MULTIUSER DETECTION FOR VECTOR MULTIPLE ACCESS CHANNELS USING GENERALIZED DECISION FEEDBACK EQUALIZATION

Wei Yu and John M. Cioffi

STAR Lab, Electrical Engineering Department, Packard Building Room 360
Stanford University, Stanford, CA 94305-9515, U.S.A.
Phone: 1-650-723-2525. FAX: 1-650-723-9251
email: {weiyu,cioffi}@stanford.edu

ABSTRACT

The multiuser detection problem for vector Gaussian multiple access channel with multiple inputs and multiple outputs (MIMO) is considered. It is shown that the capacity region for the multiple access channel may be characterized and the optimal multiuser transmitter spectrum may be obtained by solving a convex programming problem. With the optimized transmitter, a practical bit-loading and multiuser detection scheme based on generalized decision feedback equalizer (GDFE) is proposed. This GDFE-based detection scheme achieves the extreme points in the multiple access channel capacity region.

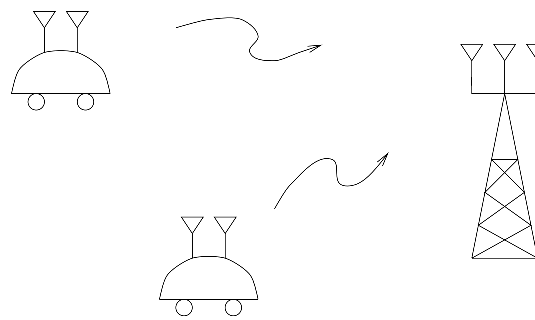

Fig. 1. A wireless channel with antenna arrays

I. INTRODUCTION

A fundamental result in multiuser information theory is the characterization of the capacity region for multiple access channels. In a multiple access channel, multiple uncoordinated transmitters can potentially communicate with a single receiver at the same total data rate as if they coordinate [1]. However, although the capacity region for multiple access channels is well-understood in terms of mutual information, its numerical characterization is not necessarily easy, nor methods to achieve the capacity necessarily apparent. Such is the case for channels whose inputs and outputs are multi-dimensional. In vector channels, the characterization of capacity region involves transmitter optimization, which, unlike the single user "waterfilling", is non-trivial. In this direction, [2] solved a special case where the multiple access channel has ISI. The general case was recently solved in [3], where it was shown that the optimal transmitter spectrum can be found using convex programming methods. With an optimal transmitter spectrum, it is then possible to design optimal coding/decoding schemes. The goal of this paper is to propose one such practical bit-loading and multiuser detection method based on decision feedback equalizer that achieves multiuser capacity.

This work was supported by a Stanford Graduate Fellowship, and in part by a gift from France Telecom.

From an information theory perspective, multiuser capacity is achieved with superposition and interference cancellation. So, it is natural to consider decision feedback detection schemes. Decision feedback equalizers (DFE) have long been used to mitigate intersymbol interference in bandlimited channels. Its generalization to multi-input multi-output (MIMO) systems was first proposed in [4], where a MIMO DFE based on the minimum mean square error (MMSE) criterion was described. Such schemes are also found in the CDMA literature [5], where interference cancellation is performed on a user-by-user basis. Recently, [6] found that this decision feedback structure, in addition to minimizing MSE, also maximizes sum capacity. As we shall soon see, this is not an accident. In this paper, we will develop a generalization of the decision feedback multiuser detection for vector channels which also minimizes MSE and achieves the extreme points in channel capacity at the same time. We choose a derivation based on the generalized decision feedback equalizer (GDFE) [7] that shows the intimate relation between MMSE and capacity-maximizing criteria. As one might expect, the optimal detector depends on the optimal transmitter spectrum.

The vector multiple access channels appear in many practical situations. In the wireless context, recent advances in antenna array technology are promising many advantages such as diversity and increased channel capacity for mobile

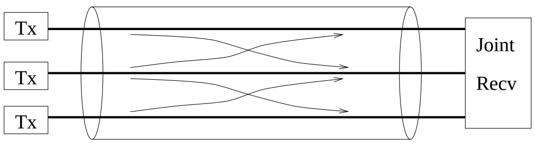

Fig. 2. Vector multiaccess channels in xDSL

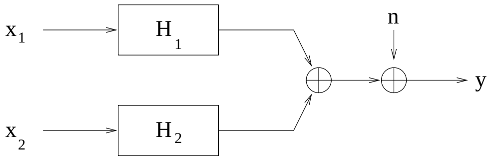

Fig. 3. A multiple access channel

networks (Fig. 1). Since each mobile transmits via a multiple antenna array, the upstream transmission scenario is a multiple access channel with vector input. Because the synchronization and perfect channel side information assumptions made in this paper, the methods proposed here are most applicable to the fixed wireless situation.

A similar situation exists for wireline access technology such as Digital Subscriber Lines (DSL) (Fig. 2). DSL is the local loop technology which brings high speed data communication to home via ordinary phone lines. The predominant noise source in copper wires is interference from adjacent lines. If coordination at the central office is feasible, the upstream transmission becomes a multiple access channel. With multiuser detection schemes, crosstalk interference is no longer regarded as noise, and dramatic data rate improvement can be achieved [8].

In the rest of the paper, we first review the transmitter optimization results for vector multiple access channels in section II. Then in section III, we propose a multiuser bitloading and a multiuser detection scheme based on GDFE, and prove that it achieves multiuser capacity. Finally, practical issues are discussed and conclusions are drawn in section IV. Throughout this paper, we make tactic assumptions that channel state information is perfectly known at both transmitter and receiver and synchronization can be achieved among the users.

II. TRANSMITTER OPTIMIZATION IN VECTOR MULTIPLE ACCESS CHANNELS

A multiple access channel is depicted in Fig. 3, where x_1, x_2 are input vector signals, y is the output vector signal, n is the vector additive Gaussian noise, H_1 and H_2 are channel responses represented as matrices so that

$$Y = H_1 x_1 + H_2 x_2 + n. \tag{1}$$

Although we have restricted ourselves to the two-user case, the development here can easily be generalized to more than two users. We have also restricted ourselves to memoryless models because channel with memory can be converted to memoryless models using block-by-block processing and guard periods. The objective is to find capacity region for this channel subject to power constraints P_1, P_2 respectively on x_1 and x_2.

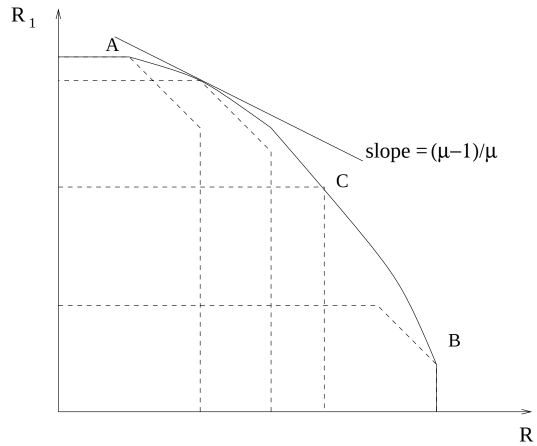

Fig. 4. Transmitter optimization in multiple access channels

The capacity region for general memoryless multiple access channels is well-known [1]. For Gaussian channels with a specific input power spectral density, the capacity region is a pentagon with one side at 45 degree slope,

$$\begin{aligned} R_1 &\leq I(x_1; y|x_2), & (2)\\ R_2 &\leq I(x_2; y|x_1), & (3)\\ R_1 + R_2 &\leq I(x_1, x_2; y). & (4) \end{aligned}$$

Since there are more than one input power spectrum density satisfying the same power constraint, the capacity region for vector Gaussian multiple access channel with power constraints is the union of pentagons over all such power spectrum densities, (Fig. 4):

$$C = \bigcup_{\substack{tr(\Sigma_1) \leq P_1,\\ tr(\Sigma_2) \leq P_2,\\ \Sigma_1, \Sigma_2 \geq 0}} \left\{ (R_1, R_2) : \begin{array}{r} R_1 \leq I(x_1; y|x_2); \\ R_2 \leq I(x_2; y|x_1); \\ R_1 + R_2 \leq I(x_1, x_2; y). \end{array} \right\} \tag{5}$$

where Σ_1 and Σ_2 are covariance matrices for input signals x_1 and x_2 respectively, (i.e., $E[x_1 \cdot x_1^*] = \Sigma_1$). The capacity region is a convex region whose boundary points can be characterized by numerically maximizing $\mu R_1 + (1-\mu) R_2$ subject to power constraints. The point (R_1, R_2) at which $\mu R_1 + (1-\mu) R_2$ is maximized is at the boundary of the capacity region where the tangent line has slope $(\mu - 1)/\mu$. Varying μ from 0 to 1 would trace through all boundary points. The following theorem makes the optimization process explicit.

Theorem 1: The capacity region for the Gaussian vector multiple access channel (1) is a convex region whose boundary points are characterized by the solution to the following set of optimization problems indexed by parameter μ ranged between 0 and 1. As $0.5 \le \mu \le 1$, the optimization problem is

$$\begin{array}{ll} \max & (1-\mu)\cdot\frac{1}{2}\log|H_1\Sigma_1H_1^*+H_2\Sigma_2H_2^*+Z| + \\ & (2\mu-1)\cdot\frac{1}{2}\log|H_1\Sigma_1H_1^*+Z| - \mu\cdot\frac{1}{2}\log|Z| \\ s.t. & tr(\Sigma_1)\le P_1, \\ & tr(\Sigma_2)\le P_2, \\ & \Sigma_1,\Sigma_2 \ge 0. \end{array} \quad (6)$$

As $0 \le \mu \le 0.5$, the optimization problem is

$$\begin{array}{ll} \max & \mu\cdot\frac{1}{2}\log|H_1\Sigma_1H_1^*+H_2\Sigma_2H_2^*+Z| + \\ & (1-2\mu)\cdot\frac{1}{2}\log|H_2\Sigma_2H_2^*+Z| - \mu\cdot\frac{1}{2}\log|Z| \\ s.t. & tr(\Sigma_1)\le P_1, \\ & tr(\Sigma_2)\le P_2, \\ & \Sigma_1,\Sigma_2 \ge 0. \end{array} \quad (7)$$

For each μ, the optimizing Σ_1 and Σ_2 are the covariance matrices for x_1 and x_2 corresponding to the boundary point of the capacity region where $\mu R_1+(1-\mu)R_2$ is maximized. Further, the optimization problem is concave in (Σ_1,Σ_2) over the set of positive semidefinite matrices pairs.

The detailed derivation of this theorem can be found in [3]. The critical observation here is that the optimization problem is concave. This follows from the fact that $\log\det(M)$ is a concave function of M in the cone of positive semi-definite matrices [9]. The composition of a concave function and a linear function is concave, so the objective is a concave function in (Σ_1,Σ_2). In addition, trace constraints are linear, therefore, the entire problem can be posed as a convex programming problem. The convexity observation is crucial because it guarantees that efficient search algorithms exist. In fact, the classical waterfilling and the recent multiuser waterfilling algorithm in [2] are particular examples of such convex optimization routines. In the particular case where $\mu = 0.5$, the problem degenerates into the $maxdet$ problem that is well studied in the field of convex optimization and for which efficient software package already exists [9]. Convex optimization algorithms are always iterative in nature, and their convergence to the global maximum is guaranteed.

III. GENERALIZED DECISION FEEDBACK EQUALIZER

Although decision feedback equalizer has long been used to combat intersymbol interference, it was not until fairly recently that minimum mean square error decision feedback equalizer (MMSE-DFE) was shown to achieve channel capacity in single-user ISI channels [10]. Traditionally, the MMSE-DFE structure includes linear time invariant filters, which operate on the entire input signal. The finite length block version was first proposed in [7], and it is called Generalized Decision Feedback Equalizer (GDFE). GDFE is not only optimal in the MMSE sense, it also achieves the mutual information between the input signal and output signal. In this section, we aim to use the same GDFE structure for multiple access channels. We will give a derivation of a multiuser decision feedback detector which simultaneously minimizes MSE and maximizes total capacity. Toward this end, we will first develop GDFE for single user, taking a slightly different route as in [7].

First, we note that information source is i.i.d. To obtain the necessary input covariance, the information source need to be shaped by a pulse filter. Such pulse shaping filter may be absorbed into the channel. So, without loss of generality, we aim to find practical coding/decoding schemes that achieve the mutual information with a white input.

The capacity region for multiple access channels with a fixed input covariance is a pentagon, i.e., the total data rate is bounded by $I(x_1,x_2;y)$. Therefore, in (1), x_1 and x_2 may be considered jointly. Define the input vector $x=[x_1x_2]^T$, and the channel response $H=[H_1H_2]$, we have,

$$y = Hx + n. \quad (8)$$

For simplicity, assume that $E[nn^*] = I$, i.e. noise is prewhitened. The goal becomes to achieve single user capacity $I(x;y)$. From Shannon's original channel capacity argument, we know that with a randomly generated code and a typicality or maximum likelihood decoder, data transmission at rates below the maximum mutual information has probability of error exponentially approaching zero. However, randomly choosing codewords from Gaussian ensembles is not practical, and typicality decoder or maximum likelihood decoder involves searching through multidimensional space which is computationally intensive. The aim is, therefore, to find practical ways to achieve $I(x;y)$, preferably on a symbol-by-symbol basis. Generalized decision feedback equalization (GDFE) is one such practical method that achieves capacity.

The development of GDFE involves three key ideas. The first idea is the following observation made in [10]:

$$I(y;x) = H(y) - H(y|x) = \frac{1}{2}\log\frac{|\Sigma_{yy}|}{|\Sigma_{y|x}|}, \quad (9)$$

$$I(x;y) = H(x) - H(x|y) = \frac{1}{2}\log\frac{|\Sigma_{xx}|}{|\Sigma_{x|y}|}. \quad (10)$$

The equivalence of $I(x;y)$ and $I(y;x)$ shows that mutual information may be achieved in two different ways. In the forward direction (9), the input signal is x, the output signal is y, and $\Sigma_{y|x}$ is just the covariance matrix of additive noise. In the backward direction (10), y is considered as the input signal, x the output signal, and the error signal e

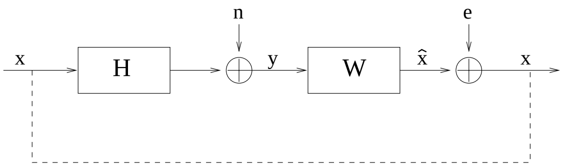

Fig. 5. Forward and backward channels with MMSE estimation

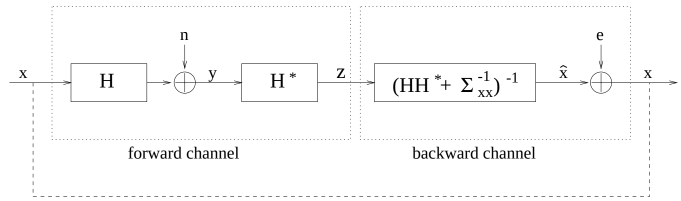

Fig. 6. Forward and backward canonical channels

is the MMSE linear estimation error of x based on y. The relationship between the forward and backward directions is illustrated in Fig. 5, where W is the MMSE linear estimation filter of x given y. The fact that the backward channel associated with minimum-mean-square-error estimation has the same channel capacity as the original channel suggests that with an MMSE filter, detecting x from $\hat{x}$ achieves the same channel capacity as detecting x from y. Therefore, an MMSE estimator is also necessarily capacity achieving. Note, at this point, we are still doing block-based processing. Component-wise decision on $\hat{x}$ directly is not the right thing to do because the MMSE error is not necessarily white, hence such detection scheme is not optimal in the sense that it does not achieve the channel capacity $\frac{1}{2}\log(|\Sigma_{xx}|/|\Sigma_{ee}|)$. Our goal is to use decision feedback to allow component-wise decision while not increasing the mean square error or decreasing mutual information. As the following development shows, such is possible with GDFE.

To facilitate component-wise detection, we need to whiten the MMSE error e, while preserving the "information content" in $\hat{x}$. Toward this end, we break up MMSE filter W in Fig. 5 into two components, creating two so-called canonical channels. The notion of canonical channel is the second key idea in GDFE. Canonical channel is useful in that it makes symbol-by-symbol detection possible, as will be apparent shortly. Let us first write down the MMSE filter W explicitly,

$$\begin{aligned} W &= \Sigma_{xy}\Sigma_{yy}^{-1} && (11)\\ &= \Sigma_{xx}H^*(H\Sigma_{xx}H^*+I)^{-1} && (12)\\ &= (H^*H+\Sigma_{xx}^{-1})^{-1}H^*, && (13) \end{aligned}$$

where (11) follows from the standard linear estimation theory, (13) follows from the matrix inversion lemma [11], which states,

$$(A+BCD)^{-1} = A^{-1} - A^{-1}B(C^{-1}+DA^{-1}B)^{-1}DA^{-1}. \quad (14)$$

Now, it is clear that W may be split into a matched filter H^*, and an estimation filter, as shown in Fig. 6. The channel from x to z is the forward canonical channel,

$$z = H^*Hx + H^*n = R_f x + n'. \quad (15)$$

The name canonical channel comes from the fact that the covariance of noise n' is the same as the channel R_f. Interestingly and perhaps surprisingly, the backward channel

$$x = (H^*H+\Sigma_{xx}^{-1})^{-1}z + e = R_b z + e \quad (16)$$

is also canonical. The following computation verifies this fact,

$$\begin{aligned} E[ee^*] &= E[(x-\hat{x})(x-\hat{x})^*]\\ &= E[(x-\Sigma_{xy}\Sigma_{yy}^{-1}y)(x-\Sigma_{xy}\Sigma_{yy}^{-1}y)^*]\\ &= \Sigma_{xx} - \Sigma_{xx}H^*(H\Sigma_{xx}H^*+I)^{-1}H\Sigma_{xx}\\ &= (H^*H+\Sigma_{xx}^{-1})^{-1}\\ &= R_b, && (17) \end{aligned}$$

where we had again used the matrix inversion lemma (14).

Recall that in order to perform optimal detection, we need to whiten the MMSE error. The third key idea in GDFE is to recognize that whitening may be done with decision feedback via Cholesky factorization. Since decision feedback is causal, the unique causal whitening filter (up to scaling) is the Cholesky factor of the error covariance, which by construction, is just R_b. Because in a canonical channel, the noise variance and the channel response are the same, the noise whitening process may be accomplished by splitting the filter R_b into a feedback configuration. Let $R_b = G^{-1}S^{-1}G^{-*}$, where S is diagonal, G is an upper triangular matrix with 1's on the diagonal. Then

$$\begin{aligned} x &= R_b z + e && (18)\\ x &= G^{-1}S^{-1}G^{-*}z + e && (19)\\ Gx &= S^{-1}G^{-*}z + Ge, && (20) \end{aligned}$$

which suggests the successive decoding structure in Fig. 7. Because G is upper triangular, x_n can be decoded first, then with its interference subtracted, x_{n-1} may be decoded, etc. In this structure, noise prior to the decision device is just Ge, which is white,

$$E[Ge(Ge)^*] = GR_bG^* = S^{-1}, \quad (21)$$

and has the same means square error as an MMSE estimator:

$$|S^{-1}| = \prod S_i^{-1} = R_b = |\Sigma_{ee}|. \quad (22)$$

Note that the minimum mean square error here is taken as the determinant of the covariance matrix instead of the usual trace. These two criteria are equivalent [11].

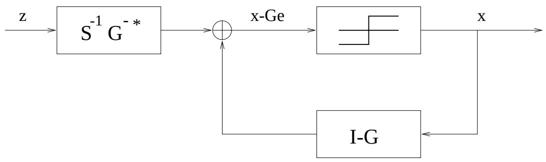

Fig. 7. Generalized decision feedback equalizer

Now, the noise has been whitened, and input x is also white by assumption, so we have converted the vector channel into a set of independent AWGN sub-channels, each with a capacity $\frac{1}{2}\log(SNR_i) = \frac{1}{2}\log(\mathcal{E}_i S_i)$, where $\mathcal{E}_i$ is the energy of the input signal in ith component. The SNR here is so-called biased SNR, which is the ratio of output signal energy to noise energy. The capacity of the vector channel is the sum of capacities of each individual sub-channels. So,

$$C = \frac{1}{2}\log\prod_i \mathcal{E}_i S_i \tag{23}$$

$$= \frac{1}{2}\log\frac{|\Sigma_{xx}|}{|\Sigma_{ee}|}, \tag{24}$$

which is the capacity of the original channel. Therefore, GDFE is able to convert a vector channel with white input into a set of parallel AWGN sub-channels in such a way that the mean-square error is the same as an MMSE estimator, and the sum capacity of the parallel channels is exactly same as the original vector channel capacity.

IV. GDFE FOR MULTIUSER DETECTION

GDFE works with vector channels with white input where components of the input vector are not coordinated. Without coordination, the single user vector channel is identical to a multiple access channel with one user at each vector component. So, it is natural to expect GDFE to achieve multiuser channel capacity also when each user has only one input dimension. This is the case we will consider first.

The capacity region for multiple access channel with scalar inputs is precisely a pentagon. Because GDFE achieve the total capacity and every point on the 45 degree slope of the pentagon has the same total capacity, the question is therefore, which point in the capacity pentagon is achieved by GDFE, or in other words, how data rates are divided among the users. To calculate individual data rates from GDFE, it is necessary to go into the details of Cholesky factorization:

$$R_b^{-1} = H^*H + \Sigma_{xx}^{-1} = G^*SG. \tag{25}$$

Let $H = [h_1 h_2 ... h_m]$, $\Sigma_{xx} = diag\{\mathcal{E}_1, \mathcal{E}_2, ...\mathcal{E}_m\}$. Then, $H^*H + \Sigma_{xx}^{-1}$ equals:

$$\begin{bmatrix} h_1^* \\ \vdots \\ h_m^* \end{bmatrix} \begin{bmatrix} h_1 \cdots h_m \end{bmatrix} + \begin{bmatrix} \frac{1}{\mathcal{E}_1} & & \\ & \ddots & \\ & & \frac{1}{\mathcal{E}_m} \end{bmatrix}. \tag{26}$$

The Cholesky factorization into G^*SG can be obtained by comparing coefficients. The process yields a unique solution, and the diagonal terms of S are obtained successively. Let

$$G = \begin{bmatrix} 1 & g_{11} & g_{21} & \cdots \\ & 1 & g_{22} & \\ & & 1 & \\ \vdots & & & \ddots \end{bmatrix}, \tag{27}$$

$S = diag\{S_1, S_2, ..., S_n\}$. Then G^*SG is:

$$\begin{bmatrix} S_1 & g_{11}S_1 & g_{21}S_1 & \cdots \\ g_{11}S_1 & g_{11}^2S_1 + S_2 & g_{21}g_{11}S_1 + g_{22}S_2 & \\ g_{21}S_1 & g_{21}g_{11}S_1 + g_{22}S_2 & g_{21}^2S_1 + g_{22}^2S_2 + S_3 & \\ \vdots & & & \ddots \end{bmatrix}. \tag{28}$$

To find the data rate on each AWGN sub-channel, it is only necessary to obtain the sub-channel SNR, which is just $\mathcal{E}_i S_i$. Comparing (26) with (28),

$$S_1 = ||h_1||^2 + \frac{1}{\mathcal{E}_1}. \tag{29}$$

So, the channel capacity of the first sub-channel is,

$$C_1 = \frac{1}{2}\log(\mathcal{E}_1 S_1) = \frac{1}{2}\log(1 + \mathcal{E}_1||h_1||^2). \tag{30}$$

This is the sub-channel capacity as if there is no interference from other sub-channels. (Note, the noise variance is scaled to $\sigma^2 = 1$.) The second sub-channel capacity is calculated from S_2:

$$\begin{aligned} C_2 &= \frac{1}{2}\log(\mathcal{E}_2 S_2) \\ &= \frac{1}{2}\log[(1 + \mathcal{E}_1||h_1||^2 + \mathcal{E}_2||h_2||^2 - \\ &\quad \mathcal{E}_1\mathcal{E}_2(||h_1||^2||h_2||^2 - (h_1 \cdot h_2)^2))/ \\ &\quad (1 + \mathcal{E}_1||h_1||^2)]. \end{aligned} \tag{31}$$

We claim that (31) is precisely the capacity of the second sub-channel when signal from the first channel is regarded as noise and signals from all other channels are perfectly cancelled. To see this, notice that output covariance due to x_i in the ith sub-channel is $E[(h_i x_i)(h_i x_i)^*] = \mathcal{E}_i h_i h_i^*$. So, if the output term includes contributions from x_1, x_2 and additive Gaussian noise n, and noise term includes n

and contribution from x_1, then the capacity of the second sub-channel can be computed as:

$$\begin{aligned} C_2' &= I(x_2; y|x_3 \cdots x_m) \\ &= \frac{1}{2}\log\frac{|\Sigma_{yy}|}{|\Sigma_{nn}|} \\ &= \frac{1}{2}\log\frac{|I+\mathcal{E}_1 h_1 h_1^* + \mathcal{E}_2 h_2 h_2^*|}{|I+\mathcal{E}_1 h_1 h_1^*|}. \end{aligned} \quad (32)$$

(31) and (32) are equal, which can be shown using the matrix identity $\det(I+XY) = \det(I+YX)$, thus verifying our claim. In fact, this argument may be repeated for every sub-channel, and it becomes apparent that the following theorem holds:

Theorem 2: For a vector multiple access channel defined by $y = Hx + n$, where $x \in R^m$ has a diagonal covariance matrix, and each component of vector x is interpreted as a single user, the GDFE structure is a capacity-achieving multiuser detector in the sense that the sub-channel for user i has its capacity equal to $C_i = I(x_i; y|x_{i+1}, ..., x_m)$, achieving a corner point in the multiple access channel capacity polyhedron. Further, components of the vector x may be ordered in $m!$ ways, achieving all $m!$ corners.

In multiuser information theory, the corner points of the multiple access channel capacity polyhedron are achieved with superposition and interference cancellation. We see that GDFE is precisely the practical method which implements interference cancellation.

Thus far, we have treated $y = Hx + n$ as a m-user multiple access channel where each component of x is a single user. In the original model, $x = [x_1 x_2]$, where x_1, x_2 are input vector signals corresponding to two users. GDFE does not make a distinction between two users with vector inputs and m users with scalar inputs. By the chain rule of conditional mutual information, the corner point of m-dimensional polyhedron corresponding to the ordering $x = [x_1 x_2]$, when combined into two users, collapses into the corner point of two-dimensional pentagon.

In a real systems design, corner points may not always be the desirable operating points. For example, it may be of interest to achieve the point with equal rates for all users. From information theory point of view, corner points are the extreme points of the convex rate region. Any other points may be achieved with time-sharing. But, with GDFE on vector channels, there is enough versatility to approach many of the middle points on the pentagon without time-sharing. The trick is to re-order the components of the vector $x = [x_1 x_2]$. The vector components may be shuffled, i.e. a vector component for user 1 is decoded first, then a component for user 2 is decoded, then back to 1, etc. Each different shuffling achieves a different middle point. In addition, even in the case where the user input signal is a scalar, artificial vector components may be created making shuffling possible. For example, the system $y = x_1 + x_2$ may be re-written as $y = x_{11} + x_{12} + x_{21} + x_{22}$, which now has two components for each user. The trade-off is that the GDFE structure becomes more complex, and it also worsens the error propagation effect. Nevertheless, achieving only the corner points is not a limitation of GDFE. Once the corner points of the multiaccess capacity region is achieved, all middle points can be easily dealt with.

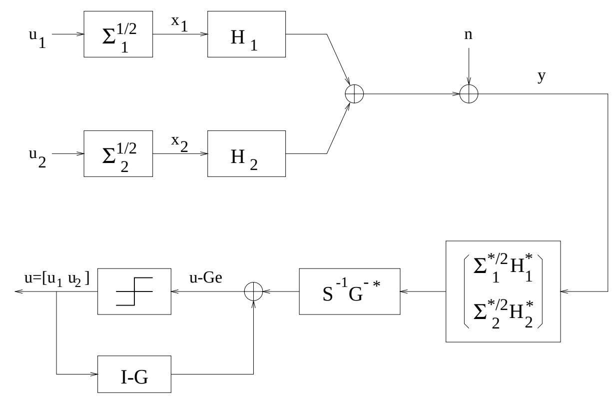

Fig. 8. A multiuser bit-loading and detection scheme

We now put everything together and present a multiuser detection scheme based on GDFE. The first task is to find the multiuser channel capacity region and the optimal input covariance matrices corresponding to the desirable point in the capacity region by solving for Σ_i in the convex optimization problems (6) and (7). The covariance matrices acts as pulse shaper to the information source and it is absorbed into the channel matrix as shown in Fig. 8. The information sources are assumed to have unit covariance matrices.

Next, determine the appropriate ordering among the components of the two users in GDFE. If the chosen rate point is a corner point, then ordering is automatic. The higher priority user occupies the lower indices in u. If the chosen rate point is a middle point, it would be necessary to shuffle the indices of the two users in order to find a point close to the desirable rate trade-off. Let the permutation matrix corresponding to the ordering be P. The system equation becomes $y = [H_1\Sigma_1^{1/2}, H_2\Sigma_2^{1/2}]Pu + n = Au + n$.

Now, design the GDFE receiver. This involves a Cholesky factorization of $(A^*A + \Sigma_{uu}^{-1}) = G^*SG$, where $\Sigma_{uu} = I$. GDFE yields a set of parallel AWGN sub-channels and the corresponding SNR for each sub-channel is computed from the Cholesky factor. Note that each user has as many sub-channels as the number of dimensions, thus the input bit sequence need to be arbitrarily divided into sub-streams. In the detector, GDFE decodes each sub-stream as an individual user and then re-assembles the decoded bits back to the original order. Because GDFE decomposes the vector channel into a set of parallel independent sub-channels, the exact same code as used in AWGN channels can now be applied to each of the sub-channels.

Therefore, the number of bits in each sub-channel (or equivalently, the input constellation size) may be determined from the sub-channel SNR using standard single-user methods. The choice of constellation size would depend on the probability of error desired and the amount of coding used. Overall, this multiuser system design is optimal in the sense that with an optimal code for AWGN single-user channels, this encoding/decoding scheme achieves the multiuser capacity.

Finally, we will comment on an important issue in decision feedback design so far tactically ignored: error propagation. The capacity result for the multiuser detector here are calculated assuming correct decisions are made in each stage. In practical systems, this is not going to be true, and errors in early stage generate errors in all subsequent stages. Traditionally, precoding [12] is used to deal with this problem by moving the decision feedback structure to the encoder. Precoding is not possible in the present setting because it requires coordination among different users. We suggest the following methods to deal with the error propagation problem. First, error correcting codes may be used to decrease error probability at early stages. A block of transmitted vectors may be coded component-wise. Decision for the first vector component is not made until an entire block of codewords is processed, allowing the error correcting code to correct early decision errors. This would lessen the error propagation effect in subsequent stages. We could also take error propagation into account in bit-loading. Suppose P_i is the error probability of ith stage assuming no error propagation. With error propagation, the first stage error probability is still P_1, but the second stage error probability becomes $P_1 + P_2$, and the last stage error probability becomes $\sum P_i$. To take this effect into account in bit-loading, we can load the last stage with raw error probability $P_m = \frac{1}{2}P_e$, where P_e is the target probability of error. Then, load the second last stage with $P_{m-1} = \frac{1}{4}P_e$, etc., and first stage with $P_1 = \frac{1}{2^m}P_e$. This scheme would guarantee that with error propagation, the probability of error is below P_e. In effect, each stage takes a loss of approximately 0.2 dB (assuming target probability of error $P_e = 10^{-6}$).

V. CONCLUSIONS

This paper addresses the problem of optimal receiver design for Gaussian multiple access channels with vector input and outputs. The underlying goal is to achieve multiuser channel capacity. To this end, we first noted that the multiuser channel capacity may be explicitly characterized by solving a set of convex optimization problems, which also yields the optimal transmitter spectral density. We then proposed a multiuser detector based on generalized decision feedback equalizer (GDFE). This GDFE-based detector decomposes the multiuser channel into a set of parallel AWGN channels, allowing AWGN codes to be used in each sub-channel. We proved that the GDFE-based detector, together with an optimized transmitter, achieves the extreme points in the multiple access channel capacity region.

ACKNOWLEDGMENT

Discussions with Wonjong Rhee have been very helpful.

VI. REFERENCES

[1] Thomas Cover and Joy A. Thomas, *Elements of information theory*, Wiley, New York, 1991.

[2] Roger S. Cheng and Sergio Verdu, "Gaussian multiaccess channels with ISI: Capacity region and multiuser water-filling," *IEEE Transactions on Information Theory*, vol. 39, no. 3, pp. 773–785, May 1993.

[3] Wei Yu, Wonjong Rhee, and John M. Cioffi, "Transmitter optimization for vector multiple access channels," Submitted to VTC'2000.

[4] Alexandra Duel-Hallen, "Equalizers for multiple input/multiple output channels and pam systems with cyclostationary input sequences," *IEEE Journal on Selected Areas in Communications*, vol. 10, no. 3, pp. 630–639, April 1992.

[5] Sergio Verdu, *Multiuser Detection*, Cambridge University Press, 1998.

[6] Mahesh K. Varanasi and Tommy Guess, "Optimum decision feedback multiuser equalization with successive decoding achieves the total capacity of the guassian multiple-access channel," in *Proc. Asiloma Conf. on Signals, Systems and Computers*, 1997, pp. 1405–1409.

[7] John M. Cioffi and G. David Forney, "Generalized decision-feedback equalization for packet transmission with ISI and Gaussian noise," in *Communications, Computation, Control and Signal Processing: a tribute to Thomas Kailath*, Arogyaswami Paulraj, Vwani Roychowdhury, and Charles D. Shaper, Eds. 1997, Kluwer Academic Publishers.

[8] George Ginis and John M. Cioffi, "Vectored VDSL," 1999, ITU contribution.

[9] Lieven Vandenberghe, Stephen boyd, and Shao-Po Wu, "Determinant maximization with linear matrix inequality constraints," *SIAM journal on matrix analysis and applications*, vol. 19, pp. 499–533, 1998.

[10] John M. Cioffi, Glen P. Dudevoir, M. Vedat Eyuboglu, and G. David Forney, "MMSE decision feedback equalizers and coding: Part I and II," *IEEE Transactions on Communications*, vol. 43, no. 10, pp. 2582–2604, October 1995.

[11] Thomas Kailath, Ali Sayed, and Babak Hassibi, *State-space Estimation*, Prentice Hall, 1999.

[12] Richard D. Wesel and John M. Cioffi, "Achievable rates for Tomlinson-Harashima precoding," *IEEE Transactions on Information Theory*, vol. 44, no. 2, pp. 824–831, March 1998.