CCIT Report #661 November 2007

A Broadcast Approach for a Single User Slowly Fading MIMO Channel†

D R A F T

Shlomo Shamai (Shitz), Avi Steiner

Department of Electrical Engineering, Technion—Israel Institute of Technology

Haifa 32000, Israel

E-mail: sshlomo@ee.technion.ac.il, savi@tiger.technion.ac.il.

May 12, 2003

Abstract

A broadcast transmission strategy for the slowly fading Gaussian multiple input multiple output (MIMO) channel is introduced. This broadcast strategy is an extension of the single input single output (SISO) broadcast approach. Perfect channel state information (CSI) is assumed known at the receiver end only. This strategy facilitates to adapt the reliably decoded rate to the actual channel state without having any feedback link to the transmitter. Transmission of layered coded information is motivated by the theory of majorization. We derive the basic equations characterizing achievable rates of the strategy. Several ad-hoc approximations to the achievable region are considered and their performance is compared with the SISO setting and the ergodic capacity. It has been demonstrated that a single layer outage approach is reasonably efficient in the MIMO setting in terms of the average reliably decoded rate. A multiple-access (MAC) broadcast approach is also applied for the MIMO case, and demonstrated to be relatively efficient.

I. Introduction

Fading channels are often used for models of wireless communications, see [1] and references therein. Slowly fading channels serve as a common model of relatively slowly varying channel

†This research has been supported by the ISF-Israel Science Foundation administered by the Israel Academy of Sciences and Humanities

characteristics. This is true for a mobile front-end moving slowly relative to the transmission rate. In such conditions a receiver can estimate the channel fading coefficients with high accuracy, and this motivates the consideration of channel state information (CSI) available perfectly at the receiver. With no delay constraints imposed and very long transmission blocks, in terms of the fading dynamics, the ergodic nature of the fading process is revealed. The achievable rate is the well studied ergodic capacity [1].

We assume however that stringent delay constraints imply that the transmission block of length L though still large (as to give rise to the notion of reliable communication [2]) is much shorter than the dynamics of the slow fading process. This scenario is approximated by assuming that $h_i = h \; \forall i = 1, 2, ..., L$. In this case the notion of capacity versus outage was introduced and discussed [2] and [1, see references therein]. In an outage strategy the transmission rate is fixed, and the information is reliably detected when the instantaneous channel realization allows, otherwise nothing gets decoded, and this is called an outage event [1]. The term outage capacity refers to the maximal achievable average rate, and can be also cast as the capacity of the appropriately defined compound channel [1].

In this paper we introduce another approach, termed the broadcast strategy. This strategy facilitates reliable transmission rates adapted to the actual channel conditions, without providing any feedback from the receiver to the transmitter. Cover in his original paper [3] suggests the use of a broadcast approach for the compound channel. Since the slowly fading channel may be viewed as a compound channel with the channel realization as the parameter of the compound channel, it is essentially what the broadcast strategy is. This strategy is useful in a variety of applications and it matches the successive refinement source coding approach [4] and later work [5]. That is, the more information rate is provided, the less average distortion is evident in the reconstructed source. An example for successive refinement of source coding [4] is image compression in which a gross description exists at first, and only later follow successive refinements of the description that further refine the image quality. An application example is progressive JPEG encoding, where additional coded layers serve to refine the image quality. In the broadcast approach, the transmitter sends layered coded information, and in view of the receiver as a continuum of ordered users, the maximum number of layers successively decoded is dictated by the fading channel realization. Thus, the channel realization implies on the received quality of the data. The broadcast approach has a practical appeal in a voice communication cellular environment, where layered voice coding is possible. Service quality then, depends on the channel realization.

The problem of layered coding suggests unequal error protection on the transmitted data, which was studied in [6, see references therein]. A related subject is the priority encoding transmission (PET). Boucheron *et. al.* [7] show that sending hierarchically organized messages over lossy packet-based networks [8], can be analyzed using the broadcast erasure channel with degraded message set, using the information spectrum approach.

This approach hinges on the broadcast channel, which was first explored by Cover [3], [9]. In a broadcast channel a single transmission is directed to a number of receivers, each enjoying possibly different channel conditions, reflected in their received signal-to-noise ratio (SNR). The Gaussian broadcast channel with a single transmit antenna coincides with the classical physically degraded Gaussian broadcast channel, whose capacity region is well known, see [9] for the deterministic case and [10], [11] in the composite or ergodic cases. However, for multiple transmit antennas the Gaussian broadcast channel is in general a non-degraded broadcast channel, for which the capacity region is not fully known [12], [13], [14], [15], [16] and cannot be reduced to an equivalent set of parallel degraded broadcast channels, as studied in [17], [10], [11].

Broadcasting for a single user implies on broadcasting of common information. Information theoretic results and challenges for broadcasting a common source are discussed in [18], and in light of endless information data transmission termed streaming in [19]. The very straightforward interpretation of single user broadcasting is the hierarchical broadcasting using multi-level coding (MLC) [20], [21], [22]. Schill et. al. [21] demonstrates the spectral efficiency of MLC with hierarchical demodulation in an additive white Gaussian noise (AWGN) channel and a fading channel. In work of Sajadieh *et. al.* [23] the fading interleaved channel is examined, with one bit of side information about the fading process. The broadcast approach is adapted, so that different rates can be decoded for channels taking these two distinct states (determined by whether the SNR is above or below a threshold value). Since the channel is memoryless the average rate $I(Y, \hat{S}, ; X)$ (where $\hat{S}$ is the partial state information) is achievable which is not the case with the broadcast approach, which seems to be unfit here, where channel state is assumed to be iid.

Liu *et. al.* [24] considers a super-position coding scheme to achieve higher transmission rate in the slowly fading channel. This work adopts the broadcast approach for the SISO channel with a finite number of receivers. The number of receivers is the number of code layers, e.g. in a single receiver case the maximal achievable rate is the outage capacity. It is evident from [24], that for the SISO channel, a few levels of code layering closely approximates the optimal strategy employing transmission of infinite code layers.

In the sequel, we consider the single transmit antenna case, where the realization of the fading parameter can be interpreted as an index (possibly continuous), which designates the SNR at the receiver of interest. This original approach first presented in [25], is elaborated here and we provide the derivation of the expressions related to the broadcast approach concept, optimal power distribution, and the associated maximal achievable average rate. The maximal achievable average rate is demonstrated for the SISO and single input multiple output (SIMO) Rayleigh fading channel, and is compared to the ergodic capacity reference. We then consider the multiple-input multiple-output (MIMO) channel, for which we suggest a sub-optimal ranking at the receiver. The ranking of channel matrices (as opposed to a vector in a SIMO case) is achieved via supermajorization ranking of the singular values of $HH^\dagger$, which imply on channel conditioned capacity ranking. We state the optimization variational problem for deriving the optimal power distribution for the MIMO broadcast strategy. The optimal solution seems not to lend itself to close form expressions, thus a sub-optimal solution using majorization is considered and demonstrated for the Rayleigh fading channel. This approach is called the 1-D approximation, and is developed for the 2x2 (two transmit and two receive antennas) channel. It suggests breaking the mutual dependency of the optimal power distribution $\rho(a, b)$ by requiring $\rho(a, b) = \rho(a)\rho(b)$. Such a representation bares two independent solutions, solved from the optimal SISO broadcast strategy. Another sub-optimal approach is based on finite level of code layering, and was suggested in [24] for the SISO scheme. Accordingly, we examine the single layer (outage) coding with and without employing majorization ranking at the reciver. A two layers coded scheme for the 2x2 channel is also studied and compared to the outage approach. We then approach the MIMO channel as a multiple-access channel (MAC). In a MAC approach for the MIMO channel, instead of performing joint encoding for all transmit antennas, each antenna has an independent encoder, thus the receiver then views a multiple-access channel. When each encoder performs layered coding, we essentially get a MAC-Broadcast strategy. This approach was first presented in [26] for the multiple-access channel, employing the broadcast approach at the receiver. Its advantage is that each transmitter views an equivalent degraded broadcast channel, and the results of the SISO broadcast strategy may be directly used. We pose an iterative algorithm for optimizing the transmit power distribution. The MAC approach is also studied for a single code layer at each transmitter. Its performance is evaluated with both successive decoding as used in the broadcast strategy, and with optimal joint decoding.

The structure of the paper is as follows. In section 2. the SISO broadcast strategy is presented, and the optimal achievable rate is outlined and demonstrated for the Rayleigh fading channel. A general MIMO broadcast strategy is presented in section 3., and a sub-optimal solution to the optimal transmit power distribution is suggested in section 4.. Finite level code layering is considered in section 5.. Section 6. presents the MAC-broadcast approach. A summary of the MIMO numerical results is then given in section 7., followed by concluding remarks.

II. SISO Broadcast Strategy

We start by describing the broadcast strategy conceptually for a single user SISO channel, first presented in [25]. A Gaussian SISO channel is also known to be physically degraded [9]. This is used in developing the broadcast strategy, which assumes an infinite number of ordered receivers. We give here some preliminaries and definitions, and then describe the broadcast strategy for the SISO channel.

A. Channel Model

Consider the following SISO channel,

$$y_i = hx_i + n_i \ , \tag{1}$$

where $\{y_i\}$ are samples of the received symbols, $\{x_i\}$ are the transmitted complex symbols. $\{n_i\}$ are the additive noise samples, which are complex Gaussian i.i.d $\mathcal{CN}(0,1)$, and h is the fading coefficient. For each realization of h there is an achievable rate. We are interested in the average achievable rate for various independent transmission blocks. Thus we present the results in terms of average performance, averaged over the distribution of h.

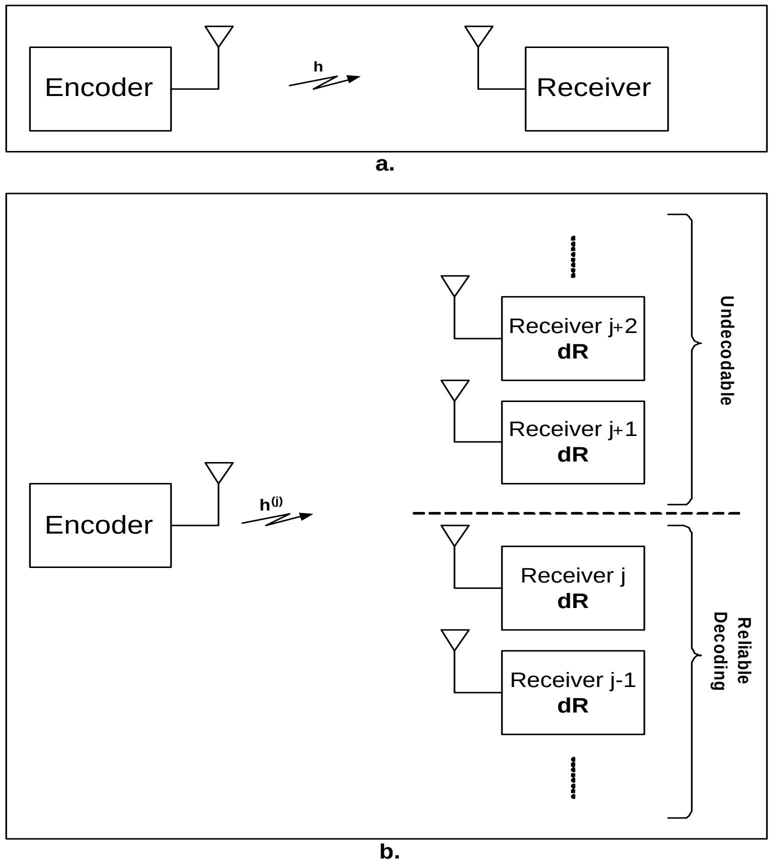

Fig. 1. a. A SISO channel with a fading parameter h. b. The equivalent SISO broadcast channel model. For a channel realization $h^{(j)}$, only receivers indexed up to j can decode their fractional rate dR.

Information theoretic considerations for this simple model were discussed in [2, and references therein], as a special case of the multi-path setting. With the h value informed to the transmitter, and with a short term power constrained (excluding power optimization in different blocks), the expected reliable rate over many block realization is given by

$$C_{erg} = E_s \log(1 + sP) \tag{2}$$

where $s \triangleq |h_i|^2$ is the random fading power, the normalized signal-to-noise ratio (SNR), following the channel model definition (1), is $P = E|x|^2$. E stands for the expectation operator (subscripts if added denote the random variables with respect to which the expectation is taken).

B. The Broadcast Strategy

The SISO channel defined in (1) is also illustrated in Figure 1a. An equivalent broadcast channel is demonstrated in Figure 1b. This figure also illustrates the broadcast approach, where the transmitter sends infinite number of layers of coded information. The receiver is equivalent to a continuum of ordered users, each decoding a code layer if channel realization allows. In general, the number of code layers(and respectively receivers) depends on the cardinality of the fading power random variable (RV). Predetermined ordering is achieved due to the degraded nature of the Gaussian SISO channel [9]. Each of the users has to decode a fractional rate, denoted dR in Figure 1b. The fractional rates dR of the different users are not equal, but depend on their receiver index. For some fading realization $h^{(j)}$, only the continuum of receivers up-to receiver j can decode their fractional rates dR. The first receiver decodes only its own dR, the second decodes initially the interference dR (information intended to the first user) and then decodes its own dR. Finally, receiver j decodes all fractional interferences up-to layer $j-1$, and then decodes its information layer dR. Hence the total achievable rate for a realization $h^{(j)}$ is the integral of dR over all receivers up-to j. This model is the general case of code layering. The broadcast approach in [25] with finite number of code layers, termed also superposition coding is presented in [24]. In finite level code layering, only a finite set of ordered receivers is required. Obviously, the approach has lower decoding complexity, however it is a broadcast sub-optimal approach.

Assume now that the fading power RV S is continuous. Then for some channel realization $h^{(j)}$ of Figure 1b, with a fading power $s^{(j)}$, the designated reliably conveyed information rate is denoted by $R(s^{(j)})$. We now drop the superscript j, and refer to s as the realization of the fading power RV S. As illustrated, the transmitter views the fading channel as a degraded Gaussian broadcast channel [9] with a continuum of receivers each experiencing a different signal-to-noise ratio specified by $s \cdot P$. The total transmitted power P is also the SNR as the fading and additive noise are normalized according to (1). The value s is therefore interpreted as a continuous index. The incremental differential rate is then given by

$$dR(s) = \log\left(1 + \frac{s\rho(s)ds}{1 + sI(s)}\right) = \frac{s\rho(s)ds}{1 + sI(s)} \tag{3}$$

where $\rho(s)ds$ is the transmit power of a layer parameterized by s, intended for receiver s, which also designates the transmit power distribution. Information streams intended for receivers indexed by $u > s$ are undetectable and play a role of additional interfering noise, denoted by $I(s)$. The interference for a fading power s is

$$I(s) = \int_{s}^{\infty} \rho(u)du, \tag{4}$$

which is also a monotonically decreasing function of s. The total transmitted power is the overall collected power assigned to all layers,

$$P = \int_{0}^{\infty} \rho(u)du = I(0). \tag{5}$$

As mentioned earlier, the total achievable rate for a fading realization s is an integration of the fractional rates over all receivers with successful layer decoding capability,

$$R(s) = \int_0^s \frac{u\rho(u)du}{1 + uI(u)}. \tag{6}$$

Average rate is achieved with sufficiently many transmission blocks, each viewing an independent fading realization. Therefore, the total average rate R_{bs} over all fading realizations is

$$R_{bs} = \int_{0}^{\infty} du\ f(u)R(u) = \int_0^\infty du(1 - F(u))\frac{u\rho(u)}{1 + uI(u)} \tag{7}$$

where $f(u)$ is the probability distribution function (PDF) of the fading power, and $F(u) = \int_0^u daf(a)$ is the corresponding cumulative distribution function (CDF).

For performance comparison, it is of interest to examine the average rate versus outage probabilities. This follows directly by replacing $F(u)$ in (7) with the conditional probability distribution function conditioned on $s \geq s_{th}$, given by

$$F_{s_{th}}(u) = Pr(s < u | s \geq s_{th}) = \frac{F(u) - F(s_{th})}{1 - F(s_{th})}, \quad s_{th} \leq u \leq \infty. \tag{8}$$

The average rate conditioned on $s \geq s_{th}$ is denoted by $R_{av}(s_{th})$ and is given by

$$R_{bs,o}(s_{th}) = (1 - F(s_{th}))^{-1} \int_{s_{th}}^{\infty} (1 - F(u)) \frac{u\rho(u)du}{1 + uI(u)}, \tag{9}$$

where evidently $R_{bs} = R_{bs,o}(0)$.

The notion of $R_{av}(s_{th})$ plays the role of the average transmission rate associated with the outage probability $F(s_{th})$. As opposed to the standard approach, here with probability $1 - F(s_{th})$ the reliably conveyed rate is not fixed but depends on actual realization of $s \geq s_{th}$. The standard capacity versus outage relation is found as a special case with $\rho(s) \to P\delta(s - s_{th})$, where $\delta(s)$ is the dirac delta function.

Optimization of $R_{bs,o}(s_{th})$ with respect to the power distribution $\rho(s)$ (or equivalently with respect to $I(u)$, $u \geq 0$) under the power constraint P (5) is of interest and can in certain cases be found by solving the associated constrained Eüler equation [27]. We turn back to the expression in (7), corresponding to $s_{th} = 0$, and explicitly write the optimization problem posed

$$R_{bs,max} \quad = \quad \max_{I(u)} \int_0^\infty du(1 - F(u)) \frac{u\rho(u)}{1 + uI(u)} \tag{10}$$

where we maximize R_{bs} (7) over the residual interference function $I(u)$. For an extremum function $I(x)$, the variation of the functional (10) is zero [27], corresponding to a proper Eüler equation, which yields the extremal solution for $I(x)$. Let us first present the functional of (10) subject to maximization

$$S(x, I(x), I'(x)) = (1 - F(x)) \frac{-xI'(x)}{1 + xI(x)}. \tag{11}$$

The necessary condition for a maximum of the integral of $S(x, I(x), I'(x))$ over x is a zero variation of the functional [27]. Correspondingly, the Eüler Equation is given by

$$S_I - \frac{d}{dx} S_{I'} = 0 \tag{12}$$

where

$$
\begin{array}{l}
S_I = (1 - F(x)) \frac{x^2 I'(x)}{(1 + x I(x))^2} \\
S_{I'} = (1 - F(x)) \frac{-x}{1 + x I(x)} \\
\frac{d}{dx} S_{I'} = \frac{x f(x)}{1 + x I(x)} + (1 - F(x)) \frac{x^2 I'(x) - 1}{(1 + x I(x))^2}
\end{array}
\tag{13}
$$

which finally simplifies from a differential equation (12) to a linear equation by $I(x)$, and has the following closed form solution

$$
I(x) = \begin{cases} \frac{1 - F(x) - x \cdot f(x)}{x^2 f(x)} & x_0 \leq x \leq x_1 \\ 0 & else \end{cases} ,
\tag{14}
$$

where x_0 is determined by $I(x_0) = P$, and x_1 by $I(x_1) = 0$. Some of the results are demonstrated in the following, for an exponentially distributed fading power.

All above derivations stand also for the SIMO channel. As long as the transmitter has a single antenna, the Gaussian channel is a degraded channel regardless of the number of receive antennas. Therefore the described broadcast strategy withholds. The number of receive antennas is related with the distribution $f(u)$ of the equivalent fading power coefficient. The contribution of increasing the number of receive antennas to the achievable average rate is demonstrated in section 2.4..

C. The Rayleigh Flat Fading Channel

We demonstrate here some results for the SISO Rayleigh flat fading channel. That is the fading power S is exponentially distributed with

$$
f(u) = e^{-u}, \quad F(u) = 1 - e^{-u}, \quad u \geq 0.
\tag{15}
$$

The optimal transmitter power distribution which maximizes R_{bs} in (10) is given by substitution of $f(u)$ and $F(u)$ from (15) into (14), and by derivation with respect to the fading power s,

$$
\rho(s) = -\frac{d}{ds} I(s) = \begin{cases} \frac{2}{s^3} - \frac{1}{s^2} & , \; s_0 \leq s \leq s_1 \\ 0 & , \; else \end{cases}
\tag{16}
$$

where s_0,is determined by $I(s_0) = P$, which simplifies here to

$$s_0 = \frac{2}{1+\sqrt{1+4P}} \quad ,$$

and s_1 by $I(s_1) = 0$, thus $s_1 = 1$. The corresponding rate $R(s)$ using (6) is

$$R(s) = \begin{cases} 0 & , \ 0 \leq s \leq s_0 \\ 2\ln(\frac{s}{s_0}) - (s - s_0) & , \ s_0 \leq s \leq 1 \\ -2\ln(s_0) - (1 - s_0) & , \ s \geq 1 \end{cases} \tag{17}$$

and the associated total average rate following (7) is

$$R_{bs} = 2E_i(s_0) - 2E_i(1) - (e^{-s_0} - e^{-1}), \tag{18}$$

where

$$E_i(x) = \int_x^\infty \frac{e^{-t}}{t} dt, \quad x \geq 0 \tag{19}$$

is the exponential integral function. The limiting behavior of R_{bs} is found to be

$$R_{bs} \approx \begin{cases} \ln \frac{P}{9.256} & , \ P \to \infty \\ \frac{1}{e}P & , \ P \to 0 \end{cases} \tag{20}$$

The ergodic capacity in this case is given by [2],

$$C_{erg} = e^{1/P} \cdot E_i(\frac{1}{P}) \approx \begin{cases} \ln \frac{P}{1.78} & , \ P \to \infty \\ P & , \ P \to 0 \end{cases} \tag{21}$$

where $E_i(x)$ is the exponential integral function defined in (19). The average achievable rate of the standard outage approach, depends on the outage probability $P_{out} = Pr\{s \leq s_{th}\} = 1 - e^{-s_{th}}$. Thus the achievable outage rate is given by

$$R_o(s_{th}) = e^{-s_{th}} \log(1 + s_{th}P). \tag{22}$$

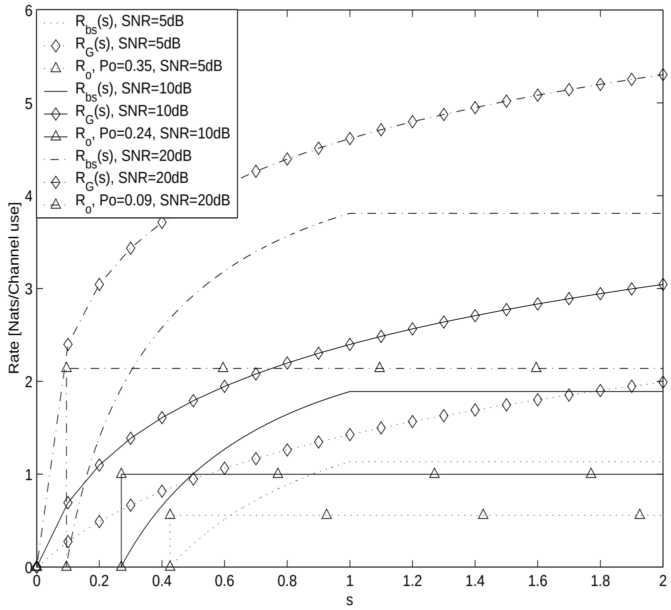

Fig. 2. SISO broadcast achievable rate $R_{bs}(s)$, outage rate R_o for an outage probability associated with s_0 (16), and the upper bound $R_G(s)$ vs. the fading realization s. Different values of SNR demonstrated.

The outage capacity is the product of maximizing the achievable outage average rate (22) with respect to the outage probability (or the fading power threshold s_{th}). This yields an outage capacity

$$R_{o,max} = e^{-s_{th,opt}} \log(1 + s_{th,opt}P). \tag{23}$$

where $s_{th,opt}$ solves the equation

$$\log(1 + s_{th,opt}P) = \frac{P}{1 + s_{th,opt}P}. \tag{24}$$

D. Numerical Results

We present here results of achievable rates for the single user SISO and SIMO Rayleigh flat fading channels under the broadcast approach. Figure 2 demonstrates the broadcast achievable rate $R_{bs}(s)$ (17) versus channel realizations s ($0 \leq s \leq 2$). It is compared to the evident upper bound

$$R_G(s) = \log(1 + sP), \tag{25}$$

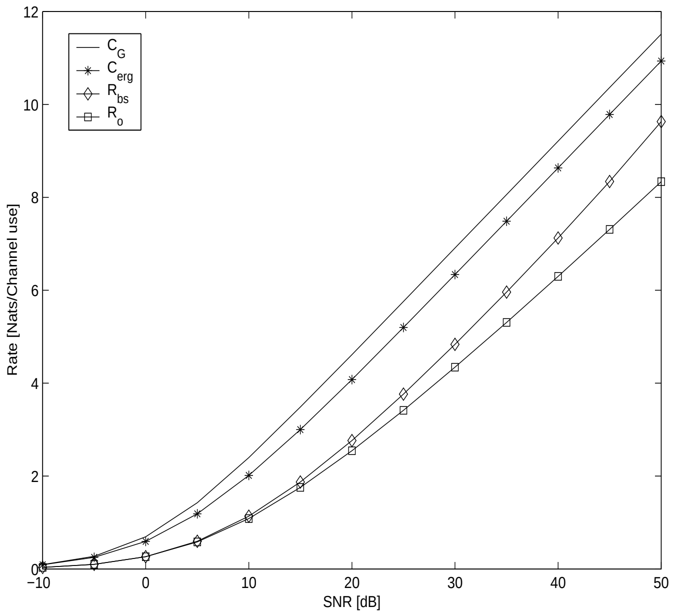

Fig. 3. SISO broadcast achievable average rate R_{bs}, outage capacity R_o, ergodic capacity C_{erg} and Gaussian channel (AWGN) upper bound C_G vs. SNR.

for SNR (P) values of 5dB, 10dB, 20dB. As an example for the capacity outage tradeoff we check the outage rate associated with $s_{th} = s_0$, where s_0 is defined below (16) and stands for the transmission threshold in the broadcast strategy. The outage probability is therefore $P_{out} = 1 - e^{-s_0}$, and the rate related to that is the standard capacity-outage approach $R_o(s_0)$ (22). The associated outage rate is compared with the broadcast achievable rate $R_{bs}(s)$ in Figure 2.

Figure 3 demonstrates the SISO broadcast achievable average rate R_{bs} (18), outage capacity R_o (23), the ergodic capacity C_{erg} (21) upper bound, and the Gaussian capacity $C_G = \log(1 + P)$ as a reference. Clearly, $R_{bs} > R_o$ as the latter is achieved by substituting $\rho(s)$ with $P\delta(s - s_{th,opt})$ in lieu of the optimized $\rho(s)$ in (6). Outage capacity is equivalent to optimized single layer coding rather than the optimized intrinsic continuum of code layers in the broadcast approach. This difference is more pronounced for high SNR. Such a comparison of the single level code-layer, and two level achievable rates is done in [24]. This comparison shows that two level code layering is already very close to the optimum R_{bs}.

The ergodic capacity in the general SIMO case, with N receive antennas, is given by (9)

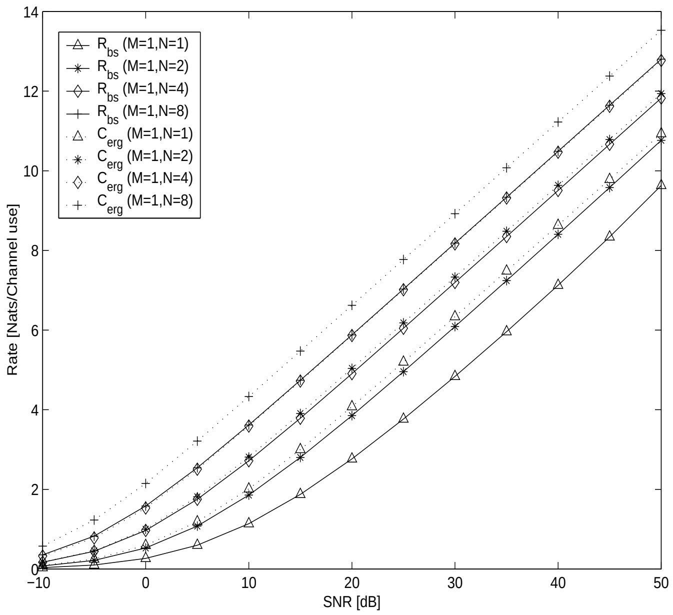

Fig. 4. SIMO broadcast achievable average rate R_{bs} and ergodic capacity C_{erg} upper bound vs. SNR, for various number of receive antennas.

in [28], which reduces to the SISO ergodic capacity for $N = 1$,

$$C_{erg} = \frac{1}{\Gamma(N)} \int_0^\infty dx \log(1 + P \cdot x) x^{N-1} e^{-x}. \tag{26}$$

The channel density is given by [29]

$$f(\lambda) = const(N) \cdot \lambda^{N-1} e^{-\lambda}, \tag{27}$$

where $const(N)$ is a normalization constant.

Figure 4 demonstrates the SIMO broadcast achievable average rates R_{bs} compared to the ergodic capacities C_{erg} upper bound vs. SNR, for various number of receive antennas. There is a noticeable gain in spectral efficiency for all cases $N = 1, 2, 4, 8$. Increasing N reduces the spectral efficiency loss of the broadcast approach with respect to the ergodic capacity. As for $N \to \infty$, the impact of fading is mitigated due to the central limit theorem, both ergodic capacity [28], and the broadcast approach yield the same spectral efficiency, approaching $\log(1 + PN)$ for $N >> 1$.

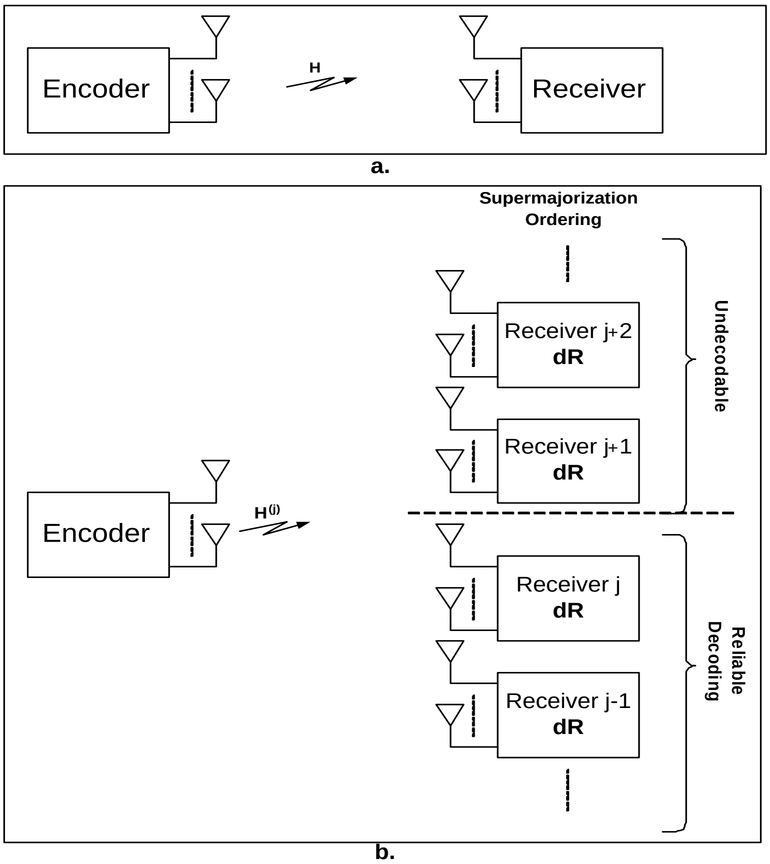

Fig. 5. a. A MIMO channel with a channel propagation matrix H. b. The equivalent MIMO broadcast channel model. A continuum of ordered receivers, with ordering dictated by supermajorization of singular values of channel norm matrix $HH^{\dagger}$. Only receivers indexed up to j decode their fractional rate dR.

III. MIMO Broadcast Strategy

We adopt here the broadcast approach [3], [25], [26] described earlier for the SISO/SIMO channels, where the receiver opts to detect the highest possible rate based on the actual realization of the propagation matrix H not available to the transmitter. In short, the better H is the higher becomes the reliably decoded information rate. Since the MIMO setting in view of infinite layer coding is equivalent to the general broadcast channel, rather than a degraded broadcast channel as in the single input case. In the sequel we demonstrate a broadcast approach, suited for this MIMO scenario. The approach suggests ordering of receivers for the broadcast approach based on supermajorization of singular values of the channel norm matrix, as illustrated in Figure 5b.

A. Channel Model

Consider the following flat fading MIMO channel with M transmit antennas and N receive antennas,

$$\mathbf{y} = H\mathbf{x} + \mathbf{n}\,, \tag{28}$$

where $\mathbf{x}$ is the input $(M \times 1)$ vector, $\mathbf{n}$ is the $(N \times 1)$ noise vector with complex Gaussian i.i.d $\mathcal{CN}(0,1)$ distributed elements. The propagation matrix $(N \times M)$ is designated by H and possesses also complex Gaussian i.i.d $\mathcal{CN}(0,1)$ distributed elements. The received $(N \times 1)$ vector is denoted by $\mathbf{y}$. We adhere to the non-ergodic case, where H is fixed throughout the code word transmission. We assume that the receiver is aware of H while the transmitter is not.

B. Receivers Ordering

B.1 Weak Supermajorization

First we introduce some partial ordering relations based on classical theory of majorization [30]. Let $\boldsymbol{\alpha} = \{\alpha_i\}$, $\boldsymbol{\beta} = \{\beta_i\}$ be two sequences of length K. Let $\{\alpha_{(i)}\}\,,\{\beta_{(i)}\}$ be the increasing ordered version of the sequences,

$$\begin{aligned} &\alpha_{(1)} \leq \alpha_{(2)} \cdots \leq \alpha_{(K)} \\ &\beta_{(1)} \leq \beta_{(2)} \cdots \leq \beta_{(K)}\,. \end{aligned} \tag{29}$$

Let $\boldsymbol{\alpha}$ be weakly supermajorized by $\boldsymbol{\beta}$, $\boldsymbol{\alpha} \prec^w \boldsymbol{\beta}$, that is

$$\sum_{i=1}^{k} \alpha_{(i)} \geq \sum_{i=1}^{k} \beta_{(i)}\,, \quad k = 1\,\ldots, K\,. \tag{30}$$

Then [30], the relation $\boldsymbol{\alpha} \prec^w \boldsymbol{\beta}$ implies that,

$$\sum_{i=1}^{K} \phi(\alpha_i) \leq \sum_{i=1}^{K} \phi(\beta_i)\,, \tag{31}$$

for all continuous decreasing convex functions $\phi(\cdot)$.

Consider now the received signal in (28), where the undetectable code layers are explicitly stated,

$$\mathbf{y} = H(\mathbf{x}_S + \mathbf{x}_I) + \mathbf{n}\,, \tag{32}$$

where $\mathbf{x}_S$ and $\mathbf{x}_I$ are decodable information and residual interference Gaussian vectors respectively, with average norms P_S and P_I correspondingly, where the total transmit power is $P = P_I + P_S$. $\mathbf{n}$ is an iid Gaussian complex vector with unit variance per component. The mutual information

$$\begin{aligned} I(\mathbf{y};\mathbf{x}_S) &= I(\mathbf{y};\mathbf{x}_S,\mathbf{x}_I) - I(\mathbf{y};\mathbf{x}_I|\mathbf{x}_S) \\ &= \log\det\left(I + \frac{P_S + P_I}{M} HH^\dagger\right) - \log\det\left(I + \frac{P_I}{M} HH^\dagger\right) \\ &= \sum_{k=1}^{\mathbf{J}} \log\left(1 + \frac{P_S\lambda_k}{1+P_I\lambda_k}\right) \\ &\triangleq C(\lambda; P_S, P_I)\,, \end{aligned} \tag{33}$$

where $\{\lambda_k\}$, $k = 1, 2 \ldots J \triangleq \min(N, M)$ designate the singular values (or eigenvalues) of the matrix $\frac{1}{M}H^\dagger H$ for $M \leq N$, or $\frac{1}{M}HH^\dagger$ for $N \leq M$, [28], and where the expression was explicitly designated by $C(\boldsymbol{\lambda}; P_S, P_I)$. Now since,

$$\phi(x) = -\log\left(1 + \frac{ax}{1+bx}\right) \tag{34}$$

is a continuous decreasing convex function of $x \geq 0$ for $a, b \geq 0$. Finally, if $\boldsymbol{\lambda} \prec^w \boldsymbol{\delta}$

$$C(\boldsymbol{\lambda}; P_S, P_I) \geq C(\boldsymbol{\delta}; P_S, P_I)\,. \tag{35}$$

C. The Broadcast Approach

We develop the broadcast approach for the MIMO channel, discussing for simplicity the case of $M = N = 2$. The signal $\mathbf{x}$ is composed of a layered double indexed data stream with indices denoted u and v. We refer to layer ordering by columns bottom up, as depicted in Figure 6, where u and v are described as a pair of indices taking integer values within the prescribed region. This is only for demonstration purposes, as indices u and v are continuous

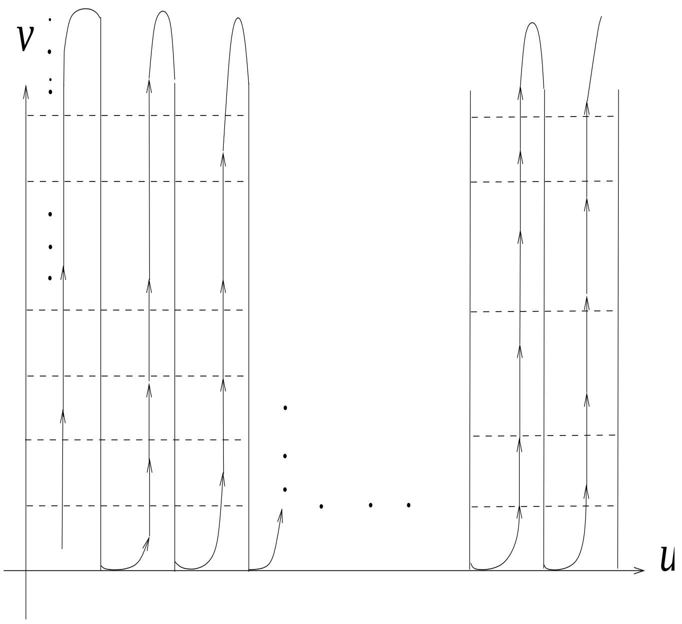

Fig. 6. Ordering of Layered Communications.

singular values of $\frac{1}{2}HH^{\dagger}$. Say u and v are associated with the **minimal** eigenvalue λ_2 and the sum of eigenvalues $\lambda_2 + \lambda_1$, respectively. Evidently, $u \geq 0$, $v \geq 2u$. Say that λ_2, λ_1 take on the set of integer values $\{0, 1, 2, 3, 4\}$, then the layered system is described by (u, v) in the order: $(0, 0)$, $(0, 1)$, $(0, 2)$, $(0, 3)$, $(0, 4)$, $(1, 2)$, $(1, 3)$, $(1, 4)$, $(2, 4)$. The actual ordering of the layers is in fact immaterial, as will be shown decoding is not done successively as in the SISO case [25], but rather according to what is decodable adhering to partial ordering.

We envisage all possible realizations of H and order them by $u = \lambda_2$, $v = \lambda_2 + \lambda_1$ where λ_2 and λ_1 are respectively the minimal and maximal eigenvalues of $\frac{1}{2}HH^{\dagger}$ (a 2×2 matrix in our case). Supermajorization ordering dictates that all streams decodable for realization H will be decodable for realization H' as long as

$$\lambda_2^{'} \geq \lambda_2, \quad \lambda_2^{'} + \lambda_1^{'} > \lambda_2 + \lambda_1 . \tag{36}$$

Thus, we visualize all possible realizations of H as channels referring to different users in a broadcast setting, and we investigate the associated rates of the users which we have ranked as in section 3.2.1., via a degraded ordering. It is evident that the current approach specifies an achievable rate region, but by no means it is claimed to be optimal. In fact we shall point out some inherent limitations of this approach. We shall investigate the current approach in the limit of an infinite array of layered codes, in parallel to the single dimensional $M = N = 1$

broadcast approach.

D. Optimal Power Distribution - Variational Problem Definition

Let $u = \lambda_2$ and $v = \lambda_1$ be the eigenvalues of $\frac{1}{2}HH^\dagger$ for some channel realization such that $v \geq u \geq 0$. Let $\rho(u, v)\, dudv$ be the power associated with the information stream indexed by (u, v), $(v \geq u)$, and featuring the incremental rate $d^2R(u, v)$. Again, for a given u and v, all rates associated with the indices $(a, b)\,,\ a \leq u,\ b \leq v$ can be decoded, as $(\lambda_2,\ \lambda_1)$ is supermajorized by $(\lambda_2 = a,\ \lambda_1 = b)$.

A natural optimization problem, in parallel to that posed and solved for the single dimensional case is to optimize the power density $\rho(u, v)$, or the related interference pattern $I(u, v)$ maximizing the average rate, under the power constraint $I(0, 0) = P$.

Let $I(u, v)$ designate the residual interference at (u, v) (of the undetected streams) $\lambda_2,\ \lambda_1$,

$$I(u, v) = P - \int_0^u da \int_a^v db\, \rho(a, b). \tag{37}$$

The associated incremental rate $d^2R(u, v)$ equation, based on (3) and (33), is then given by

$$\begin{aligned} d^2R(u, v) &= \log\left(1 + \tfrac{u\rho(u,v)\, dudv}{1+uI(u,v)}\right) + \log\left(1 + \tfrac{v\rho(u,v)\, dudv}{1+vI(u,v)}\right) \\ &= \tfrac{u\rho(u,v)\, dudv}{1+uI(u,v)} + \tfrac{v\rho(u,v)\, dudv}{1+vI(u,v)} \end{aligned} \tag{38}$$

The power density is the second order derivative of the residual interference function (37),

$$\rho(u, v) = -\frac{\partial^2}{\partial uv} I(u, v) \triangleq I_{uv} \tag{39}$$

and the incremental rate may be expressed as

$$d^2R(u, v, I, I_{uv}) = -\frac{uI_{uv}(u, v)dudv}{1 + uI(u, v)} - \frac{vI_{uv}(u, v)dudv}{1 + vI(u, v)}. \tag{40}$$

The accumulated reliable rate decoded at (u, v) is

$$R(u, v) = \int_0^u \int_a^v d^2R(a, b). \tag{41}$$

The expected rate is then given by

$$R_{av} = \int_0^\infty \int_0^\infty f(u,v)\, R(u,v)\,, \tag{42}$$

where $f(u,v)$ designates the joint density distribution function of the ordered eigenvalues of $\frac{1}{2}HH^\dagger$, random variables u and v. For Gaussian H with iid components the joint density function of λ_2, λ_1 is given by, [28]

$$f_{\lambda_2,\lambda_1}(u,v) = 16\, e^{-2v-2u}(v-u)^2, v \geq u \geq 0 \tag{43}$$

The optimal expected rate is a product of optimal selection of the power distribution $\rho(u,v)$. Power distribution directly implies the residual interference function $I(u,v)$ (37) and (39) , hence the optimization of R_{av} can be over $I(u,v)$.

$$R_{av}^{max} = \max_{I(u,v)} \int_0^\infty da \int_0^\infty db f(a,b) \int_0^a du \int_u^b dv R_F(u,v,I,I_{uv}) \tag{44}$$

where $f(a,b)$ is defined by (43), $R_F(u,v,I,I_{uv}) \triangleq \frac{d^2R(u,v,I,I_{uv})}{dudv}$ from (40) which depends on the interference function $I(u,v)$ and the power density $I_{uv}(u,v)$ from (37) and (39) correspondingly. The maximization of R_{av} (42) with respect to the functional $I(u,v)$ is a variational problem, which is developed in Appendix A. Consequently, the optimization problem may be stated in a form of partial differential equation (PDE),

$$S_I + \frac{\partial^2}{\partial uv} S_{I_{uv}} = 0. \tag{45}$$

where

$$S(a,b,I,I_{ab}) \triangleq (1 + F(a,b) - F(a) - F(b)) \cdot R_F(a,b,I,I_{ab}),$$

and S_I is the partial derivative with respect to the function $I(u,v)$. $S_{I_{uv}}$ is the partial derivative with respect to the function I_{uv}. I_{uv} is the second order partial derivative of $I(u,v)$ with respect to u and v.

The necessary condition for extremum (45) is given, in Appendix A, in terms of a non-linear second order PDE, and does not appear to have any straight forward analytical solution.

Therefore, we demonstrate, in the following, a single dimension approximation to the optimal solution.

This approximation is termed the 1-D approximation, where the maximization problem is solved by assuming independency between the random variables u and v. As this assumption is never true, the sub-optimal solution may depend on the variables (u, v) assignment. We demonstrate three variable transformations for (u, v) in Appendix B. We explicitly state there the joint CDF, the marginal distributions and the incremental rate $R_F(u, v, I, I_{uv})$, which depend on the variable transformations. It should be noted that the solution of the optimization problem stated in (45) is invariant to the selection of variable transformation of the ordered eigenvalues.

Another sub-optimal approach, not assuming majorization at the receiver is based on finite level of code layering, and was suggested in [24]. In this approach the transmitted data consists of a finite number of code layers, and the number of decodable layers depends on the channel realization, this method is presented in section 5..

IV. The MIMO 1-D Approximation

We suggest here to approximate the optimal $\rho(a, b)$ by breaking the mutual dependency of the variables a and b requiring

$$\rho(a, b) = \rho(a)\rho(b). \tag{46}$$

Such a representation (46) bares two independent solutions for $I(a)$ and $I(b)$ as $\rho(a) = -\frac{d}{da}I(a)$ and $\rho(b) = -\frac{d}{db}I(b)$ respectively. Hence it is desired to adhere two separate 1-D problems. Each channel is slowly fading, one is $F(a)$ distributed and the other $F(b)$ given by (B.2) and (B.3) respectively. In a SISO channel with a fading distribution (B.2) the achievable average rate denoted by $R_{T,a}$ is given in (7) and equivalently for $F(b)$ the total rate is denoted $R_{T,b}$.

We use the general result of (14) to explicitly determine the residual interference function for three variable transforms suggested in Appendix B. For all three variable transforms the channel distribution $F(a)$ is similar since $u = \lambda_2$. Therefore, the residual interference function is also common, and its solution using (14) and (B.2) is

$$I(a) = \begin{cases} \frac{1}{4}a^{-2} - a^{-1} & a_0 \leq a \leq a_1 \\ 0 & else \end{cases} \tag{47}$$

where a_0, a_1 are determined by $I(a_0) = \sqrt{SNR}$, $I(a_0) = 0$ respectively, and

$$\rho(a) = -\frac{d}{da}I(a) = \begin{cases} -\frac{1}{2}a^{-3} + a^{-2} & a_0 \leq a \leq a_1 \\ 0 & else \end{cases} \tag{48}$$

this is exactly the scalar Rayleigh channel solution as the distribution density of a is also exponential (see [25] for the slowly fading SISO channel).

We similarly solve the Eüler equation for $I(b)$ and $\rho(b)$ for the pairs $(a, b) = (\lambda_2, \lambda_1)$, $(a, b) = (\lambda_2, \lambda_1 + \lambda_2)$ and $(a, b) = (\lambda_2, \lambda_1 - \lambda_2)$. This is done using (14) and the marginal distribution $F(b)$ for each pair given by (B.3), (B.6) and (B.10) respectively. Finally, the residual interference function is the product of the single dimensional interference functions $I(a, b) = I(a)I(b)$. The achievable average rate $R_F(a, b, I, I_{ab})$ can then be computed for each pair of variable transform using (40), (B.7) and (B.11) for the respective variable transform.

V. MIMO Finite Level Code Layering

Finite level of code layering was suggested in [24] for the single transmit antenna case. In this approach the transmitted data consists of a finite number of code layers, and the number of decodable layers depends on the channel realization.

A. Single State Channel Approximation

We consider now the achievable rates with an outage approach, which will also enhance the understanding of the limitations of the majorization approach discussed earlier. In this scheme we transmit in a constant rate $R_{m1,c}$, and assume nothing is decoded on the outage events.

This setting is described by assuming two delta functions for the power density

$$\rho(\lambda_1, \lambda_2) = \delta(\lambda_1 - L_1)\delta(\lambda_2 - L_2) \tag{49}$$

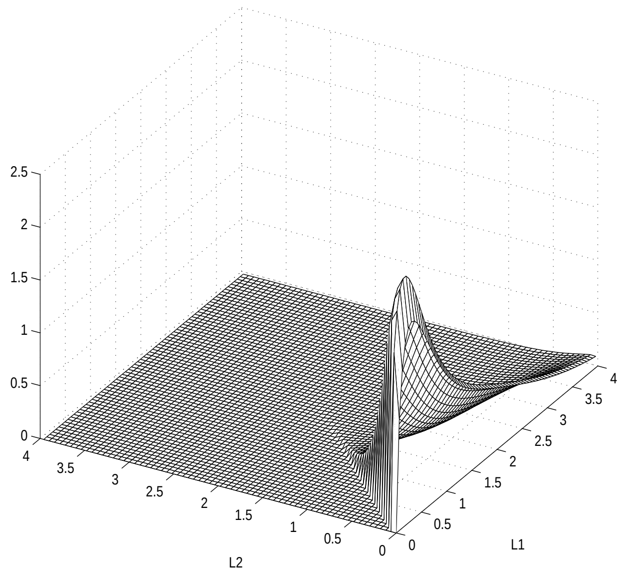

Fig. 7. $R_{x,av}(\lambda_1, \lambda_2)$, for SNR=15 dB, where L1 denotes λ_1 and L2 denotes λ_2

where $L_1 \geq L_2$ are some non-negative constants, and λ_1, λ_2 are associated with the maximal and minimal singular values of the realization of $\frac{1}{2}HH^{\dagger}$. The associated, constant transmit rate $R_{m1,c}$ is then

$$R_{m1,c} = \log(1 + PL_1) + \log(1 + PL_2). \tag{50}$$

The transmitter sends a single stream at rate $R_{m1,c}$, with a total power P. This stream will be correctly decoded, if (L_1, L_2) is weakly supermajorized [30] by the (λ_1, λ_2) ($(\lambda_1, \lambda_2) \prec^w (L_1, L_2)$), where $\lambda_1 \geq \lambda_2$. The associated probability of that is

$$P_{major} = Pr\left\{\lambda_2 \geq L_2, \lambda_2 + \lambda_1 \geq L_2 + L_1\right\}. \tag{51}$$

The average rate for some (L_1, L_2) is then

$$R_{m1} = P_{major} \cdot R_{m1,c}. \tag{52}$$

It is optimized over (L_1, L_2) to achieve a maximal rate in the proposed setting, given the joint density of (λ_1, λ_2) by (43). The explicit expression of P_{major} is

$$\begin{aligned} P_{major} &= \int\limits_{L_2}^{\frac{L_1+L_2}{2}} da \int\limits_{L_2+L_1-a}^{\infty} db f_{\lambda_2,\lambda_1}(a,b) + \int\limits_{\frac{L_1+L_2}{2}}^{\infty} da \int\limits_{a}^{\infty} db f_{\lambda_2,\lambda_1}(a,b) = \\ &= \left(1 + \tfrac{2}{3}\left(L_1 - L_2\right)\left(2\left(L_1 - L_2\right)^2 - 3L_1 + 3L_2 + 3\right)\right) e^{-2L_1 - 2L_2} \end{aligned} \tag{53}$$

Figure 7 demonstrates R_{m1} for the different selections of threshold pairs (L_1, L_2), and a fixed SNR.

It is clear that the majorization requirement $\lambda_1 \geq L_1$ and $\lambda_1 + \lambda_2 \geq L_1 + L_2$ guarantees reliable decoding, however it is a sufficient but not necessary condition. If this requirement is not met, it does NOT imply that this rate is undecodable. This stands in contrast to the scalar (SISO) case.

It would be interesting to see the loss in SNR from the majorization requirement. The Outage Capacity is the maximal achievable average rate under the single state constraint, when the transmitter sends data in a fixed rate $R_{m1,c}$. The conveyed rate is limited by the mutual information For channel realization (λ_1, λ_2),

$$R_{o1,c} \leq \log(1 + P\lambda_1) + \log(1 + P\lambda_2)$$

For some rate $R_{o1,c}$, the probability of successfully decoding the stream is

$$P_{o1} = Pr\left\{R_{o1,c} \leq \log(1 + P\lambda_1) + \log(1 + P\lambda_2)\right\}.$$

Using the joint density of (λ_1, λ_2) (43), we practice change of variables $R_k = \log(1 + Pk)$, for $k = \lambda_1, \lambda_2$. The new density function is given by

$$f_{R_b,R_a}(r_1, r_2) = \frac{16}{P^4} \exp\left\{-2/P\left(e^{r_1} + e^{r_2} - 2\right)\right\}\left(e^{r_1} - e^{r_2}\right)^2 e^{r_1 + r_2} \tag{54}$$

where the factor $e^{r_1+r_2}/P^2$ is the Jacobian. Thus the probability (54) is given by

$$P_{o1} = \int\limits_0^\infty dr_2 \int\limits_{\max\{R_{x1}-r2, r2\}}^\infty dr_1 f_{R_b,R_a}(r_1, r_2) \tag{55}$$

The consequential average rate for some $R_{o1,c}$ is

$$R_{o1} = P_{o1} \cdot R_{o1,c}. \tag{56}$$

which is optimized over $R_{o1,c}$ to achieve maximal average rate R_{o1}.

B. The 2-State Channel Approximation

Let the transmit power P consist of two streams of power P_1, P_2 and with associated rates R_1, R_2 respectively. This setting is associated with a double impulse (extension of the single impulse as in equation (49)).

Let region $\mathcal{A}$ be the region where R_1 can be decoded reliably, treating R_2 as interference, and similarly region $\mathcal{B}$, as the region where R_2 can be decoded treating R_1 as an interference. Region $\mathcal{C}$, is the MAC region, where both R_1 and R_2 can be decoded. Naturally, the regions, as defined below, are not disjoint.

Region $\mathcal{A}$, is then defined for some channel realization (λ_1, λ_2) $(\lambda_1 \geq \lambda_2)$ that satisfy

$$R_1 \leq \log(1 + \frac{P_1\lambda_1}{1 + P_2\lambda_1}) + \log(1 + \frac{P_1\lambda_2}{1 + P_2\lambda_2}). \tag{57}$$

Region $\mathcal{B}$ is defined by by (λ_1, λ_2) that satisfy

$$R_2 \leq \log(1 + \frac{P_2\lambda_1}{1 + P_1\lambda_1}) + \log(1 + \frac{P_2\lambda_2}{1 + P_1\lambda_2}). \tag{58}$$

Region $\mathcal{C}$ where both R_1 and R_2 are decoded (interpreted as a multiple access channel with two users), is defined (λ_1, λ_2) that satisfy:

$$\begin{aligned} R_1 &\leq \log(1 + P_1\lambda_1) + \log(1 + P_1\lambda_2), \\ R_2 &\leq \log(1 + P_2\lambda_1) + \log(1 + P_2\lambda_2), \\ R_1 + R_2 &\leq \log(1 + P\lambda_1) + \log(1 + P\lambda_2) \end{aligned} \tag{59}$$

Now let Pr_1, Pr_2 and Pr_3 be the following probabilities:

$$\begin{aligned} Pr_1 &= Pr\left\{\mathcal{A} \bigcap \mathcal{B}^c \bigcap \mathcal{C}^c\right\} \\ Pr_2 &= Pr\left\{\mathcal{B} \bigcap \mathcal{A}^c \bigcap \mathcal{C}^c\right\} \\ Pr_3 &= Pr\left\{\mathcal{C}\right\} \end{aligned} \tag{60}$$

where X^c denotes the complementary of X. The average rate is given by:

$$R_{o2} = Pr_1R_1 + Pr_2R_2 + Pr_3(R_1 + R_2) \tag{61}$$

Evidently R_{o2} is optimized over R_1, R_2 and P_1, $(P_1 + P_2 = P)$. This two-state channel approximation was shown to compensate the loss of the outage approach in the SISO setting, as can be seen in figure 8 of [24].

It was observed in numerical optimization, that for the MIMO channel this approach has only a small gain over the outage approach from the previous sub-section.

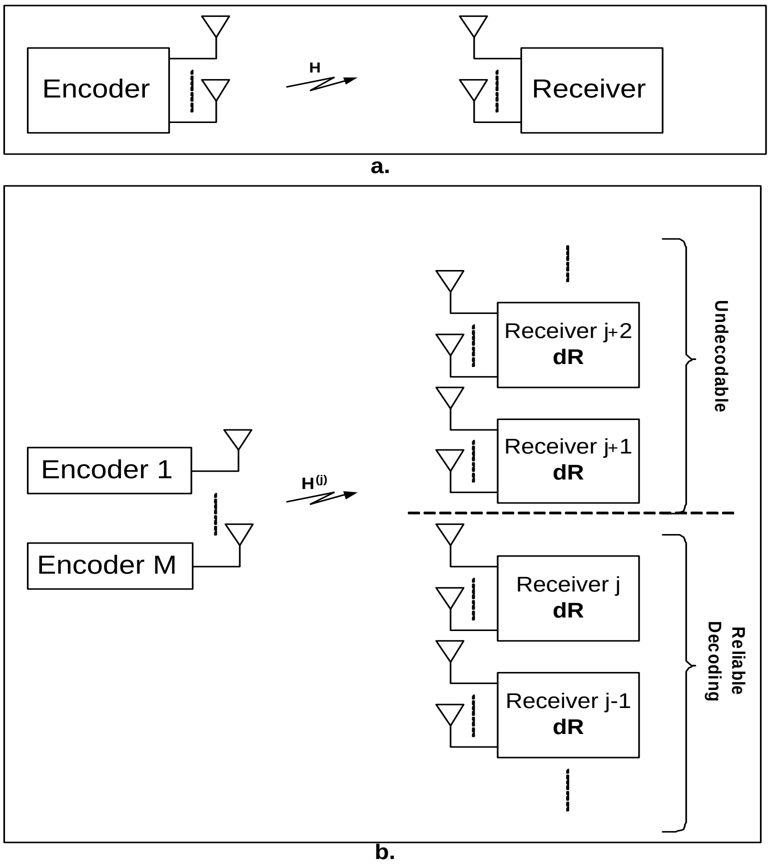

Fig. 8. a. A MIMO channel with a channel propagation matrix H. b. The equivalent MIMO (MAC) broadcast channel model. The transmitter includes independent encoders. Ordering of receiver is follows from the SISO broadcast approach.

VI. MAC-Broadcast Approach

We extend the broadcast approach here by reducing the coding complexity. Instead of performing joint encoding for all transmit antennas, each antenna has an independent encoder, as illustrated in Figure 8. The receiver then views a multiple-access channel.

This approach was first presented in [26] for the multiple-access channel, employing the broadcast approach at the receiver. Its advantage is that each transmitter views an equivalent degraded broadcast channel, and the results of the SISO broadcast strategy may be directly used here. In the following we present a broadcast strategy for the MAC or MIMO channel. We suggest an iterative algorithm for optimizing the power distribution. In the work of Liu *et. al.* [31], which also incorporates the MAC approach of [26], the authors suggest an iterative algorithm for computing the achievable rate using MMSE based decoding, for some transmit power distribution, without optimizing the transmit power distribution. The MAC approach is also studied with a single code layer at each transmitter, denoted as MAC-Outage. It is considered with both successive decoding as proposed in the broadcast strategy, and with optimal joint decoding. Numerical results of achievable rates are then demonstrated in the next section.

A. *MAC-Outage*

In this subsection we discuss a MAC-outage approach. In contrast to the MAC-Broadcast each encoder has a single code layer here. The decoder attempts first to decode both streams adhering to the optimal joint detection. If that fails it tries to decode either of the streams, treating the other stream as an interfering signal, while optimally preprocessing for the channel response H. We use the term joint detector to characterize this detection procedure, which resembles the strongest user detector concept, first introduced in [32]. Later on, we bring the relevant mutual information expressions that govern the performance of both the joint detector and a successive decoding strategy.

A.1 Optimal Detection

Consider the MIMO channel defined in (28). We specify the received signal again, for the case of $M = N = 2$:

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} h_{11}x_1 + h_{12}x_2 + n_1 \\ h_{21}x_1 + h_{22}x_2 + n_2 \end{bmatrix} \tag{62}$$

where h_{ij} are the elements of the channel propagation matrix H.

Each antenna is assumed to transmit independently at rate R and power $P/2$. Thus the SNR at each receive antenna is P. We define now the three decoding regions, which are not disjoint. When trying to decode just one stream (if joint detection has failed), the receiver decodes user I as if the other user is an additive Gaussian interferer, using an MMSE decoder [33], [34]. The actual achievable rate by a single user decoder depends on the distribution of symbols transmitted by the interferer. However, as the channel inputs are power-constrained, the minimax distribution of the interferer is Gaussian [34]. Hence, a single-user transmitting at rate R and power $P/2$ in presence of a single interferer will be decoded if

$$R \leq I(\mathbf{y}; x_1) \tag{63}$$

where $I(\mathbf{y}; x_1)$ is the mutual information of the received vector and the transmitted information x_1. The decoding region of user I as defined by (63) is denoted by $\mathcal{A}$. Region $\mathcal{A}$ includes all possible channel realizations H for which the mutual information $I(\mathbf{y}; x_1)$ is greater than the transmission rate of user I. Similarly, a second case is defined, the condition for decoding user II successfully, when considering user I as an interferer, and using the MMSE decoder is

$$R \leq I(\mathbf{y}; x_2), \tag{64}$$

and decoding region of user II is denoted by $\mathcal{B}$. In the third case, which in fact the decoder attempts first, the receiver tries to decode both users jointly. The condition for decoding both users, is the same as in a multi access channel (MAC) with two users. Hence the decoding region $\mathcal{C}$ is

$$\begin{aligned} \mathcal{C}_1 :& \quad 2R \leq I(\mathbf{y}; \mathbf{x}) \\ \mathcal{C}_2 :& \quad R \leq I(\mathbf{y}; x_1|x_2) \ , \\ \mathcal{C}_3 :& \quad R \leq I(\mathbf{y}; x_2|x_1) \end{aligned} \tag{65}$$

and is defined as the intersection of all three subregions $\mathcal{C} = \{\mathcal{C}_1 \cap \mathcal{C}_2 \cap \mathcal{C}_3\}$.

The average achievable rate is

$$R_{mac1,opt} = (P_A + P_B + 2P_C)R \tag{66}$$

where the disjoint probabilities of correct decoding in each region is defined by

$$\begin{aligned} P_A &= Pr\{\mathcal{A} \cap \mathcal{B}^c \cap \mathcal{C}^c\} \\ P_B &= Pr\{\mathcal{B} \cap \mathcal{A}^c \cap \mathcal{C}^c\} \\ P_C &= Pr\{\mathcal{C}\} \end{aligned} \tag{67}$$

where X^c denotes the complementary of X.

The explicit expressions of mutual information in each region are detailed in the following. For region $\mathcal{A}$,

$$I(\mathbf{y}; x_1) = \log\det\Lambda_{\mathbf{y}} - \log\left\{1 + \frac{P}{2}(|h_{12}|^2 + |h_{22}|^2)\right\} = \log\det\{\Lambda_{\mathbf{y}}\} - \log(1 + \frac{P}{2}\Lambda_{H,22}) \quad (68)$$

where $\Lambda_{\mathbf{y}} = I + \frac{P}{2}\Lambda_H$, and Λ_H is the covariance matrix of the channel propagation coefficients. Similarly, we derive $I(\mathbf{y}; x_2)$

$$I(\mathbf{y}; x_2) = \log\det\Lambda_{\mathbf{y}} - \log\left\{1 + \frac{P}{2}(|h_{21}|^2 + |h_{11}|^2)\right\} = \log\det\{\Lambda_{\mathbf{y}}\} - \log(1 + \frac{P}{2}\Lambda_{H,11}). \quad (69)$$

For region $\mathcal{C}$ the expressions of mutual information are given by

$$\begin{array}{ll} I(\mathbf{y}; \mathbf{x}) = & \log\det\Lambda_{\mathbf{y}} \\ I(\mathbf{y}; x_1|x_2) = & \log\left\{1 + \frac{P}{2}(|h_{21}|^2 + |h_{11}|^2)\right\} = \log(1 + \frac{P}{2}\Lambda_{H,11}) \\ I(\mathbf{y}; x_2|x_1) = & \log\left\{1 + \frac{P}{2}(|h_{12}|^2 + |h_{22}|^2)\right\} = \log(1 + \frac{P}{2}\Lambda_{H,22}) \end{array} . \quad (70)$$

The maximal achievable average rate is achieved by optimizing R_{mu} (66) over the transmission rate R.

Interestingly, Λ_H is a complex Wishart matrix [29], which is an Hermitian random matrix $HH^\dagger$, where H consists of iid complex Gaussian elements. The distribution of its elements is [29]

$$p(\Lambda_H) = \frac{1}{2^{MN}\tilde{\Gamma}_M(N)}\exp(-\frac{1}{2}tr(\Lambda_H))(\det\Lambda_H)^{N-M} \quad (71)$$

where the complex multivariate gamma function is defined by

$$\tilde{\Gamma}_M(N) = \pi^{M(M-1)/2}\prod_{i=1}^{M}\Gamma(N-i+1)$$

In [29] (and also in (71)) the elements of H are iid complex Gaussian distributed, where the real and imaginary parts are each distributed $N(0,1)$. Therefore the distribution of the Wishart matrix Λ_H with complex elements distributed $CN(0,1)$ each, for $M = N = 2$ is given by

$$p(\Lambda_H) = \exp(-tr(\Lambda_H)), \quad (72)$$

If we further specify the elements of Λ_H

$$\Lambda_H = \begin{bmatrix} w_1 & w_{12} \\ w_{21} & w_2 \end{bmatrix}$$

and recall that the determinant of $\Lambda_{\mathbf{y}}$ is $w_1 \cdot w_2 - w_{12} \cdot w_{21}$, with $w_{12} = w'_{21}$, it turns out that the PDF of Λ_H depends solely on three real variables:

$$p_{\Lambda_H}(w_1, w_2, w_3) = \begin{cases} \exp(-tr(w_1 + w_2)) & , \ w_1 w_2 \geq w_3 \\ 0 & , \ otherwise \end{cases} \tag{73}$$

where $w3 \triangleq w_{12} w_{21}$. Note that Λ_H can be defined as either $\Lambda_H = HH^\dagger$ or $\Lambda_H = H^\dagger H$. In both cases Λ_H is a Wishart matrix and $\det(I + \frac{P}{2} HH^\dagger) = \det(I + \frac{P}{2} H^\dagger H)$, therefore it's preferable to use $\Lambda_H = H^\dagger H$ and keep consistency with definitions of (68)-(70).

A.2 Successive Decoding

Consider here as well the MIMO channel defined in (28). The transmission scheme is identical to the one described in sub-section 6.1.1.. That is each antenna transmits independently at rate R and power $P/2$. Three decoding regions for successive decoding are specified in the following. In the first case, the receiver decodes user I as if the other user is an additive Gaussian interferer. A single-user transmitting at rate R and power $P/2$ in presence of a single interferer will be decoded in the region $\mathcal{A}$ specified in (63). Similarly, the region $\mathcal{B}$ for decoding user II successfully, when considering user I as an interferer is given by (64).

A rate $2R$ may be achieved, when both users are successfully decoded. In contrast to the optimal joint detection region specified in (65), here the receiver tries to decode a user, and when if successful decodes the other user cancelling the first user. The implied region $\mathcal{C}$ of successive decoding is

$$\begin{gathered} \{R \leq I(\mathbf{y}; x_1) \cap R \leq I(\mathbf{y}; x_1 | x_2)\} \\ \cup \\ \{R \leq I(\mathbf{y}; x_2) \cap R \leq I(\mathbf{y}; x_2 | x_1)\} \end{gathered} \tag{74}$$

The average achievable rate, denoted by $R_{mac1,bs}$, is specified in (66), where P_C in (66) is the probability of the new region $\mathcal{C}$ from (74). The expressions of the mutual information, are specified also in (68)-(70).

B. Two User Broadcast Approach

We now adapt the single user broadcast strategy to fit a two-user multiple-access fading channel. Consider the channel model described by (28), where M designates the number of transmit antennas, or the number of users in the MAC equivalent. We restrict the study to $M = 2$, and outline the implications of generalization to M-users.

As in the SISO broadcast strategy, each user comprises a continuum of parallel transmissions with the power distributions $\rho_m(s)ds$, $m = 1, 2$. Let

$$I_m(s) = \int_s^\infty \rho_m(u)du, \quad m = 1, 2. \tag{75}$$

The receiver performs MMSE Decision feedback equalization (DFE) front end for each user. That is it does MMSE accounting for the residual interference of the other user, and assuming also cancellation of his decoded layers. This results in a new normalized SIR denoted by ω_m, which is equivalent to the fading power in the SISO case. This ω_m defines the residual interference $I(\omega_m)$ to be accounted for by the MMSE receiver of the other user. Thus when decoding one user by an MMSE-DFE receiver, it considers the undecoded layers of the other user as interference.

We now formalize the MMSE-DFE decoding rules. The MMSE resulting signal to interference ratio (SIR) [34] for users $m = 1, 2$, for some residual interference from the other user, is given by

$$\begin{aligned} SIR_1 &= \tfrac{1}{[1+HQ_1H^\dagger]^{-1}_{1,1}} - 1 \\ SIR_2 &= \tfrac{1}{[1+HQ_2H^\dagger]^{-1}_{2,2}} - 1 \end{aligned} \tag{76}$$

where H is the channel propagation matrix, $Q_1 \triangleq diag\{P_1, I_2(\omega_2)\}$, $Q_2 \triangleq diag\{I_1(\omega_1), P_2\}$, and P_m is the original transmit power of user $m = 1, 2$. The normalized SIR of user m, is simply a normalization of the SIR by the transmit power of user m,

$$\begin{aligned} \omega_1 = SIR_1/P_1 = \tfrac{1}{P_1[1+HQ_1H^\dagger]^{-1}_{1,1}} - 1/P_1 \\ \omega_2 = SIR_2/P_2 = \tfrac{1}{P_2[1+HQ_2H^\dagger]^{-1}_{2,2}} - 1/P_2 \end{aligned} \tag{77}$$

The simultaneously achievable rate of the users $R_1(\omega_1, \omega_2)$, $R_2(\omega_1, \omega_2)$ respectively depend now on both normalized SIRs ω_1 and ω_2 through (77),

$$R_m = \int_0^{\omega_m} \tfrac{s\rho_m(s)ds}{1+sI_m(s)} \quad m = 1, 2 \tag{78}$$

where $\rho_m(s)$ is the transmit power distribution defined earlier. The expected rates per user are then

$$R_{mac-bs,m} = E_H[R_m(\omega_m)] \quad m = 1, 2$$

which is the average of instantaneous achievable rates R_m for all channel realizations H. Using (78) for the instantaneous rate and performing variable transform on the RVs H to ω_m, the expected rate per user is

$$R_{mac-bs,m} = \int_0^\infty (1 - F_{\omega_m}(u)) \frac{u\rho_m(u)du}{1 + uI_m(u)}, \quad m = 1, 2 \tag{79}$$

where $F_{\omega_m}(s)$ is the CDF of the ω_m, that is

$$F_{\omega_m}(\nu) = Pr(\omega_m(H) \leq \nu) = \int\int_{\mathcal{R}(\nu)} f_H. \tag{80}$$

where $\mathcal{R}(\nu)$ stands for the associated region in which the RVs of H lie. The functions $I_1(u)$, $I_2(u)$ can be optimized as to maximize the expected rates, or the total expected throughput $R_{mac-bs,1} + R_{mac-bs,2}$ as is discussed in the following subsection. In parallel to the single user case, also here expected rates per outages can be considered by replacing the original probability distribution of the fading powers $F_{\omega_m}(u)$ by $F^{\mathbf{s_{th}}}_{\omega_m}(u)$ similar to (8), where now $F^{\mathbf{s_{th}}}_{\omega_m}(u)$ designates the conditional distribution function of ω_m, conditioned on the event $\omega_m \notin \mathbf{s_{th}}$, where $\mathbf{s_{th}}$ denotes now the a region $0 \leq \omega_m \leq s_{th}$, which is associated with an outage probability.

We pose an algorithm for optimizing the maximal average rate R_{mac-bs}, in the symmetric case where $R_{mac-bs,m} = R_{mac-bs}$ for $m = 1, 2$. The following algorithm presented for the two-user setting, is straightforwardly extended to the M-user setting. Assuming the users are transmitting with power $P/2$ each, due to symmetry $I_m(u) = I(u)$ for $m = 1, 2$. The algorithm's main steps are as follows

1. *Initialization* - $I^0(u)$ is the interference function initialized, possibly to the optimized single-user function satisfying (14). The following steps are repeated iteratively. The iteration index is denoted i_{it}.
2. Solve $\omega_m(H)$ for $m = 1, 2$, by the set of equations defined in (77).
3. Find the associated CDF of $F_{\omega_m}(\nu)$ for $m = 1, 2$, using (80).
4. Calculate $I^{i_{it}}(u)$, using (14), and by replacing the previous $F_{\omega_m}(\nu)$ with the one computed in the previous step. The new $I^{i_{it}}(u)$ is used in the next iteration, starting at step No. 2.

This procedure is iteratively repeated until convergence to $I^{\infty} \triangleq I^*(u)$. Equation (77) does not in general guarantee a unique solution. However all solutions here are jointly monotonic. That is if $\omega_1' > \omega_1$, so is $\omega_2' > \omega_2$. The stable solution to be selected is the minimal pair. This ensures a unique solution.

VII. MIMO - Numerical Results

In the following, the average achievable rates of the various suggested methods are compared in Figure 9, the ergodic capacity upper bound for the SISO, SIMO, MIMO is illustrated as a performance reference measure. The MIMO approach is tested in the 1-D approximation with different set of indices. It turns out that only for the set $u = \lambda_2$ and $v = \lambda_1 - \lambda_2$ the ad-hoc MIMO setting has some gain over the SISO setting. A comparison the MIMO approach to the SIMO, SISO using the 1-D approximation was performed. The best performance was evident for the selection of $u = \lambda_2$ and $v = \lambda_1 - \lambda_2$, as demonstrated by R_{ma} in Figure 9. It can be seen that for high SNR, the MIMO majorization based approach is inferior to both SIMO and SISO schemes. It may be concluded that all proposed schemes, using two transmit antennas and two receive antennas in the 1-D approximation fail to introduce the expected gain. Reference SIMO, SISO ergodic capacities are also given in (26).

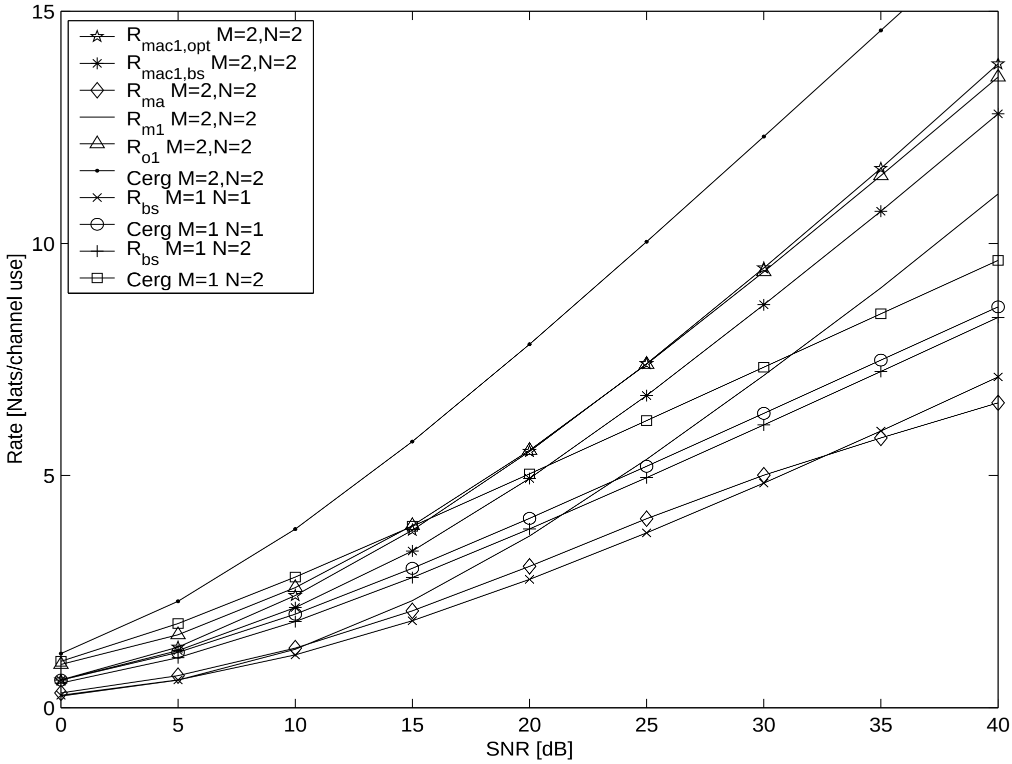

Fig. 9. Comparison a MIMO channel: $R_{mac1,opt}$ - average achievable rate in the single layer MAC outage approach (optimal joint detection). $R_{mac1,bs}$ - average achievable rate in the single layer MAC-broadcast approach (successive decoding). R_{ma} - average achievable rate via the single dimensional approximation for the optimal power distribution enforcing majorization. R_{o1} - outage capacity. R_{m1} - outage achievable average rate when enforcing majorization ordering at the receiver. R_{bs} - optimal achievable average rate in the broadcast approach using a single transmit antenna (SIMO, SISO results are demonstrated). The ergodic capacity C_{erg} is also presented as upper bound for each of the settings.

Figure 9 also shows the calculation results of R_{m1} and R_{o1} and R_{mac} vs. SNR, compared to the ergodic capacity of this MIMO setting. For $M = N = 2$ the ergodic capacity [28] is

$$C_{erg} = \int_0^\infty dx \log(1 + \frac{P}{2}x)(1 + (1 - x)^2)e^{-x}. \tag{81}$$

The optimized $R_{mac1,opt}$ (66) is higher than the outage capacity R_{o1} of the $M = N = 2$ setting, for high SNR. As can be inferred from Figure 9, the loss of $R_{mac1,opt}$ relative to the ergodic capacity C_{erg} for high SNR is 6.4 dB, the outage capacity R_{o1} has an additional loss of 0.6 dB, and the majorization limited outage capacity has an additional ~ 6 dB loss.

The large penalty paid due to majorization is clearly demonstrated in figure 9, which emphasizes the advantage of the independent antenna coding for high SNR in the MAC outage approach. The slopes of the outage and MAC curves though close seem to be different, which point to the asymptotic advantage of the MAC approach. Intuitively, the MAC approach allows three levels of decoding, i.e. a level where both streams (users) are detected (rate $2R$), and a second level where only one of the streams (users) is decoded correctly (rate R), and a third where neither streams can be detected (outage event). As opposed to the classical outage approach, where there are two levels of decoding, i.e. either the received signal can be correctly decoded, or in an outage event nothing is decoded. It can be noted that the rate $R_{mac1,opt}$ achieved with optimal joint detection at the receiver, a gain of about 2 dB over the $R_{mac1,bs}$ achieved with successive decoding, inherent in the MAC-broadcast approach.

VIII. Conclusions

The MIMO broadcast approach, as described here, demonstrates an achievable region which by no means is claimed to be optimal. The central reason for sub-optimality is the ordering of information layers required by the supermajorization which, at this point is not proved to maintain optimality. Thus, the very idea of layering in the fashion done here may imply inherent suboptimality, as also observed through some ad-hoc approximations.

As opposed to the single dimensional broadcast approach to fading channels, where the layers were decoded up to a certain point (index) determined by the actual fading realization, the relation of fading and indexing in MIMO is different. Namely, the layers indexed now in an array (u, v) are decoded as dictated by the decodable set that is all indices u, v which are associated with eigenvalues, such that, those eigenvalues supermajorize the actually realized eigenvalues. This facilitates different selection of the index coordinate system, as was demonstrated here for $N = M = 2$. In the single dimensional case the supermajorization property follows trivially.

We have demonstrated here the MIMO broadcast approach for $M = N = 2$. The generalization to any (M, N) is evident, as the supermajorization ranking holds in general. The variational problem associated with maximization of the average rate for $M = N = 2$ has no straight-forward solution. Some approximations for the realizing power distribution was used. By requiring independency in each dimension (46), we get two separate optimization problems. The results of this approximation show a degradation from the SISO approach.[36]

Another ad-hoc method is the outage approach demonstrated here for $M = N = 2$, in this setting one decodes all information if the channel realization is better then a predetermined threshold. We extend the outage approach to the double impulse distribution, suggesting transmission of two streams, each decoded above a different predetermined threshold. The numerically optimized rates of the different methods demonstrate the large penalty paid by the sufficient but not necessary demand for decoding of supermajorization. Moreover the fact that the double impulse gains only marginally over a single impulse and only at very high SNR, indicates on the relative efficiency of the outage approach in the MIMO case as compared to the SISO case. This is due to the less randomness (hardening) associated with the MIMO case, and at the extreme of large number of received and transmitted antennas, keeping their ratio constant, the capacity crystallizes and becomes deterministic, equaling to the ergodic rate, which makes the this approach optimal. This is true for the MAC-outage approach as well, and is an immediate artifact of the results in [28], [34] and [35].

We adapted the SISO broadcast approach for the MIMO channel, by requiring independent encoding at each transmit antenna, similar to the multiple-access channel, and employing the broadcast approach at the receiver. Its advantage is that each transmitter views an equivalent degraded broadcast channel, and the results of the SISO broadcast strategy can be directly used. An iterative algorithm for optimizing the transmit power distribution was suggested for this setting. We have presented the MAC approach with a single code layer at each transmitter, considered with both successive decoding as used in the broadcast strategy, and with optimal joint decoding. It was demonstrated that broadcast strategy may suffer a non-negligible penalty due to its inherent successive decoding. Solving in general the optimal transmit power distribution is a subject for further research.

The relative small loss of the successive decoding as compared to the optimal joint decoding in the MAC setting supports the application of the MAC broadcast approach as discussed in section 6.2.. Further the observation that every point in the dominant rate face in the MAC M user case, can be also achievable by successive cancellation where each user mimics at most a two user signal [36] bolsters the argument that marginal degradation is expected in the broadcast approach as compared to the joint detection of the layered system. See also [37] for application to a faded ergodic environment.

The presented broadcast approaches, namely, majorization, and all MIMO approaches extend in a direct manner to the M transmit and N receive antennas case. Specifically, for the MAC broadcast approach, the extension is outlined in Appendix C.

Appendix

Appendix A
Rate optimization - variational problem

Let us define a functional of interest, which is the average rate in (42) for some residual interference function $I(u,v)$,

$$J[I] = \int_0^\infty da \int_0^\infty db f(a,b) \int_0^a du \int_u^b dv R_F(u,v,I,I_{uv}). \tag{A.1}$$

For simplicity, we write $G(u,v) \triangleq \int_0^a du \int_u^b dv R_F(u,v,I,I_{uv})$, and the joint cumulative distribution $F(a,b) \triangleq \int_0^a du \int_u^b dv f(u,v)$. Moreover, for compactness of the presentation the joint density function will be denoted as $f(a,b) \triangleq F_{ab}$. Now we can rewrite equation (A.1)

$$J[I] = \int_0^\infty da \int_0^\infty db F_{ab} G(a,b) \tag{A.2}$$

Applying integration in parts is developed here in the next few steps

$$\begin{aligned} J[I] &= [F(a,b)G(a,b)]_{a,b=0}^\infty - \int_0^\infty da \int_0^\infty db\,(FG_a b + F_a G_b + F_b G_a) = \\ &[F(a,b)G(a,b)]_{a,b=0}^\infty - \int_0^\infty da \int_0^\infty db F G_{ab} \\ &- \int_0^\infty db\,[FG_b]_{a=0}^\infty + \int_0^\infty da \int_0^\infty db F G_{ab} - \\ &- \int_0^\infty da\,[FG_a]_{b=0}^\infty + \int_0^\infty da \int_0^\infty db F G_{ab} = \\ &= \int_0^\infty da \int_0^\infty db\,(1 + F(a,b) - F(a) - F(b)) \cdot R_F(a,b,I,I_{ab}) \end{aligned} \tag{A.3}$$

where $F(a)$ and $F(b)$ denote the marginal cumulative distributions of a and b respectively. Now that the functional contains less integrals, we can define the integrand, which is also a function of I as

$$S(a,b,I,I_{ab}) \triangleq (1 + F(a,b) - F(a) - F(b)) \cdot R_F(a,b,I,I_{ab}).$$

According to the selection of transformation for variables a and b, our functional $J[I]$ is defined over an area where $b \geq a \geq 0$, and therefore

$$J[I] = \int_0^\infty da \int_a^\infty dbS(a, b, I, I_{a,b}) \tag{A.4}$$

We are interested in $I(u, v)$, which maximizes $J[I]$. One can replace I by some varied funtion $I(u, v) + h(u, v)$. Let $h(u, v)$ be an arbitrary function which has first and second derivatives in $\mathcal{R}$ (the region of integration (A.4)), and vanishes on the boundary Γ of $\mathcal{R}$. Then if $I(u, v)$ belongs to the domain of the definition of (A.4), so does $I(u, v) + h(u, v)$.

$$\Delta J = J[I + h] - J[I] = \int \int_R S(u, v, I + h, I_{uv} + h_{uv}) - S(u, v, I, I_{uv})dudv \tag{A.5}$$

It follows by using the first order representation of the Taylor's theorem, that the actual variation of $J[I]$ [27] is

$$\delta J = \int \int_R S_I h + S_{I_{uv}} h_{uv} dudv \tag{A.6}$$

where S_I is the partial derivative of $S(u, v, I, I_{uv})$ with respect to I. The integral over $S_{I_{uv}} h_{uv}$ may be simplified by considering the properties of h. Using the rules of partial derivatives

$$\begin{aligned} S_{I_{uv}} h_{uv} &= \tfrac{\partial}{\partial u}(S_{I_{uv}} h_v) - (\tfrac{\partial}{\partial u} S_{I_{uv}}) h_v = \\ &= \tfrac{\partial}{\partial u}(S_{I_{uv}} h_v) - \tfrac{\partial}{\partial v}[(\tfrac{\partial}{\partial u} S_{I_{uv}}) h] + (\tfrac{\partial^2}{\partial uv} S_{I_{uv}}) h. \end{aligned} \tag{A.7}$$

Further on, using Green's theorem

$$\int \int_R (\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}) dxdy = \oint_\Gamma (Pdx + Qdy)$$

and substituting P, Q with the two left arguments in the second line of (A.7) we get

$$\int \int_R \left(\frac{\partial}{\partial u}(S_{I_{uv}} h_v) - \frac{\partial}{\partial v}[(\frac{\partial}{\partial u} S_{I_{uv}}) h] \right) dudv = \oint_\Gamma (S_{I_{uv}} h_v) dv + ((\frac{\partial}{\partial u} S_{I_{uv}}) h) du. \tag{A.8}$$

The integral along Γ is zero, since $h(u, v)$ and its derivatives vanish on Γ. Consequentially, the variation may be expressed by an integral of a product of h and a function of S

$$\delta J = \int \int_R \left(S_I + \frac{\partial^2}{\partial uv} S_{I_{uv}} \right) \cdot h \ dudv. \tag{A.9}$$

According to Theorem 2 in section 3.2 of [27], a necessary condition for $J[I]$ to have an extremum for $I = I(u,v)$ is that $\delta J = 0$. Furthermore, using the lemma (21) on p. 22 of [27] we get

$$S_I + \frac{\partial^2}{\partial uv} S_{I_{uv}} = 0. \tag{A.10}$$

In order to give explicit expressions to S_I and $S_{I_{uv}}$, we need to compute $F(a,b)$, $F(a)$ and $F(b)$. Let us recall the joint density of the ordered eigenvalues $(\lambda_1 \geq \lambda_2 \geq 0)$ of $\frac{1}{2}HH^\dagger$ (43). In the next section we explicitly state the joint CDF, the marginal distributions and the incremental rate.

Appendix B
Eigenvalues pairs variable transformations

We explicitly state the joint CDF, the marginal distributions and the incremental rate $R_F(u,v,I,I_{uv})$, for three variable transformations. Naturally, there are many more, however the main intension in these transformations is to show the difference in achievable rates of the corresponding sub-optimal solutions, which depend on variable transformations.

A. $(u,v) = (\lambda_2, \lambda_1)$

The region of majorization is demonstrated in figure 10. The darkened area in figure 10 represents all the layers (streams) of information decodable. The joint cumulative distribution function of (43) is

$$F_{\lambda_2,\lambda_1}(a,b) = 1 + \left(4\,(b+a)^2 + 2 + 8b - 16ba\right) e^{-2b-2a} - e^{-4a} - \left(4b^2+2\right) e^{-2b} \tag{B.1}$$

The marginal distribution of a is thus given by

$$F_{\lambda_2}(a) = \lim_{b\to\infty} F(a,b) = 1 - e^{-4a} \tag{B.2}$$

and the marginal distribution of b is

$$F_{\lambda_1}(b) = F(a,b)|_{a=b} = 1 + e^{-4b} - (4b^2+2)e^{-2b}. \tag{B.3}$$

The incremental rate $R_F(u,v,I,I_{uv})$ for $(u,v) = (\lambda_2, \lambda_1)$ is given in (40).

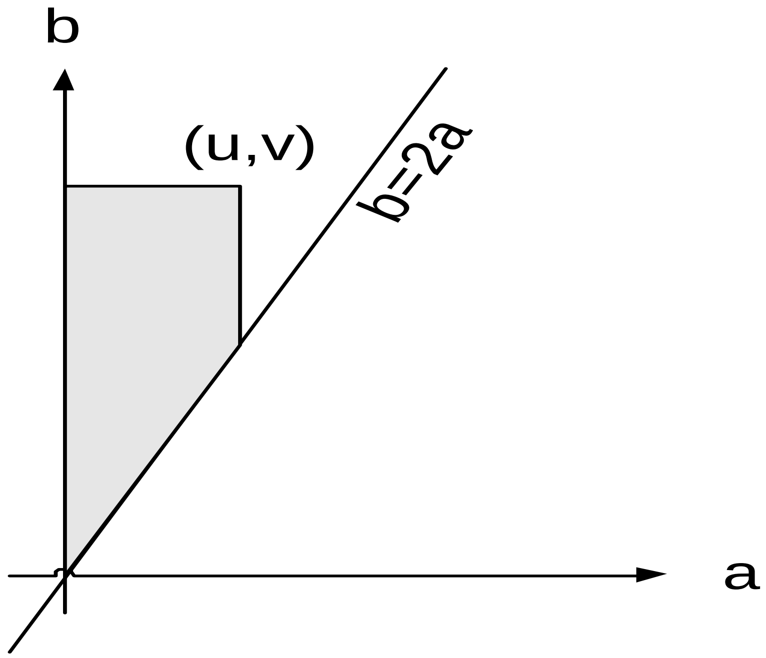

Fig. 10. The majorized area (dotted) for the selection of indices $a = \lambda_2$ and $b = \lambda_1 + \lambda_2$, where $(\lambda_1 \geq \lambda_2)$.

B. $(u, v) = (\lambda_2, \lambda_1 + \lambda_2)$

The region of majorization for this case is demonstrated in figure 11. Here also, the darkened area represents all the layers (streams) of information decodable. In this case, the density function is

$$f_{\lambda_2,\lambda_1+\lambda_2}(a, b) = 16e^{-2b}(b - 2a)^2, \tag{B.4}$$

and the associated joint cumulative distribution is

$$F_{\lambda_2,\lambda_1+\lambda_2}(a, b) = 1 - e^{-4a} + e^{-2b}\left(-8b^2a - 8ab - 4a + 16a^2b + 8a2 - \tfrac{32}{3}a^3\right) \tag{B.5}$$

The marginal distribution of a is unchanged from the previous transform, and is given by (B.2).

$$F_{\lambda_1+\lambda_2}(b) = F(a, b)|_{a=b/2} = 1 - \left(2b^2 + 2b + \frac{4}{3}b^3 + 1\right)e^{-2b}. \tag{B.6}$$

The incremental rate is now

$$R_F(u, v, I, I_{uv}) = -\frac{uI_{uv}(u, v)dudv}{1 + uI(u, v)} - \frac{(v - u)I_{uv}(u, v)dudv}{1 + (v - u)I(u, v)}. \tag{B.7}$$

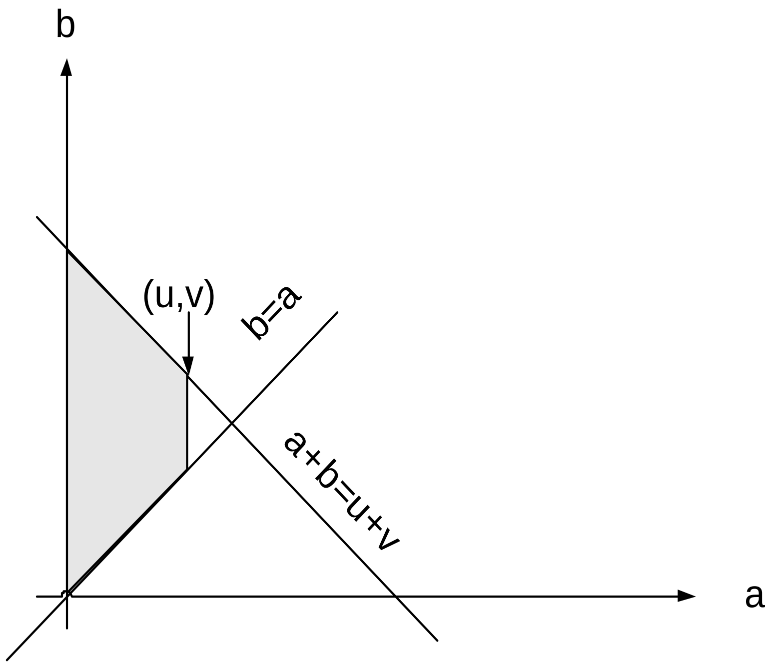

Fig. 11. The majorized area (dotted) for the selection of indices $a = \lambda_2$ and $b = \lambda_1$, where $(\lambda_1 \geq \lambda_2)$.

C. $(u, v) = (\lambda_2, \lambda_1 - \lambda_2)$

Figure 12 demonstrates the region of majorization for this case. The density function is

$$f_{\lambda_2,\lambda_1-\lambda_2}(a, b) = 16e^{-2b-4a}b^2 \tag{B.8}$$

and the associated joint cumulative distribution is

$$F_{\lambda_2,\lambda_1-\lambda_2}(a, b) = 1 - e^{-4a} - (2b^2 + 2b + 1)e^{-2b}(1 - e^{-4a}) \tag{B.9}$$

The marginal distribution of a is stays (B.2) in this case as well.

$$F_{\lambda_1+\lambda_2}(b) = F(a, b)|_{a\to\infty} = 1 - (2b^2 + 2b + 1)e^{-2b}. \tag{B.10}$$

The incremental rate is then

$$R_F(u, v, I, I_{uv}) = -\frac{uI_{uv}(u, v)dudv}{1 + uI(u, v)} - \frac{(v + u)I_{uv}(u, v)dudv}{1 + (v + u)I(u, v)}. \tag{B.11}$$

Appendix C

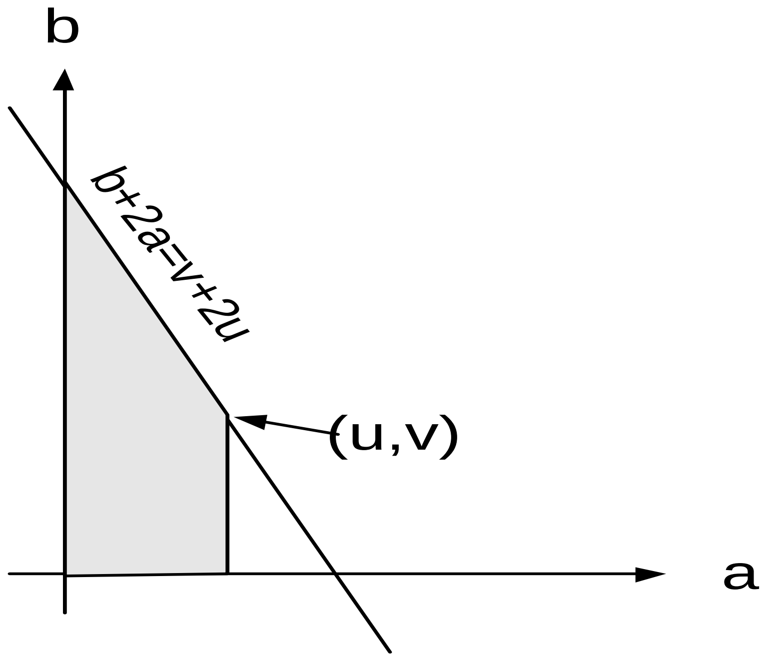

Fig. 12. The majorized area (dotted) for the selection of indices $a = \lambda_2$ and $b = \lambda_1 - \lambda_2$, where $(\lambda_1 \geq \lambda_2)$.

K-User MAC-Broadcast

The two user MAC-Broadcast approach extends straightforwardly to the general M-user case, with N receive antennas. Let now $m = 1, 2, ..., M$ in (75), where $\rho_m(s)$ stands for the power distribution of the m-th user - all subjected to the same average power constraint, SNR. The strategy induces a set of M-nonlinear equations of the MMSE SIR of each user,

$$SIR_m = \frac{1}{[1 + HQ_mH^\dagger]^{-1}_{m,m}} - 1 \tag{C.1}$$

where H is the $N \times M$ channel propagation matrix, $Q_m \triangleq diag\{I_1(\omega_1), I_2(\omega_2), .., P_m, ..., I_M(\omega_M)\}$, and P_m is the original transmit power of user m. The normalized SIR of user m, is simply a normalization of the SIR by the transmit power of user m,

$$\omega_m = SIR_m/P_m = \frac{1}{P_m[1 + HQ_mH^\dagger]^{-1}_{m,m}} - 1/P_m \tag{C.2}$$

which replaces (77). The achievable rates associated with user m, and the average achievable rates for each user can be directly taken from (78)-(80), using the new normalized SIRs from (C.2) above.

The total expected throughput is then

$$R_{mac-bs,T} = \sum_{m=1}^{M} R_{mac-bs,m}. \tag{C.3}$$

The extension of the transmit power density optimization algorithm to the M-user setting is also straightforward.

References

[1] E. Biglieri, J. Proakis, and S. Shamai (Shitz), "Fading channels: Information theoretic and communication aspects," *IEEE Transactions on Information Theory*, vol. 44, pp. 2619–2692, October 1998.

[2] L. Ozarow, S. Shamai (Shitz), and A. Wyner, "Information theoretic considerations for cellular mobile radio," *IEEE Trans. Veh. Tech.*, vol. 43, pp. 359–378, May 1994.

[3] T. Cover, "Broadcast channels," *IEEE Transactions on Information Theory*, vol. 18, pp. 2–14, January 1972.

[4] W. H. R. Equitz and T. M. Cover, "Successive refinement of information," *IEEE Trans. Inform. Theory*, vol. 37, pp. 269–279, April 1991.

[5] B. Rimoldi, "Successive refinement of information: Characterization of the achievable rates," *IEEE Trans. Inform. Theory*, vol. 40, pp. 253–259, January 1994.

[6] M. Trott, "Unequal error protection codes: Theory and practice," *IEEE Inform. Theory Workshop*, p. 11, June 9-13 1996.

[7] M. R. S. S. Boucheron, "About priority encoding transmission," *IEEE Trans. Inform. Theory*, vol. 46, pp. 609–705, March 2000.

[8] M. R. S. S. Boucheron, "Priority encoding transmission," *IEEE Trans. Inform. Theory*, vol. 42, pp. 1737–1744, November 1996.

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

[10] D. Tse, "Optimal power allocation over parallel Gaussian broadcast channels.," *Proceedings of International Symposium on Information, Ulm Germany, http://degas.eecs.berkeley.edu/d̃tse/pub.html*, p. 27, June 1997.

[11] L. Li and A. Goldsmith, "Capacity and optimal resource allocation for fading broadcast channels: Part I: Ergodic capacity and part II: Outage capacity," *IEEE Trans. Inform. Theory*, vol. 47, pp. 1083–1127, March 2001.

[12] P. Viswanath and D. Tse, "Sum capacity of the multiple antenna broadcast channel," *Proceedings of the IEEE International Symposium on Information Theory, Lausanne, Switzerland*, p. 497, June 30-July 5 2002.

[13] S. Vishwanath, N. Jindal, and A. Goldsmith, "Duality, achievable rates and sum-rate capacity of Gaussian mimo broadcast channel," *submitted to IEEE Trans. on Information Theory*, August 2002.

[14] G. Kramer, S. Vishwanath, S. Shamai (Shitz), and A. Goldsmith, "Information-theoretic issues concerning broadcasting," *Proc. Workshop on Sig. proc. for wireless comm. (DIAMACS 2002)*, Rutgers University, NJ, October 2002.

[15] G. Caire and S. Shamai (Shitz), "On the achievable throughput of a multi-antenna Gaussian broadcast channel," *to appear in IEEE Trans. on Information Theory*, submitted Novermber 2002.

[16] W. Yu and J. Cioffi, "The sum capacity of a Gaussian vector broadcast channel," *Proceedings of the IEEE International Symposium on Information Theory, Lausanne, Switzerland*, p. 498, June 30-July 5 2002.

[17] A. E. Gamal, "Capacity of the product and sum of two unmatched broadcast channels," *Prob. Peredachi, Infor.*, vol. 16, pp. 3–23, January-March 1980.

[18] S. Sesia, G. Caire, and G. Vivier, "Broadcasting a common source: Information thoeretic results and system challenges," *Proceedings of the IEEE International Symposium on Information Theory, Monte Verita, Switzerland, http://www.isi.ee.ethz.ch/winterschool/*, February 24-27 2003.

[19] N. Shulman and M. Feder, "Source broadcasting with unknown amount of receiver side information," *Proceedings of ITW2002, Bangalore, India*, October 20-25 2002.

[20] P. Schramn, "Multilevel coding with independent decoding on levels for efficient communications on static and interleaved fading channels," *Proceedings of IEEE PIMRC'97*, pp. 1196–1200, Helsinki, Findland 1997.

[21] D. Schill and J. Huber, "On hierarchical signal constellations for the Gaussian broadcast channel," *Proc. International Conference on Telecommunications (ICT'98), Porto Carras, Greece*, pp. 34–38, 21-25, June 1998.

[22] D. Schill, D. Yuan, and J. Huber, "Efficient broadcasting using multilevel codes," *IEEE Inform. Theory Workshop (ITW'99), Metsovo, Greece*, p. 72, June 27 - July 1 1999.

[23] M. Sajadieh, F. R. Kschischang, and A. Leon-Garcia, "Analysis of two-layered adaptive transmission systems," *IEEE, 46th Vehicular Technology Conference (VTC'96), Atlanta, Georgia*, pp. 1771–1775, April 28-May 1 1996.

[24] Y. Liu, K. Lau, C. Takeshita, and M. Fitz, "Optimal rate allocation for superposition coding in quasi-static fading channels," *IEEE Int. Symp. Inform. Theory (ISIT'02), Lausanne, Switzerland*, p. 111, June 30 - July 5 2002.

[25] S. Shamai (Shitz), "A broadcast strategy for the Gaussian slowly fading channel," *IEEE Int. Symp. Inform. Theory (ISIT'97), Ulm Germany*, p. 150, June 29–July 4 1997.

[26] S. Shamai (Shitz), "A broadcast approach for the multiple-access slow fading channel," *IEEE Int. Symp. Inform. Theory (ISIT'00), Sorrento, Italy*, p. 128, June 25-30 2000.

[27] I. Geldfand and S. Fomin, *Calculus of Variations.* Prentice Hall, New York, 1963.

[28] E. Telatar, "Capacity of multi-antenna Gaussian channels," *European Transactions on Telecommunications*, vol. 10, pp. 585–595, November 1999.

[29] A. Edelman, "Eigenvalues and condition numbers of random matrices," *PhD thesis dissertation*, May 1989.

[30] A. Marshall and I. Olkin, *Inequalities: Theory of Majorization and Its Applications.* Academic Press, New York, 1979.

[31] Y. Liu, K.N.Lau, and T. Chen, "An iterative siso strategy for mimo broadcast channels," *to appear in the IEEE International Symposium on Information Theory*, June 2003.

[32] S. Shamai (Shitz) and S. Verdu, "Decoding only the strongest CDMA users," *Codes Graphs and Systems a Celebration of the Life and Career of G.D. Forney Jr. on the occasion of the Sixtieth Birthday.*, pp. 217–228, 2002.

[33] S. Verdú, *Multiuser Detection.* New-York: Cambridge Univ. Press, 1998.

[34] S. Verdú and S. Shamai (Shitz), "Spectral efficiency of cdma with random spreading," *IEEE Transactions on Information Theory*, vol. 45, pp. 622–640, March 1999.

[35] T. Marzetta, B. Hochwald, and B. Hassibi, “Space-time autocoding: Arbitrary reliable communication in a single fading interval,” *IEEE Transactions on Information Theory*, vol. 47, pp. 2761–2781, November 2001.

[36] B. Rimoldi and R. Urbanke, “A rate splitting approach to the gaussian multiple-access channel,” *IEEE Transactions on Information Theory*, vol. 42, pp. 364–375, March 1996.

[37] B. Rimoldi, “Rdma for multiple access channels: An optimum asynchronous low complexity technique, st. peterburg, russia,” *7th Joint Swedish-Russian Int. Work. Inform. Theory*, pp. 196–199, June 1995.

[38] T. Cover, “An achievable rate region for the broadcast channel,” *IEEE Transactions on Information Theory*, vol. 21, pp. 399–404, July 1975.

[39] T. Cover, “Comments on broadcast channels,” *IEEE Transactions on Information Theory*, vol. 44, pp. 2524–2530, October 1998.

[40] B. Hajek and M. Pursley, “Evaluation of an achievable rate region for the broadcast channel,” *IEEE Transactions on Information Theory*, vol. 25, pp. 36–46, January 1979.

[41] A. J. Viterbi, “Very low rate conventional codes for maximum theoretical performance of spread-spectrum multiple-access channels,” *IEEE Journal on Selected Areas in Communications*, vol. 8, pp. 641–649, May 1990.

[42] A. Seeger, “Hierarchical channel coding for Rayleigh and free fading,” *Globecom ’97, Phoenix, Arizona, USA*, pp. 208–212, November 3-8 1997.

[43] M. Sajadieh, F. Kschischang, and A. Leon-Garcia, “Analysis of two-layered adaptive transmission systems,” *IEEE 46th Vehicular Technical Conference*, pp. 1771–1775, April 28 - May 1 1996.

[44] P. Whiting and E. Telatar, “Coding schemes for slowly varying channels,” *Bell-Laboratories, Lucent, Technical Memo, in preparation.*

[45] A. Lozano, J. Kovacevic, and M. Andrews, “Quantized frame expansions in a wireless environment,” *Bell-Laboratories, Lucent, Technical Memo, in preparation.*

[46] R. Gallager, “An inequality on the capacity region of multiple access multipath channels,” *Communications and Cryptography: Two Sides of One Tapestry, Boston, MA: Kluwer*, pp. 129–139, January-March 1994.

[47] J. Lim and S. Gelfand, “Performance analysis of hirarchical coded modulation system,” *Globecom’00*, 27 November - 1 December 2000.

[48] A. Seeger, “Broadcast communication on fading channels using hierarchical coded modulation,” *Globecom’00*, 27 November - 1 December 2000.

[49] M. Speth and A. Seeger, “Wireless transmission using hierarchical modulationand coding: Implications for system design,” *3rd ITG Conference on Source and Channel Coding, Munich, Germany*, pp. 193–198, 1-19 January 2000.

[50] S. Gadkari and K. Rose, “Embedded codes for combined source-channel coding,” *34th Annual Allerton Conference on Communication, Control and Computing, Allerton House, Monticello. Illinois*, pp. 74–83, October 2-4 1996.

[51] G. Caire and S. Shamai (Shitz), “On the achievable throughput of a multi-antenna Gaussian broadcast channel,” *38th Annual Allerton Conference on Communication, Control and Computing, Allerton House, Monticello. Illinois October 4-6, 2000, and the Proceedings of the IEEE International Symposium on Information Theory*, p. 147, June 24-29 2001.

[52] S. V. N. Jindal and A. Goldsmith, “Duality, dirty paper coding, and capacity for multiuser wireless channels,”

International Symposium on Information Theory, July 2002, and in Information, Coding and Mathematics, proceedings of Workshop honoring Prof. Bob McEliece on his 60th Birthday Eidotrs (M. Blaum, P.C. Farell and H.C.A. van Tilborg), Kluwer Academic Press, London 2002, ISBN: 1-4020-7079-9 See also: S. Vishwanath, N. Jindal, and A. Goldsmith, "On the Capacity of Multiple Input Multiple Output Broadcast Channels," ICC 2002., pp. 239–256.

[53] W. H. R. Equitz and T. M. Cover, "Successive refinement of information, attendum," *IEEE Trans. Inform. Theory*, vol. 39, pp. 1465–1466, July 1993.

[54] G. H. Golub and C. F. van Loan, *Matrix Computations*. 1989 The John Hopkins University Press, 2 ed., 1990.