose right-hand side grows unboundedly with P indicating that there is hope for $I(\mathbf{X};\mathbf{Y})$ to grow unboundedly with P.

Conversely, when $K \geq L$, and given the full-rank of $\mathbf{X}$,

$$I(\mathbf{Y};\mathbf{H}|\mathbf{X}) = \sum_{n=1}^{N} \left(L \log_2 \mathsf{SNR}_n + \mathbb{E}\left[\log_2 \det\left(\mathbf{X}^\dagger \mathbf{G}_n \mathbf{X}\right)\right]\right) + o(1) \tag{22}$$

and (16) becomes, in the high-power regime,

$$I(\mathbf{X};\mathbf{Y}) \leq -\sum_{n=1}^{N} \mathbb{E}\left[\log_2 \det\left(\mathbf{X}^\dagger \mathbf{G}_n \mathbf{X}\right)\right] + o(1) \tag{23}$$

from which the claimed upper bound in (14) follows. ■

In light of (5), C_∞^{UB} remains bounded even if $K, N \to \infty$ with $L < K$. In sufficiently large systems, therefore, there is no hope for C_∞ to grow without bound with P.

Note that the rank constraint on $\mathbf{X}$ is very mild, accommodating every signaling strategy utilized in wireless communication. In particular, if the signal sequence transmitted by each user is IID complex Gaussian, then C_∞^{UB} in (14) can be expressed using [19, Prop. 4] in a form that is closed although not particularly convenient to work with when K and L are large. A more compact analytical handle on C_∞^{UB} can be obtained by resorting to large-dimensional results in random matrix theory, with the added advantage that the expressions then hold for IID signals corresponding to arbitrary modulation types (not limited to Gaussian). This approach leads to the following proposition.

Proposition 2 *Consider the uplink of a cellular system subject to block-fading with $L < K$. Then,*

$$C_\infty^{\mathsf{UB}} \approx \frac{1}{K} \sum_{n=1}^{N} \left[\log_2 \frac{e\, a_n}{L} - \frac{1}{L} \sum_{k=1}^{K} \log_2(1 + a_n g_{nk}) \right] \tag{24}$$

with each a_n the nonnegative solution to

$$\sum_{k=1}^{K} \frac{g_{nk}}{g_{nk} + 1/a_n} = L \tag{25}$$

and with (24) becoming exact as $K, L \to \infty$.

Proof : If each user transmits an IID sequence, then the entries of $\mathbf{X}$ are IID and the asymptotic analysis in [20], [21] can be applied to Proposition 1. Couched in the notation of this paper, we have that [20, Section VI-A]

$$-\frac{1}{L} \log_2 \det\left(\mathbf{X}^\dagger \mathbf{G}_n \mathbf{X}\right) \approx \log_2(e) - \log_2 \left(\sum_{k=1}^{K} \frac{g_{nk}}{1 + a_{nk}} \right) - \frac{1}{L} \sum_{k=1}^{K} \log_2(1 + a_{nk}) \tag{26}$$

where the approximation becomes exact as $K, L \to \infty$, with $K \geq L$, and where a_{nk} is the nonnegative solution to

$$a_{nk} = \frac{L\, g_{nk}}{\sum_{j=1}^{K} \frac{g_{nj}}{1 + a_{nj}}}. \tag{27}$$

Defining

$$a_n = \frac{L}{\sum_{j=1}^{K} \frac{g_{nj}}{1 + a_{nj}}} \tag{28}$$

we can write $a_{nk} = a_n g_{nk}$ and (26) becomes

$$-\frac{1}{L} \log_2 \det\left(\mathbf{X}^\dagger \mathbf{G}_n \mathbf{X}\right) \approx \log_2(e) + \log_2 \left(\frac{a_n}{L} \right) - \frac{1}{L} \sum_{k=1}^{K} \log_2(1 + a_n g_{nk}) \tag{29}$$

with a_n the nonnegative solution to

$$\sum_{k=1}^{K} \frac{g_{nk}}{g_{nk} + 1/a_n} = L. \tag{30}$$

■

In certain special cases such as Example 1, Propositions 1 and 2 simplify significantly.

Corollary 1 *If the system is isotropic in the sense that the set of gains from the transmitters looks the same from the vantage of each receiver, i.e., every set $\{g_{nk}\}$ for $n = 1, \ldots, N$ can be reordered into a common set $\{g_k\}$, then we can define a unique matrix $\mathbf{G}$ containing such set on its diagonal and rewrite (14) as*

$$C_\infty^{\mathsf{UB}} = -\frac{N}{K} \frac{1}{L} \mathbb{E}\left[\log_2 \det\left(\mathbf{X}^\dagger \mathbf{G} \mathbf{X}\right)\right]. \tag{31}$$

In turn, Proposition 2 then specializes to

$$C_\infty^{\mathsf{UB}} \approx \frac{N}{K} \left[\log_2 \frac{e\, a}{L} - \frac{1}{L} \sum_{k=1}^{K} \log_2(1 + a g_k) \right] \tag{32}$$

with the unique a being the nonnegative solution to

$$\sum_{k=1}^{K} \frac{g_k}{g_k + 1/a} = L. \tag{33}$$

To conclude this section, we verify the accuracy of the asymptotic approximations that have been presented for finite $K = N$ and for L small enough that (31) can be computed numerically.

Example 2 *Consider the uplink of a square fragment of the system in Example 1 having 20×20 cells, such that $K = 20 \times 20 \times 3 = 1200$, and let $L = 100$. All the sectors of all the cells in the system cooperate fully. The distance-decay exponent is $\gamma = 3.8$ and $Q|_{\text{dB}} = 20$ dB. Monte-Carlo evaluation of (31) gives $C_\infty^{\mathsf{UB}} = 5.183$ b/s/Hz/user whereas (32) gives $C_\infty^{\mathsf{UB}} = 5.181$ b/s/Hz/user.*

IV. Discussion

In this section we discuss implications, possible extensions, and caveats of the results.

A. Numerical Example

We now apply Corollary 1 to the full system in Example 1 and for values of L in the range of practical interest.

Example 3 *Re-consider Example 2 for $K, N \to \infty$. For $L = 20,000$, $C_\infty^{\mathsf{UB}} = 11.86$ whereas for $L = 1000$, $C_\infty^{\mathsf{UB}} = 7.98$.*

Let us now try to gauge, from the values of C_∞^{UB} obtained in Example 3, values for $\{\mathsf{SIR}_n\}_{n=1}^N$ in the representation in Relationship 3. Because of the isotropy, every user operates at the same spectral efficiency and $\mathsf{SIR}_n = \mathsf{SIR}$, $n = 1, \ldots, N$. The question then becomes: if one wants to assume perfect CSI at the receivers, which SIR should be inserted into Relationship 3 in order to reproduce the results in Example 3? With perfect CSI at the receivers, complex Gaussian signals are capacity-achieving and the uplink capacity of Relationship 3 with full cooperation equals, for $P \to \infty$,

$$C_\infty = \frac{1}{K} \log_2 \det \left(\mathbf{I} + \mathsf{SIR}\, \mathbf{H}\mathbf{H}^\dagger \right). \tag{34}$$

The structure of $\mathbf{H}$ satisfies the conditions of [20, Thm. 5] and thus, for $K = N \to \infty$,

$$C_\infty \overset{\text{a.s.}}{\to} 2 \log_2 \left(\frac{1 + \sqrt{1 + 4\,\mathsf{SIR}}}{2} \right) - \frac{\log_2(e)}{4\,\mathsf{SIR}} \left(\sqrt{1 + 4\,\mathsf{SIR}} - 1 \right)^2. \tag{35}$$

Solving for the SIR that equates (35) with the values for C_∞^{UB} in Example 3 we obtain

$$\begin{array}{ll} \mathsf{SIR} = 39.96\,\text{dB}, & L = 20,000 \text{ (pedestrian)} \\ \mathsf{SIR} = 28.02\,\text{dB}, & L = 1000 \text{ (vehicular)} \end{array} \tag{36}$$

which are upper bounds, not necessarily tight, on the actual SIR. Although with shadow fading and randomized user locations these values will change, they are indicative: even if an entire large system were to be operated as one cluster and with full cooperation, interference brings about a fundamental performance ceiling that corresponds to values of SIR within the range of interest in high-SNR analysis.

B. Dependencies and Extensions

The spectral efficiency ceiling depends exclusively on:

- The coherence L, which relates to the degree of mobility.
- The normalized gains $\{g_{nk}\}$, which quantify the degree of signal connectivity among users, and for which we coin the term *geometry profile*.

If $\{g_{nk}\}$ is highly skewed for a given n, then most of the power received by BS n corresponds to a few nearby users. Intuition then says that, given their relative strength, the fading of these mobiles and the overlaying signals could be determined and most of the received power should be rendered useful. The asymptotic results confirm this intuition: with fewer than L nonnegligible terms, C_∞^{UB} can be arbitrarily large and thus a sustained increase of the spectral efficiency with P is feasible. Alternatively, if the geometry profile for a given n contains a myriad minute terms, rather than a few strong ones, each of the constituent signals is simply too weak relative to the aggregate rest. Intuitively, this should give rise to a bulk of residual interference that is fundamentally undecodable, and which is substantial in the aggregate even if each of the terms is by itself small. Again, the asymptotic results confirm this intuition: if $g_{nk} = 1/K$ for $k = 1, \ldots, K$, then, as K grows without bound for fixed L, the interference becomes overwhelming and C_∞^{UB} vanishes.

Since the gains $\{g_{nk}\}$ are normalized, they are scale independent. *Cell size is therefore immaterial* in terms of the geometry profile and the amount of residual interference. Note also that the schedulers that determine which user(s) in each cell are allocated to a given signaling resource play a significant role in establishing the geometry profile. Subject to latency and quality-of-service constraints, the schedulers can therefore shape the geometry profile. Finally, power control – which plays an important role in the uplink – can be applied to modify the geometry profile, but the inclusion of *power control does not change the nature any of the results in Section III*.

For the downlink, the roles of transmission and reception are reversed relative to the uplink. The corresponding noncoherent performance can be upper-bounded by allowing all the receivers to cooperate (which is extremely optimistic in the downlink), in which case the uplink derivations carry over and a performance ceiling is observed. Tighter upper bounds might be obtained by removing the premise of receiver cooperation while considering cooperative downlink transmission strategies other than IID signaling. We conjecture that regardless of the particular assumptions, downlink behavior also abides by Relationship 2 irrespective of the size of the cooperation clusters and that the spectral efficiency ceilings and power cross-over points are generally in the same range, all else being equal.

V. Conclusions

This paper has shown that saturation of the spectral efficiency at sufficiently high SNR is unavoidable in large systems, and that the spectral efficiency behaves as illustrated in Fig. 1(b). Thus, systems should be modeled according to Relationships 2–3, not Relationship 1. Namely, the high-power regime for user n, characterized by $\mathsf{SNR}_n \gg 1$, splits into two regimes:

- The DoF regime where $\mathsf{SNR}_n \ll \mathsf{SIR}_n$. In this regime, the external interference is negligible relative to the noise and the spectral efficiency grows approximately linearly with $\log(P)$ according to the number of DoF, which remains a valid notion.
- The saturation regime where SNR_n is comparable to or greater than SIR_n. In this regime, the spectral efficiency chokes as it approaches C_∞. The notion of DoF becomes meaningless.

The transition between these regimes takes place, for user n, at $\mathsf{SNR}_n \approx \mathsf{SIR}_n$.

The points made in this paper do not nullify the benefits of cooperation, rather, it shows simply that cooperation does have some fundamental limitations that cannot be overcome just by faster backhaul overhead sharing, more sophisticated signal processing, or any other technological advance. Nevertheless, in many contexts, the value of $P_{\rm sat}$ should be large enough to allow significant gains to be observed in the DoF regime. That is, cooperation can help approach C_∞ faster and, more importantly, it can increase the value of C_∞. But the ultimate capacity is in fact bounded as a function of P.

More work is needed to solidify the meaning of these observations for real communication systems. For example, some WiFi or enterprise femtocell networks may be naturally clustered, spread out from one another, and further isolated by walls, pushing the value of $P_{\rm sat}$ up to a large enough value that the saturation effect is not actually observed in practice. In the context of coherent detection, it remains open to find the optimal cluster size for a given configuration, trading off channel estimation accuracy with overhead [14]. Other cooperative settings, for example in decentralized or relay networks, can also be considered. General quantifications of $P_{\rm sat}$ and C_∞ in terms of system parameters would be a further welcome contribution.

ACKNOWLEDGMENTS

The work of Angel Lozano is supported by the European Project FET 265578 "HIATUS". The work of Robert W. Heath Jr. is supported by the Army Research Lab Grant W911NF-10-1-0420 and the Office of Naval Research Grant N000141010337.

REFERENCES

[1] G. J. Foschini, K. Karakayali, and R. A. Valenzuela, "Enormous spectral efficiency of isolated multiple antenna links emerges in a coordinated cellular network," *IEE Proceedings - Communications*, vol. 153, no. 4, pp. 548–55, Aug. 2006.

[2] O. Simeone, O. Somekh, H. V. Poor, and S. Shamai, "Local base station cooperation via finite-capacity links for the uplink of linear cellular networks," *IEEE Trans. Info. Theory*, vol. 55, no. 1, pp. 190–204, Jan. 2009.

[3] J. G. Andrews, H. Claussen, M. Dohler, S. Rangan, and M. C. Reed, "Femtocells: Past, present, and future," *To appear, IEEE Journal on Sel. Areas in Comm.*, Apr. 2012.

[4] H. Weingarten, Y. Steinberg, and S. Shamai (Shitz), "The capacity region of the Gaussian multiple-input multiple-output broadcast channel," *IEEE Trans. on Info. Theory*, vol. 52, no. 9, pp. 3936–3964, Sept. 2006.

[5] V. R. Cadambe and S. A. Jafar, "Interference alignment and the degrees of freedom for the K user interference channel," *IEEE Trans. Inform. Theory*, vol. 54, no. 8, pp. 3425–3441, Aug. 2008.

[6] M. Maddah-Ali, A. Motahari, and A. Khandani, "Communication over MIMO X channels: Interference alignment, decomposition, and performance analysis," *IEEE Trans. Inform. Theory*, vol. 54, no. 8, pp. 3457–3470, Aug. 2008.

[7] M. K. Karakayali, G. J. Foschini, and R. A. Valenzuela, "Network coordination for spectrally efficient communications in cellular systems," *IEEE Wireless Communications*, vol. 13, no. 4, pp. 56–61, Aug. 2006.

[8] D. Gesbert, S. Hanly, H. Huang, S. Shamai, O. Simeone, and Wei Yu, "Multi-cell MIMO cooperative networks: A new look at interference," *IEEE Journal on Sel. Areas in Communications*, vol. 28, no. 9, pp. 1380 –1408, Dec. 2010.

[9] A. El Gamal, Sreekanth V, and V. V. Veeravalli, "Degrees of freedom (DoF) of locally connected interference channels with coordinated multi-point (CoMP) transmission," *arXiv:1109.1604v1*, Sept. 2011.

[10] Further advancements for E-UTRA physical layer aspects (Release 9), *3GPP TR 36.814*, 2010.

[11] A. Gorokhov, "Coordinated joint transmission in WWAN," *IEEE Communication Theory Workshop (CTW'10), available at "http://www.ieee-ctw.org/2010/mon/Gorokhov.pdf"*, May 2010.

[12] R. Irmer, H. Droste, P. Marsch, M. Grieger, G. Fettweis, S. Brueck, H.-P. Mayer, L. Thiele, and V. Jungnickel, "Coordinated multipoint: Concepts, performance, and field trial results," *IEEE Communications Magazine*, vol. 49, no. 2, pp. 102 –111, Feb. 2011.

[13] S.A. Ramprashad and G. Caire, "Cellular vs. network MIMO: A comparison including the channel state information overhead," in *IEEE International Symposium on Personal, Indoor and Mobile Radio Communications*, Sept. 2009, pp. 878 –884.

[14] A. Lozano, R. W. Heath, and J. G. Andrews, "The limits of cooperation in pilot-assisted wireless communication," in *IEEE Intl. Symposium on Information Theory (ISIT'12)*, Cambridge, MA, submitted 2012.

[15] R. H. Etkin, D. N. C. Tse, and H. Wang, "Gaussian interference channel capacity to within one bit," *IEEE Trans. on Inform. Theory*, vol. 54, no. 12, pp. 5534–5562, 2008.

[16] S. A. Jafar and S. Vishwanath, "Generalized degrees of freedom of the symmetric Gaussian K user interference channel," *IEEE Trans. on Inform. Theory*, vol. 56, no. 7, pp. 3297–3303, 2010.

[17] C. Suh, M. Ho, and D. N. C. Tse, "Downlink interference alignment," *IEEE Trans. on Communications*, vol. 59, no. 9, pp. 2616–2626, Sept. 2011.

[18] N. Jindal and A. Lozano, "A unified treatment of optimum pilot overhead in multipath fading channels," *IEEE Trans. on Communications*, vol. 58, no. 10, pp. 2939–2948, Oct. 2010.

[19] A. Lozano, A. M. Tulino, and S. Verdu, "High-SNR power offset in multiantenna communication," *IEEE Trans. Inform. Theory*, vol. 51, no. 12, pp. 4134–4151, Dec. 2005.

[20] A. M. Tulino, A. Lozano, and S. Verdú, "Impact of antenna correlation on the capacity of multiantenna channels," *IEEE Trans. Inform. Theory*, vol. 51, no. 7, pp. 2491–2509, Aug. 2005.

[21] J. W. Silverstein and Z. D. Bai, "On the empirical distribution of eigenvalues of a class of large dimensional random matrices," *Journal Multivariate Analysis*, vol. 54, pp. 175–192, 1995.