A THROUGHPUT ANALYSIS FOR OPPORTUNISTIC BEAMFORMING WITH QUANTIZED FEEDBACK

Jose Lopez Vicario
Centre Tecnològic de Telecomunicacions de Catalunya
SPAIN
jose.vicario@cttc.es

Roberto Bosisio, Umberto Spagnolini
Politecnico di Milano
ITALY
rbosisio@elet.polimi.it
spagnoli@elet.polimi.it

Carles Anton-Haro
Centre Tecnològic de Telecomunicacions de Catalunya
SPAIN
carles.anton@cttc.es

ABSTRACT

Opportunistic schemes employing multiple beams have recently attracted significant interest due to the capability to achieve both multiuser diversity and spatial multiplexing gain at limited feedback load. In this paper we explore the effect of the feedback quantization on the performance of these schemes. We derive a closed-form expression of the system throughput with and without feedback quantization. By doing so, we analytically assess the impact of the number of terminals and the restriction in the feedback bandwidth. It is shown that most of the throughput is preserved by using a few quantization bits.

I. INTRODUCTION

Multi-user diversity (MUD) concepts, first introduced by Knopp and Humblet in [1], rely on the assumption that different users in a wireless multi-user system experience independent fading processes. In those circumstances, the downlink aggregated cell throughput of a Single-Input Single-Output (SISO) multi-user system can be maximized by scheduling in each time slot the user with the most favorable channel conditions [2]. To do so, only partial CSI (i.e., SNRs) channel is to be estimated by the terminals and reported back to the Base Station (BS) over a feedback channel. Going one step further, in [3] the authors showed that most of the MUD can still be extracted when partial CSI is quantized with a very low number of bits. The analysis was conducted for a SISO system.

In a context of Multiple-Input Multiple-Output (MIMO) Broadcast Channels (BC) it was recently shown [4] that Dirty Paper Coding (DPC) is the capacity-achieving technique. However, DPC has two main drawbacks: its computational and implementation complexity due to successive encoding and decoding processes and, also, the need for perfect channel state information (CSI) at the transmitter. Fortunately, transmit Zero-Forcing (ZF) methods constitute a less computationally-intensive alternative with an identical growth rate of the sum-rate for an increasing number of active users [5]. Still, ZF methods require perfect transmit CSI which in FDD systems is difficult to obtain. Multi-user random beamforming schemes [6], on the contrary, merely require partial CSI at the transmitter, mostly SINR measurements for each transmit beam. Hence, those schemes have emerged as a viable alternative to DPC and ZF, in particular in the asymptotic case of a growing number of users since, in that region, the sum-rate exhibits the same growth rate as ZF and DPC do [6].

In this paper, we show how in a MIMO opportunistic beamforming context most of the MUD gains can be extracted with quantized versions of the measured SINRs. In this way, we extend the previous work by the authors in [3] to a more general case. In particular, we derive closed-form expressions of the aggregated throughput for such communications scenario in the presence of adaptive modulation and coding (AMC). Throughput expressions are obtained both for systems with quantized and non-quantized (analog) partial CSI. Besides, we propose a non-uniform quantization law (on the basis of post-scheduling SINR statistics) which provides substantial gains with respect to the uniform case to finally prove that penalties associated to quantization are small even for a very reduced number of quantization levels.

II. SIGNAL MODEL

Consider the downlink of a cellular system with one Base Station (BS) equipped with \( M \) antennas, and \( K \) single-antenna Mobile Stations (MS). In order to serve multiple users in the same time-slot, a random (and linear) precoding matrix [6] is used at the transmit side. For a given time slot, we construct a pre-coding matrix \( \mathbf{W} = [\mathbf{w}_1, \mathbf{w}_2, ..., \mathbf{w}_M] \), whose columns \( \mathbf{w}_i \in \mathbb{C}^{M \times 1} \), \( i = 1..M \), are random orthonormal vectors drawn according to an isotropic distribution [7]. Each of those vectors is then used to transmit data to the users experiencing the highest SINRs. The received signal at the \( k \)-th MS can be written as:

\[
r_k = \mathbf{h}_k^T \mathbf{W} \mathbf{s} + n_k
\]

where in the above expression the time index has been dropped for the ease of notation, \( \mathbf{h}_k \in \mathbb{C}^{M \times 1} \) is the channel vector gain between the BS and the \( k \)-th MS, for which each component is assumed to be independent and identically distributed, circularly symmetric Gaussian random variable with zero mean and unit variance (\( \mathbf{h}_k \sim \mathcal{CN}(0, \mathbf{I}_M) \)), \( \mathbf{s} \in \mathbb{C}^{M \times 1} \) is the symbol vector, and \( n_k \in \mathbb{C} \) denotes additive Gaussian noise (AWGN) with zero mean and variance \( \sigma^2 \). The active users in the system are assumed to undergo independent Rayleigh fading processes. Further, we consider quasi-static fading, i.e., the channel response remains constant during one time-slot and it changes to a new independent realization in the subsequent one.

Concerning channel state information, we assume perfect CSI knowledge at the receive side (MSs), and the avail-
ability of a low-rate error- and delay-free feedback channel to convey partial CSI to the transmitter. Finally, the total transmit power, $P_t$, is constant and evenly distributed among active beams, i.e., $\mathbb{E}\{s_i^Hs_i\} = P_t$. Then, we can define $\rho = \frac{P_t}{2}$ as the average transmit SINR.

III. POST-SCHEDULING SINR STATISTICS

According to the signal model presented in the previous section, the received signal for user $k$ when using beamformer $i$ for transmission can be re-written (recall Eq. (1)) as:

$$r_k = h_k^Tw_is_i + \sum_{j=1}^M h_j^Tw_js_j + n_k$$  \hspace{1cm} (2)

where $s_j$ stands for the symbol transmitted with beam $j$. Notice that the last two terms in the above expression correspond to the interference-plus-noise contribution and, hence, the corresponding SINR amounts to:

$$\text{SINR}_{k,i} = \frac{|h_k^Tw_i|^2}{M/\rho + \sum_{j \neq i} |h_j^Tw_i|^2} = \frac{z}{M/\rho + y}$$

Since we assume all users experience i.i.d Rayleigh fading and the beamformers are orthonormal to each other, $z$ and $y$ become independent chi-square random variables, $z \sim \chi^2_{2}$ and $y \sim \chi^2_{2M-2}$ [6]. Bearing this in mind, the CDF and pdf of the SINR can be expressed as:

$$F_{\text{SINR}}(\gamma) = 1 - e^{-\frac{\gamma}{M/\rho + y}}$$  \hspace{1cm} (3)

$$f_{\text{SINR}}(\gamma) = e^{-\frac{\gamma}{M/\rho + y}} \left( \frac{M}{\rho} (1 + \gamma) + M - 1 \right)$$

The scheduling process is organized in a slot-by-slot basis following a max-SINR (greedy) rule. That is, for beam $i$, the scheduler selects the active user $k_i^*$ satisfying:

$$k_i^* = \arg \max_{k = 1:K} \{ \text{SINR}_{k,i} \}$$

where it is assumed that a different user is selected for each beam$^2$. Since all users experience i.i.d Rayleigh fading, the CDF of the post-scheduling SINR, $F_{\text{SINR}}^\gamma (\gamma)$, i.e. the SINR experienced by the scheduled user can be readily expressed in terms of Eq. (3) as:

$$F_{\text{SINR}}^\gamma (\gamma) = (F_{\text{SINR}}(\gamma))^{K} = \left( 1 - e^{-\frac{\gamma}{M/\rho + y}} \right)^K$$

Finally, by simply differentiating the above expression the pdf expression results:

$$f_{\text{SINR}}^\gamma (\gamma) = K \cdot e^{-\frac{\gamma}{M/\rho + y}} \left( \frac{M}{\rho} (1 + \gamma) + M - 1 \right) \times \left( 1 - e^{-\frac{\gamma}{M/\rho + y}} \right)^{K-1}$$  \hspace{1cm} (4)

These expressions will be used in subsequent sections for the computation of the aggregated throughput.

IV. AGGREGATED THROUGHPUT WITH ANALOG FEEDBACK

For practical systems scenarios with a limited number of AMC modes and realistic coding methods, link-layer throughput provides a closer idea on the actual data rates than capacity and sum-rate metrics. Therefore, in the sequel we derive a closed-form expression for the aggregated (system) throughput. For a given modulation scheme, indexed by variable $m$, the aggregated throughput can be expressed as:

$$\eta \approx \mathbb{E} \left\{ \sum_{i \in M} b_m \left( 1 - PER_m \left( \max_{1 \leq k \leq K} \text{SINR}_{k,i} \right) \right) \right\}$$

$$= M \mathbb{E} \left\{ b_m \left( 1 - \text{SER}_m \left( \max_{1 \leq k \leq K} \text{SINR}_{k,i} \right) \right)^L \right\}$$

$$= M b_m \int_0^\infty (1 - \text{SER}_m(\gamma))^L f_{\text{SINR}}^\gamma (\gamma) d\gamma$$  \hspace{1cm} (5)

where $L$ stands for the number of symbols in the burst, $b_m$ is the number of bits per symbol and SER$_m$ denotes Symbol Error Rate$^2$. As shown in [8] the SER for M-QAM modulation schemes (and also for BPSK) can be approximated by:

$$\text{SER}_m(\gamma) \approx b_m 0.2 e^{-1.6 \frac{\gamma}{\gamma_m}} = \alpha_m e^{-\beta_m \gamma}$$  \hspace{1cm} (6)

where $\alpha_m$ and $\beta_m$ are constellation-dependent parameters. Note that, by using such SER expressions we implicitly assume not only the noise component but also the overall inter-user interference to be Gaussian-distributed. Indeed, symbol constellations do not fulfill this condition but in cases where the number of transmit beams ($M$) is high we can invoke the central limit theorem [9]. Besides, the proposed scheduler is aimed at finding the MSs which maximize the resulting SINR or, equivalently, minimize inter-user interference. Therefore, the relative weight of the interference term in Eq.2 is expected to be low, in particular when the number of users to choose from is high (or, of course in the low-SNR regime). Figure 1.a illustrates the validity of the Gaussian approximation: even in the $M = 2$-beam case, the approximation is quite accurate for SNRs below 10 dB.

Going back to the derivation of the throughput expressions, the proposed adaptive modulation mechanism selects for each beam $i$ the constellation size satisfying:

$$m_i = \arg \max_{m \in A_m} b_m (1 - \alpha_m e^{-\beta_m \gamma_{k^*,i}})^L$$

where $\gamma_{k^*,i}$ stands for the SINR corresponding to the scheduled user on beam $i$ ($\gamma_{k^*,i} = \max_{1 \leq k \leq K} \text{SINR}_{k,i}$). From the above expression, it is straightforward to obtain the corresponding AMC thresholds, $\gamma_{k,m}$. As an example Fig. 1.b shows the AMC thresholds for a system with a number of modulation schemes given by the ordered set $M = \{\text{BPSK, QPSK, 16-QAM}\}$.$^3$ Consequently,

$^2$For mathematical tractability, we will restrict ourselves to uncoded transmissions.

$^3$In an opportunistic beamforming context, the potentially low SINRs seldom support constellation sizes larger than 16-QAM.
In this section, we are interested in adapting the aggregated throughput expression to the (realistic) case where the SINRs conveyed over the feedback channel are quantized versions of the analog ones. We start by defining $Q = \{q_1, q_2, \ldots, q_{2^{L_q}}\}$ as the set of quantization levels. After an arbitrary user $k$ identifies the beam $i^*$ with the highest SINR, $\gamma_{k,i^*} = \max_{1 \leq i \leq M} \text{SINR}_{k,i}$, it is quantized according to the following rule:

$$Q(\gamma_{k,i^*}) = q_j, \quad \text{if } \gamma_{q_j} \leq \gamma_{k,i^*} < \gamma_{q_{j+1}}$$

where $q_{j+1} - q_j = 1$. The different SINR thresholds associated with the quantization levels. Next, an $L_q = L_q + L_d$-bit message is sent over the feedback channel, with $L_d = \log_2(M)$ bits devoted to encode the selected beam $i^*$.

Now, we define $A_{k,i}$ as the event that user $k$ is selected for transmission on beam $i$ by the scheduler. Borrowing some results from [3], we know that the probability of the event $A_{k,i}$ conditioned on the fact that $\gamma_{k,i}$ belongs to the quantization level $j$ can be expressed as

$$\text{Prob}(A_{k,i} | \gamma_{k,i} \in q_j) = \frac{\text{F}_{\text{SINR}}(\gamma_{q_j}+1)}{\text{F}_{\text{SINR}}(\gamma_{q_j} + 1) - \text{F}_{\text{SINR}}(\gamma_{q_j})} K$$

Therefore the throughput share corresponding to user $k$ on beam $i$ turns out to be:

$$\eta_{k,i} = \sum_{j=1}^{\text{card}(Q)} \text{Prob}(A_{k,i} | \gamma_{k,i} \in q_j) b_{m_j} \times \int_{\gamma_{q_j}}^{\gamma_{q_{j+1}}} (1 - \alpha_m e^{-\beta_m \gamma}) L \text{F}_{\text{SINR}}(\gamma) d\gamma$$

where for each quantization level $j$, we assume the modulation scheme is selected according to the quantized value of $\gamma_{k,i}$, that is:

$$m_j = m \iff \gamma_{q_j} \leq \gamma_{k,i} < \gamma_{q_{j+1}}$$

In a homogeneous scenario, $\text{Prob}(A_{k,i}|\gamma_{k,i} \in q_j)$ does not depend on $k$ or $i$ and, hence, we can rewrite the average throughput expression as:

$$\eta = \sum_{k=1}^{K} \sum_{i=1}^{M} \eta_{k,i} = B K \eta_{k,i} = M \sum_{j=1}^{\text{card}(Q)} b_{m_j} \sum_{l=0}^{L_{q_j}} \left( \frac{L}{l} \right) \times \left( -\alpha_m \right)^l \frac{\text{F}_{\text{SINR}}(\gamma_{q_{j+1}})}{\text{F}_{\text{SINR}}(\gamma_{q_{j+1}}) - \text{F}_{\text{SINR}}(\gamma_{q_j})} K \times \int_{\gamma_{q_j}}^{\gamma_{q_{j+1}}} \frac{M}{\rho} (1 + \gamma) + M - 1 \times e^{-\beta_m \gamma} \left( 1 + \gamma \right)^{M-1} d\gamma$$

The integral term in the above expression resembles that of Eq.9. Therefore, one can follow again the procedure detailed in the Appendix to finally obtain the aggregated throughput with quantized feedback:

\[\text{This quantization rule results into a conservative, but reliable, assignment of AMC modes at the transmitter.}\]
The 17th Annual IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC’06)

So far, nothing has been said concerning the quantization law (either uniform, non-uniform, etc.). We can shed some light into that issue by bearing in mind that, according to the adopted max-SINR scheduling rule, only the highest SNIRs are relevant in terms of multi-user diversity exploitation gains [10] and link throughput. Therefore, we should avoid uniform quantization laws (possibly in-between some limiting values) and, instead, go for a non-uniform one with smaller quantization intervals in the high-SINR region. By doing so, we minimize the clipping rate, that is, the probability that the quantized metric exceeds the highest quantization threshold), and its negative effects on system performance (further details on clipping rate effects can be found e.g. in [11]). In summary and as depicted in Fig. 2, the quantization thresholds could be given by the inverse of either the post-scheduling CDF function or the pre-scheduling, i.e. individual, CDFs. Intuitively, a quantization law tailored to the post-scheduling SINR should give better results since this is directly related with the scheduling rule. Computer simulation results in the next section confirm this extent. In summary, the SINR thresholds related to the different quantization levels are selected as follows:

\[ \gamma_{q_j} = F_{\text{SINR}}^{-1} \left( \frac{j - 1}{2L_q} \right) \text{ for } j = 1..\text{card}(Q) \]

with \( \gamma_{q_1} = 0 \) and \( \gamma_{\text{card}(Q)+1} = \infty \). As a final remark, notice this is a cross-layer quantization law since quantization levels in the physical layer depend on the number of admitted users which, ultimately, is decided by access control mechanisms in the link layer.

VI. QUANTIZATION LAW

Consider a system with a number of active users in the range \( K = 10..100 \) transmitting data packets with \( L = 10 \) symbols in each and \( M = 3 \) active beams.

Several conclusions can be drawn from Fig. 3. To start with, one can clearly appreciate the accuracy of the (approximate) closed-form expressions of the aggregated throughput derived in the previous sections. Apart from that, it becomes apparent that the post-scheduling based criterion significantly outperforms its pre-scheduling counterpart for the whole range of users. For an increasing number of active user, the gap between both curves becomes wider. This is due to an increased clipping rate (the SINR of the scheduler user is potentially higher when \( K \) becomes larger) which penalizes much more the quantization law based on pre-scheduling statistics. Last, when adopting a quantization law based on the post-scheduling CDF we can clearly see that most of the MUD can be efficiently capture with as few as \( L_q = 4 \) or \( L_q = 2 \) quantization bits.

The curves depicted in Fig. 4 provide some more insights on the impact of quantization on system performance. First, the higher the SNR the larger the impact of quantization since, in this case, the system becomes interference-limited and SINR fluctuations are larger (i.e. higher clipping rate). As commented above, this is true in particular for scenarios with a high number of users. However, most of the MUD can be effectively captured for a very low number
of bits. More precisely, the proposed quantization law attains 81.68% and 91.30% of the analog throughput by just using 2 or 3 quantization bits, respectively.

In summary, we have shown that opportunistic beamforming with multiple beams can be an appropriate strategy in systems with considerably restrictions in the feedback channel. In several situations, most of the multi-user gains can be extracted with only 2 quantization bits.

VIII. APPENDIX

In order to derive a closed-from expression of the system throughput, one should solve the following expression:

\[
\eta = M \sum_{m=1}^{\text{card}(M)} b_m \int_{\gamma=\gamma_{th,m}}^{\gamma_{th,m+1}} \left(1 - \alpha_m e^{-\beta_m \gamma}\right)^L f_{\text{SNR}}(\gamma) d\gamma
\]

By plugging (4) into (8) and using the binomial expansion, the following expression results:

\[
\begin{align*}
\eta = & M \sum_{m=1}^{\text{card}(M)} b_m K \sum_{l=0}^{L} \binom{L}{l} (-\alpha_m)^l \sum_{k=0}^{K-1} \binom{K-1}{k} (-1)^k \\
& \times \int_{\gamma=\gamma_{th,m}}^{\gamma_{th,m+1}} \frac{M}{\rho} \left(1 + \gamma\right)^{M-1} e^{-\gamma \left(\beta_m + \frac{M}{\rho} (k+1)\right)} (1+\gamma)^{k(M-1)+M} d\gamma
\end{align*}
\]

The integral in the above equation can be re-written as:

\[
\begin{align*}
& \int_{\gamma=\gamma_{th,m}}^{\gamma_{th,m+1}} \frac{M}{\rho} \left(1 + \gamma\right)^{M-1} e^{-\gamma \left(\beta_m + \frac{M}{\rho} (k+1)\right)} (1+\gamma)^{k(M-1)+M} d\gamma \\
& = \frac{M}{\rho} \int_{\gamma=\gamma_{th,m}}^{\gamma_{th,m+1}} e^{-\gamma \left(\beta_m + \frac{M}{\rho} (k+1)\right)} (1+\gamma)^{k(M-1)+M-1} d\gamma \\
& \quad + (M-1) \int_{\gamma=\gamma_{th,m}}^{\gamma_{th,m+1}} e^{-\gamma \left(\beta_m + \frac{M}{\rho} (k+1)\right)} (1+\gamma)^{k(M-1)+M} d\gamma
\end{align*}
\]

Since the two integrals in Eq. (9) are of the type:

\[
\int_{t=u}^{\infty} e^{-at} (1+t)^n dt = \int_{t=u}^{\infty} e^{-at} (1+t)^{n-1} dt - \int_{t=u}^{\infty} e^{-at} (1+t)^n dt
\]

the problem is reduced to solve the following integral:

\[
\int_{t=u}^{\infty} e^{-at} (1+t)^{m} dt
\]

By using the change of variables \( x = (1+t)a \), the integral in the above equation can be easily solved by resorting to the identity [12, Eq. 8.350.2]:

\[
\int_{t=u}^{\infty} e^{-at} (1+t)^{m} dt = e^a a^{n-1} \Gamma(1-n, (1+u)a)
\]

where \( \Gamma(\alpha, x) \) stands for the complementary incomplete gamma function \( \Gamma(\alpha, x) = \int_{x}^{\infty} e^{-t \left(\alpha-1\right)} dt \) [12, Eq. 8.350.2]. Finally, by using Eq. (11) in Eq. (10) one can obtain the integrals in Eq. (9) and verify that Eq. (7) holds.

REFERENCES


