CCIT Report #671 November 2007

Downlink macro-diversity with limited backhaul capacity

O. Simeone[1], O. Somekh[2], H. V. Poor[2] and S. Shamai (Shitz)[3]

[1]CWCSPR, New Jersey Institute of Technology,

Newark, New Jersey 07102, USA

[2]Department of Electrical Engineering, Princeton University,

Princeton, NJ 08544, USA

[3]Department of Electrical Engineering, Technion, Haifa, 32000, Israel.

e-mail: osvaldo.simeone@njit.edu

(*) corresponding author

Abstract

Multicell processing in the form of joint encoding for the downlink of a cellular system is studied under the realistic assumption that the base stations (BSs) are connected to a central unit via finite-capacity links (finite-capacity backhaul). Three scenarios are considered that present different trade-offs between global processing at the central unit and local processing at the base stations and different requirements in terms of codebook information (CI) at the BSs: 1) local encoding with CI limited to a subset of nearby BSs; 2) mixed local and central encoding with only local CI; and 3) central encoding with oblivious cells (no CI). Three transmission strategies are proposed that provide achievable rates for the considered scenarios. Performance is evaluated in asymptotic regimes of interest (high backhaul capacity and extreme signal-to-noise ratio, SNR) and further corroborated by numerical results. The major finding of this work is that central encoding with oblivious cells is a very attractive option for both ease of implementation and performance, unless the application of interest requires high data rate (i.e., high SNR) and the backhaul capacity is not allowed to increase with the SNR, in which case some form of CI at the BSs becomes necessary.

I. Introduction

Multicell processing, sometimes alternatively labeled as "distributed antenna systems" or "base station cooperation", has by now become a very active area of research for both academia and industry (see [1] for a recent review). The technology prescribes joint encoding or decoding of different base stations' (BSs) signals in an infrastructure (cellular or hybrid) network for downlink or uplink, respectively. The rationale is that joint processing (encoding or decoding) in such networks is feasible since, in general, a high-capacity backbone is available to connect the BSs.

Traditionally, analysis of the performance of multicell processing has been carried out under the assumption that all the BSs in the network are connected to a central processor via links of unlimited capacity. In this case, the set of BSs effectively acts as a multiantenna transmitter (downlink) or receiver (uplink) with the caveat that the antennas are geographically distributed over a large area [1]. Since the assumption of unlimited-capacity links to a central processor is quite unrealistic for large networks, more recently, there have been attempts to alleviate this condition by considering alternative models. In [2], [3], and [4] a model is studied in which only a subset of neighboring cells is connected to the same central unit for joint processing. In [5] [6] (uplink) and [7] (downlink) a topological constraint is imposed in that there exist links only between adjacent cells, and message passing techniques are implemented in order to perform joint decoding or encoding. Finally, reference [8] focuses on the uplink and assumes that the links between all the BSs and a central processor have finite capacity (*finite-capacity backhaul*).

In this paper, we study a cellular system with finite-capacity backhaul as in [8]. In [8], the uplink of this model was studied in two scenarios: (*i*) the BSs are oblivious to the codebooks used by the mobile stations (MSs) so that decoding is exclusively performed at the central processor; and (*ii*) the BSs are aware of the codebooks used by the local and the nearby MSs. Here, we focus on the downlink and consider three scenarios that, similarly to [8], present different requirements in terms of codebook information (CI) at the BSs and different trade-offs between global processing at the central unit and local processing at the base stations:

1) *Local encoding with cluster codebook information (CI)*: in this first scenario, encoding is performed exclusively at the base stations, which are informed by the central processor (over finite-capacity links) about the messages to be transmitted (and possibly about additional information). In order to allow sophisticated encoding techniques such as dirty paper coding (DPC) [15], in addition to the local codebook, every base station is assumed to have available the encoding functions from a number of adjacent cells, similarly to case (*ii*) of [8] (we refer to this situation as "cluster CI")[1];
2) *Mixed central and local encoding with local CI*: here we assume that each BS is only aware of its own codebook (local CI). Moreover, in order to enable a better handling of inter-cell interference, we allow encoding to take place not only at the base stations, as in

[1] It should be remarked that, when employing DPC, encoding is performed with a more sophisticated encoding strategy than simple look-up on a table of codewords on the basis of the transmitted message. The transmitted signal is in fact a function of the interference sequence to be cancelled. Therefore, a more appropriate term for what we refer to as codebook information (CI) would be encoding function information. We choose the first for simpliciy but this distinction should be kept in mind.

the previous case, but also at the central unit;

3) *Central encoding with no CI (oblivious BSs)*: here encoding takes place exclusively at the central unit and base stations are oblivious to all the codebooks employed in the system, as in case (*i*) studied in [8] for the uplink.

Achievable rates are derived for all three scenarios by proposing three basic transmission schemes. It is noted that the third scenario (oblivious BSs) is a special case of the second (local CI), so that any transmission scheme designed for oblivious CI (and corresponding achievable rate) is also applicable to the local-CI scenario. Performance comparison is carried out in different regimes of interest such as high-backhaul capacity and extreme signal-to-noise ratio (SNR). The performance analysis, corroborated by numerical results, sheds light into the roles of central/ local processing, on one hand, and CI, on the other, as a function of the system parameters.

II. System model

We study the downlink of a cellular system modelled as in Fig. 1, where M cells are arranged in a linear geometry, and one terminal is active in each cell (as for intra-cell TDMA) and is located at the border between successive cells. In this case, each active terminal, say the mth, receives signals from the local mth BS and the previous, $(m-1)$th, BS. This framework is a variation of the Wyner model [9] and has been studied in [2] and later [10] in terms of sum-rate for the case where there are no restrictions on the backbone connecting the BSs. Deviating from this ideal condition, here we assume that each BS is connected to a central processor via a finite-capacity link of capacity C (bits/ channel use), as in [8]. The model is further characterized by a single parameter to account for intercell interference, namely the power gain $\alpha^2 \leq 1$ (in [2] it was $\alpha^2 = 1$). Accordingly, the signal received at the mth MS is given by

$$Y_m = X_m + \alpha X_{m-1} + Z_m, \tag{1}$$

where X_m is the symbol transmitted at a given discrete time by the mth BS with power constraint $E[|X_m|^2] = P$ and the noise Z_m is a white proper complex Gaussian process with unit power. We remark that we will be interested in asymptotic results where the number M of cells is large, and we refer to [1] for a thorough discussion on the validity of this assumption. Moreover, we focus on Gaussian (nonfaded) channels for simplicity. Finally, we assume that each MS has available CI of the local transmission only, thus ruling out sophisticated joint decoding techniques at the MSs (see, e.g., [12]).

Messages $\{W_m\}_{m=1}^M$ to be delivered to the respective mth MS are generated randomly and uniformly in the set $\{1, 2, ..., 2^{nR}\}$ at the central processor (see Fig. 1), where R (bits/ channel use) is the common rate of all the messages (*per-cell rate*). Using standard definitions, we will say that a per-cell rate R is achievable if there exists a sequence of codes (i.e., encoders and decoders) with codewords of length n such that the probability of having at least one decoding error in the system vanishes as $n \rightarrow \infty$, i.e., $\Pr[\cup_m\{\hat{W}_m \neq W_m\}] \rightarrow 0$, where $\hat{W}_m$ is the estimated message at the mth MS.

III. Reference results

In this section, we review an upper bound on the per-cell rate that can be easily derived from a result presented in [2] for $\alpha = 1$, and later extended by [10] to any $\alpha \leq 1$[2].

Proposition 1 (upper bound): The per-cell capacity of the system is upper bounded by

$$R_{ub} = \min\left\{C,\ \log_2\left(\frac{1+(1+\alpha^2)P+\sqrt{1+2(1+\alpha^2)P+(1-\alpha^2)^2P^2}}{2}\right)\right\}. \quad (2)$$

Proof: This result follows by considering a cut-set bound for two cuts, one dividing the central processor from the BSs and one the BSs from the MSs. For the second cut, it is noted that the system is equivalent to the infinite-capacity backbone case for which the per-cell capacity has been derived in [2] and [10]. ■

It is relevant to notice that upper bound (2) remains valid even if we allow multiple MSs to be simultaneously active in each cell (and P is the per-cell power constraint), as it follows

[2]Notice that this result was not given in this form in [10] but can be easily derived from Lemma 3.5 therein.

easily from [9] and duality arguments [2]. Therefore, whenever achievable rates will be shown in the following to attain (2) in specific regimes, optimality should be intended not only under the restriction of intra-cell TDMA strategies but also for the general case where more MSs can be scheduled at the same time (with a total per-cell power constraint).

For future reference, two further observations on the upper bound (2) are in order. First, it is interesting to study the low-SNR behavior, in the sense of [13]. Accordingly, the minimum energy per bit for reliable communication $E_b/N_{0\min}$, and the corresponding slope of the spectral efficiency [13] are easily shown to be given by

$$\frac{E_b}{N_{0\,\min,ub}} = \frac{\log_e 2}{1+\alpha^2} \quad , \quad S_{0,ub} = \frac{2(1+\alpha^2)^2}{1+4\alpha^2+\alpha^4}. \tag{3}$$

This result shows that the power gain with respect to a single-link (interference-free) Gaussian channel (for which $E_b/N_{0\min} = \log_e 2$) due to multicell processing can be quantified in the low-SNR regime by the factor $(1+\alpha^2) \geq 1$ (and the slope $S_{0,ub}$ is a decreasing functions of α^2). A second observation concerns the following question: how fast need the backhaul capacity C grow with increasing P in order to guarantee the optimal multiplexing gain of a system with unlimited backhaul capacity? Recalling that the maximum multiplexing gain of a multiantenna broadcast channel with channel state information at the transmitter equals the number of transmit antennas (assuming there is at least one user per cell) [11], it easily follows that the optimal multiplexing gain of the per-cell rate (2) rate is one and that, in order to achieve it, the capacity C needs to grow as $C \sim \log_2 P$. In the following, this requirement in terms of capacity C will be compared with that of practical transmission schemes.

IV. Local encoding and cluster CI

As anticipated in Sec. I, we will consider different scenarios with respect to the encoding capability of each BS. In this section, we investigate the case in which encoding is performed locally at each BS. In other words, no encoding is carried out at the central unit, whose only function is to deliver different subsets of messages $\{W_m\}_{m=1}^{M}$ to each BS. Under this assumption, we derive achievable rates based on a transmission scheme first proposed in [14]. Moreover, we comment on the performance in the asymptotic regimes of large backhaul capacity, and extreme SNR, with respect to the upper bound (2).

The considered transmission scheme is inspired by the *sequential* DPC scheme of [14] and works as follows. Every mth BS knows its encoding function and the encoding functions of the J BSs preceding it (i.e., BSs $m-i$ with $i=1,...,J$). At the beginning of the transmission block, each BS receives from the central processor $J+1$ messages $\{W_{m-i}\}_{m=0}^{J}$, that is, the local message and the messages of the J preceding BSs. The basic idea is now that, based on these J additional messages and the knowledge of the corresponding encoding functions, the mth BS can perform DPC over these messages and cancel the inter-cell interference achieving the single-user (interference-free) rate $\log_2(1+P)$.

As pointed out in [14], in order to implement the sequential DPC scheme correctly, we need to "turn off" every $(J+2)$th BS (e.g., BSs $J+2$, $2(J+2),...$) and consider the clusters of $J+1$ BSs in between silent BSs. Let us focus on any cluster and index the participating BSs as $m'=0,1,...,J+1,$ where $m'=0$ corresponds to a silent BS, $m'=1$ is the leftmost BS in a cluster and so on. Due to the inter-cell interference structure in the model at hand (recall Fig. 1), the BS with $m'=1$ in each cluster can achieve single-user rate without performing any DPC. The second BS ($m'=2$) instead needs to perform DPC on the signal transmitted by $m'=1,$ which can be done, since the second BS can reconstruct this signal knowing the encoding function and the corresponding message of the BS with $m'=1.$ Proceeding, the third BS ($m'=3$) will need to DP-code over the signal transmitted by the second, which requires, according to the discussion above, to reconstruct both the signal transmitted by the first and that transmitted by the second. From this brief description, it is clear that clustering is necessary because the signal transmitted by the m'th BS in each cluster depends in fact on the signals transmitted by all the preceding $m'-1$ BSs within the cluster due to the successive DPC encodings. It follows then from our assumptions that the maximal cluster size is exactly $J+1$.

Finally, it should be noted that since we are interested in equal per-cell rates, it is necessary to perform equal-time time-sharing between $J+2$ cluster configurations so that each BS is silent in one and only one configuration (see details in the proof below). The following proposition states the rate achievable with this strategy.

Proposition 1 (scheme 1): Assuming that every mth BS knows its own encoding function and the encoding function of the J BSs preceding it (cluster CI), the following rate is achievable with local encoding:

$$R_1 = \min\left\{\frac{2C}{J+2}, \left(1-\frac{1}{J+2}\right)\log_2(1+P)\right\}. \tag{4}$$

Proof: We consider equal-time time-sharing among $J+2$ cluster configurations so that in the jth configuration ($j=1,...,J+2$), we silence cells $(J+2)+j-1, 2(J+2)+j-1, ...$ This way, each BS occupies all the $J+2$ positions $m'=0,1,...,J+1$ in a cluster, one for each configuration. Rate splitting is then performed so that a given message W_m is split into $J+1$ messages with equal rate R' ($R=R'(J+1))$ to be transmitted during the $J+1$ configurations where the mth BS is not silent. It is easy to see that, since each BS occupies all the $J+2$ positions in a cluster and that m' messages need to be delivered by the central processor when the BS occupies position m' (see discussion above), the backhaul links to all the BSs are equally utilized and the constraint on the backhaul capacity becomes

$$C \geq R'\sum_{m=0}^{J+1} m = \frac{R'}{2}(J+1)(J+2) = \frac{R}{2}(J+2). \tag{5}$$

Moreover, from the fact that each BS is active in $J+1$ out of the overall $J+2$ configurations, we have the following further constraint on the rate:

$$R \leq \frac{J+1}{J+2}\log_2(1+P). \tag{6}$$

From (5) and (6), rate (4) easily follows. ■

Remark 1: An alternative scheme could be devised that exploits the transmission power of the "silent" cell ($m' = 0$) in each cluster. This could be done by sending the message of the first BS ($m' = 1$) to the "silent" BS ($m' = 0$) on the corresponding finite-capacity link in order to allow the latter to cooperate via coherent power combining with the first BS. Notice that the transmission of the "silent" cell does not affect the feasibility of the transmission scheme since this transmission depends only on the message of the next BS. Following the same steps as in the proof above, the rate achievable by this scheme is easily derived to be

$$\min\left\{\frac{2C(J+1)}{(J^2+3J+4)}, \frac{J}{(J+2)}\log_2(1+P) + \frac{1}{(J+2)}\log_2(1+(1+\alpha)^2P)\right\}. \quad (7)$$

Performance comparison of this rate with (4) depends on the operating regime of interest, and will not be further considered here since it would not alter meaningfully the main conclusions.

A. Performance in asymptotic regimes

In the limit of a large backhaul capacity $C \to \infty$, for fixed cluster size $J+1$, scheme 1 at hand achieves rate $R_1 \to (1 - 1/(J+2))\log(1+P)$ and is therefore limited by the loss in multiplexing gain (see also below) that follows from the need to silence a fraction $1/(J+2)$ of the BSs [14]. However, assuming that parameter J can be optimized, then using an asymptotically large cluster size $J \to \infty$ so that $2C/J > \log(1+P)$, we see that for $C \to \infty$, scheme 1 is able to achieve the single-link capacity: $R_1 \to \log_2(1+P)$, which is noted to be smaller than the upper bound R_{ub} in (2). We will see in the remark below that this is not a limit of local processing *per se*, since a transmission scheme based on local processing can be devised that achieves the upper bound R_{ub} when $C \to \infty$.

Consider now the regime of large power $P \to \infty$. In this case, the performance is limited by the backhaul capacity and we have $R_1 \to 2C/(J+2)$, which, if we allow optimization of the cluster size, becomes $R_1 \to R_{ub} = C$ (for $J = 0$, that is each cluster consists of only one active cell[3]). Letting C increase with power P, we can then ask as in Sec. III how fast capacity

[3]This corresponds to the Inter-Cell-Time-Sharing (ICTS) strategy [1]; see also discussion in the next section.

C should grow in order to be able to achieve the maximal multiplexing gain. As pointed out above, for any finite J, the maximal multiplexing gain of scheme 1 is $1 - 1/(J+2) < 1$, and, from (4), achieving this rate scaling requires the backhaul capacity C to grow as $C \sim (J+1)/2 \cdot \log_2 P$. Comparing this result with the optimal multiplexing gain of the upper bound (see Sec. III), we see that local encoding entails here a loss in terms of multiplexing gain that can be made arbitrarily small by increasing the cluster size J at the expense of a proportionally stricter requirement on the scaling of backhaul capacity C.

Finally, we obtain the low-SNR characterization for R_1 as

$$\frac{E_b}{N_{0\,\min}} = \frac{\log_e 2}{1 - \frac{1}{2+J}} \quad , \quad S_0 = 2\left(1 - \frac{1}{2+J}\right). \tag{8}$$

Comparing this result with (3), we see that in the low-SNR regime the proposed local processing-based scheme falls short of achieving the performance of the upper bound since it fails to take advantage of the inter-cell channel gains α^2, being designed to cancel inter-cell interference. However, by selecting a sufficiently large J it is clear that the single-user performance $E_b/N_{0\min} = \log_e 2$, and $S_0 = 2$, can be achieved.

Remark 2 (joint cluster-DPC): Here, we describe briefly a transmission scheme based on local processing that, unlike scheme 1, is able to achieve the upper bound R_{ub} in the regime of unlimited backhaul capacity ($C \rightarrow \infty$). The idea is to cluster the BSs as in the previously discussed scheme by silencing one every $(J+2)$th BS, send all the messages to be delivered within the cluster to all the participating BSs, and then perform joint DPC for the messages in the cluster at each BS. This contrasts with the previous scheme where sequential DPC was carried out. Notice that this scheme requires a modification with respect to the previous framework in that every BS within a cluster needs to be informed about the encoding functions of all the $J+1$ BSs within the same cluster (instead of the preceding BSs). This implies, once time-sharing is taken into account as explained above, that knowledge of $2J+1$ encoding function is required at each node (instead of $J+1$ as in the case of sequential DPC). As can be easily inferred from the results in [2] which leverage the uplink-downlink duality of [16], the rate achievable by this scheme is

$$\tilde{R}_1 = \min\left\{\frac{C}{J+1}, \frac{1}{J+2}\min_{tr(\mathbf{\Upsilon})\leq 1/P}\max_{tr(\mathbf{P})\leq 1}\log_2\frac{|\mathbf{\Upsilon}+\mathbf{HPH}^H|}{|\mathbf{\Upsilon}|}\right\}, \tag{9}$$

with the $(J+1)\times(J+1)$ channel matrix defined as

$$\mathbf{H} = \begin{bmatrix} 1 & 0 & \cdots & & 0 \\ \alpha & 1 & 0 & \ddots & \\ 0 & \alpha & \ddots & 0 & \vdots \\ \vdots & 0 & \ddots & & 0 \\ 0 & & \ddots & \alpha & 1 \end{bmatrix} \tag{10}$$

and $\mathbf{P} = \mathrm{diag}([P_1 \cdots P_{J+1}]) \succeq \mathbf{0}$ and $\mathbf{\Upsilon} = \mathrm{diag}([\gamma_1 \cdots \gamma_{J+1}]) \succ \mathbf{0}$ being diagonal matrices. As can be concluded from (9) and the results in [2], for $C \to \infty$, $J \to \infty$ and $C/J \to \infty$, we have $\tilde{R}_1 \to R_{ub}$. However, it will be shown in Sec. VII, that for relatively small values of C, rate (9) is generally smaller than (4). Finally, it is easy to see that this scheme has the same limitations in terms of multiplexing gains as scheme 1 above and that its requirement in terms of scaling of capacity C is more demanding ($C \sim (J+1)^2/(J+2)\cdot\log_2 P$). The results discussed in this section suggest that this limitation in the multiplexing gain is due to local processing and could be overcome by global processing at the central unit. This is indeed confirmed in the next sections.

V. MIXED LOCAL AND CENTRAL ENCODING WITH LOCAL CI

In this section, we consider a second scenario that alleviates two main practical problems of the local-encoding scenario studied in the previous section, namely: (*i*) the large computational complexity associated with the multiple DPC encodings to be carried out at each BS (it is noted that each BS needs to calculate a number of DPC encodings on the order of the cluster size J); and (*ii*) the need for each BS to be aware of the codebooks of J other BSs. Towards this goal, here we assume that the central unit has encoding capabilities and that each BS is aware only of its own codebook (local CI). As in the previous section, we derive an achievable rate under the said assumption and then study its characterization in asymptotic regimes of interest with respect to the upper bound (2).

It should be mentioned right away that rate

$$R_{ICTS} = \min\{C,\ 1/2 \log_2(1+P)\} \tag{11}$$

can be straightforwardly achieved under the assumption of local CI by turning off one of every two BSs and using single-user codes for the active BSs (which now see interference-free channels). Notice that this corresponds to the scheme presented in the previous section with $J = 0,$ and that it follows the Inter Cell Time Sharing (ICTS) approach of [1]. Moreover, in this case, no encoding is carried out at the central processor. It should also be noted that any rate achievable under the assumptions of oblivious cells, studied in the next section, can also be achieved in the less restrictive case of local CI studied here (recall discussion in Sec. I).

In order to improve on R_{ICTS}, we consider the following transmission scheme (to be referred to as scheme 2). As far as the first BS is concerned, the central processor simply sends message W_1 and the BS uses a regular Gaussian codebook transmitting the sequences of n symbols $\mathbf{X}_1$. The central unit then quantizes $\mathbf{X}_1$ using a proper Gaussian quantization codebook with 2^{nR_q} codewords, producing the sequence of n symbols $\hat{\mathbf{X}}_1$. This is delivered, along with the local message W_2, on the limited-capacity link towards the second BS. The latter transmits its message W_2 by performing DPC over the quantized signal $\hat{\mathbf{X}}_1$. The procedure is repeated in the same way for the successive BSs (notice that the central unit must reproduce the transmitted signal $\mathbf{X}_m$, which is possible given that the central unit knows messages, encoding functions and quantization codebooks). Notice that in order to satisfy the capacity constraint on the backhaul links, the quantization rate must satisfy $R_q + R \leq C$. The following proposition quantifies the rate achievable with this scheme.

Proposition 3 (scheme 2): Assuming that every mth BS knows only its own encoding function (local CI), the following rate is achievable with mixed local and central encoding:

$$R_2 = \begin{cases} C & \text{if } C \leq \log_2\left(1+\frac{P}{1+\alpha^2 P}\right) \\ R_2' & \text{otherwise} \end{cases} \tag{12}$$

where

$$R_2' = \log_2\left(1 - \frac{2^C}{\alpha^2 P} + \sqrt{1 + \frac{2^{C+1}}{\alpha^2}\left(2+\frac{1}{P}\right) + \frac{2^{2C}}{\alpha^4 P^2}}\right) - 1 \tag{13}$$

for $\alpha > 0$ and $\log_2(1+P)$ for $\alpha = 0$.

Proof: See Appendix-A.

It is noted that condition $C \leq \log_2(1 + P/(1+\alpha^2 P))$ in (12) corresponds to the case where a rate C, which upper bounds the performance as per (2), can be achieved by simple single-user encoding and decoding in each cell, whereby inter-cell signals are treated as interference. Also notice that it can be easily proved that rate R_2' (13) is a continuous function of α for $\alpha \geq 0$.

A. Performance in asymptotic regimes

From (12), we can derive the asymptotic performance of the proposed scheme. For $C \to \infty$, we have $R_2 \to \log_2(1+P) < R_{ub}$ (as for R_1 and $\tilde{R}_1$), which corresponds to perfect interference pre-cancellation via DPC.

For $P \to \infty$, we have

$$\lim_{P\to\infty} R_2 = \min\left(C,\ \log_2\left(1+\sqrt{1+\frac{2^{C+2}}{\alpha^2}}\right) - 1\right), \tag{14}$$

which is a non-increasing function of α and reduces to C when $\alpha = 0$. It is noted that the second term of (14) is dominant for $\alpha^2 \geq 1/(2^C - 1)$, in which case R_2, unlike R_1, is asymptotically (with P) smaller that the upper bound C. In particular, with $\alpha^2 = 1$ and increasing C, the rate $R_2 \to C/2$ for $P \to \infty$.

We now turn to the analysis of the multiplexing gain when the capacity C is allowed to increase with the power P. By substituting $C = r\log_2 P$ in (4), it can be seen that the multiplexing gain with this choice is given by $\min(r/2, 1)$ so that the optimal multiplexing gain of 1 can be achieved by having $C \sim 2\log_2 P$. This contrasts with the case of local processing studied in the previous section where the optimal multiplexing gain was not achievable.

Finally, the low-SNR characterization is given by

$$\frac{E_b}{N_{0\,\min}} = \log_e 2 \quad , \quad S_0 = \frac{2}{1+2\alpha^2 2^{-C}}. \tag{15}$$

where we see that single-user performance in terms of $E_b/N_{0\min}$ is achieved, similarly to the case treated in the previous section, whereas the same can be said for the slope only as $C \to \infty$ (see also the discussion above).

VI. Central encoding with no CI

Here, we study the case of oblivious BSs (no CI) investigated in [8] for the uplink of the channel at hand. In particular, we assume that encoding is exclusively performed at the central unit and that the BSs are not aware of *any* codebook in the system. We consider the following transmission scheme. The central unit performs joint DPC as in the previous case with the caveat that it assumes a smaller signal-to-noise ratio $\tilde{P}$ (see the sketch of the proof for details):

$$\tilde{P} = \frac{P}{\frac{1+(1+\alpha^2)P}{2^C-1}+1}, \tag{16}$$

and a power constraint

$$E[|\tilde{X}_m|^2] = \frac{P}{1+\frac{1}{2^C-1}}, \tag{17}$$

producing the sequences of n symbols $\{\tilde{\mathbf{X}}_m\}_{m=1}^M$. Similarly to the previous section, each $\tilde{\mathbf{X}}_m$ is quantized using a proper Gaussian quantization codebook with 2^{nC} codewords, producing the sequence of n symbol $\hat{\mathbf{X}}_m$. Finally, each sequence $\hat{\mathbf{X}}_m$ is communicated to the mth BS on the limited-capacity link and transmitted by the BS (i.e., $\mathbf{X}_m = \hat{\mathbf{X}}_m)$. As shown in the proof, this choice satisfies the power constraint $E[|X_m|^2] = P$.

Proposition 4 (scheme 3): Assuming that the BSs are oblivious (no CI), the following rate is achievable with central encoding:

$$R_3 = \log_2\left(\frac{1+(1+\alpha^2)\tilde{P}+\sqrt{1+2(1+\alpha^2)\tilde{P}+(1-\alpha^2)^2\tilde{P}^2}}{2}\right). \tag{18}$$

Proof: See Appendix-B.

A. Performance in asymptotic regimes

In absence of constraints on the backhaul, $C\rightarrow\infty$, unlike R_1 and R_2, the scheme proposed above achieves the upper bound (2) $R_3 \rightarrow R_{ub}$ (since $\tilde{P}\rightarrow P$).

Moreover, for $P\rightarrow\infty$ we have

$$\lim_{P\rightarrow\infty} R_3 = C-1+\log_2\left(1+\sqrt{1-\frac{4\alpha^2}{(1+\alpha^2)^2}(1-2^{-C})^2}\right), \tag{19}$$

which is larger than $C-1$ but generally smaller than the upper bound $R_{ub}=C$ (for $P\rightarrow\infty$) unless $\alpha=0$.

As far as the multiplexing gain (with capacity C scaling with P) is concerned, substituting $C = r\log_2 P$ in (18), it can be seen that the multiplexing gain with this choice is given by $\min(r,1)$, so that the optimal multiplexing gain of 1 can be achieved by having $C\sim\log_2 P$. This shows again that central encoding is instrumental in achieving the optimal multiplexing and, compared with scheme 2, presents a reduction by a factor 2 in the required scaling for capacity C.

Finally, the low-SNR characterization is given by

$$\frac{E_b}{N_0}_{\min} = \frac{E_b}{N_0}_{\min,ub}\cdot\frac{1}{(1-2^{-C})} \quad , \quad S_0 = S_{0,ub}\cdot\frac{1}{1+S_{0,ub}\frac{2^{-C}}{1-2^{-C}}}. \tag{20}$$

This result shows that the power loss due to finite capacity backhaul can be quantified in a simple way in the low-SNR regime by $(1-2^{-C})$, which, accordingly to the discussion above, tends to zero for $C\rightarrow\infty$. It is remarked that, interestingly, the low-SNR performance (20) of the scheme at hand coincides with the uplink transmission strategy of [8].

VII. NUMERICAL RESULTS AND DISCUSSION

In the previous sections, we discussed the asymptotic behavior of the proposed techniques, which has shed some light on the performance trade-offs of different assumptions in terms of local/ central processing and CI. In this section, we further investigate the regime of finite capacity C and power P.

Dependence on the backhaul capacity C: Fig. 2 shows the rates achievable by local processing and cluster CI (R_1 and $\tilde{R}_1$ with optimized J, and R_{ICTS}), by mixed processing and local CI (R_{ICTS}, R_2 and R_3) and by central processing and no CI (R_3) versus the backhaul capacity C for $P = 10dB$ and $\alpha = 1$[4]. The optimal cluster-size J is, as expected from the discussion in Sec. IV-A, increasing with the capacity C (not shown). It is seen that if C is large enough, and for relatively small to moderate values of P (see next figure), scheme 3, which performs central processing with oblivious cells, is to be preferred. Moreover, if central processing is not feasible for limitations at the central unit, it is seen that for sufficiently small values of C ($C < 30$), scheme 1 (R_1) is generally advantageous over $\tilde{R}_1$, even though the latter is asymptotically ($C \to \infty$) optimal. For this reason in the following we shall not consider $\tilde{R}_1$. Also notice that while scheme 2 and scheme 3 attain the respective asymptotic values for $C \simeq 10$, convergence is much slower for schemes based on no central processing.

Dependence on the power P: Fig. 3 shows the same achievable rate discussed above versus the power P for $C = 6$ and $\alpha = 1$. Here, the optimal cluster-size J for scheme 1 (R_1) is, as expected from the discussion in Sec. IV-A, decreasing with the power P. For small-to-moderate power P, as discussed in the previous example, the preferred scheme is scheme 3 for

[4]What we report is actually an upper bound on $\tilde{R}_1$ obtained by setting $\mathbf{\Upsilon} = 1/(P(J+1))\mathbf{I}$ in (9):

$$\tilde{R}_1 \leq \min\left\{\frac{C}{J+1}, \frac{1}{J+1}\max_{tr(\mathbf{P})\leq 1}\log\left|\mathbf{I}+P(J+1)\mathbf{H}\mathbf{P}\mathbf{H}^H\right|\right\}, \tag{21}$$

which can be easily solved by numerical tools for convex optimization. This choice has no consequences in our discussion since it is enough to give evidence to the negative conclusion about the performance of $\tilde{R}_1$ discussed in the text. Notice that in order to solve the original problem (9), one could employ the numerical technique in [16].

its capability of performing joint DPC via central processing. However, as the power increases, we know from the asymptotic analysis that CI, either local (as in ICTS) or cluster (as in scheme 1), plays the leading role. This is confirmed by Fig. 3, where it is clearly shown that R_1 and R_{ICTS} become advantageous over R_3 for $P > 30dB$.

Dependence on the inter-cell power gain α: The impact of the inter-cell power gain α^2 is shown in Fig. 4. Scheme 1 is designed to cancel the inter-cell interference and thus its performance does not depend on α (see also discussion in Sec. IV-A on the low-SNR regime). Moreover, while scheme 2 suffers from increasing α^2 due to the enhanced noise level caused by quantization of the adjacent-cell transmission signal, scheme 3, similarly to the upper bound R_{ub}, is able to exploit the extra signal path due to a larger α^2.

A. Discussion

Our analysis and the numerical results above have shown that by appropriately selecting the encoding schemes and corresponding system parameters (e.g., the cluster size J for scheme 1): (*i*) the upper bound R_{ub} can be achieved in all the three considered scenarios listed in Sec. I in the regime of large backhaul capacity ($C \to \infty$)[5]; (*ii*) in the regime of large power ($P \to \infty$) with fixed capacity C, achieving the upper bound $R_{ub} = C$ is only possible if some form of CI is available at the BSs (as for the scenarios in Sec. V and VI); (*iii*) allowing the capacity C to increase with power P, the optimal multiplexing gain of 1 is achieved only by schemes 2 and 3 that employ central encoding, and, moreover, in order to attain this, the capacity C needs to scale as $2\log_2 P$ for scheme 2 and $\log_2 P$ for scheme 3; and (*iv*) for finite C and P, low-SNR analysis and numerical results have shown that all the considered schemes fall short of achieving the upper bound. Moreover, for relatively small to moderate values of P, scheme 3, which performs central processing with oblivious cells, is to be preferred, whereas for higher values of P (and finite capacity C), transmission schemes that can leverage CI generally allow

[5]It is noted that $R_2 \to \log(1+P)$ for $C \to \infty$, but, as explained in Sec. V, R_3 is achievable also under the less restrictive assumptions of local CI and $R_3 \to R_{ub}$ for $C \to \infty$.

the system to achieve better performance.

Remark 3 (extension to regular Wyner model): As a final remark, we would like to comment on the extension of this work to the regular linear Wyner model (where every MS receives from the local cell and two adjacent cells) [9]. It is readily noted that schemes 1 (R_1) and 2, that employ the causal structure of the channel to enable DPC, are not directly applicable, whereas scheme 3 (and also $\tilde{R}_1$) can be applied not only to a regular Wyner model but, more generally, to any vector broadcast channel with limited backhaul.

VIII. CONCLUSIONS

This paper has studied the performance of multicell processing for the downlink of a cellular system under the realistic assumption that the base stations are connected to a central processor via finite-capacity (typically wired) links. We have studied three scenarios characterized by different trade-offs between central and local encoding, and codebook information (CI) at the BSs. The main conclusion is that central processing, even with BSs oblivious to all the codebooks used in the system (i.e., no CI), is the preferred choice for small-to-moderate SNRs, or when the backhaul capacity is allowed to increase with the SNR. On the other hand, for high SNR and fixed backhaul capacity, a system with oblivious cells is limited by the quantization noise, and knowledge of the codebooks at the BSs becomes the factor dominating the performance. Therefore, in this scenario, transmission schemes characterized by local CI or cluster CI coupled with local processing allow to achieve better performance than central processing with oblivious cells.

An interesting issue is the assessment of possible duality results between uplink and downlink channels with limited-capacity backhaul under different assumptions concerning CI and central/ local processing. In this paper, we have provided a downlink transmission scheme that offers the same low-SNR performance as the uplink strategy of [8] for oblivious base stations and the Wyner model, but the general problem remains open (see [17] for a recent duality result concerning relay-enhanced uplink and downlink channels).

IX. APPENDIX

A. Appendix-A: Proof of Proposition 3

Following the discussion in Sec. V, the signal $\mathbf{X}_{m-1}$ is quantized as $X_{m-1} = \hat{X}_{m-1} + Z_{q,m-1}$, where $Z_{q,m-1}$ is a sequence of i.i.d. complex Gaussian random variables with zero mean and variance σ_q^2, independent of $\hat{X}_{m-1}$, which models the quantization error. In order to send the quantized signal $\hat{X}_{m-1}$ to the mth BS, the following condition must be satisfied (subscript "2" is dropped from the rate for simplicity of notation):

$$R_q = C - R \geq I(\hat{X}_{m-1}; X_{m-1}) = \log_2 \left(\frac{P}{\sigma_q^2} \right), \tag{22}$$

where $C - R \geq 0$ is the excess capacity on the mth link (recall that message W_m must be transmitted as well). From the previous equation, we can conclude that the variance of the quantization error is

$$\sigma_q^2 = \frac{P}{2^{C-R}}. \tag{23}$$

The mth BS performs DPC over the quantized codeword $\hat{X}_{m-1}$. In order to derive the rate achieved by DPC, we can write the received signal at the mth MS (1) as

$$\begin{aligned} Y_m &= X_m + \alpha X_{m-1} + Z_m = \\ &= X_m + \alpha \hat{X}_{m-1} + \tilde{Z}_m, \end{aligned} \tag{24}$$

where $\tilde{Z}_m$ is proper Gaussian with power $1 + \alpha^2 \sigma_q^2$ and is independent of X_m and $\hat{X}_{m-1}$. Therefore, we have that the achievable rate with the scheme at hand satisfies

$$R \leq \log_2 \left(1 + \frac{P}{1 + \alpha^2 \sigma_q^2} \right) = \log_2 \left(1 + \frac{P}{1 + \frac{\alpha^2 P}{2^{C-R}}} \right). \tag{25}$$

From (25), if $C \leq \log_2 (1 + P/(1 + \alpha^2 P))$, rate $R = C$, which corresponds to the upper bound (2) is clearly achievable. Otherwise, we can consider (25) with equality and solve the corresponding fixed-point equation. This leads to (12)-(13).

B. Appendix-B: Proof of Proposition 4

Quantization is performed at the central unit as $\hat{X}_m = \tilde{X}_m + Z_{q,m}$, where $\tilde{X}_m$ and $Z_{q,m}$ are independent sequences of i.i.d. complex Gaussian random variables with zero mean and variances $P/\left(1+1/(2^C-1)\right)$ (due to the power constraint (17)) and σ_q^2, respectively ($Z_{q,m}$ models the quantization error). In order to send the quantized signal $\hat{X}_m$ to the mth BS, the following condition must be satisfied:

$$C \geq I(\hat{X}_m; \tilde{X}_m) = \log_2\left(1 + \frac{P}{\sigma_q^2\left(1+\frac{1}{2^C-1}\right)}\right), \tag{26}$$

so that, taking (26) with equality, we have

$$\sigma_q^2 = \frac{P}{2^C}. \tag{27}$$

The signal transmitted by each BS is the quantized sequence $X_m = \hat{X}_m$, which satisfies the power constraint $E[|X_m|^2] = P/(1+1/(2^C-1)) + P/2^C = P$. The signal received at each MS reads

$$Y_m = \tilde{X}_m + \alpha\tilde{X}_{m-1} + \tilde{Z}_m, \tag{28}$$

with $\tilde{Z}_m = Z_m + \alpha Z_{q,m} + \alpha Z_{q,m-1} \sim \mathcal{CN}(0, 1+(1+\alpha^2)\sigma_q^2)$, independent of $\tilde{X}_m$ and $\tilde{X}_{m-1}$. From the previous equation we see that the system can be seen as a modified Wyner model in the sense of (1) with enhanced noise due to quantization. The corresponding SNR is

$$\tilde{P} = \frac{E[|\tilde{X}_m|^2]}{1+(1+\alpha^2)\sigma_q^2} = \frac{P}{\frac{1+(1+\alpha^2)P}{2^C-1}+1} \tag{29}$$

The result then follows from application of the upper bound (2).

ACKNOWLEDGMENT

The research was supported in part by a Marie Curie Outgoing International Fellowship and the NEWCOM++ network of excellence both within the 6th European Community Framework Programme, by the U.S. National Science Foundation under Grants CNS-06-25637, CNS-06-26611, and ANI-03-38807, and also by the REMON Consortium.

REFERENCES

[1] O. Somekh, O. Simeone, Y. Bar-Ness, A. Haimovich, U. Spagnolini and S. Shamai, "An information theoretic view of distributed antenna processing in cellular systems," in *Distributed Antenna Systems: Open Architecture for Future Wireless Communications*, Auerbach Publications, CRC Press.

[2] O. Somekh, B.M. Zaidel and S. Shamai (Shitz), "Spectral efficiency of joint multiple cell-site processors for randomly spread DS-CDMA systems," to appear in *IEEE Trans. Inform. Theory*.

[3] P. Marsch and G. Fettweis, "A framework for optimizing the downlink performance of distributed antenna systems under a constrained backhaul," in *Proc. European Wireless Conf. (EW 2007)*, April 1-4, Paris, France, 2007.

[4] P. Marsch and G. Fettweis, "A Framework for Optimizing the Uplink Performance of Distributed Antenna Systems under a Constrained Backhaul," in *Proc. IEEE International Conference on Communications (ICC 2007)*, Jun. 24-28, Glasgow, Scotland, 2007.

[5] E. Atkas, J. Evans and S. Hanly, "Distributed decoding in a cellular multiple access channel," in *Proc. IEEE International Symposium on Inform. Theory (ISIT 2004)*, p. 484, Jun. 27-Jul. 2, Chicago, Illinois, 2004.

[6] O. Shental, A.J. Weiss, N. Shental and Y. Weiss, "Generalized belief propagation receiver for near optimal detection of two-dimensional channels with memory," in *Proc. Inform. Theory Workshop* (ITW 2004), San Antonio, Texas, Oct. 24-29, 2004.

[7] B. L. Ng, J. S. Evans and S. V. Hanly, "Distributed downlink beamforming in cellular networks," in *Proc. IEEE International Symposium on Inform. Theory (ISIT 2007),* Nice, France, June 24-29, 2007.

[8] A. Sanderovich, O. Somekh and S. Shamai (Shitz), "Uplink macro diversity with limited backhaul capacity," in *Proc. IEEE International Symposium on Inform. Theory (ISIT 2007)*, Nice, France, June 24-29, 2007.

[9] A. D. Wyner, "Shannon-theoretic approach to a Gaussian cellular multiple-access channel," *IEEE Trans. Inform. Theory*, vol. 40, no. 6, pp. 1713-1727, Nov. 1994.

[10] S. Jing , D. N. C. Tse, J. B. Soriaga, Jilei Hou, J. E. Smee and R. Padovani, "Downlink macro-diversity in cellular networks," in *Proc. IEEE International Symposium on Inform. Theory (ISIT 2007)*, Nice, France, June 24-29, 2007.

[11] N. Jindal, “MIMO broadcast channels with finite-rate feedback,” *IEEE Trans. Inform. Theory*, vol. 52, no. 11, pp. 5045-5060, Nov. 2006.

[12] J. Jiang, Y. Xin and K. Garg, "Interference channels with common information," submitted [arXiv:cs/0610107v1].

[13] S. Verdú, "Spectral efficiency in the wideband regime," *IEEE Trans. Inform. Theory*, vol. 48, no. 6, pp. 1319-1343, June 2002.

[14] A. Lapidoth, S. Shamai (Shitz) and M. Wigger, "A linear interference network with local side-information," in *Proc. International Symposium on Inform. Theory (ISIT 2007)*, Nice, France, June 24-29, 2007.

[15] M. H. M. Costa, "Writing on dirty paper," *IEEE Trans. Inform. Theory*, vol. IT-29, no. 3, pp. 439-441, 1983

[16] Wei Yu and Tian Lan, "Transmitter optimization for the multi-antenna downlink with per-antenna power constraints," *IEEE Trans. Signal Processing*, vol. 55, no. 6, pp. 2646-2660, June 2007.

[17] S. A. Jafar, K. S. Gomadan and C. Huang, "Duality and rate optimization for multiple access and broadcast channels with amplify-and-forward relays," *IEEE Trans. Inform. Theory*, vol. 53, no. 10, pp. 3350-3370, Oct. 2007.

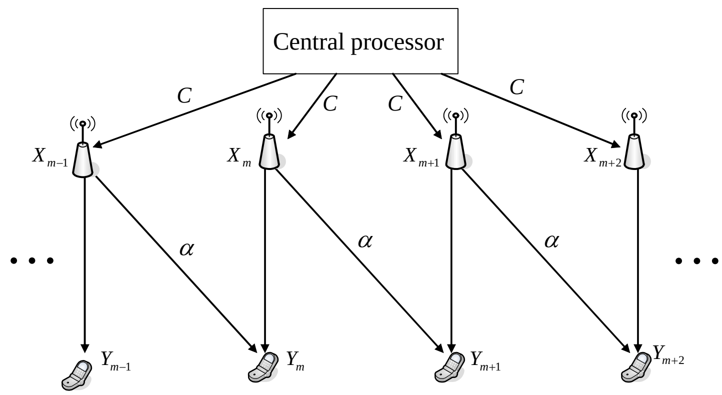

Fig. 1. Linear cellular model of interest characterized by users on the borders between successive cells and finite-capacity links between a central unit processor (that generates the messages to be delivered to each user) and the base stations.

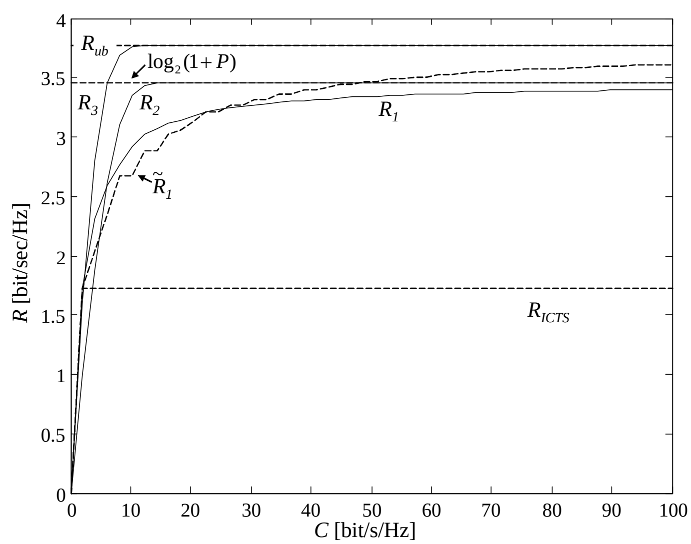

Fig. 2. Rates achievable with local processing and cluster CI (R_1, $\tilde{R}_1$ and R_{ICTS}), with mixed processing and local CI (R_{ICTS}, R_2 and R_3) and with central processing and no CI (R_3) versus C for $P = 10dB$ and $\alpha = 1$.

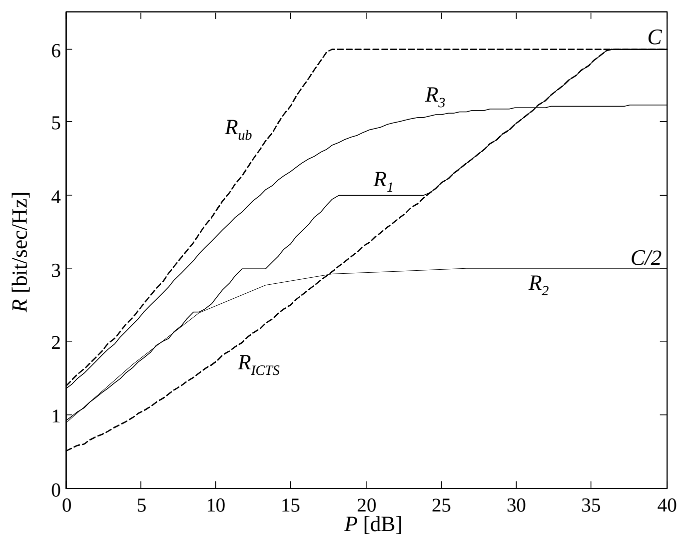

Fig. 3. Rates achievable with local processing and cluster CI (R_1 and R_{ICTS}), with mixed processing and local CI (R_{ICTS}, R_2 and R_3) and with central processing and no CI (R_3) versus P for $C = 6$ and $\alpha = 1$.

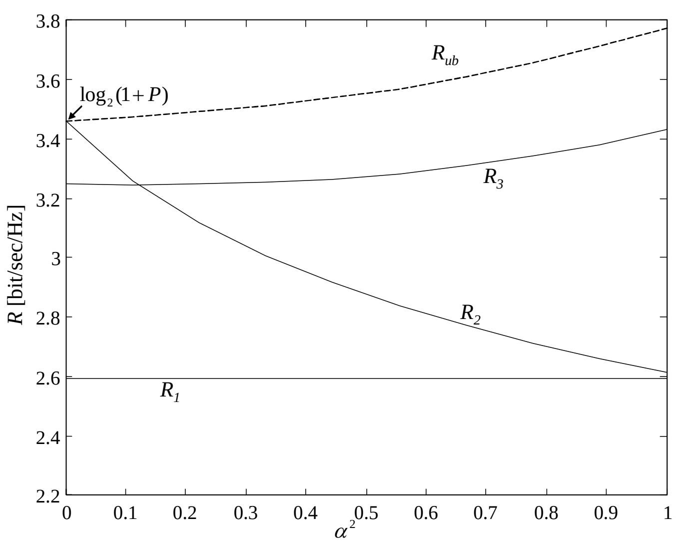

Fig. 4. Rates achievable with local processing and cluster CI (R_1), with mixed processing and local CI (R_2 and R_3) and with central processing and no CI (R_3) versus α^2 for $C = 6$ and $P = 10dB$.