Optimum Decision Feedback Multiuser Equalization with Successive Decoding Achieves the Total Capacity of the Gaussian Multiple-Access Channel

Mahesh K. Varanasi, SENIOR MEMBER, IEEE, and Tommy Guess
ECE Dept., University of Colorado, Boulder, CO 80309-0425, USA.

Abstract— **The complex Gaussian Multiple-Access Channel (GMAC) $\underline{Y} = \mathbf{A}\underline{X} + \underline{N}$ is considered. The transmitters send information independently with a power constraint so that $\underline{X}$ has a product distribution with $E[|X_k|^2] \leq P_k$. It is known that multiuser codes exist that will achieve any rate-tuple in the capacity region of the GMAC provided that an optimum joint decoder is used [1] [2] [3]. However, little progress has been made in multiuser coding, and moreover, optimum joint decoding would be too complex. In this paper, we restrict the decoder to be of a successive decoding type with equalization. Such a decoder is parameterized by feedforward and feedback equalization vectors. The optimum successive decoder (OSD) is obtained by maximizing the mutual information for each user over those vectors. The key result of this paper is that the OSD achieves the total capacity of the GMAC at any vertex of the capacity region. With the OSD, each transmitter can use a single-user code independently of the other users. The complexity of equalization is also only linear in the number of users. For the conventional GMAC, $Y = \sum_{i=1}^{M} X_i + N$, where A is a row vector with unit elements, the OSD is degenerate and involves no equalization. It reduces to the well-known successive decoder for that channel as described in [4] [5].**

Furthermore, for the particular case of the uncoded channel, the OSD reduces to a new decision feedback multiuser detector. This detector is optimum in the sense that it maximizes Signal-to-Interference Ratio for each user.

I. The Gaussian Multiple-Access Channel (GMAC)

Consider the Gaussian multiple-access channel

$$\underline{Y} = \mathbf{A}\underline{X} + \underline{N}. \tag{1}$$

For the sake of generality, we allow all deterministic and random quantities to be complex-valued. The particular case of the real-valued GMAC models baseband channels and the complex GMAC models passband channels. For the latter, the complex random vectors can be restricted to being circularly symmetric or *proper* [6]. The GMAC is a vector channel where the information symbols of the M independent active users are represented by the elements of the vector $\underline{X} = [X_1 \cdots X_M]^{\mathrm{T}}$, with the superscript T denoting the matrix-transpose operation. $\underline{X}$ has a product distribution, and moreover, the power of the k^{th} user in units of energy per channel use, is constrained so that $E[|X_k|^2] \leq P_k$. We let $\mathbf{P}$ be a diagonal matrix with P_k as its k^{th} diagonal element. $\mathbf{A}$ is a $K \times M$ matrix with complex-valued elements. The complex Gaussian random noise vector $\underline{N}$ has zero-mean and a covariance matrix $E[\underline{N}\underline{N}^{\dagger}] = \mathbf{N}$ (this is denoted as $N \sim \mathcal{N}(\underline{0}; \mathbf{N})$), and it is assumed to be statistically independent of $\underline{X}$.

This paper was presented in part at the Intl. Symp. Inform. Th., Ulm, Germany, in July 1997. This work was supported by NSF grants NCR-9406069 and NCR-9706591.

The particular case where $\mathbf{A}$ is a square identity matrix results in M decoupled single-user channels. This case models Orthogonal Waveform or Time-Division Multiple-Access (OWMA and TDMA) channels [7]. On the other extreme is the conventional real GMAC [8, Sec. 14.1.2] given as $Y = \sum_{i=1}^{M} X_i + N$ with a product distribution on $\underline{X}$, power constraints $E[X_k^2] \leq P_k$, and with N being a zero-mean Gaussian random variable $E[N^2] = \sigma^2$. This channel is a special case of the GMAC with a real-valued scalar output, and the matrix $\mathbf{A}$ being a row vector of unit elements. It models the Identical Waveform Multiple-Access (IWMA) channel [7].

The non-trivial instances of $\mathbf{A}$ arise in channels of practical interest such as the Correlated Waveform Multiple-Access (CWMA) channel (or CDMA) and in Bandwidth Efficient Multiple-Access (BEMA) communications [7]. There are also distortive channels in which the transmitters cannot entirely control the signals that arrive at the receiver. In such channels, it may be impossible to force $\mathbf{A}$ to be an identity matrix or a row vector of identical elements.

Following the work of [1] [2], the capacity region of the GMAC with power constraints was obtained in [3] for the real GMAC and was later extended to the complex case in [9]. It is therefore known that there exist multiuser codes that, together with optimum joint decoding at the receiver, will achieve any rate-tuple within the capacity region. However, good multiuser codes haven't been found, and even if they were to be found, optimum joint decoding would most likely be too complex.

II. Successive Decoding with Equalization

A successive decoder decodes users on a per-user basis in an arbitrary but fixed order (here, we assume without loss of generality that the users are decoded in the increasing order of their indices). It is parametrized by feedforward and feedback equalization vectors. The feedforward equalizers are described by the set of K-length vectors $\{\underline{F}_k\}_{k=1}^{M}$ whereas the feedback equalizers are described by the sets of K-length vectors $\{\underline{B}_{k1}, \underline{B}_{k2}, \ldots, \underline{B}_{kk}\}_{k=1}^{M-1}$. In decoding the k^{th} user, the interference from the already decoded "past" users (i.e., users $1, \cdots, k-1$) is mitigated by subtracting from $\underline{Y}$ a linear combination of the k^{th} user's feedback vectors. The coefficients of the linear combination are the decoded and re-encoded symbols of those past users. Following that, an inner product of the resulting vector and the feedforward vector of the k^{th} user is computed as an input to the k^{th} user's decoder.

In particular, the first user is decoded based on Z_1 which is obtained by taking the inner product between $\underline{F}_1$ and the matched filter output vector $\underline{Y} = \underline{Y}_1$ so that $Z_1 =< \underline{F}_1, \underline{Y}_1 > \triangleq \underline{F}_1^\dagger \underline{Y}_1$. The decoded output is re-encoded to produce $\hat{X}_1$. The second user is similarly decoded based on Z_2, where Z_2 is now obtained as the inner product between $\underline{F}_2$ and a vector $\underline{Y}_2$ which in turn is obtained by subtracting the vector $\underline{B}_{11}\hat{X}_1$ from $\underline{Y}$. Therefore, $\underline{Y}_2 = \underline{Y} - \underline{B}_{11}\hat{X}_1$ and $Z_2 =< \underline{F}_2, \underline{Y}_2 >$. In general, the k^{th} user is decoded based on $Z_k =< \underline{F}_k, \underline{Y} - \sum_{j=1}^{k-1} \underline{B}_{k-1j}\hat{X}_j >$. The function of the feedforward equalization through $\underline{F}_k$ is to minimize the effects of multiple-access interference from the as yet undecoded ("future") users $k+1$ to M, and the residual interference after subtraction, if any, from the already decoded ("past ") users 1 to $k-1$. The function of feedback equalization through the feedback vectors $\{\underline{B}_{k-1j}\}_{j=1}^{k-1}$ is to mitigate the interference contributed by the past users. Since we are concerned with capacity, we assume that all re-encoded symbols are error-free. Note that Z_k can hence be expressed as

$$Z_k = \underline{F}_k^\dagger \underline{A}_k X_k + \sum_{j=k+1}^{M} \underline{F}_k^\dagger \underline{A}_j X_j \quad (2)$$

$$+ \sum_{j=1}^{k-1} \underline{F}_k^\dagger (\underline{A}_j - \underline{B}_{k-1j}) X_j + \underline{F}_k^\dagger \underline{N}. \quad (3)$$

We assume that the users' symbols are distributed according to the product distribution that maximizes the capacity region so that they are independent zero-mean Gaussian random variables with covariance matrix $\mathbf{P}$. Therefore, from the viewpoint of the decoder for the k^{th} user, Z_k is effectively the output of a single-user Gaussian channel. Note that near-capacity achieving coding and decoding schemes such as TCM [10] and Turbo coding [11] are understood best for such channels.

A. The Optimum Successive Decoder

In this section, we seek the successive decoder that maximizes the mutual information $I(X_k; Z_k)$ between X_k and Z_k for each $k = 1, 2, \cdots, M$. We therefore define the optimal feedforward and feedback vectors as follows:

$$\underline{F}_k^{\text{opt}}, \{\underline{B}_{k-1j}^{\text{opt}}\}_{j=1}^{k-1} \in \arg \max_{\underline{F}_k, \{\underline{B}_{k-1j}\}_{j=1}^{k-1}} I(X_k; Z_k). \quad (4)$$

The maximum mutual information thus obtained is denoted as

$$C_k^{osd} = \max_{\underline{F}_k, \{\underline{B}_{k-1j}\}_{j=1}^{k-1}} I(X_k; Z_k). \quad (5)$$

Before we obtain the optimum successive decoder and its associated maximum mutual informations, we introduce the following notation and definitions.

Notation: For any $1 \le k \le M$, let $\mathbf{A}_{(k)}$ denote the sub-matrix of $\mathbf{A}$ formed by retaining only columns k to M so that $\mathbf{A}_{(k)} = [\underline{A}_k \cdots \underline{A}_M]$. Moreover, let $\mathbf{P}_{(k)}$ denote the principal sub-matrix of $\mathbf{P}$ formed by indices k to M, that is $\mathbf{P}_{(k)} = \text{diag}(P_k, \ldots, P_M)$.

Define the following sets of matrices:

$$\mathbf{K}_{(k)} = \mathbf{A}_{(k)} \mathbf{P}_{(k)} \mathbf{A}_{(k)}^\dagger + \mathbf{N} \quad (6)$$

$$\mathbf{H}_{(k)} = \mathbf{K}_{(k)} - P_k \underline{A}_k \underline{A}_k^\dagger. \quad (7)$$

The next lemma records two important relationships between the above-defined matrices. The proof is omitted.

Lemma 1: For the matrices defined in (6), the following identities hold:

$$\underline{A}_k^\dagger \mathbf{H}_{(k)}^{-1} = \left(1 - P_k \underline{A}_k^\dagger \mathbf{K}_{(k)}^{-1} \underline{A}_k\right)^{-1} \underline{A}_k^\dagger \mathbf{K}_{(k)}^{-1} \quad (8)$$

$$1 + P_k \underline{A}_k^\dagger \mathbf{H}_{(k)}^{-1} \underline{A}_k = \left(1 - P_k \underline{A}_k^\dagger \mathbf{K}_{(k)}^{-1} \underline{A}_k\right)^{-1}. \quad (9)$$

The next theorem gives explicit formulae for the optimum equalization vectors and the associated maximum mutual informations.

Theorem 1 (Optimum Successive Decoder) The feedback equalization vectors of the OSD are given as

$$\underline{B}_{k-1j}^{\text{opt}} = \underline{A}_j \qquad 2 < k \le M \quad 1 \le j \le k-1, \quad (10)$$

and the feedforward equalization vectors admit the expressions

$$\underline{F}_k^{\text{opt}} = \alpha_k \mathbf{H}_{(k)}^{-1} \underline{A}_k, \quad (11)$$

for any $\alpha_k \neq 0$. Furthermore, the maximum mutual information (in nats per channel use) is given by

$$C_k^{osd} = \log\left(1 + P_k \underline{A}_k^\dagger \mathbf{H}_{(k)}^{-1} \underline{A}_k\right). \quad (12)$$

Proof: For the k^{th} user, we have from (2) that

$$I(X_k; Z_k) = \log(1 + \gamma_k), \quad (13)$$

where γ_k is the signal-to-interference (SIR) ratio defined as

$$\gamma_k = \frac{|\underline{F}_k^\dagger \underline{A}_k|^2 P_k}{\mathcal{I}_k + \sum_{j=k+1}^{M} |\underline{F}_k^\dagger \underline{A}_j|^2 P_j + \underline{F}_k^\dagger \mathbf{N} \underline{F}_k} \quad (14)$$

where $\mathcal{I}_k$ is the residual interference from past users and is equal to $\sum_{j=1}^{k-1} \left|\underline{F}_k^\dagger (\underline{A}_j - \underline{B}_{k-1j})\right|^2 P_j$. Since the capacity is monotonically increasing in γ_k, we can equivalently maximize the SIR. The numerator is not a function of the feedback vectors, and no matter what the feedforward vectors are, the denominator is minimized when the residual interference is set equal to zero. Hence, $\underline{B}_{k-1j}^{opt} = \underline{A}_j$ for all $1 \le j \le k-1$. When the feedback vectors are so chosen, the expression for SIR reduces to

$$\gamma_k = \frac{\underline{F}_k^\dagger (\underline{A}_k \underline{A}_k^\dagger P_k) \underline{F}_k}{\underline{F}_k^\dagger \mathbf{H}_{(k)} \underline{F}_k}. \quad (15)$$

To maximize the SIR, we rewrite the problem in terms of a Rayleigh-Ritz parameterization of an eigenvalue. To this end, consider the Cholesky factorization to decompose the positive-definite matrix $\mathbf{H}_{(k)}$ as $\mathbf{H}_{(k)} = \mathbf{C}_k^\dagger \mathbf{C}_k$, where $\mathbf{C}_k$ is lower triangular. Letting $\underline{V} = \mathbf{C}_k \underline{F}_k$, an invertible transformation, we equivalently seek to maximize

$$\gamma_k = \frac{\underline{V}^\dagger (\mathbf{C}_k^\dagger)^{-1} (\underline{A}_k \, \underline{A}_k^\dagger P_k) \mathbf{C}_k^{-1} \underline{V}}{\underline{V}^\dagger \underline{V}}. \quad (16)$$

The above expression is maximized for any $\underline{V}$ which is an eigenvector of the largest eigenvalue of the matrix $(\mathbf{C}_k^\dagger)^{-1}(\underline{A}_k \, \underline{A}_k^\dagger P_k)\mathbf{C}_k^{-1}$. This is a rank-one positive-semidefinite matrix so its only non-zero eigenvalue is also its maximum eigenvalue. That eigenvalue is therefore equal to its trace. Hence,

$$\gamma_k^{\text{opt}} = \operatorname{tr}\left((\mathbf{C}_k^\dagger)^{-1}(\underline{A}_k \, \underline{A}_k^\dagger P_k)\mathbf{C}_k^{-1}\right) \quad (17)$$

$$= P_k \underline{A}_k^\dagger \mathbf{C}_k^{-1} (\mathbf{C}_k^\dagger)^{-1} \underline{A}_k \quad (18)$$

$$= P_k \underline{A}_k^\dagger \mathbf{H}_{(k)}^{-1} \underline{A}_k, \quad (19)$$

where we have used the fact that $\operatorname{tr}(\mathbf{CD}) = \operatorname{tr}(\mathbf{DC})$ whenever both products are defined. By inspection, we see for any $\alpha_k \neq 0$ that $\underline{V} = (\mathbf{C}_k^\dagger)^{-1} \underline{A}_k$ is an eigenvector corresponding to this non-zero eigenvalue. Pre-multiplying this by $\mathbf{C}_k^{-1}$ yields the desired expression for $\underline{F}_k^{\text{opt}}$. The expression for the maximum mutual information C_k^{osd} is a direct consequence of (19). ■

III. The Representation Problem

In this section, we address the problem of finding simple matrix representations of the feedforward and feedback equalizers of the OSD obtained in Theorem 1. Let $\mathbf{A}_{(\overline{k})}$ denote the sub-matrix of $\mathbf{A}$ formed by the first $k-1$ columns of $\mathbf{A}$, i. e., $\mathbf{A}_{(\overline{k})} = [\underline{A}_1 \cdots \underline{A}_{k-1}]$. And similarly define $\underline{X}_{(\overline{k})}$ to be the sub-vector of $\underline{X}$ given by $[X_1 \cdots X_{k-1}]^{\mathrm{T}}$. Then the OSD produces

$$Z_k = (\underline{F}_k^{\text{opt}})^\dagger (\underline{Y} - \mathbf{A}_{(\overline{k})} \underline{X}_{(\overline{k})}), \quad (20)$$

which allows us to write

$$\underline{Z} = \mathbf{F}\underline{Y} - \mathbf{B}\underline{X}, \quad (21)$$

where the feedforward and feedback matrices $\mathbf{F}$ and $\mathbf{B}$ of the OSD are given as follows: the k^{th} row of $\mathbf{F}$ is $(\underline{F}_k^{\text{opt}})^\dagger$ and $\mathbf{B}$ is the strictly lower-triangular part of $\mathbf{FA}$ which we denote as $\mathbf{B} = \mathcal{L}(\mathbf{FA})$ for brevity. The implementation of the OSD can be realized as in Figure 1. $\mathbf{B}_{ij}$ denotes the $(i,j)^{th}$ element of the feedback matrix $\mathbf{B}$.

We state next the main result of this section. It gives simple matrix formulae for $\mathbf{F}$ and $\mathbf{B}$. The proof is given in [12] but it is omitted here due to space restrictions.

Theorem 2 (Representation) Let $\mathbf{L}^\dagger \mathbf{L}$ be a Cholesky decomposition of $\mathbf{A}^\dagger \mathbf{N}^{-1} \mathbf{A} + \mathbf{P}^{-1}$ with $\mathbf{L}$ being lower triangular, then

$$\mathbf{F} = \mathbf{D}\operatorname{diag}(\mathbf{L}^{-1})(\mathbf{L}^\dagger)^{-1}\mathbf{A}^\dagger \mathbf{N}^{-1} \quad \text{and} \quad \mathbf{B} = \mathbf{D}\operatorname{diag}(\mathbf{L}^{-1})\mathcal{L}(\mathbf{L}), \quad (22)$$

where $\operatorname{diag}(\mathbf{C})$ of square matrix $\mathbf{C}$ is a diagonal matrix whose diagonal elements are those of $\mathbf{C}$, and $\mathbf{D}$ is the diagonal matrix whose k^{th} diagonal element is $\alpha_k^* P_k^{-1}(1 - P_k \underline{A}_k^\dagger \mathbf{K}_{(k)}^{-1} \underline{A}_k)^{-1}$.

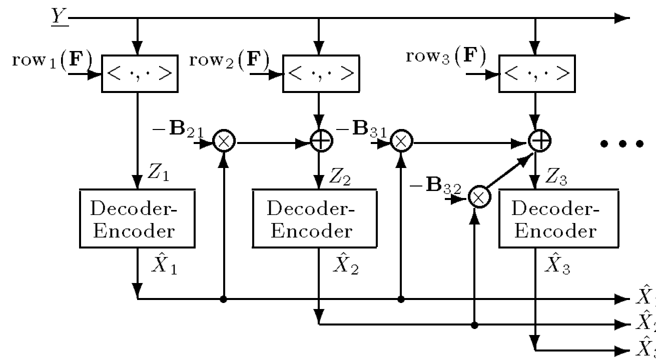

Fig. 1. Optimum Successive Decoder for the GMAC

IV. MSIR Decision Feedback Multiuser Detection

In this section, we consider the particular case of the uncoded channel. Here, the optimum successive decoder reduces to a decision feedback multiuser *detector*. In particular, the decoder-encoder blocks are replaced by the minimum distance decision rule

$$\hat{X}_k = \arg_{X \in \mathcal{F}_k} |Z_k - (\underline{F}_k^{opt})^\dagger \underline{A}_k X|^2 \quad (23)$$

where $\mathcal{F}_k$ is a QAM alphabet of the k^{th} user. Since the optimum successive decoder was obtained by maximizing mutual information, and mutual information is monotonic with signal-to-interference ratio, we call this multiuser detector the Maximum Signal-to-Interference Ratio (MSIR) decision feedback multiuser detector.

This brief detour shows, for the first time, an important connection between the multiuser detection (cf. [13] [14] [15]) and information theoretic approaches to multiple-access channels.

V. The OSD Achieves Total Capacity

From [1] [2] or [8, Sec. 14.3.6], we know that the capacity region of the GMAC is given by the convex hull of the regions

$$\bigcap_{\mathcal{J} \subseteq \{1,\ldots,M\}} \left\{ (R_1, \ldots, R_M) : 0 \leq \sum_{j \in \mathcal{J}} R_j \leq I(\underline{X}_{\mathcal{J}}; \underline{Y} | \underline{X}_{\overline{\mathcal{J}}}) \right\} \quad (24)$$

over all product distributions of $\underline{X}$ on $\mathbb{C}^M$ that satisfy $E\{|X_k|^2\} \leq P_k$ for each k. Here we use the notation $\underline{X}_{\mathcal{J}}$ to denote the sub-vector of $\underline{X}$ formed by retaining only those elements whose indices are in $\mathcal{J}$, and by $\overline{\mathcal{J}}$ we mean the complement of $\mathcal{J}$ with respect to the set $\{1, 2, \ldots, M\}$. Define the matrix $\mathbf{S}$ as

$$\mathbf{S} = \mathbf{I} + \mathbf{P}^{1/2}\mathbf{A}^\dagger \mathbf{N}^{-1} \mathbf{A} \mathbf{P}^{1/2}. \quad (25)$$

Let $\mathbf{S}_{\mathcal{G}\,\mathcal{H}}$, for $\mathcal{G}, \mathcal{H} \subseteq \{1, \ldots, M\}$, denote the sub-matrix of $\mathbf{S}$ formed by deleting the rows contained in $\bar{\mathcal{G}}$ and the columns contained in $\bar{\mathcal{H}}$.

Lemma 2: With $\underline{Y} = \mathbf{A}\underline{X} + \underline{N}$ and $\underline{X} \sim \mathcal{N}(\underline{0}; \mathbf{P})$ and $\underline{N} \sim \mathcal{N}(\underline{0}; \mathbf{N})$, the mutual information $I(\underline{X}; \underline{Y})$ and the conditional mutual information $I(\underline{X}_{\mathcal{J}}; \underline{Y} | \underline{X}_{\overline{\mathcal{J}}})$ admit the following formulas:

$$I(\underline{X}; \underline{Y}) = \log|\mathbf{S}| \quad \text{and} \quad I(\underline{X}_{\mathcal{J}}; \underline{Y} | \underline{X}_{\overline{\mathcal{J}}}) = \log|\mathbf{S}_{\mathcal{J}\mathcal{J}}| \quad (26)$$

where $|\mathbf{C}|$ denotes the determinant of the square matrix $\mathbf{C}$.

The following theorem was proved in [3] for the real GMAC and was extended to the complex case in [9].

Theorem 3 (Capacity Region of GMAC) The capacity region of the GMAC is given by

$$\bigcap_{\mathcal{J}\subseteq\{1,\ldots,M\}} \left\{ (R_1,\ldots,R_M):\ 0 \leq \sum_{j\in\mathcal{J}} R_j \leq \tfrac{1}{2}\log\left(|\mathbf{S}_{\mathcal{J}\mathcal{J}}|\right) \right\}. \tag{27}$$

It was shown in [3] and [9, Sec. 6.2] that the product distribution for $\underline{X}$, wherein $\underline{X}$ is a complex, proper, zero-mean Gaussian random vector with $\mathbf{P}$ as its covariance matrix, results in a region that contains the regions defined by *every* other valid product distribution[1]. Note from Lemma 2 that the capacity region given by Theorem 3 is the polyhedral region of (24) evaluated for $\underline{X} \sim \mathcal{N}(\underline{0};\mathbf{P})$.

A. Vertices of the Capacity Region of the GMAC

The *vertex* of the polyhedral region described by (24) for a given permutation of the M users, say $\{i_1,\ldots,i_M\}$, is the point in the region where, for telescoping sets defined as

$$\mathcal{G}_M = \{i_M\}, \mathcal{G}_{M-1} = \{i_{M-1}, i_M\}, \cdots, \mathcal{G}_1 = \{i_1, i_2, \ldots, i_M\}, \tag{28}$$

the following M out of the $2^M - 1$ constraining inequalities of the region are satisfied with equality:

$$\begin{aligned} R_{i_M} &= I(\underline{X}_{\mathcal{G}_M};\underline{Y}|X_{\overline{\mathcal{G}}_M}) && (29)\\ R_{i_M}+R_{i_{M-1}} &= I(\underline{X}_{\mathcal{G}_{M-1}};\underline{Y}|X_{\overline{\mathcal{G}}_{M-1}}) && (30)\\ &\vdots \\ \sum_{j=1}^{M} R_{i_j} &= I(\underline{X};\underline{Y}). && (31)\end{aligned}$$

Note that all vertices (there is one for each permutation of the user indices) satisfy the condition that the inequality for the sum rate of all the users is satisfied with equality. They are thus said to lie on the *dominant face* of the polyhedral region. Since there are M factorial permutations of the set of users, there are up to M factorial distinct vertices of this type.

In the case of the GMAC where the polyhedral region defined by (27) is the capacity region , the vertices correspond to rate-tuples whose sum is the maximum sum rate (or *total capacity*) of the channel. Whenever no pair of users are orthogonal to each other, there will be M factorial distinct vertices.

In Figure 2, we illustrate a three-user capacity region for the GMAC. Pictured are two vertices A and B which lie on the dominant face. For A, the following three of seven inequalities are satisfied with equality,

$$\begin{aligned} R_2 &= I(X_2;\underline{Y}|X_1,X_3) && (32)\\ R_2+R_3 &= I(X_2,X_3;\underline{Y}|X_1) && (33)\\ R_1+R_2+R_3 &= I(\underline{X};\underline{Y}) && (34)\end{aligned}$$

[1] Unlike the real-valued GMAC, this maximal product distribution for the complex-valued channel need not be unique [6].

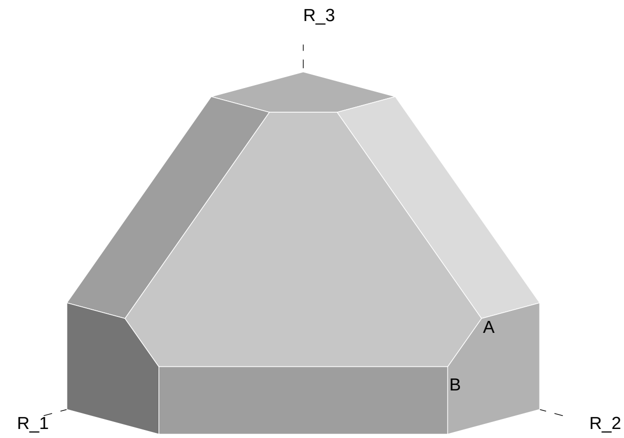

Fig. 2. A Three-User Capacity Region for the GMAC

so that A corresponds to the permutation $(1,3,2)$. Similarly, note that vertex B corresponds to the permutation $(3,1,2)$.

B. The Rate-Tuple of the OSD

We begin with the definition of the Schur complement of $\mathbf{S}_{\overline{\mathcal{G}}\overline{\mathcal{G}}}$ given as

$$\mathbf{S}^{(\mathcal{G})} = \mathbf{S}_{\mathcal{G}\mathcal{G}} - \mathbf{S}_{\mathcal{G}\overline{\mathcal{G}}}(\mathbf{S}_{\overline{\mathcal{G}}\overline{\mathcal{G}}})^{-1}\mathbf{S}_{\overline{\mathcal{G}}\mathcal{G}}. \tag{35}$$

We will have occasion to make use the following two lemmas. The first one gives a factorization of the determinant of any positive-definite matrix [16, Sec. 0.8.5] and the next one gives closed form expressions for certain key mutual information quantities.

Lemma 3: For any $M\times M$ positive-definite matrix $\mathbf{C}$, and any two non-empty sets $\mathcal{G}$, $\mathcal{H} \subseteq \{1,\ldots,M\}$,

$$|\mathbf{C}_{\mathcal{G}\cup\mathcal{H}\,\mathcal{G}\cup\mathcal{H}}| = |\mathbf{C}_{\mathcal{H}\mathcal{H}}|\left|(\mathbf{C}_{\mathcal{G}\cup\mathcal{H}\,\mathcal{G}\cup\mathcal{H}})^{((\mathcal{G}\cup\mathcal{H})\setminus\mathcal{H})}\right|, \tag{36}$$

where the set $\mathcal{J}\setminus\mathcal{H}$ denotes the set with all elements in $\mathcal{J}$ that are not in $\mathcal{H}$.

Lemma 4: With $\underline{Y} = \mathbf{A}\underline{X} + \underline{N}$ and $\underline{X} \sim \mathcal{N}(\underline{0};\mathbf{P})$ and $\underline{N} \sim \mathcal{N}(\underline{0};\mathbf{N})$, and with the sets $\mathcal{H} \subseteq \mathcal{G} \subseteq \{1,\cdots,M\}$, the mutual information $I(\underline{X}_{\mathcal{G}};\underline{Y})$ and the conditional mutual information $I(\underline{X}_{\mathcal{H}};\underline{Y}|\underline{X}_{\mathcal{G}\setminus\mathcal{H}})$ admit the following formulas:

$$\begin{aligned} I(\underline{X}_{\mathcal{G}};\underline{Y}) &= \log\left|\mathbf{S}^{(\mathcal{G})}\right| && (37)\\ I(\underline{X}_{\mathcal{H}};\underline{Y}|\underline{X}_{\mathcal{G}\setminus\mathcal{H}}) &= \log\left|(\mathbf{S}_{\mathcal{H}\cup\overline{\mathcal{G}}\,\mathcal{H}\cup\overline{\mathcal{G}}})^{(\mathcal{H})}\right|. && (38)\end{aligned}$$

Theorem 4 (OSD Achieves Total Capacity) The OSD achieves the rate-tuples corresponding to the vertices of the capacity region.

Proof: We will prove the theorem for the vertex corresponding to the permutation $\{1,2,\cdots,M\}$. It will be shown that the OSD derived in Section 1 that decodes users according to the increasing order of their indices achieves that vertex. Any other vertex can be achieved by the OSD that decodes users in the order dictated by the permutation corresponding to that vertex.

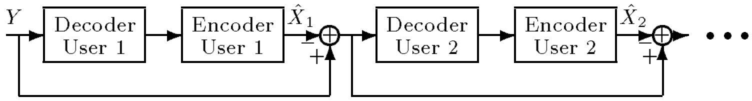

Fig. 3. Successive Decoder for the Conventional Gaussian Multiple-Access Channel

The vertex of interest is characterized by the following M equalities:

$$C_M = I(X_M; \underline{Y} | X_{M-1}, \ldots, X_1) \quad (39)$$
$$C_M + C_{M-1} = I(X_M, X_{M-1}; \underline{Y} | X_{M-2}, X_{M-3}, \ldots, X_1) \quad (40)$$
$$\vdots$$
$$\sum_{i=1}^{M} C_i = I(\underline{X}; \underline{Y}). \quad (41)$$

Equivalently, we see that using the chain rule for mutual information, this vertex is also characterized by the equalities

$$C_k = I(X_k; \underline{Y} | X_{k-1}, \ldots, X_1) \quad 1 \leq k \leq M. \quad (42)$$

Define the telescoping sets $\mathcal{G}_k = \{k, \ldots, M\}$. Setting $\mathcal{H} = \{k\}$ and $\mathcal{G} = \mathcal{G}_k$ in Lemma 4, another equivalent characterization of the vertex of interest is also the equalities:

$$C_k = \log | (\mathbf{S}_{\mathcal{G}_k \mathcal{G}_k})^{(\{k\})} | \quad 1 \leq k \leq M. \quad (43)$$

In order to show that the OSD derived in Section 1 achieves the above vertex, we need only show that the expressions for the rates C_k^{osd} achieved by the OSD given in equation (12) do indeed coincide with the rates C_k in (43). It is easy to see that $\mathbf{S}_{\mathcal{G}_k \mathcal{G}_k} = \mathbf{I} + \mathbf{P}_{(k)}^{1/2} \mathbf{A}_{(k)}^{\dagger} \mathbf{N}^{-1} \mathbf{A}_{(k)} \mathbf{P}_{(k)}^{1/2}$, whence the inverse of this matrix can be expressed via the Woodbury identity as $\mathbf{I} - \mathbf{P}_{(k)}^{1/2} \mathbf{A}_{(k)}^{\dagger} (\mathbf{A}_{(k)} \mathbf{P}_{(k)} \mathbf{A}_{(k)}^{\dagger})^{-1} \mathbf{A}_{(k)} \mathbf{P}_{(k)}^{1/2}$. The Schur complement $(\mathbf{S}_{\mathcal{G}_k \mathcal{G}_k})^{(\{k\})}$ in turn is the reciprocal of the first diagonal element of that inverse which can therefore be expressed as $(1 - P_k \underline{A}_k^{\dagger} \mathbf{K}_{(k)}^{-1} \underline{A}_k)^{-1}$. By a result in Lemma 1, the latter expression is equal to $1 + P_k \underline{A}_k^{\dagger} \mathbf{H}_{(k)}^{-1} \underline{A}_k$ which in turn coincides with C_k^{osd} in (12). The proof is complete. ∎

VI. Successive Decoding for the Conventional GMAC

As a corollary of Theorems 1 and 4, we have the following result for the conventional GMAC.

Corollary 1: For the conventional GMAC, $Y = \sum_{k=1}^{M} X_k + N$, the optimum successive decoder is degenerate where the optimum feedforward and feedback vectors are scalars and all equal (to unity), i.e., $\underline{F}_k^{opt} = 1$ and $\underline{B}_{k-1j}^{opt} = 1$. It achieves the total capacity of that channel at a vertex.

The OSD for the conventional GMAC therefore requires no equalization. It coincides with the well-known successive decoder for the conventional GMAC (cf. [4] [5] and [8]). This successive decoder is shown in Figure 3.

The fact that the successive decoder achieves the vertices of the capacity region of the conventional GMAC has also been known since [4] [5]. That result can now be understood as a particular case of Theorem 4. Moreover, the results of this paper explain *why* the successive decoder achieves the vertices of the capacity region of the conventional GMAC.

VII. Conclusions

This paper provides a systematic approach to multiuser equalization and information theory of the Gaussian multiple-access channel. It is shown that optimum decision feedback equalization together with successive (single-user) decoding allows us, without loss of total capacity, to side-step the multiuser coding (and joint optimal decoding) problem. Independent single-user coding and the optimum succesive decoder are shown to be sufficient to achieve the total capacity of the channel at the vertices of the capacity region.

Moreover, the optimum successive decoder reduces to the Maximum Signal-to-Interference Ratio multiuser decision feedback *detector* for the particular case of the uncoded channel. This result shows for the first time a key connection between the multiuser detection and information theoretic approaches to multiple-access channels.

References

[1] R. Ahlswede, "Multi-way Communication Channels," *Proc. of the 2nd IEEE Int. Symp. Info. Transmission*, U.S.S.R., 1971.

[2] H. Liao, "Multiple-Access Channels," Ph.D. dissertation, EE Dept., Univ. Hawaii, Honolulu, Hawaii, 1972.

[3] S. Verdu, "Capacity Region of Gaussian CDMA Channels: The Symbol-Synchronous Case," *Proc. 24th Allerton Conf Commun., Contr. and Comput.,* Allerton, IL, Oct. 1986, pp. 1025-1034.

[4] A. D. Wyner, "Recent Results in the Shannon Theory," *IEEE Trans. Inform. Theory*, vol. IT-20, No. 1, pp. 2-10, Jan. 1974.

[5] T. Cover, "Some Avances in Broadcast Channels" in *Advances in Communication Systems Theory and Applications*, A. Balakrishnan and A. Viterbi (eds.), Academic Press, New York, 1975.

[6] F. D. Nesser and J. L. Massey, "Proper Complex Random Processes with Applications to Information Theory," *IEEE Trans. Inform. Theory*, vol. IT-39, no. 4, pp. 1293-1302, Jul. 1993.

[7] M. K. Varanasi and T. Guess, "Bandwidth Efficient Multiple-Access via Signal Design for Decision Feedback Receivers," Proc. Communication Theory Mini-Conference, IEEE-Globecom, Phoenix, AZ Nov. 1997.

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

[9] M. Ruph, *Coding for CDMA Channels and Capacity* , Hartung-Gorre Verlag Konstanz, Ph.D. Thesis, Swiss Federal Institute of Technology, Zurich, Switzerland, 1994.

[10] E. Biglieri, D. Divsalar, P. J. McLane, and M. K. Simon, *Introduction to Trellis-Coded Modulation with Applications*, Macmillan Publishing Co., New York, 1991.

[11] C. Berrou, A. Glavieux and P. Thitimajshima, "Near Shannon Limit Error-Correcting Coding: Turbo Codes," *Proc ICC'93*, Geneva, Switzerland, pp. 1064-1070, May 1993.

[12] M. K. Varanasi, *Lecture Notes on Advanced Digital Communication, ECEN 5002*, ECE Dept., University of Colorado at Boulder, 1997.

[13] S. Verdu, "Multiuser Detection," in *Advances in Statistical Signal Processing: Signal Detection*, H. V. Poor and J. B. Thomas, eds., JAI Press, 1993, pp. 369-410.

[14] S. Moshavi, "Multi-User Detection for DS-CDMA Communications," *IEEE Communications Magazine*, Oct. 1996, pp. 124-136.

[15] M. K. Varanasi, "Optimizing Symmetric Energy and Permuting Users for Decision Feedback Multiuser Detection to User-Wise Outperform Linear Multiuser Detection," *Proc. Conf. Info. Sc. & Sys.*, pp. 492-497, Johns Hopkins University, MD, March 1997.

[16] R. A. Horn and C. R. Johnson, *Matrix Analysis*, Cambridge University Press, Melbourne, Australia, 1993.